
Characterisation of SS4 and MRC in starch granule initiation

Jiawen Chen

*A thesis submitted to the University of East Anglia for the degree of
Doctor of Philosophy*

University of East Anglia

John Innes Centre

December 2022

Registration no. 100241112

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Starch is the main storage carbohydrate in plants, important for maintaining plant growth. Starch granules in the leaves of most green plants look similarly disc-shaped, but there is a huge diversity in storage starch granule morphology across species and tissues. Starch granule initiation determines the size, shape and number of granules that form within a plastid, but we do not yet know its mechanism. Several granule initiation proteins have been discovered in Arabidopsis, with Starch Synthase 4 (SS4) at the centre, as the only enzyme. One of the direct biochemical interactors of SS4 is Myosin Resembling Chloroplast protein, MRC.

Here, I explored the biochemistry and function of SS4 and MRC, to better understand their roles during granule initiation in Arabidopsis and wheat. Wheat endosperm contains a bimodal size distribution of small, spherical B-type granules, and large, flattened A-type granules. I characterised a previously unknown role for MRC in repressing the initiation of B-type granules during early wheat endosperm development, resulting in a higher volume percentage of B-type granules in mature *mrc* endosperm than in wild type. However, wheat *mrc* leaves had fewer granules per chloroplast than wild type, suggesting tissue-specific protein functions. Additionally, wheat SS4 and MRC can interact with each other along the entire SS4 protein. However, Arabidopsis *in planta* studies suggest that the SS4-MRC interaction might be transient. Size exclusion chromatography demonstrates that Arabidopsis and wheat SS4 exist at high molecular weight in leaves, and *in vitro*. Interestingly, negative stain transmission electron microscopy of the recombinant AtSS4 reveals densities that are consistent with an AtSS4 dimer, suggesting that stable AtSS4 dimers could assemble into more transient homooligomeric complexes, or the AtSS4 dimer might have an extended shape. Together, this work advances our knowledge on the biochemical context of SS4 function, with potential for further structural characterisation of SS4.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

Abstract.....	ii
List of abbreviations.....	viii
List of figures.....	xii
List of tables.....	xv
Acknowledgements.....	xvi
Chapter 1 – Introduction.....	1
1.1 Starch is the major storage carbohydrate in plants	1
1.1.1 The structure of starch.....	2
1.1.2 The diversity of starch granule morphologies	6
1.2 Mechanism of starch synthesis and degradation	9
1.2.1 Starch synthesis requires starch synthases (SS), starch branching enzymes (SBE) and debranching enzymes (DBE)	9
1.2.2 Starch granule initiation determines the number and morphology of starch granules in a plastid.....	12
1.2.3 Starch degradation requires phosphorylation and dephosphorylation	14
1.3 Starch synthases (SSs) have specific roles in polymer elongation during starch synthesis.15	
1.4 SS4 is crucial for Arabidopsis starch granule initiation	17
1.4.1 SS4 is involved in granule initiation, influencing the number and morphology of starch granules.....	17
1.4.2 The different phenotypes in <i>ss4</i> depend on many environmental and metabolic factors	20
1.4.3 The role of SS4 in generating a pre-amylopectin primer and <i>de novo</i> primer hypotheses	22
1.4.4 SS4 domains are involved in protein interactions and localisation	24
1.5 MRC is a direct interactor of SS4, important for starch granule initiation	29
1.6 Protein complexes as a general feature in starch formation	30
1.7 Starch granule initiation in crops	32
1.7.1 Granule initiation orthologs in crops	32
1.7.2 Wheat as a model to study granule initiation in leaves and endosperm	35
1.8 Biotechnology prospects for modifying starch granules	40
1.9 Aims of the thesis.....	43
Chapter 2 – Materials and methods	44
2.1 Plant material.....	44
2.1.1 <i>Triticum turgidum</i> for <i>mrc</i> phenotyping	44
2.1.2 <i>Triticum turgidum</i> for RNA sequencing	45
2.1.3 <i>Arabidopsis thaliana</i>	45
2.1.4 <i>Nicotiana benthamiana</i>	46

2.2 RNA-seq analysis of Kronos wheat developing endosperm	46
2.2.1 RNA extraction and sequencing.....	46
2.2.2 TPM value calculation and differential expression	46
2.3 Wheat bioinformatics.....	47
2.3.1 Publicly available endosperm expression data	47
2.3.2 <i>TtMRC</i> sequence analysis.....	47
2.4 Wheat phenotyping.....	48
2.4.1 Starch purification from mature and developing wheat endosperm	48
2.4.2 Total starch quantification from wheat leaves and endosperm.....	48
2.4.3 Grain morphometrics.....	49
2.4.4 Leaf granule image analysis	49
2.4.5 Coulter counter analysis of wheat endosperm starch	51
2.4.6 Analysis of amylopectin structure and amylose content.....	51
2.5 Light microscopy	52
2.5.1 Transmission light microscopy.....	52
2.5.2 Confocal microscopy	53
2.6 Plant material electron microscopy	53
2.6.1 Transmission electron microscopy	53
2.6.2 Scanning electron microscopy	54
2.7 Transmission electron microscopy of proteins	54
2.7.1 Negative staining	54
2.7.2 Imaging conditions	54
2.7.3 Data processing and NS-TEM image analysis.....	55
2.8 Cloning.....	56
2.8.1 Gateway cloning.....	56
2.8.2 In-Fusion cloning/Gibson Assembly	58
2.8.3 Plasmids	60
2.9 Transformation of organisms.....	61
2.9.1 <i>E.coli</i> transformation.....	61
2.9.2 <i>Agrobacterium</i> transformation.....	61
2.9.3 Bacterial glycerol stocks for storage	62
2.9.4 <i>Nicotiana benthamiana</i> transient transformation	62
2.9.5 <i>Arabidopsis thaliana</i> stable transformation by floral dipping	62
2.10 Generation of Arabidopsis transgenic lines	63
2.10.1 Screening by antibiotic resistance	63
2.10.2 Seed sterilisation.....	64
2.11 Generation of Arabidopsis lines expressing <i>TaSS4</i> and <i>TaMRC</i>.....	64
2.11.1 Crossing.....	64

2.11.2 Genotyping.....	65
2.12 Biochemistry methods.....	67
2.12.1 Crude protein extraction from Arabidopsis leaves and wheat leaves.....	67
2.12.2 Protein quantification.....	68
2.12.3 SDS-PAGE and Western Blot.....	68
2.12.4 <i>In vitro</i> recombinant protein expression in <i>E. coli</i>	69
2.12.5 <i>In vitro</i> starch synthase activity assay.....	72
2.12.6 Size Exclusion Chromatography (SEC).....	73
2.12.7 Acetone precipitation.....	74
2.13 Protein-protein interaction analysis.....	74
2.13.1 Co-immunoprecipitation (co-IP) in <i>N. benthamiana</i>	74
2.13.2 Yeast two-hybrid.....	75
2.14 Affinity purification Mass spectrometry.....	76
2.14.1 Affinity purification from Arabidopsis.....	76
2.14.2 Preparation of samples for mass spectrometry.....	76
2.14.3 Mass spectrometry data analysis.....	77
2.15 Statistics and data visualisation.....	77
Chapter 3 – The distinct roles of MRC in wheat leaves and endosperm.....	78
3.1 Introduction.....	78
3.2 Results.....	80
3.2.1 The wheat orthologs of MRC are encoded on chromosomes 6A and 6D.....	80
3.2.2 Expression analysis during early grain development reveals distinct patterns of starch granule initiation genes.....	84
3.2.3 Loss of MRC does not affect the growth of wheat plants, or grain development.....	87
3.2.4 Loss of MRC greatly alters starch granule size distributions in the mature endosperm.....	93
3.2.5 Loss of MRC results in the early initiation of B-type granules.....	99
3.2.6 <i>TaMRC</i> promotes granule initiation in wheat leaves.....	102
3.3 Discussion.....	108
3.3.1 A novel role for MRC in repressing B-type granule initiation during endosperm starch synthesis.....	108
3.3.2 MRC has tissue-specific roles in promoting or repressing granule initiation.....	111
3.3.3 Exploring the biochemical basis of the contrasting roles of MRC between organs.....	112
3.3.4 MRC as a gene target for biotechnological modification of starch granule size.....	113
Chapter 4 – Biochemical interactions between SS4 and MRC.....	114
4.1 Introduction.....	114
4.1.1 SS4 and MRC are conserved in land plants.....	114
4.1.2 The direct interaction between SS4 and MRC is of particular interest among SS4-associated proteins.....	115

4.2 Results	117
4.2.1 <i>TaSS4</i> can partially complement <i>ss4</i> in Arabidopsis, but <i>TaMRC</i> cannot complement <i>mrc</i>	117
4.2.2 <i>AtSS4</i> and <i>AtMRC</i> colocalise to puncta in the chloroplast.....	121
4.2.3 <i>SS4</i> and <i>MRC</i> interact in co-immunoprecipitation.....	125
4.2.4 <i>TaSS4</i> and <i>TaMRC</i> interact in yeast two-hybrid	129
4.3 Discussion	131
4.3.1 Both the N-terminus and C-terminus of <i>TaSS4</i> are sufficient for protein interactions.	131
4.3.2 The granule initiation protein interaction networks are different in Arabidopsis and wheat	132
4.3.3 <i>AtMRC</i> is not necessary but sufficient for directing the punctate localisation of <i>AtSS4</i>	133
Chapter 5 – <i>SS4</i> dimers could be part of a large protein complex	135
5.1 Introduction	135
5.2 Results	137
5.2.1 <i>SS4</i> is part of a high molecular weight assembly in Arabidopsis leaves and wheat leaves	137
5.2.2 Development of transgenic Arabidopsis lines	140
5.2.3 Affinity purification mass spectrometry of <i>AtSS4</i> -HA in Arabidopsis leaves	143
5.2.4 Optimisation of recombinant <i>SS4</i> and <i>MRC</i> purification from <i>E. coli</i>	146
5.2.5 The <i>SS4</i> protein alone forms high molecular weight assemblies <i>in vitro</i>	154
5.2.6 Structural prediction of <i>SS4</i> proteins suggests association of the N-termini	156
5.2.7 Negative staining TEM reveals potential dimeric assemblies of <i>AtSS4 in vitro</i>	160
5.3 Discussion	164
5.3.1 <i>SS4</i> could assemble into dimers and possibly higher order complexes	164
5.3.2 <i>MRC</i> may be a transient component of the <i>SS4</i> granule initiation complex.....	166
5.3.3 Complex formation as a method of substrate concentration	167
5.3.4 Complex formation and substrate concentration in different species.....	168
Chapter 6 - General Discussion	170
6.1 The role of starch granule initiation proteins can vary between species and tissues	171
6.2 Coordination of stable and transient protein interactions could facilitate the mechanisms of the granule initiation complex	173
6.3 How can we consolidate the punctate localisation of granule initiation proteins with the granule initiation complex?	176
References	178
Appendices.....	188
Appendix 1: List of Primers	189
Appendix 2: gBlocks synthesised	192
Appendix 3: Standard media components	195

