Analysis of the effect of cell wall protein AGP31 on seed storage protein CRU3 translation in Arabidopsis

Arun Krishnaraj

Bio331L, Fall 2020

**Abstract**

Arabinogalactan protein 31 (AGP31) is a cell wall protein in Arabidopsis that aids the regulation of regular plant development through interaction with signaling hormones. Cruciferin 3 (CRU3) is a seed storage protein (SSP) involved in Arabidopsis seedling development and maturation. AGP31 null-mutant line (6134) seedlings are known to experience transient upregulation of CRU3 mRNA during embryo maturation compared to wild type seedlings (Ler). We examine CRU3 protein presence in green and root samples across AGP31 variants to better understand the role of cell wall proteins in SSP translational regulation. Bradford assay of purified protein samples showed somewhat low protein yields from tissue of 1.29-2.15 mg/g and higher protein concentrations for 6134 samples. Gel dye performed on SDS-PAGE of protein samples showed acceptable protein separation and intactness. Antibody signaling on PVDF blots showed stronger CRU3 antigenic bands for green-part samples and near equal bands between AGP31 variants. CRU3 bands have lower molecular weight than expected in all samples. From this experiment, CRU3 protein is suggested to be regulated by alternate pathways, contributing to equivalent antigenic bands and unexpected band size.

**Introduction**

The predictable developmental stages of plants and accumulation of photosynthetic energy in storage systems has been historically utilized and expanded for agricultural and human health benefit. The accumulation of energy and nutrients into plant seeds has been modified to be more favorable for human consumption through agriculture; edible seeds include most grains, legumes, and nuts, which have grown to contribute significantly to human dietary practices. Understanding how plants respond to external stimuli to regulate energy storage systems enables more effective agricultural practices and could enable genetic modifications to produce more desirable crops. The mechanisms by which seed storage proteins, the primary energy storage form of seeding plants, are regulated is not fully known. Additional understanding of seed storage protein regulation may enable improvements to dietary value or availability of common seed foodstuffs such as rice, millet, and corn (Lin et al., 2013).

Regular plant development and response to environmental stimuli are mediated by several plant hormones such as abscisic acid, jasmonates, cytokines, and auxin. These hormones are able to affect changes in gene and protein regulation throughout the plant by diffusion transport and interaction with a number of signaling pathways. While exact mechanism details are incomplete, regulatory hormones are thought to interact with cell wall proteins to modify intracellular processes. Plant signaling hormones are known to exhibit significant crosstalk in the determination of eventual physiological response. Additionally, large amounts of functional redundancy, and feedback regulation are present in hormone effector pathways, further allowing effective regulation if certain pathways are inactivated (Wolters and Jürgens, 2009).

Plant seeds accumulate seed storage proteins during development, which allow for storage of materials later used in germination and seedling growth. Seed storage proteins (SSPs) may vary between plants but are ubiquitously accumulated in high levels in specific seedling tissues at key points in development (Li et al., 2007). The primary SSP in Arabidopsis are cruciferins, which exist as 12S globulin proteins prior to cleavage into α (20 kD) and β (30 kD) subunits (Heath et al., 1986). These subunits are eventually assembled into hexamers for material storage but can be seen as distinct bands when stained on SDS gel. Cruciferin proteins undergo a variety of post-translational modifications induced by signaling pathways, leading to differing molecular weights based on cellular conditions (M Mehdy, personal communication). Cruciferin 3 (CRU3) is known to experience regulation by plant hormones during regular seedling development. Arabinogalactan protein 31 (AGP31) is a cell wall protein known to play a role in the regulation of CRU3 mRNA transcription in seedlings (Liu, 2007) (M Mehdy, personal communication). While it is known that AGP31 null-mutant seedlings (6134) experience upregulation of CRU3 mRNA compared to wild-type lines, whether these mRNAs are later modified to produce normal levels of CRU3 protein is unknown. Considering the important role of CRU3 in seedling development, I expect to see no difference in CRU3 protein across AGP31 variants. We tested this hypothesis by purifying protein samples from AGP31 variants and separating by SDS-PAGE to observe the presence of bands characteristic of CRU3 subunits; CRU3 was then selectively quantified by fluorescent antibody detection. From this experiment, we find that 6134 seedlings experience no increase in CRU3 protein during seed maturation, suggesting that CRU3 translation is regulated by a mechanism other than AGP31. Additionally, we observe that samples containing seedling leaf portions show higher CRU3 protein than samples containing root portions.

**Results**

***Phylogenetic Analysis***

Prior to analyzing variation of CRU3 protein across AGP31 variants, we would like to investigate AGP31 protein homology. The neighbor-joining analysis tree provides evidence of transferability of findings to other protein mutant species and identifies potential plant species of interest for further study.

Diagram

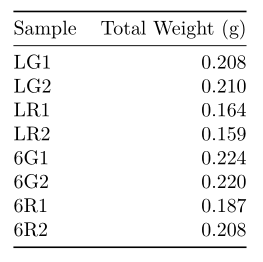
Description automatically generated

**Figure 1.** The neighbor-joining tree is generated from the alignment of full-length amino acid sequences of the homologs of AtAGP31. The protein sequences of various genes were collected from the NCBI website, and evolutionary analyses were conducted using MEGAX. The numbers given to the left of nodes are bootstrap value, which represents the phylogenetic confidence of the tree topology. Bootstrap values, less than 50, are not shown on the tree.

The large bootstrap values obtained indicate that AGP31 homologous protein is very likely to exist in other species. HyPRP1 is very similar to *Arabidopsis thaliana* AGP proteins, being found to share a most-recent ancestor in nearly all trials. We also note the proximity of the two Arabinogalactan proteins, indicating the high similarity of AtAGP31 to AtAGP30.

***Biological Materials***

Our goal was to investigate whether cruciferin proteins occur at a higher level in AGP31-mutant line (6134) than wild-type (Ler). We have from prior semesters that cruciferin mRNA is upregulated in 6134 seedlings around 13 days of growth. Ler and 6134 mutant line Arabidopsis were grown on MS media and harvested at 11 days of growth; seedlings were separated at stem-root junctions to generate green and root samples. It is known that the green segment of the seedlings tends to contain higher concentration of seed storage proteins (SSPs) than the root portions. Tissue samples were ground and combined to obtain larger masses for each treatment.

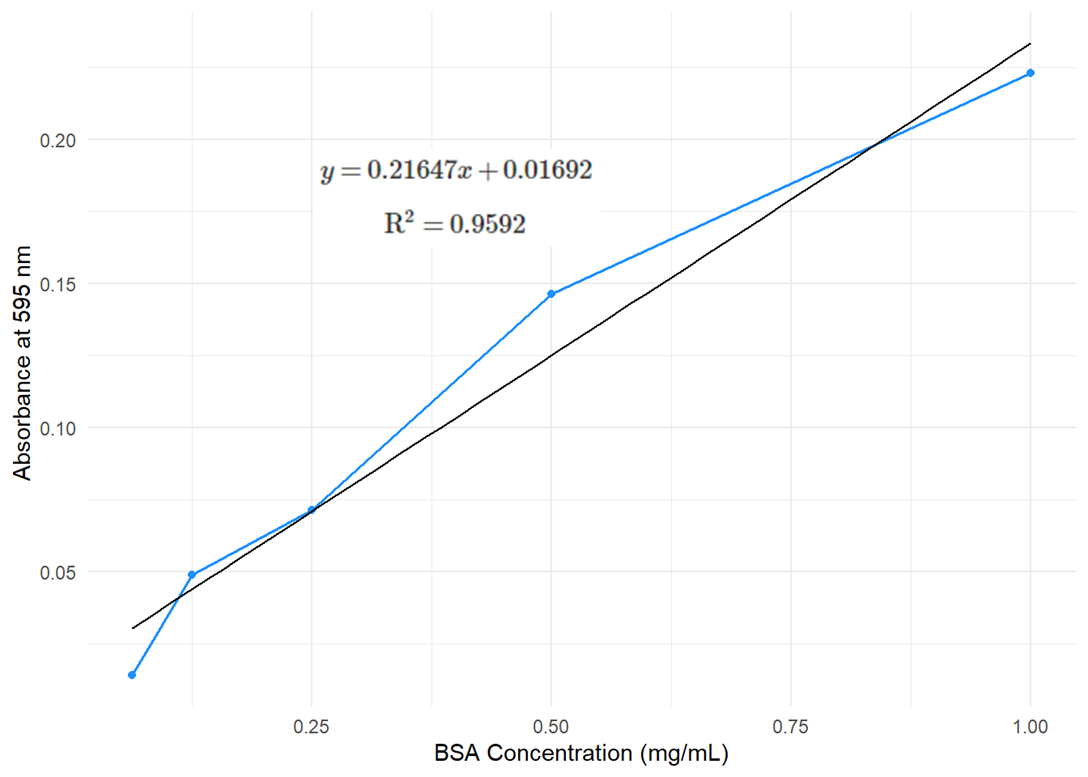


**Table 1.** Sample tissue weights are measured for treatment groups after seedling separation. Replicates are constructed to be similar weight, permitted by seedling number.