Appendix 4: <i>Aegilops speltoides</i> MRC gene model.....	196
Appendix 5: Calibration of HMW standards on Superose 6 Increase 10/300	199
Appendix 6: Arabidopsis SEC immunoblot replicates	201

List of abbreviations

AD – activation domain
ADP – adenosine diphosphate
ADPG - ADP-glucose (adenosine diphosphoglucose)
AGPase - ADP-glucose pyrophosphorylase
AMY3 - α -amylase 3
ANOVA – analysis of variance
AP-MS – affinity purification mass spectrometry
arc3 - *accumulation and replication of chloroplast 3*
At – *Arabidopsis thaliana*
ATP - adenosine triphosphate
BAM - β -amylase
Bd – *Brachypodium distachyon*
BD – binding domain
bp – base pair
BiFC – bimolecular fluorescence complementation
BGC1 – B-Granule Content 1
CBM – carbohydrate binding module
cDNA – complementary DNA
CDS – coding sequence
Co-IP – co-immunoprecipitation
Col-0 – Columbia 0
CTP – chloroplast transit peptide
CV – column volume
cv. - cultivar
DBE – Debranching Enzyme
DEG – differentially expressed genes
DP – degree of polymerisation
dpa – days post anthesis
DPE1/2 - Disproportionating Enzyme 1/2
DNA – deoxyribonucleic acid
DTT - dithiothreitol
EDTA - ethylenediaminetetraacetic acid
EMS - Ethyl methanesulfonate
ESV1 – Early Starvation 1

F6P - fructose-6-phosphate
FBN - Fibrillin
FLO6 – Flourey Endosperm 6
G1P - glucose-1-phosphate
G6P - glucose-6-phosphate
GBSS - Granule Bound Starch Synthase
gDNA – genomic DNA
GFP – Green Fluorescent Protein
GS – Glycogen Synthase
GT1/5 – glycosyltransferase 1/5
GWD - Glucan, Water Dikinase
HMW – high molecular weight
HPAEC-PAD - High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection
Hv – Hordeum vulgare
IP-MS - immunoprecipitation mass spectrometry
IP - immunoprecipitation
ISA1/2/3 – Isoamylase 1/2/3
JIC – John Innes Centre
LB – Lysogeny Broth
LDA – Limit Dextrinase
LDH - Lactate Dehydrogenase
LESV – Like ESV
LiOAc - lithium acetate
LLPS – liquid-liquid phase separation
LM – light microscopy
LSF2 – LIKE SEX FOUR 2
MES - 2-(N-morpholino)ethanesulfonic acid
MFP1 - MAR-Binding Filament Protein 1
MOS - maltooligosaccharide
MRC - Myosin Resembling Chloroplast Protein
MS – Murashige and Skoog or mass spectrometry
NAD⁺ - nicotinamide adenine dinucleotide
NADH - nicotinamide adenine diphosphate hydride
NS-TEM – negative stain transmission electron microscopy
Os – Oryza sativa

p35S - cauliflower mosaic virus 35S promoter
PAE – predicted aligned error
PCR – polymerase chain reaction
PDB – Protein Data Bank
PEP - Phosphoenolpyruvate
PGI – Phosphoglucose Isomerase
PGM – Phosphoglucomutase
PHS1 – Plastidial Alpha-glucan Phosphoryase 1
Pi – orthophosphate
PII1 – Protein Involved In Starch Initiation 1
PK – pyruvate kinase
pLDDT – local distance difference test
PPi – inorganic pyrophosphate
PTST1/2/3 – Protein Targeting To Starch 1/2/3
PWD – Phosphoglucan, Water dikinase
RFP – Red Fluorescent Protein
RNA – ribonucleic acid
RNA-seq – RNA sequencing
rpm – revolutions per minute
SBE – Starch Branching Enzyme
SDS-PAGE - sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC – size exclusion chromatography
SEM - scanning electron microscopy or standard error of the mean
SEX1/4 – Starch Excess 1/4
SS – Starch Synthase
SSG4/6 – Substandard Starch Grain 4/6
Ta – Triticum aestivum
TBS - tris-buffered saline
TBST - tris-buffered saline with Tween20
TEM – transmission electron microscopy
TGW – thousand grain weight
TILLING – Targeting Induced Local Lesions In Genomes
TPM – transcripts per million
Tt – Triticum turgidum
UBQ – ubiquitin 10
UTR – untranslated region

UV - ultraviolet

WT – wild type

Y2H – yeast two-hybrid

YFP – Yellow Fluorescent Protein

Zm – *Zea mays*

List of figures

Figure 1.1 The hierarchal organisation of starch granule structure.

Figure 1.2. Overview of starch synthesis and degradation in a leaf chloroplast.

Figure 1.3 The three processes of starch synthesis.

Figure 1.4. SS4 has conserved domains and interacts with various proteins.

Figure 1.5. The granule initial proteins are important for determining starch granule number and morphology.

Figure. 1.6. Summary of known *AtSS4* interactions.

Figure 1.7. The granule initial proteins localise to distinct puncta in the chloroplast.

Figure. 1.8. Influence of starch structure on starch physicochemical properties.

Figure 2.1 Schematic of the generation of Arabidopsis transgenic lines.

Figure 2.2 Schematic of the generation of Arabidopsis lines expressing *TaSS4* and *TaMRC*.

Figure 3.1. *MRC* homeologs in wheat are encoded on chromosomes 6A and 6D, with a disruption of the 6B homeolog.

Figure 3.2. Phylogenetic analysis of wheat *MRC* sequences.

Figure 3.3. *MRC* is expressed in wheat leaves and endosperm during early grain development.

Figure 3.4 Mutations in *MRC* do not affect plant growth or the number and size of grains.

Figure 3.5 Mutations in *MRC* have variable effects on total starch in mature grains.

Figure 3.6. Endosperm starch in mature grains of *mrc* mutants have altered granule size distribution.

Figure 3.7. Endosperm starch granules of mature grains in three wheat *mrc* mutants are similar in shape compared to wild type.

Figure 3.8 Chain length distribution and amylose content of the *mrc* mutants.

Figure 3.9. In developing endosperm, granule number increases in *mrc-1* compared to WT despite having similar starch content.

Figure 3.10. The *mrc-1* mutant initiates B-type granules earlier in grain development than the WT.

Figure 3.11. Chloroplasts of *mrc-1* mutants have fewer and larger granules than the WT.

Figure 3.12. Loss of MRC has variable effects on the total end of day (ED) leaf starch content in wheat leaves

Figure 3.13. Model of MRC function in wheat developing endosperm.

Figure 4.1 Pairwise amino acid sequence alignment between wheat and Arabidopsis SS4 and MRC shows high sequence similarity between Arabidopsis and wheat orthologs.

Figure 4.2 *TaSS4* can partially complement Arabidopsis *ss4*, but *TaMRC* cannot complement *mrc*.

Figure 4.3 Localisation patterns of expression constructs in *Nicotiana benthamiana*.

Figure 4.4 Co-localisation of *AtSS4* and *AtMRC* using transient expression in *Nicotiana benthamiana*.

Figure 4.5 Co-immunoprecipitation (co-IP) of *AtSS4*-HA and *AtMRC*-RFP using transient expression in *N. benthamiana*.

Figure 4.6 Co-immunoprecipitation (co-IP) of *TaSS4* and *TaMRC* using transient expression in *N. benthamiana*.

Figure 4.7 Yeast two-hybrid of pairwise interactions between *TaSS4* and *TaMRC*.

Figure 5.1. SS4 exists at high molecular weight in size exclusion chromatography (SEC) fractions of Arabidopsis and wheat leaf total soluble protein extracts.

Figure 5.2. The development of transgenic Arabidopsis lines expressing tagged *AtSS4* and *AtMRC*.

Figure 5.3. Three independent stable lines of the *ss4/pAtSS4:AtSS4*-HA lines show complementation of the *ss4* phenotype to varying degrees.

Figure 5.4. *AtSS4*-HA is pulled down at various abundances in transgenic lines.

Figure 5.5. Plasmid maps of SS4 recombinant expression constructs.

Figure 5.6. *In vitro* purification of SS4 proteins produces full length SS4 proteins but also degradation products and contamination.

Figure 5.7. Recombinant *AtMRC* is soluble but can precipitate.

Figure 5.8. *In vitro* starch synthase activity assays of purified recombinant SS4 proteins using maltotriose as a substrate.

Figure 5.9. The SS4 protein alone forms high molecular weight assemblies *in vitro*.

Figure 5.10 Structural predictions of SS4, MRC and SS4 dimerization.

Figure 5.11 Protein disorder prediction of *AtSS4*, *TaSS4*, *AtMRC* and *TaMRC*.

Figure 5.12 Negative staining TEM reveals potential dimeric assemblies of *AtSS4*.

Figure 5.13. 3D density map of the *AtSS4* 2D classes show a size that is consistent with a dimer of *AtSS4*.

Figure 5.14. Hypothesised mechanisms of granule initiation by *SS4*.

List of tables

Table 1.1. Diversity in granule morphology and composition among major starch crops.

Table 1.2. Examples of gene targets used for altering starch structure and physicochemical properties in crops.

Table 2.1 BP reaction components.

Table 2.2 LR reaction components.

Table 2.3 Plasmids used in this thesis.

Table 2.4 Antibiotics used in this study and their working concentrations.

Table 2.5 Bradford assay standard curve preparation with BSA.