A relatively uniform weight of tissue samples was obtained for each sample, with slight variations owing to limitations on number of biological replicates and seedlings available.

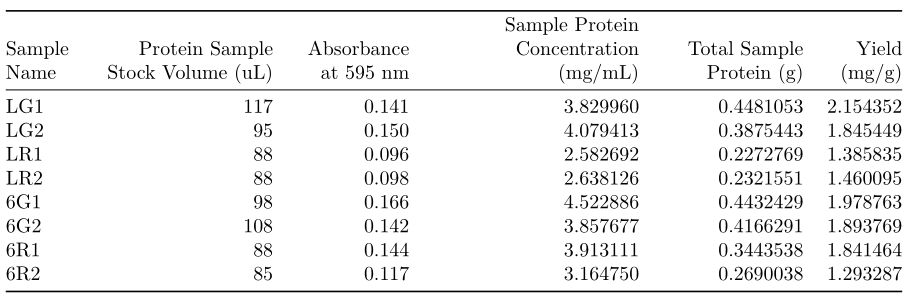
***Protein Purification***

Our goal for this section was to isolate protein from seedling tissue samples, while taking steps to reduce protein degradation or contamination. We used a commercial extraction buffer to obtain protein samples, which were then validated for quality through Bradford Assay. Using the absorbance values for a reference set of known BSA dilutions, we can then calculate sample protein concentrations and yields.



**Figure 2.** BSA standardization curve is constructed from samples containing 30 µL BSA standard stock. Linear regression formula uses MSE and reports adjusted-R2.

The high explanatory power of the linear regression indicates that the BSA curve can be accurately used to calculate sample protein concentrations and amount.