Table 2.6 Self-made SDS-PAGE gel components.

Table 2.7 List of primary and secondary antibodies used.

Table 3.1. Pairwise comparisons of number of tillers per plant.

Table 3.2. Pairwise comparisons of number of grains per plant.

Table 3.3. Pairwise comparisons of Thousand Grain Weight (TGW).

Table 3.4. Pairwise comparisons of grain size.

Table 3.5. Pairwise comparisons of leaf granule per chloroplast in young wheat leaves.

Table 3.6. Pairwise comparisons of granule size in young wheat leaves.

Table 3.7. Summary of leaf starch quantification experiments.

Table 5.1 Interaction partners of AtSS4-HA identified in an AP-MS of *ss4/pAtSS4:AtSS4-HA* lines.

Table 5.2 Calculations of average *k_{cat}* values for *in vitro* starch synthase activity assays of recombinant AtSS4 (A) and TaSS4 (B).

Acknowledgements

I would like to thank my supervisor, David Seung for all his guidance, patience, and enthusiasm over the past four years. During this PhD I have felt uncertain about many things, but he always took the time to discuss my questions in detail, however small, and filled discussions of new ideas with such energy that I came out of every meeting feeling inspired and ready to tackle new challenges. He has not only been a great support in my scientific journey but has also gone above and beyond to help me in my general career development. It has been a privilege to start in this lab from the beginning, and to see it grow so wonderfully over the years, heading into diverse and exciting directions.

I would also like to thank my secondary supervisor Alison Smith, for all her insight, support, and interesting discussions during my supervisory team meetings, and for all her thoughtful comments on the MRC manuscript. To my other supervisory team members: I am grateful to Matthias Thalmann, for his sharp observations and for being a great colleague. A big thanks to Michael Webster for all his guidance and discussions on the NS-TEM work over the past year, this has been such an exciting development in this project, and I feel very lucky to have the chance to learn some of these techniques.

I would like to thank Xiaoqi Feng for giving me the chance to do a great rotation in my first year, where I learned a lot of new concepts and techniques, and for staying on my supervisory team afterwards. Thanks to Toby for guiding me during the rotation and introducing me to phase separation. I am also grateful to Matt Moscou for supervising me during my first rotation, for the warm welcome into his group and for the introduction into the world of bioinformatics. I learned a lot during both of these rotations.

I would like to thank Dave Lawson and Jake Richardson for all their help with the NS-TEM in my project, I look forward to continuing with this work and learning more. Thank you to Gerhard Saalbach and Carlo Martins for helping me process several rounds of IP-MS samples. Thanks to Richard Hughes for his help on getting the SEC set up and for all the troubleshooting since, and to Marina Franceschetti for her advice on Strep-tag purifications. A big part of my PhD journey has also been trying to understand statistics and data visualisation. I am grateful to George Savva for his clear courses on R and for answering my questions, to James Brown for helping me with my questions at the beginning, to Matthew Hartley and Pirita Paajanen from the Informatics team for running R workshops during the pandemic, and to Martin Vickers for help on the wheat RNA-seq analysis.

I have really enjoyed being part of the Seung lab, and working together with such a welcoming, warm group of people. I thank all past and present members of the lab for their engaging discussions, daily help in the lab, fun lunches and cake breaks, and brilliant poems. To Alex, for all his help on the RNA-seq project. To Brendan, for his great help on cloning and Y2H while I have been writing, for tagging my wheat plants when I was away, and for all other help around the lab. To Daniel, for developing and showing me the starch synthase activity assay. To Rose, for the always enthusiastic discussions. To Stanislav, for his great help on pushing forward the Y2H work. To Erica, for all her work on the MRC project and for being a great colleague and friend from the start. To Lara, for her always interesting questions, for always being there for me and for the careful reading of this thesis.

Thanks to all the co-authors on the wheat MRC manuscript who I haven't yet named elsewhere – To Andy for the detailed bioinformatics analysis. To Elaine for all the microscopy sectioning and embedding on this project and on many other projects. To Robin, Fred and Andreas for the amylose content and amylopectin chain length analysis. To Kendall for getting this project started, and Cristobal for helping with the generation of the original wheat lines.

I also wouldn't have gotten here without valuable experiences gained in my undergraduate research projects. For this, I would like to thank Alistair McCormick for guiding me in my first summer lab project, which really sparked my interest into research, Justin Goodrich for a great project in my final undergraduate year, and Mark Banfield for the introduction into the world of protein purification, which has helped a great deal during my PhD. For this, I also want to thank Freya for being a great teacher.

I am grateful to Steph Bornemann for welcoming me into the JIC rotation programme, for the memorable journal clubs, and for his genuine care and guidance during that first year. Thanks to the John Innes Foundation for funding this programme. I want to thank all of my rotation gang – Jenny, Anna, Thomas and Lauren, it has been amazing to share these few years with you, and you are all an inspiration. A special thanks to Lauren for her help getting me set up doing Y2H, and of course for being a great housemate over the past few years.

Thanks to all my other friends in Norwich who have been with me along the way – all the dinners, walks, games and chats with you have made this a truly wonderful time. And finally, thanks to my family – my mum, dad and brother. Thank you for always being there for me and supporting me, for making me a freezer full of food to take with me every time we see each other, and for all your love.

Chapter 1 – Introduction

1.1 Starch is the major storage carbohydrate in plants

Green plants are distinct from animals in their ability to perform photosynthesis, converting light energy into chemical energy by fixing carbon from the air into sugars and starch. Because of this, plants are a crucial source of energy for animals, including humans who consume foods with high starch content as a major source of carbohydrates, such as wheat, rice, maize and potato. Due to its unique physicochemical properties, starch is also important as a raw material for many industrial purposes such as in binding agents, adhesives, coating agents and biodegradable plastics. Because starch is so widely used, it is important to understand the mechanism of how starch is synthesised in plants.

Plants photosynthesise during the day, using sugars for immediate energy provision and using starch to store energy for when photosynthesis is not possible, when starch is then broken down into sugars. Some starch breakdown does also happen during the day, but we still have much to learn on the exact balance between starch synthesis and degradation (more in section 1.2.3). Starch is made within specialised organelles called plastids, which include chloroplasts and amyloplasts. A portion of daily fixed carbon is converted to transitory starch in chloroplasts of leaves for use during the night (Seung & Smith, 2019). For longer term energy storage, plants make storage starch in amyloplasts of seeds or storage organs for mobilisation during germination or sprouting. Starch is also found in amyloplasts in roots and functions in gravitropism (Zhang *et al*, 2019). Transitory and storage starch are important for sustaining the growth and development of plants. Both are synthesised as semi-crystalline insoluble granules within plastids, but they each have their own roles within plant metabolism.

The important role of transitory starch in preventing night-time starvation is demonstrated by a starchless *Arabidopsis* mutant (*pgm*) (Caspar *et al*, 1985), which grows normally under constant light but is greatly impaired in its growth under short days, as it does not have enough energy to sustain the long night. There is a fine balance between carbon partitioning into sucrose for use during the day, or into starch for use during the night. Sucrose is synthesised in the cytosol, starch in the chloroplast, and these processes need to be coordinated so that fixed carbon is partitioned in a way that is favourable for plant growth.

One hypothesis is that leaf starch is produced as an overflow for photosynthesis, made when more than enough sucrose is available. Although this could be true in some species, it does not

seem to be the mechanism in *Arabidopsis*, where starch metabolism has been extensively studied. Under low light conditions, *Arabidopsis* still actively partitions carbon into starch (Streb & Zeeman, 2012). The balance between sucrose and starch production appears to be regulated by several feedback and feed-forward loops (Stitt & Zeeman, 2012). The length of the day and night has a large influence on partitioning carbon to sucrose or starch, with long nights causing more carbon partitioning to starch during the day, restricting growth and causing accumulation of sucrose that is not exported (Mengin *et al*, 2017).

Storage starch is synthesised in non-photosynthetic amyloplasts of seeds and storage organs using sucrose imported from the leaves (Smith & Zeeman, 2020). Amyloplasts differ from chloroplasts both structurally and biochemically (Sadali *et al*, 2019). They do not contain thylakoid membranes and some species have leaf-specific (chloroplast) and endosperm-specific (amyloplast) isoforms of starch biosynthesis proteins (Tetlow & Emes, 2014). Cereal grains are ~70% starch by dry weight (Hannah, 2007), and plants that are impaired in their storage starch synthesis can have shrunken (Tuncel *et al*, 2014) or inviable grains (Abe *et al*, 2014). Starch synthesised in the pollen is also important for pollen viability (Zhang *et al*, 2022).

1.1.1 The structure of starch

Starch granules consist of two kinds of glucose polymers - amylopectin, which is composed of α -1,4-linked glucose chains with frequent α -1,6-linked branches; and amylose, which is composed of mostly unbranched α -1,4-linked glucose chains (Figure 1.1A). Most of the starch granule is composed of amylopectin. Amylopectin chains assemble in a hierarchical manner (Smith & Zeeman, 2020) (Figure 1.1A): first, the branch placement in amylopectin forms clusters of linear chains, which allows adjacent chains to assemble into double helices. These helices pack together in an ordered fashion to form crystalline lamellae. Amorphous lamellae form where the branch points are, in between the crystalline lamellae. These crystalline and amorphous lamellae alternate with a periodicity of 9 – 10 nm and form the semi-crystalline granule matrix (Smith & Zeeman, 2020). This matrix can grow radially to form a starch granule. In some storage starches, a less crystalline region can be observed in the centre of a granule, called a hilum (Figure 1.1A). This is thought to reflect the initiation point.

There is still some debate as to how the amylopectin matrix is arranged (Nakamura & Kainuma, 2022). The traditional model is the cluster model (Nakamura & Kainuma, 2022) (Figure 1.1A) where clusters of amylopectin helices are arranged in tandem, and both short and long chains form a cluster together. The second model is the building block backbone model (Tetlow &

Bertoft, 2020), where long chains form the backbone and short chains compose the building blocks that arrange amylopectin in crystalline lamellae (Figure 1.1B). In the cluster model, short and long chains are both present in crystalline and amorphous lamellae, while in the backbone model, the long chains are present in the amorphous lamellae and the short chains are mostly present in the crystalline lamellae. This means that in the cluster model, all chains are formed in the same direction but in the backbone model the long and short chains are formed perpendicular to each other (Figure 1.1B). Most of the literature adheres to the cluster model, and there has been no evidence yet to disprove this model (Nakamura & Kainuma, 2022). These different models still exist after many years of starch research because it is difficult to visualise polymer arrangement within a semi-crystalline starch granule. In the coming years, advancements in imaging techniques might allow this to be resolved.