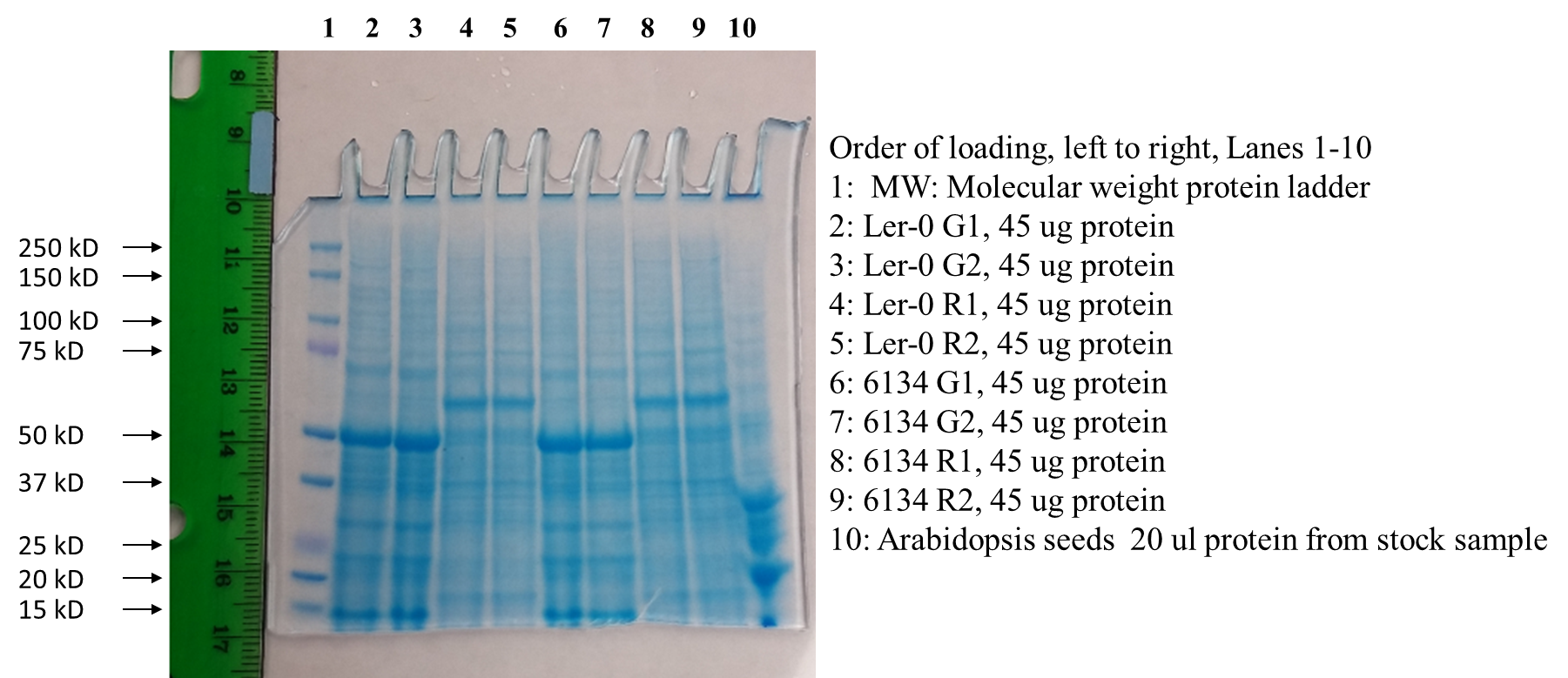


**Table 2.** Aliquots containing 5 µL of sample are measured for absorbance at 595 nm. Protein concentration is predicted for each sample using linear regression formula presented in Figure 2. Total protein weight is found using volume of sample stock, and yield is calculated using initial tissue weight presented in Table 1.

While the average protein concentration is near 3.6 mg/mL, green-part samples had significantly higher concentrations than root-part samples. Additionally, protein concentration was larger in 6134 mutant samples compared to Ler samples. The total amount of protein obtained and yield from samples is somewhat variable, owing to large variations in volume of protein sample stock. The protein samples were determined to be pure enough to proceed with the experiment.

***SDS-PAGE***

Our goal for this portion of the experiment was to separate our protein samples by molecular weight and verify intactness. We were also interested in verifying CRU3 protein presence in each sample. SDS-PAGE was performed on protein samples, and a replicate gel was treated with Coomassie blue dye to identify of proteins in gel lanes.

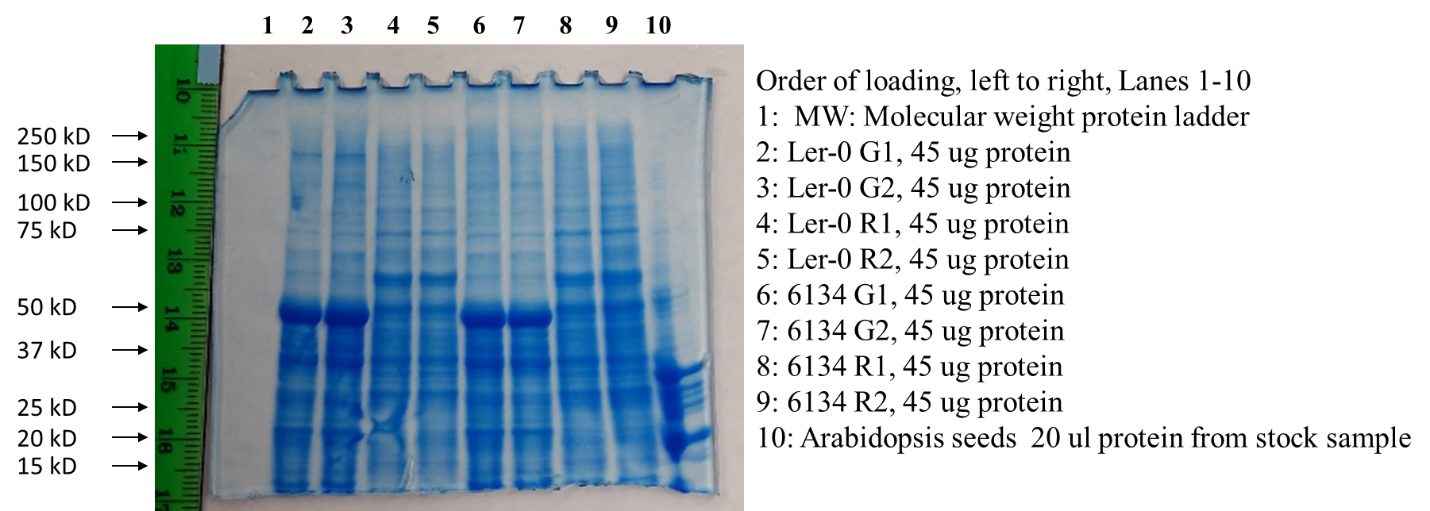


**Figure 3.** SDS-PAGE gel replicate 3 was stained with Coomassie Blue dye. Gel lanes contain equal masses of sample protein and Bio-Rad Precision Plus MW standard is included in lane 1.

We see clear bands overall and minimal smearing, suggesting that protein separation was successful, and samples are highly intact. Protein staining intensity is relatively consistent between genetic lines. Green-part and root-part samples show strongly staining bands in differing positions, especially past 75 kD. Staining in the 20-35 kD region seems to occur more strongly in green-part samples.

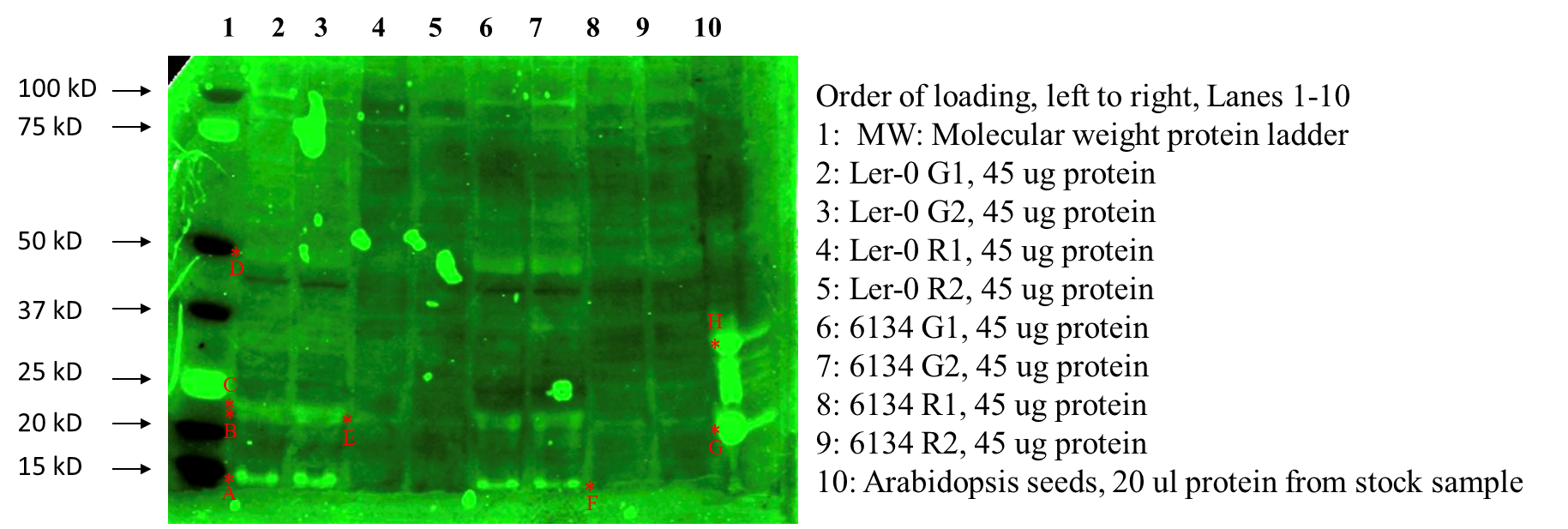
***Western Blotting/ Antibody Detection***

After generating protein separation gels via SDS-PAGE, and validating presence of proteins near size of interest, we would like a method of specifically quantifying the amount of CRU3 protein in each sample. To this end, we employed Western blotting and antibody detection. SDS-PAGE gel replicates were transferred to PVDF membranes, after which staining was performed to verify protein transfer.