On a higher level of organisation than the amylopectin matrix, some storage starches have growth rings. These rings are organised with a periodicity on a scale of hundreds of nanometres, which are superimposed over the alternating lamellae. The precise mechanism of ring formation is unknown. One proposed theory is that they reflect periods of daily fluctuation in day- and night-time starch synthesis rates, but the evidence for this has been inconsistent (Smith & Zeeman, 2020). Leaf starches usually do not have growth rings, because of their small size and fast turnover (Zeeman *et al*, 2002).

Surprisingly, starch granules are only estimated to have 20 – 50% crystallinity, suggesting the rest of the granule is in an amorphous state (Lopez-Rubio *et al*, 2008). Amylose is deposited inside the starch granule, where it is thought to occupy the amorphous regions of the starch granule, where it can form single helices (Figure 1.1A). However, the exact location of amylose within the granule is uncertain, as amorphous regions do not only form between the crystalline lamellae (Seung, 2020). It could for instance also be located between the higher order growth rings mentioned above. There is a large variation in amylose content among starches from different species and organs. Arabidopsis leaf starch contains approximately 5 – 10 % amylose, whereas storage starches typically contain 15 – 35% amylose (Seung, 2020; Seung & Smith, 2019). Amylose is not essential for the formation of the semi-crystalline granule matrix and starch granules. Mutants without amylose in various species have no growth defects and natural populations of Arabidopsis with amylose-free starch have been found with no apparent detriment to fitness (Seung *et al*, 2020).

The semi-crystalline amylopectin matrix is crucial for the formation of an insoluble starch granule. In animals, fungi and bacteria, carbohydrates are stored in the form of glycogen. Glycogen particles are about 20 - 60 nm in diameter, and glycogen is also made up of a glucose polymer composed of α -1,4-linked and α -1,6-linked chains. However, glycogen has more branch points

than amylopectin, arranged in a uniform rather than discontinuous pattern, and it does not form secondary structures (Zeeman *et al.*, 2010). Glycogen is soluble because this branched structure prevents the higher-order crystalline organisation seen in amylopectin.

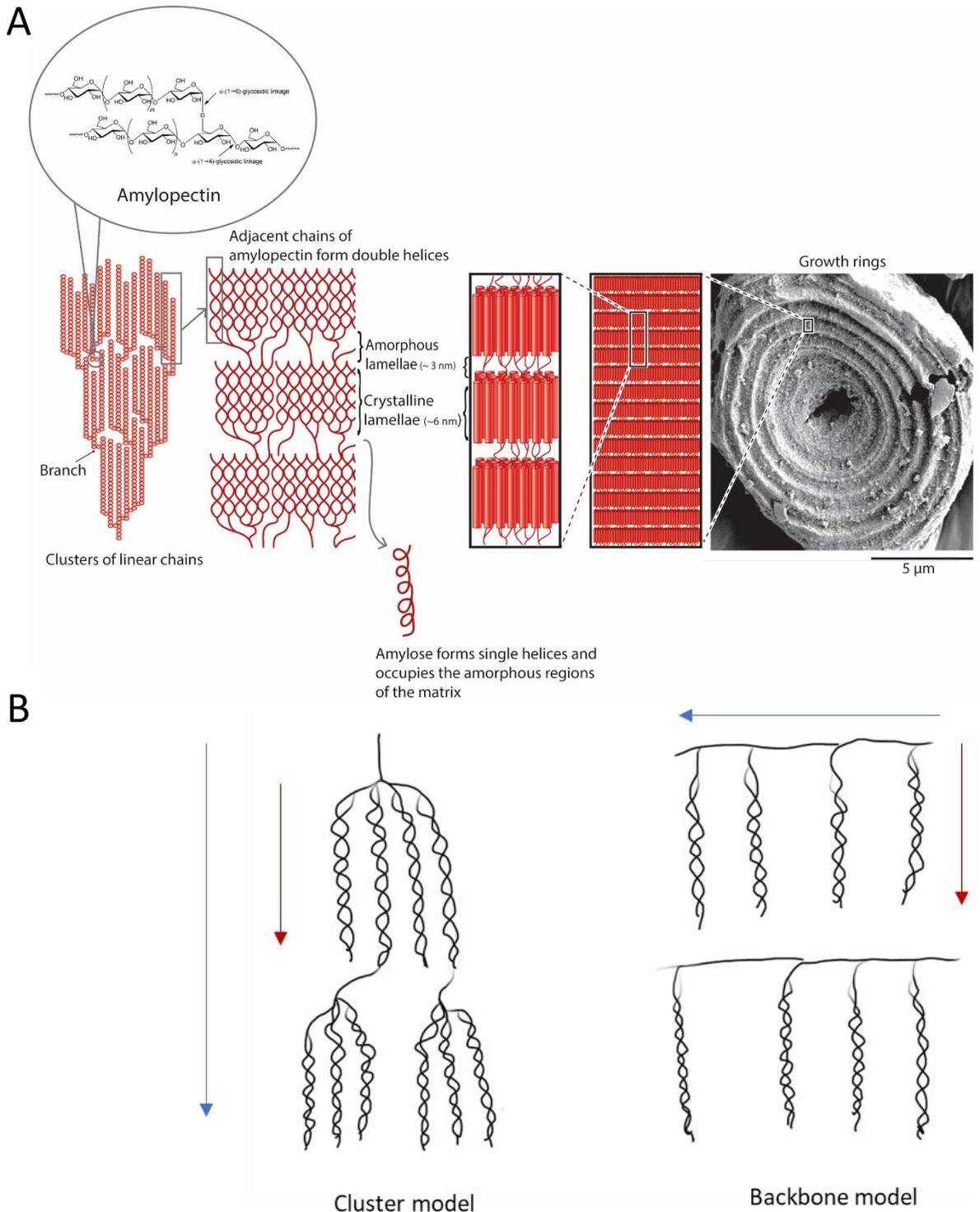


Figure 1.1 The hierarchal organisation of starch granule structure. A) An amylopectin molecule consists of clusters of linear glucan chains, connected by α -1,6-linked branches. Adjacent linear chains arrange into double helices, which pack together to form the crystalline lamellae. The α -1,6-linked branches sit in the amorphous lamellae that form between the crystalline lamellae. The amorphous regions are also where linear amylose chains are located, which can form single helices. The total repeat distance of a crystalline and amorphous unit is 9 – 10 nm, forming the semi-crystalline matrix of a starch granule. The last picture shows an inner face of a maize endosperm granule, etched to show the growth rings, with each ring spanning tens of the semi-crystalline repeats. A hilum can be seen in the middle. This figure is adapted from (Smith & Zeeman, 2020) with permission from the publisher, with elements from (Streb & Zeeman, 2012) and (Zeeman *et al*, 2010). **B)** Simplified diagrams of the amylopectin matrix models: the cluster model (Nakamura & Kainuma, 2022) and the building block backbone model (Tetlow & Bertoft, 2020). Blue arrows indicate the direction of growth of the long chains, and red arrows indicate the direction of growth of the short chains.

1.1.2 The diversity of starch granule morphologies

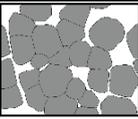
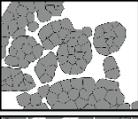
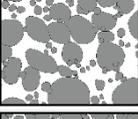
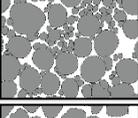
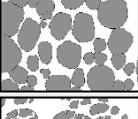
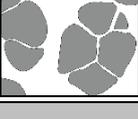
Transitory and storage starch granules tend to have different morphologies within the same plant. The best studied leaf starch granules are from *Arabidopsis*. At the end of the day, each chloroplast typically contains 5 – 7 granules in mature leaves, quantified using focused ion beam-scanning electron microscopy (FIB-SEM) (Crumpton-Taylor *et al*, 2012). However, this is variable and the number of granules per chloroplast increases with chloroplast size (Crumpton-Taylor *et al*, 2012). Another recent study using higher resolution microscopy techniques measured about 12 granules per chloroplast on average (Burgy *et al*, 2021). Starch granules form throughout the chloroplast and within pockets between the thylakoid membranes, rarely aggregated to one region (Burgy *et al*, 2021; Seung *et al*, 2018). Leaf starch granules tend to be similarly shaped across studied species – they are disc shaped and about 2 μm in diameter (Seung & Smith, 2019). Conversely, there is enormous diversity in granule morphology (size and shape) in the storage starch of different species (Jane *et al*, 1994). It is not known if there is a biological reason for this large diversity. Three general types of storage starches have been observed: compound, simple and bimodal (a version of simple) (Table 1.1).

Grasses are especially diverse in their granule morphology and there are grasses that have starch granules in each of these three categories. Rice is a well-studied compound granule forming species. These granules form from multiple granule initiations within one amyloplast, which produces polyhedral granulae that push against each other and fill up the amyloplast space (Matsushima *et al*, 2015). The granulae do not completely fuse together, which could be due to a septum-like structure separating the individual granules (Yun & Kawagoe, 2010), and these granulae separate during starch purification. Compound granules may be the ancestral granule morphology in the grass family, as they are present in early diverging species of the Poaceae (Matsushima *et al*, 2013). While they are a common morphology type in the endosperm of grass species, it is not certain how prevalent compound granules are in other storage tissues like tubers. A compound granule-like structure has been reported in sweet potato and eddoe tubers (Matsushima, 2015) (Table 1.1), but this has not been well studied.

By contrast to compound granules, simple granules form from a single initiation per amyloplast, such as in maize (Myers *et al*, 2011). In the Triticeae (including barley and wheat, see Table 1.1), the simple granules have a bimodal size distribution of one large, flattened A-type granule within one amyloplast that forms during early grain development and smaller, round B-type granules that form later in endosperm development in protrusions from the amyloplast called stromules (Howard *et al*, 2011; Langeveld *et al*, 2000; Parker, 1985).