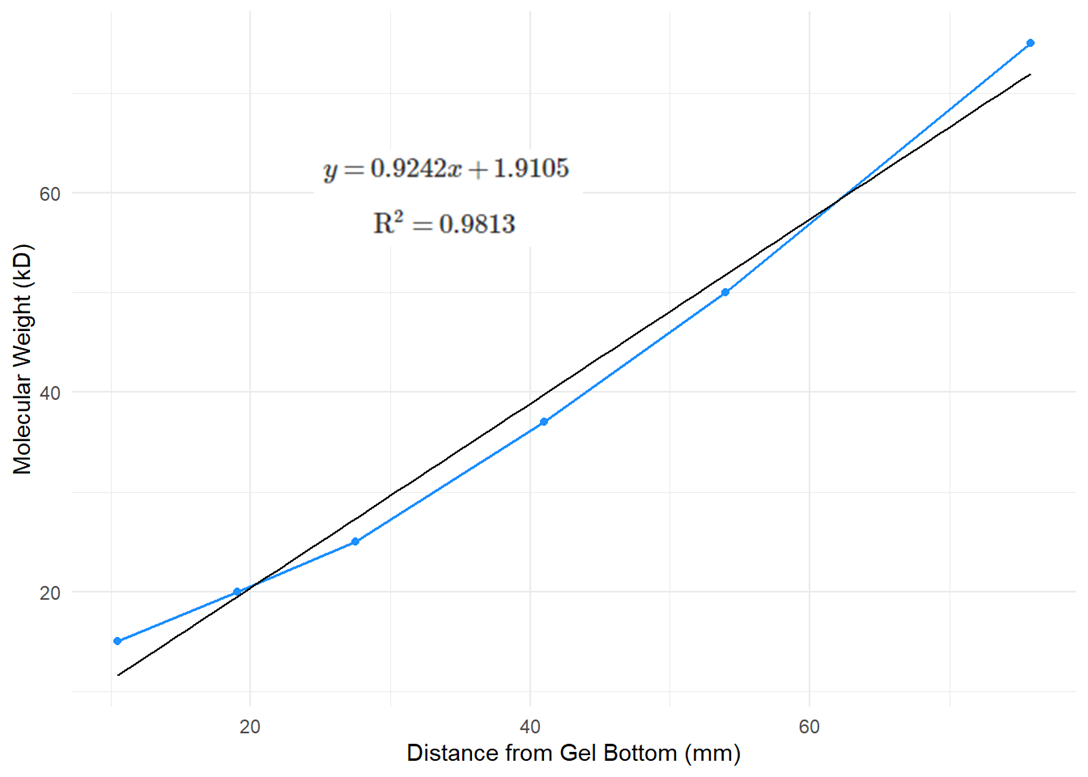
**Figure 4.** SDS-PAGE gel replicate 1 is stained with Coomassie Blue following protein transfer to PVDF membrane. Gel lanes had been loaded with equal masses of sample protein.

The dye bands of the SDS gel after transfer appear closer to the surface, indicating that the protein samples have moved out of the gel to the blotting membrane. This transfer result was deemed sufficient to proceed with antibody probing. The secondary antibody fluorescence method was used to detect CRU3 in the generated blots, the following Western blot was obtained after incubation.

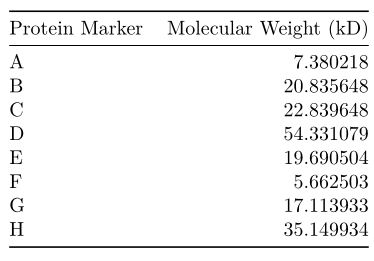


**Figure 5.** Western Blot is generated from SDS-PAGE gel replicate 1. Signal amplification is accomplished by secondary-antibody fluorescence detection. Mouse anti-legumin and Goat anti-mouse antibodies are used as primary and secondary antibodies.

The blot image shows irregular green areas, indicating nonspecific antibody binding to PVDF gel. Lane 10 produces a smear-like band indicative of lane overloading. Bands near 20 kD are more prevalent in green-part and Arabidopsis seed samples, shown by markers B, C, E. Bands near 15 kD are more prevalent in green-part samples, shown by markers A, and F. Antigenic binding near 50 kD is stronger for green-part samples, shown by marker D. CRU3 proteins of interest are expected to occur in the region 37-15 kD. Ler green-part samples have slightly stronger antigenic bands than 6134 green-part samples in this region. No appreciable difference in band staining can be seen between genetic lines for root-part samples. Bands H and G appear almost exactly where CRU3 subunits are expected to occur and occur at higher MW than corresponding bands in seedling samples.