Overall, the sizes of storage starches range from about 3 μm in diameter for barley/wheat B-type granules and rice granulae to up to 92 μm in diameter for some yams, but even within one plant tissue there is a huge variation in size (Table 1.1).

Table 1.1. Diversity in granule morphology and composition among major starch crops. This table was made together with Erica Hawkins and appears in (Chen *et al.*, 2021). *x, y, and z dimensions are shown where information is available. ? = no information available. References (Ref.): A) (Seung & Smith, 2019) B) (Jane *et al.*, 1999) C) (Bertoft, 2017) D) (Jane *et al.*, 1994) E) (Seung, 2020) F) (Matsushima *et al.*, 2010) G) (Saccomanno *et al.*, 2017) H) (Zhu, 2015) I) (Bull *et al.*, 2018) J) (Matsushima, 2015) K) (Li *et al.*, 2018)

	Granule type	Number per plastid	Diameter (μm)*	Shape		Apparent amylose content (%)	Ref.
Cereals							
Maize endosperm <i>Zea mays</i>	simple	single	5-20	polyhedral		20-29	A, B, C, D
Rice endosperm <i>Oryza sativa</i>	compound	multiple	granulae: 3-8 compound: 10-20	polyhedral granulae assembled into round compound granules		6-29	B, C, D, E, F
Barley endosperm <i>Hordeum vulgare</i>	simple bimodal	single A-type granule, multiple B-type granules	A type: x/y: 15-32, z: 6-10 B-type: 2-3	disc-shaped A-type granules; spherical B-type granules		22-27	B, C, D
Wheat endosperm <i>Triticum spp.</i> (e.g., <i>T. aestivum</i>)	simple bimodal	single A-type granule, multiple B-type granules	A type: x/y: 22-36, z: 6-10 B-type: 2-3	disc-shaped A-type granules; spherical B-type granules		17-34	B, C, D
Sorghum endosperm <i>Sorghum bicolor</i>	simple bimodal	?	large: 10-30 small: 5	irregular round and polyhedral		22-30	C, D
Oat endosperm <i>Avena sativa</i>	compound with small simple granules	multiple	compound: 22 simple: 5	polyhedral granulae assembled into compound granules and round simple granules		18-29	C, G
Tubers/Roots							
Potato tuber <i>Solanum tuberosum</i>	simple	single	x: 15-75 y: 12-60	round, ellipsoid		25-36	A, B, C, D
Yam tuber <i>Dioscorea spp.</i> (e.g., <i>D. alata</i> , <i>D. japonica</i>)	simple	single (<i>D. japonica</i>)	2-92 (<i>D. alata</i>)	round, ellipsoid, lenticular, polyhedral (<i>D. alata</i>)		18-41 (<i>D. alata</i>)	H
Cassava storage root <i>Manihot esculenta</i>	simple (possibly some compound)	?	5-25	ellipsoid with irregular edges		18-23.5	B, D, E, I,
Sweet potato storage root <i>Ipomoea batatas</i>	simple or compound	?	5-25	polyhedral		19-20	D, J
Other							
Pea seed embryo <i>Pisum sativum</i>	simple	?	x/y: 10-45, z: 10-27	mostly ellipsoid, some disc-shaped, with a central indentation		26-33	A, D, E
Banana/plantain fruit <i>Musa spp.</i>	simple bimodal		large: 5-50 small: 0.5-5	irregularly shaped discs, ellipsoid		21	D, K

1.2 Mechanism of starch synthesis and degradation

For the following sections, I will focus on the biochemical pathways of starch metabolism in *Arabidopsis* leaves, as the molecular mechanisms and genes responsible have been best studied in this species.

1.2.1 Starch synthesis requires starch synthases (SS), starch branching enzymes (SBE) and debranching enzymes (DBE)

The pathway of starch synthesis in the chloroplast starts from the Calvin cycle intermediate fructose-6-phosphate (F6P). This is converted to glucose-6-phosphate (G6P) in a reversible reaction by phosphoglucose isomerase (PGI). Phosphoglucomutase (PGM) then converts G6P to glucose-1-phosphate (G1P), also in a reversible reaction. Finally, ADP-glucose pyrophosphorylase (AGPase) converts G1P and adenosine triphosphate (ATP) to ADP-glucose (ADPG) and inorganic pyrophosphate (PPi). This last step is considered the first committed step of starch synthesis, as this reaction is made irreversible by the hydrolysis of PPi to orthophosphate (Pi). Figure 1.2 provides an overview of the starch synthesis process (blue arrows).

AGPase activity is allosterically regulated and is important for determining the flux of carbon into starch in *Arabidopsis* in saturating light, depending on the energy demands of the plant (Smith & Zeeman, 2020). An alternative route for ADP-glucose synthesis exists in the cytosol by sucrose synthase, but mutants lacking all sucrose synthases maintain wild-type levels of ADPG and starch, demonstrating that they do not contribute to starch synthesis in plastids (Funfgeld *et al*, 2022). ADPG forms the substrate for starch synthases (SSs), which add a glucosyl unit from ADPG onto a growing glucan chain on amylopectin/amylose, or onto short maltooligosaccharides (MOSs) (SSs will be described in detail in section 1.3).

Amylopectin formation also requires the action of multiple isoforms of starch branching enzymes (SBEs) and debranching enzymes (DBEs, isoamylases ISA1 and ISA2 in *Arabidopsis*). Branching enzymes cut segments of 6 or more glucose residues from existing chains and attach them onto another chain (or the same chain) by creating new α -1,6-glycosidic linkages. Debranching enzymes remove misplaced branches to ensure correct structures for crystallisation and granule formation. The importance of this process is demonstrated by the accumulation of a soluble glucan called phytoglycogen in *Arabidopsis isa1* and *isa2* mutants, since the aberrant polymer branching pattern does not allow assembly into insoluble granules (Delatte *et al*, 2005). Expression of *Arabidopsis* starch synthesis proteins in a heterologous yeast system also

demonstrated the importance of ISA proteins, as insoluble glucans could only be made in the presence of ISA1/ISA2 (Pfister *et al*, 2016). Additionally, two starch-bound non-enzymatic proteins, EARLY STARVATION 1 (ESV1) and LIKE ESV (LESV), are proposed to act in the organisation of the amylopectin matrix. This indirectly affects the accessibility of starch granules to both synthesis and degradation enzymes and influences granule morphology (Feike *et al*, 2016; Singh *et al*, 2022).

Arabidopsis SSs, SBEs and DBEs (ISAs) together with bacterial AGPase were sufficient to assemble an insoluble starch-like structure when expressed in yeast cells (Pfister *et al.*, 2016). The yeast system also showed that specific SS/SBE ratios were important for fine-tuning glucan structure, with a higher SS/SBE ratio producing longer, more insoluble glucans, and a lower ratio producing more soluble glycogen-like structures (Pfister *et al*, 2022). These studies demonstrate the core role of these proteins in starch synthesis, but finetuning of protein stoichiometries and additional non-enzymatic proteins are likely needed to produce starch granules that better resemble those found in plants.