**Figure 6.** MW standardization is constructed from MW ladder bands between 15 kD and 75 kD, with distance measured digitally. Linear regression formula uses MSE and reports adjusted-R2.



**Table 3.** Molecular weight is constructed for each protein marker using linear regression formula presented in Figure 6. Protein markers refer to bands indicated in Figure 5.

Markers B, C, and E are near 20 kD, while markers A and F are near 6 kD. Lane 10 markers H and G appear near 35 and 17 kD respectively. Marker D appears near 55 kD across all samples.

**Discussion**

Cruciferin proteins play an important role in seed development and regular stress response in plants. While it is known that AGP31-mutant line 6134 shows upregulation of CRU3 mRNAs during seedling development, the experiment performed indicates the near equivalence of CRU3 protein in 6134 and Ler samples. Additionally, CRU3 protein seems to occur in higher amount in green-part samples, supporting prior results on the accumulation of SSPs in pre-cotyledon structures. Our experiment indicates that 6134 CRU3 mRNA translation is modified by some alternate mechanism to result in equivalent protein level.

Phylogenetic analysis of Arabidopsis AGP31 using NCBI protein sequences indicates significant homology to Arabinogalactan proteins in other species, as well as to proline-rich proteins. The relative homology of AGP31 to other cell-wall proteins thought to be important in SSP transcription regulation may be useful for further research on interactions of cell-wall signaling proteins and SSP mRNA or protein production.

Biological replicate seedlings were grown and combined to obtain roughly equivalent tissue samples. While separation of green and root parts occurred based on known CRU3 variation, root samples were somewhat smaller. Ideally, samples could have been overall increased in size, which would result in larger amounts of protein available for analysis, thereby increasing gel or antibody binding strength and clarifying results. Additionally, while we originally planned to harvest seedlings at day 13, the closure of the lab due to campus public-health policies forced harvesting at day 11. This difference may account for some of the observed variation in protein sample concentrations. Further refinements to this experiment may include several measurement periods across the seed development stages to allow quantification of CRU3 protein rate of change.

Following sample protein extraction with commercial extraction buffer, sample protein concentrations were found using Bradford Assay. The high R2 of the linear regression indicates that the assay is highly linear for the indicated range of protein concentrations. Protein concentration and absorbance was found to be larger for green-part samples. Sample protein concentration was also found to be larger for 6134 green and root part samples than those of the Ler line. Total sample protein values also suggest that significantly more protein occurs in 6134 root samples than Ler counterparts. The yields across samples is highly variable, largely owing to the variability in initial sample tissue weights. Yields are also low compared to prior studies (M Mehdy, personal communication). This could have been a result of the difference in harvesting time mentioned prior or due to some technical error replicated in all samples. Despite this issue, the consistency between sample replicates was deemed sufficient to continue with analysis. Further iterations of this study could include more sample replicates, so that outliers may be removed for a more statistically significant result, as well as potentially use different extraction buffer. Additionally, the differing amount of BSA and sample volume used in aliquots may have introduced some slight error in calculations. Ideally, enough initial tissue sample would be available to use equal amounts of protein volume in concentration determination.

SDS-PAGE gels showed good protein separation and distinct bands between green and root part samples after staining. Green part samples showed a strong band at 50 kD, while root part samples had a strong band at 60 kD. These bands are thought to be Rubisco and size differences can be attributed to variable enzyme structures across plant tissue types (Andersson, 2008). The higher staining in the 20-30 kD region in green samples is expected to result from higher amount of CRU3 subunit proteins. The difference in staining near 15 kD is expected to result from differing amounts of albumin SSP subunits. While it seems that there is more protein in the size of interest in green part samples, dye staining is non-specific, and in order to quantify the presence of CRU3 subunits we require additional analysis.