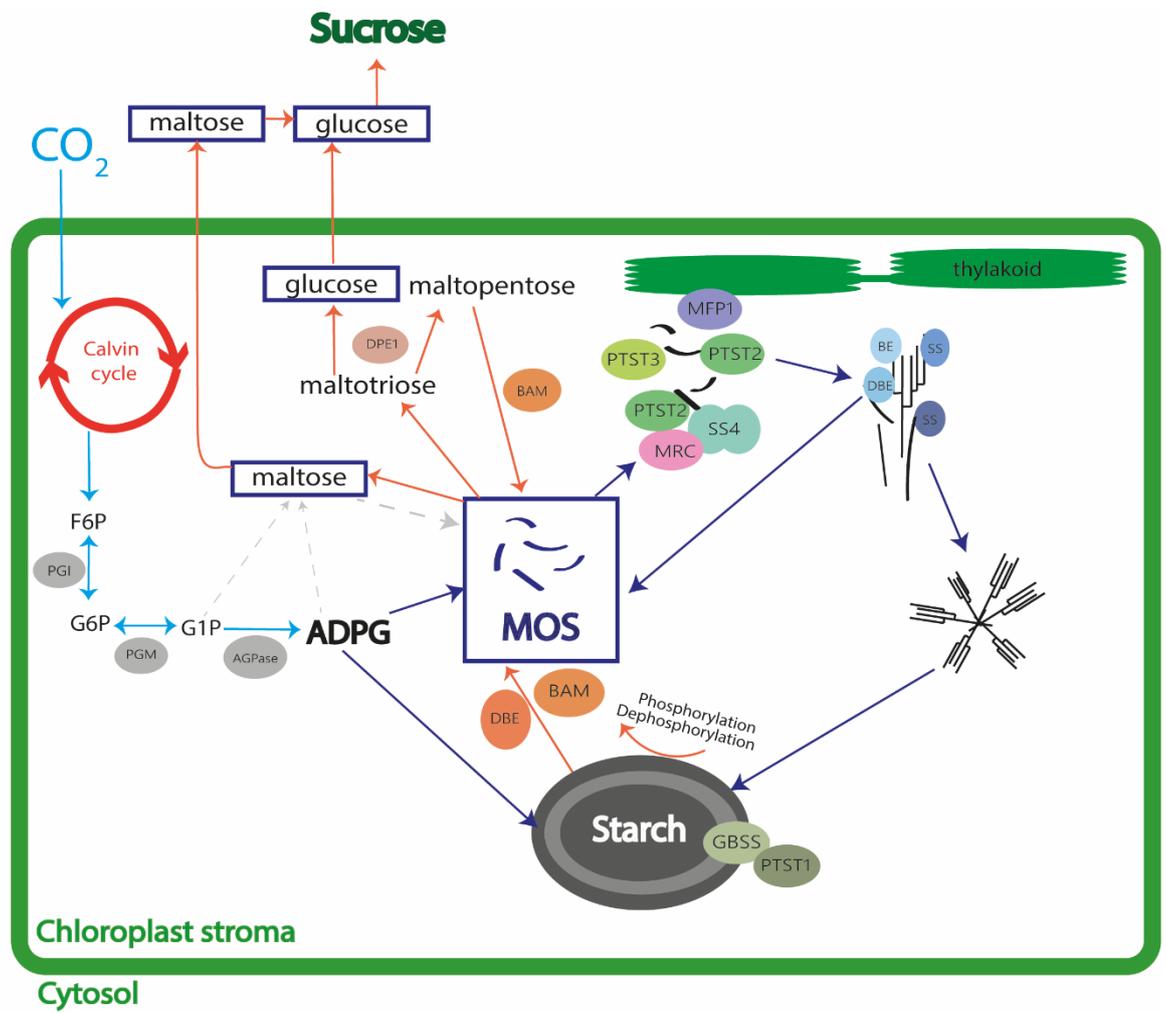


Figure 1.2. Overview of starch synthesis and degradation in a leaf chloroplast. *Synthesis, in blue arrows:* Carbon is fixed from CO_2 in the Calvin cycle. The Calvin cycle intermediate fructose-6-phosphate (F6P) is converted to glucose-6-phosphate (G6P) by phosphoglucose isomerase (PGI). G6P is converted to glucose-1-phosphate (G1P) by phosphoglucomutase. G1P is converted to ADP-glucose (ADPG) by ADPG pyrophosphorylase (AGPase). ADPG acts as a glucosyl donor for synthesis of short malto-oligosaccharides (MOS) and of growing starch granules by starch synthases (SS) (not pictured in diagram at this step). The very first glucosyl-acceptor is unknown, and might be maltose, indicated by thick dashed grey lines. This requires *de novo* synthesis of maltose, with hypothesised pathways indicated by narrow dashed grey lines. Once the first MOSs have been made, they act as glucan acceptors and are primed for elongation by a hypothesised granule initial protein complex, consisting of SS4, MRC, PTST2, PTST3, and thylakoid-bound MFP1 (see main text). The elongated glucan chains are further acted on by branching enzymes (BEs), debranching enzymes (DBEs) and various starch synthases (SSs) to form amylopectin, which grows radially to form a starch granule. More MOSs are also generated by the trimming activity of DBEs. GBSS and PTST1 synthesise amylose on the starch granule. *Degradation, in orange arrows:* Phosphorylation and dephosphorylation on the crystalline granule surface makes amylopectin chains accessible to degradation by DBEs and β -amylases (BAMs). Long amylopectin chains are degraded to short MOSs, including maltose and maltotriose. Two maltotrioses are converted to glucose and maltopentose by disproportionating enzyme 1 (DPE1). Maltopentose is further degraded by BAMs, and glucose and maltose are exported to the cytosol, where they are converted to sucrose.

1.2.2 Starch granule initiation determines the number and morphology of starch granules in a plastid

Starch synthesis can be divided into three distinct processes (Figure 1.3): the primary *de novo* starch granule initiation, the secondary granule initiation from existing glucans (these two together form the process of starch granule initiation), and amylopectin growth (Malinova *et al*, 2018; Seung & Smith, 2019). The primary *de novo* starch granule initiation would mainly occur in new chloroplasts, where no pre-existing glucan substrates are present yet. The secondary granule initiation follows this, by glucan elongation when at least some glucans have been made within the plastid, from new glucan synthesis or as products of starch degradation. Conversely, after granules are initiated, amylopectin growth would be the primary process, by polymer biosynthesis on the surface of existing granules. The secondary granule initiation and amylopectin growth have previously been named the initiation and amplification processes, respectively (Nakamura, 2015). However, the subtle distinction between the primary and secondary granule initiation has not been clearly defined in most of the literature.

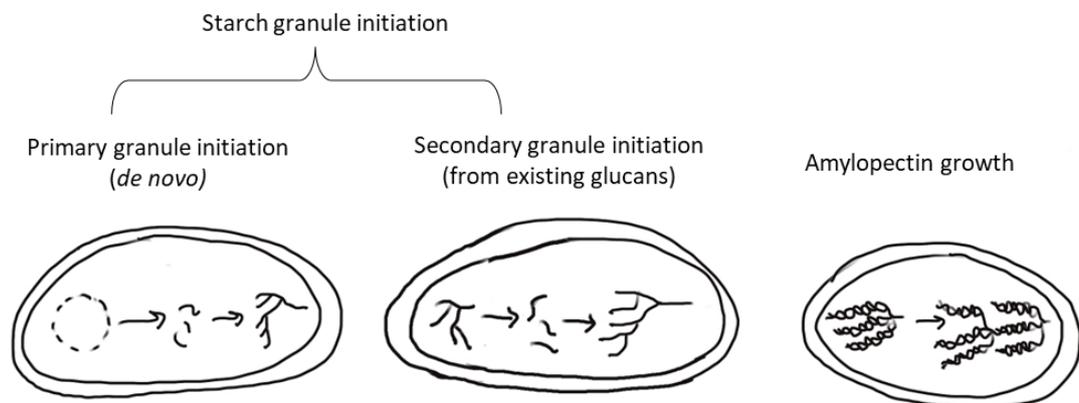


Figure 1.3 The three processes of starch synthesis. In the primary *de novo* granule initiation, no glucan substrates are present in a new chloroplast, and glucans are likely initiated through *de novo* maltose formation, but the exact mechanism is still unknown. In the secondary granule initiation, the formation of a new granule happens in the presence of existing glucans and/or starch granules, which provide substrates for elongation. In amylopectin growth, polymer biosynthesis happens on existing amylopectin chains or starch granules and mainly happens in mature plastids.

The mechanism of initial glucan crystallisation is still unknown, but some hypotheses have been proposed. Granule initiation could be formed by homogenous nucleation, where short glucan chains would be in constant motion and some parts would align in an organised state, eventually

forming a nucleation point (Crumpton-Taylor, 2010; Ziegler *et al.*, 2005). Another option is heterogenous nucleation, where something else would stabilise the glucans, such as proteins or the formation of more complex glucan structures (Crumpton-Taylor, 2010). In both these mechanisms, starch could be spontaneously crystallising out from liquid-liquid phase separated droplets (coacervates) formed by the glucan precursors in a high-density cellular environment (Crumpton-Taylor, 2010; Crumpton-Taylor *et al.*, 2012; Ziegler *et al.*, 2005). The rate at which this happens would depend on the protein composition and abundance in the cellular environment, which could be under genetic control (Crumpton-Taylor *et al.*, 2012).

Starch granule initiation has also been compared to spherulite formation (small, semicrystalline bodies within linear polymers), as heating and cooling of starch results in spherulites (Ziegler *et al.*, 2005). Spherulites are similar to starch granules in that they have a low-density region in the centre, resembling the less crystalline hilum core seen in storage starch granules, and they similarly display a Maltese cross when viewed under polarised light (Ziegler *et al.*, 2005). Although these experiments provide some insight into the physical properties of starch, they likely do not reflect the *de novo* formation of starch granules from short glucans, as spherulite formation results from reformation of crystalline structure from pre-formed amylopectin.

Starch granule initiation likely sets the correct number, size, and shape of granules, with other factors also influencing this process. In recent years, several proteins have been discovered that control the process of granule initiation, and they will be discussed in more detail further in this chapter (sections 1.4, 1.5). A single initiation point could result in the hilum in the amorphous core of one starch granule, and multiple initiations would result in compound granules, which have multiple hila. However, a recent study in *Arabidopsis* chloroplasts demonstrates that initiation events might not always be as straightforward as one initiation event per granule. Burgy *et al.* (2021) saw that numerous initiation events in young leaves form small granules that can coalesce into a single granule during chloroplast development. This could also partially explain previous observations that immature leaves contain more granules per chloroplast than mature leaves (Crumpton-Taylor *et al.*, 2012). However, it is not known if all granules form by coalescence, such as in other species or in amyloplasts. The mature *Arabidopsis* leaf starch granules are shaped like ovoid flattened discs due to the anisotropic growth of the granules during amylopectin elongation. It has been suggested that these non-random patterns of glucan deposition at the granule surface may be a general method of determining specific granule morphologies (Burgy *et al.*, 2021).

Environmental factors and physical constraints of the chloroplast could also be important for the patterns of granule initiation and for determining granule morphology. The number of starch granules per stromal volume is constant throughout a day, but environmental changes such as

light can change the volume of the granules and the absolute number of granules increases with chloroplast size (Crumpton-Taylor *et al.*, 2012). Granule initiation is coordinated with plastid division (Crumpton-Taylor *et al.*, 2013) and likely also with the plastid membranes, with starch granules distributed throughout the chloroplast (Seung *et al.*, 2018). The stromal pockets in which Arabidopsis starch granules form demonstrate the relevance of the thylakoid structure in granule initiation (Burgy *et al.*, 2021; Seung *et al.*, 2018). These pockets may also influence granule morphology, since Arabidopsis mutants defective in thylakoid organisation had disrupted granule morphology (Esch *et al.*, 2022). The mutants formed smaller granules than wild type, with an uneven surface.

1.2.3 Starch degradation requires phosphorylation and dephosphorylation

Figure 1.2 provides an overview of the starch degradation process in relation to starch synthesis, in orange arrows. Starch mobilisation at night first requires the phosphorylation of glucose residues at the crystalline granule surface, to provide amylolytic enzymes access to amylopectin chains for degradation. Phosphorylation is carried out by glucan, water dikinase (GWD, also called STARCH EXCESS1 (SEX1)) and phosphoglucan, water dikinase (PWD). This allows access for β -amylases (BAMs) which cleave α -1,4-linkages from the non-reducing ends of glucan chains to release maltose, and debranching enzymes specifically involved in starch degradation (ISA3 and limit dextrinase (LDA)) which cleave α -1,6-linkages. There is also one α -amylase in Arabidopsis chloroplasts, α -amylase 3 (AMY3), which cleaves internal α -1,4-linkages (Yu *et al.*, 2005) and can release both linear and branched glucans (Seung *et al.*, 2013). However, AMY3 is not necessary for starch degradation under regular conditions, as *amy3* mutants are not impaired in their starch degradation (Yu *et al.*, 2005).

Dephosphorylation of the glucans by the phosphatases SEX4 and LIKE SEX FOUR 2 (LSF2) is needed for complete degradation since the BAMs cannot cleave past the phosphates. Eventually, the actions of these enzymes produce maltose and maltotriose. Maltose can be directly exported from the chloroplast to the cytosol, where it can be converted to sucrose and utilised. Maltotriose is further processed by disproportionating enzyme 1 (DPE1), which converts two maltotrioses to glucose and maltopentose (which can be degraded further by BAMs). Glucose can also be exported to the cytosol for further metabolism.

In mature Arabidopsis leaves, the starch degradation is finetuned according to the daylength, so that starch is almost completely broken down by the end of the night, at a constant rate (Graf *et al.*, 2010). However, there will still be residual starch leftover at the end of the night so that during

a diurnal cycle, there is constant growth and shrinkage of starch granules, rather than the generation of completely new granules at the start of every day/night cycle (Burgy *et al.*, 2021; Crumpton-Taylor *et al.*, 2012). In *Arabidopsis*, degradation is not likely limited by the surface area, as the starch degradation rate is linear throughout the night, but rather by the granule number (Crumpton-Taylor *et al.*, 2012). The structure of the granules themselves is also important for a controlled rate of degradation, as the phytoglycogen degradation rate is exponential rather than linear (Delatte *et al.*, 2005).

Crucially, starch synthesis and degradation enzymes are all present together in the chloroplast throughout the diel cycle. Although there is a net synthesis of starch during the day, it is uncertain what the activity of degradation enzymes is at the same time. It was previously thought that starch degradation does not occur simultaneously with starch synthesis in *Arabidopsis* leaves (Zeeman *et al.*, 2002), but recent studies suggest that some starch degradation does in fact happen during the day, with starch becoming more susceptible to degradation with time in the light (Fernandez *et al.*, 2017). The rate of starch degradation is dependent on the time to dawn, and is under feedback inhibition of trehalose 6-phosphate (Ishihara *et al.*, 2022).

1.3 Starch synthases (SSs) have specific roles in polymer elongation during starch synthesis

Starch synthases (SSs) can add new glucosyl residues to the non-reducing ends of growing amylopectin and amylose, or on free MOSs that arise from degradation or *de novo* synthesis. In *Arabidopsis*, there are six isoforms of starch synthases. Their C-terminus contains the catalytic domain with glycosyltransferase activity, consisting of glycosyltransferase 5 (GT5) and glycosyltransferase 1 (GT) subdomains (Pfister & Zeeman, 2016). The C terminus is highly conserved among SSs and prokaryotic glycogen synthases, but the N-terminus of SSs is variable between isoforms (Pfister & Zeeman, 2016) (Figure 1.4A). The granule bound starch synthase (GBSS) synthesises amylose and is located within the starch granule, targeted there by PROTEIN TARGETING TO STARCH 1 (PTST1) (Seung *et al.*, 2015). There are five other soluble starch synthases in *Arabidopsis*— SS1, SS2, SS3, SS4, and SS5. SS1, SS2 and SS3 have known roles in amylopectin synthesis. SS4 is involved in starch granule initiation (see section 1.4). SS5 was only recently discovered, and its role is still unclear – it is not catalytically active but does play a role in starch granule initiation (Abt *et al.*, 2020). An SS6 isoform has been identified in potato but is not present in *Arabidopsis* (Helle *et al.*, 2018).

GBSS and SS1 – SS5 are evolutionarily conserved proteins (Abt *et al.*, 2020; Helle *et al.*, 2018). Not much research has been done on SS6 yet, but it does not seem to be as conserved as the other isoforms (Abt *et al.*, 2020; Helle *et al.*, 2018). In phylogenetic trees, these seven SSs can be divided into two groups based on their sequence similarity. Group A consists of GBSS, SS1, SS2 and group B contains SS3, SS4, SS5, SS6 (Abt *et al.*, 2020; Ball & Morell, 2003; Helle *et al.*, 2018; Leterrier *et al.*, 2008). SS1 – SS5 all have orthologs in both monocots and dicots. SS1, SS2, SS4, SS5 have orthologs in the bryophyte *Physcomitrella patens*, and SS1 – SS5 all have orthologs in the basal angiosperm *Amborella trichopoda* (Abt *et al.*, 2020).

Starch synthase isoforms have specific preferences for the length of glucan substrates, and these different activities are likely what make the branched starch structure so unique (Smith & Zeeman, 2020). These substrate preferences are demonstrated by mutant analyses in Arabidopsis and other crops, as well as *in vitro* for barley isoforms (Cuesta-Seijo *et al.*, 2016) and in a heterologous yeast system for Arabidopsis isoforms (Pfister *et al.*, 2016). In amylopectin synthesis, Arabidopsis SS1 preferentially produces short chains (degree of polymerisation (DP) = 6 – 10), SS2 produces intermediate chains (DP = 11- 25) and SS3 also produces intermediate chains (Pfister *et al.*, 2016). These functions all influence the chain length distribution of amylopectin, and thus the overall amylopectin structure. However, it should be noted that there are overlapping chain length preferences among starch synthase isoforms (Szydowski *et al.*, 2011). Defects in any of these SSs can also have knock-on effects on the activity of other enzymes such as other SSs and SBEs, by affecting their substrate availability.

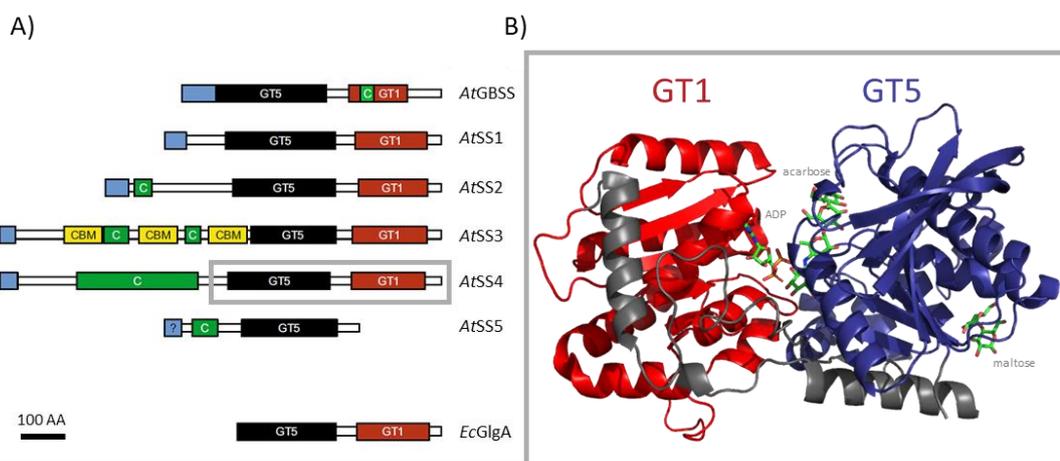


Figure 1.4. SS4 has a conserved catalytic domain. **A)** The predicted domain structure of the different starch synthase classes in Arabidopsis (*At*) and glycogen synthase from *E.coli* (*EcGlgA*). Light blue boxes are plastidial transit peptides, yellow boxes are carbohydrate binding modules of family 53 (CBM) (Gamez-Arjona & Merida, 2021), green boxes are coiled-coil domains (C), black boxes are glycosyltransferase 5 domains (GT5), red boxes are glycosyltransferase 1 domains (GT1). This figure is adapted from (Pfister & Zeeman, 2016) which is under a CC-BY license (see (Pfister & Zeeman, 2016) for details on domain prediction). The grey box around the GT domains in SS4 indicates the approximate region of the crystal structure on the right. **B)** Crystal structure of the Arabidopsis SS4 C-terminus (Nielsen *et al.*, 2018) which was modelled with ADP and acarbose in the catalytic site, and a maltose on a surface binding site. The GT1 domain is indicated in red and GT5 domain is indicated in blue, with linker domains in grey (Lu *et al.*, 2018). This image was made in PyMol with the structure with Protein Data Bank (PDB) code 6GNE.

1.4 SS4 is crucial for Arabidopsis starch granule initiation

1.4.1 SS4 is involved in granule initiation, influencing the number and morphology of starch granules

Starch synthase 4 (SS4) has been studied in detail in *Arabidopsis thaliana* and is thought to be especially important in granule initiation, creating precursors for the other starch synthesis enzymes to process. Arabidopsis SS4 was first characterised by (Roldan *et al.*, 2007). Knockout mutants (*ss4*) in both Columbia 0 (Col-0) and Wassilewskija ecotypes display the same granule initiation phenotypes, which influence the number and morphology of starch granules and plant growth. These different phenotypes could reflect distinct roles of SS4, or these roles could be connected.

1.4.1.1 SS4 ensures coordination of starch granule initiation in a regulated manner

First, *ss4* mutants have altered numbers of starch granules per chloroplast. In mature leaves, many *ss4* chloroplasts contain no granules, and the ones that do, form only one granule per chloroplast compared to the multiple granules per chloroplast in the wild type (WT) (Figure 1.5A) (Roldan *et al.*, 2007). In the youngest leaves of the rosette, *ss4* plants do not form any starch (Crumpton-Taylor *et al.*, 2013), whereas the WT has starch in all rosette leaves.

The second aspect of the *ss4* phenotype is that *ss4* mutants have smaller and paler rosettes compared to WT (Roldan *et al.*, 2007), being the only SS single mutant to have a growth phenotype. This is likely due to the accumulation of ADP-glucose (Crumpton-Taylor *et al.*, 2013; Ragel *et al.*, 2013). Together with the reduced starch granule number phenotype, this suggests that a lack of SS4 blocks the first step of the starch initiation pathway, with the other SSs unable to use ADP-glucose as a substrate without the product of SS4. The sequestering of adenylates into the ADP-glucose pool results in a severe imbalance of the adenylate distribution to ATP and ADP, leading to defects in photophosphorylation and resulting in oxidative stress, causing the distinct growth phenotype of *ss4*.

Thirdly, SS4 is also required for the anisotropic growth of Arabidopsis leaf starch granules (Burgy *et al.*, 2021). The shape of granules in *ss4* is spherical rather than lenticular, with new glucans deposited uniformly around the growing granule. The *ss4* granules are also larger than WT granules (Roldan *et al.*, 2007).