Blots treated with secondary antibody fluorescence detection show higher antigenic activity at 50 kD, 20-25 kD, and 15 kD for green part samples. The Arabidopsis seed antigenic bands at markers G and H are thought to be the CRU3 subunits β and α due to their sizes of 35 and 17 kD respectively. Taking this result into account, we believe markers B, C, and E to reflect the positions of CRU3 β subunit with sample MW near 20 kD. This interpretation would also suggest that markers A and F reflect the positions of the CRU3 α subunit near 6 kD. These molecular weights are unexpected for the CRU3 α and β subunits and could be due to alternative splicing of CRU3 (Wan and Ross, 2007) or targeted protein degradation (Zhang et al., 2017) (Higashi et al., 2006); while the known increase in CRU3 subunit mRNA in 6134 seedlings suggests that the processing effect is proteomic, this experiment provides no definitive explanation for observed CRU3 subunit size variation. The legumin primary antibody is known to detect Arabidopsis cruciferins well and is not expected to bind to the albumin subunits present near 7 kD (M Mehdy, personal communication). Further experiments could include more replicates to understand if CRU3 subunit modification is occurring, as well as utilize multiple time points to visualize changes in CRU3 subunit size; this experiment replication would allow suggestion of likely mechanism of CRU3 subunit modification based on temporal antigen banding patterns. Based on availability, alternate antibody may be used to potentially generate more specific binding patterns. I would also recommend more gel replicates be used, as technical errors like nonspecific binding and lane overload complicate blot imaging by washing out antigenic binding in treatment lanes. Strongly antigenic bands near 50 kD are expected to reflect Rubisco in sample. The 50 kD bands in leaf samples are consistent with known patterns of Rubisco variation across plant tissue (Andersson, 2008). Overall, this result suggests that CRU3 protein is nearly equal for AGP31 variants across cell-type samples, and that 6134 CRU3 mRNA upregulation does not result in CRU3 protein accumulation in 11-day seedlings. As discussed, this result is somewhat modified by the low amount of protein sample available for use in antibody analysis and the lack of measurement at differing time points. Expanding this experiment to have more tissue sample available and measuring CRU3 protein presence at multiple time points would allow interpretation of changes in CRU3 protein for 6134 and Ler samples. Potential expansions to this work could apply terminomic or proteomic techniques to identify the contributions of targeted degradation and alternative splicing to the observed CRU3 subunit sizes. Additionally, samples from cell lines mutant for several cell wall proteins thought to regulate developmental transcription may be used to determine interaction effects in protein translation.

**References**

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiology and Biochemistry*, *46*(3), 275–291. https://doi.org/10.1016/j.plaphy.2008.01.001

Heath, W. (1986). Analysis of storage proteins in normal and aborted seeds from embryo-lethal mutants of Arabidopsis thaliana. Planta, 169(3), 304–312. https://doi.org/10.1007/BF00392124

Higashi, Y., Hirai, M.Y., Fujiwara, T., Naito, S., Noji, M., Saito, K. (2006). Proteomic and transcriptomic analysis of Arabidopsis seeds: molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. The Plant Journal : for Cell and Molecular Biology, 48(4), 557–571. https://doi.org/10.1111/j.1365-313X.2006.02900.x

Li, Q., Wang, B., Xu, Y., Zhu, Y. (2007). Systematic Studies of 12S Seed Storage Protein Accumulation and Degradation Patterns during Arabidopsis Seed Maturation and Early Seedling Germination Stages. *BMB Reports*, *40*(3), 373–381. <https://doi.org/10.5483/BMBRep.2007.40.3.373>

Lin, Y., Pajak, A., Marsolais, F., McCourt, P., Riggs, C., Yang, H. (2013). Characterization of a cruciferin deficient mutant of Arabidopsis and its utility for overexpression of foreign proteins in plants. *PloS One*, *8*(5), e64980–e64980. <https://doi.org/10.1371/journal.pone.0064980>

Liu, C. (2007). *Characterization and functional analysis of arabinogalactan protein 31 in Arabidopsis*.

Wan, L., Ross, A., Yang, J., Hegedus, D., Kermode, A. (2007). Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant Arabidopsis thaliana (thale cress) seeds. *Biochemical Journal*, *404*(2), 247–256. https://doi.org/10.1042/bj20061569

Wolters, H., Jürgens, G. (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. *Nature Reviews. Genetics*, *10*(5), 305–317. https://doi.org/10.1038/nrg2558

Zhang, H., Gannon, L., Hassall, K., Deery, M., Gibbs, D., Holdsworth, M., Hoorn, R., Lilley, K., Theodolou, F. (2018). N‐terminomics reveals control of Arabidopsis seed storage proteins and proteases by the Arg/N‐end rule pathway. *The New Phytologist*, *218*(3), 1106–1126. https://doi.org/10.1111/nph.14909