Crumpton-Taylor *et al.* (2013) proposed that the primary role of SS4 is in determining the granule number in young leaves, with the effects on granule morphology and ADP-glucose accumulation being secondary. This priming of granule formation is coordinated with the chloroplast volume. *Arabidopsis arc3* (*accumulation and replication of chloroplast 3*) mutants have a defect in chloroplast division and therefore have larger chloroplasts, but the number of granules per volume of stroma remains the same (Crumpton-Taylor *et al.*, 2012). The larger chloroplasts in *ss4arc3* mutants allow more frequent granule initiation than *ss4* mutants (Crumpton-Taylor *et al.*, 2013), but these mutants still have defects in the granule number and morphology. Without SS4, the precise timing and frequency of granule initiation is stochastic. The centre of *ss4* starches imaged by transmission electron microscopy (TEM) contained cavities, which could be abnormal hilum structures (Roldan *et al.*, 2007), supporting the role of SS4 in priming starch synthesis.

The *ss4* phenotype is distinct from *ss1*, *ss2* and *ss3*, as these other mutants do not have a change in granule number or morphology. Also, the starch of other *ss* mutants have changes in amylopectin structures, whereas *ss4* mutant starch does not have a significant difference in the amylopectin chain length distribution (Szydlowski *et al.*, 2009). This all indicates a specific role for SS4 in priming starch granule initiation, and thereby ensuring correct coordination of the number and size of granules. New initiations are still able to form in *ss4*, but it is much less than in WT and does not happen in clusters within the chloroplast as is typically the case in WT (Burgy *et al.*, 2021).

1.4.1.2 SS3 can initiate some granules in the absence of SS4

SS4 is not needed for starch granule initiation *per se*, but rather to ensure it happens in the correct way. The initiations that do still form are likely due to SS3 activity, as *ss3ss4* double mutants contain almost no starch (12% of WT levels with many empty chloroplasts) (Ragel *et al.*, 2013; Szydlowski *et al.*, 2009). SS3 likely also has a role in priming, but it is only partially able to initiate granules in the absence of SS4, as the initiations in *ss4* are low in frequency and stochastic. Single *ss3* mutants also do not have the defects seen in *ss4* (Zhang *et al.*, 2005).

Zymogram analysis suggested that it is possible that SS3 could even elongate glucans using ADP-glucose alone as a substrate (Szydlowski *et al.*, 2009) but *in vitro* assays showed none of the *Arabidopsis* starch synthases have autoglucosylation activity (Brust *et al.*, 2013). Also, barley SS3 did not display any activity when glucose and ADP-glucose were provided as substrates, arguing against the hypothesis that ADP-glucose alone can act as a substrate for SS3 (Cuesta-Seijo *et al.*, 2016).

In the *ss3ss4* double mutants, the ADP-glucose accumulation and growth phenotype are exacerbated, which further indicates that SS3 and SS4 are involved in the same pathway. When ADP-glucose production is eliminated by introducing a *pgm* mutation in the *ss3ss4* background, the growth and photosynthetic capacity is restored in *ss3ss4pgm* (Ragel *et al.*, 2013), confirming that the accumulation of ADP-glucose is related to the growth phenotype.

1.4.1.3 Are the different *ss4* phenotypes linked?

It is unclear if there is a connection between the role of SS4 in controlling granule number and morphology. Malinova *et al.* (Malinova *et al.*, 2017; Malinova *et al.*, 2018) suggested that the spherical granule shape of *ss4* might be connected to the reduced number, but this is not seen in other mutants. *mrc* and *ptst2* mutants (Figure 1.5, section 1.4.4.1) have reduced numbers of granules per chloroplast, but they are not spherical. Also, *ss4* plants expressing *Agrobacterium* glycogen synthase with self-priming activity (Lu *et al.*, 2018) and *ss4amy3* mutants (see section 1.4.2) (Seung *et al.*, 2016) have spherical granules but have more granules per chloroplast than *ss4*. The introduction of *Agrobacterium* glycogen synthase into *ss4* also lowered ADP-glucose levels but could not restore normal granule morphology. This partial complementation supports the hypothesis that the absence of SS4 specifically is responsible for the aberrant granule morphology, and this may be separate from the role of SS4 in controlling the number of initiations (Crumpton-Taylor *et al.*, 2013; Lu *et al.*, 2018). However, it cannot be ruled out that some of the starch granule number and morphology phenotypes in *ss4* arise from pleiotropic effects due to the altered redox state from ADP-glucose accumulation.

The exact connection between lack of SS4, a high ADP-glucose level, and fewer starch granules in *Arabidopsis* is still unclear. *ss4amy3* double mutants (Seung *et al.*, 2016), *ptst2* mutants (more in section 1.4.4.1) (Seung *et al.*, 2017) and *mrc* mutants (more in section 1.5) (Seung *et al.*, 2018) all accumulate ADP-glucose and have fewer granules per chloroplast than WT (Figure 1.5). Double mutants of *dpe2/phs1* (DPE2 – disproportionating enzyme 2, involved in conversion of maltose from starch degradation to sucrose, PHS1- plastidial alpha-glucan phosphorylase, possible roles in both starch synthesis and degradation) contain one granule per chloroplast and have only a moderate increase in ADP-glucose (Malinova *et al.*, 2017; Malinova *et al.*, 2018). Also, continuous light caused an increase in granule number per chloroplast in *dpe2/phs1* and *ss4* mutants, but no decrease in ADP-glucose levels (Malinova *et al.*, 2017).

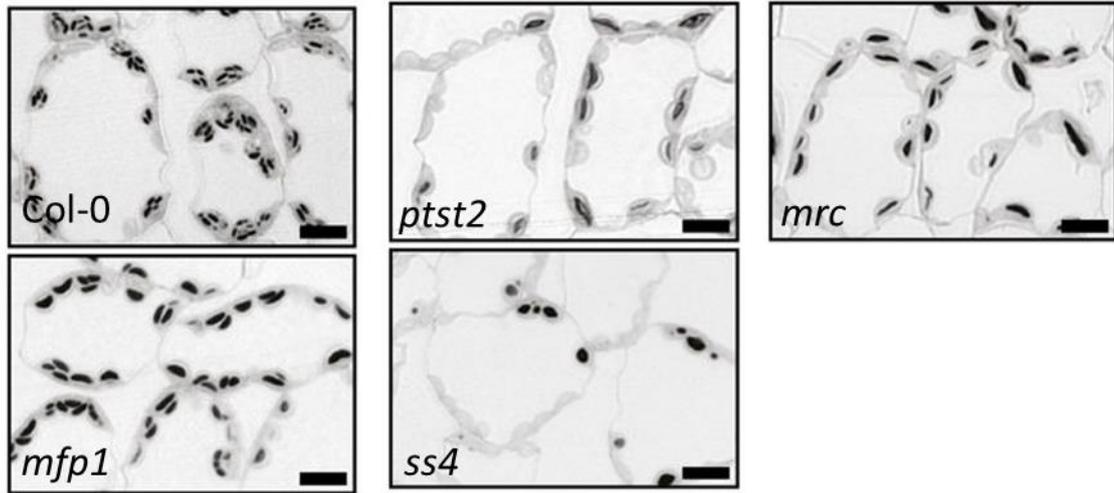


Figure 1.5. The granule initial proteins are important for determining starch granule number and morphology. Light microscopy images of Arabidopsis leaf sections from starch granule initiation mutants, showing the starch granules in mesophyll cell chloroplasts at the end of the day. The genotypes are Columbia 0 (Col-0) wild type, and mutants of *ptst2* (*protein targeting to starch 2*), *mrc* (*myosin resembling chloroplast protein*), *mfp1* (*mar-binding filament protein 1*), *ss4* (*starch synthase 4*). Bars = 10 μ m. This image is adapted from (Seung *et al.*, 2018), which is under a CC-BY license.

1.4.2 The different phenotypes in *ss4* depend on many environmental and metabolic factors

Starch granule initiation is sensitive to changes in the environment and disturbances at various steps in the pathway. Although *ss4* mutants have a severe granule initiation phenotype, this sensitivity also makes the *ss4* phenotype subject to change. For instance, the total amount of starch at the end of the illuminated period is variable in *ss4* mutants and may reflect sensitivity to environmental conditions. Although Roldan *et al.* (2007) measured reduced total starch at the end of the day in *ss4* compared to WT, Seung *et al.* (2018) did not measure any difference, suggesting that the severity of this phenotype, and perhaps total starch measured in general, is a variable trait due to the sensitivity of starch granule initiation to environmental conditions.

The total starch content at the end of the dark period does seem to be consistently increased in *ss4* compared to WT, and the rate of starch synthesis and degradation are both decreased in *ss4*, indicating a defect in starch turnover (Roldan *et al.*, 2007; Seung *et al.*, 2017; Seung *et al.*, 2018). Roldan *et al.* (2007) speculate that one reason for the reduced turnover could be a less accessible granule surface area for starch degradation enzymes, due to the large volume to surface area ratio in the single *ss4* granules. Other starch granule initiation mutants (*mrc*, see section 1.5) with a single large granule do not show such obvious defects in starch turnover, but the rounder starch granules in *ss4* could have an additional effect due to an even larger volume to surface area ratio. The accessibility of the polysaccharides on the granule surface influencing starch turnover is an

intriguing hypothesis, but is difficult to separate from other effects such as the total starch content, the specific effect of reduced numbers of granules, or altered metabolic state due to ADP-glucose accumulation.

The reduced turnover could also be influencing the slow growth in *ss4* (Roldan *et al.*, 2007), but these are probably minor effects in comparison with the growth defects due to ADP-glucose accumulation. Although Roldan *et al.* (2007) found *ss4* mutants grown in continuous light have the growth phenotype restored to near WT levels, Ragel *et al.* (2013) found that *ss4* and *ss3ss4* mutants in continuous light still did not grow as well as WT, indicating a lack of starch degradation cannot be the only reason for the reduced growth. In continuous light, little starch is made or used in the first place, so minimal resources are diverted to the starch synthesis pathway, and this would mean starch turnover is not really needed. Interestingly, when *ss4* plants are grown in continuous light, up to 3 granules per chloroplast could be formed (Malinova *et al.*, 2017) so these phenotypes must be considered in the context of environmental conditions and carbon flux.

The developmental stage and tissue also influence the *ss4* phenotype. As mentioned in section 1.4.1, SS4 mainly plays a role in young leaves, with the reduced starch granule phenotype in mature leaves being a knock-on effect from reduced initiations early in leaf development (Crumpton-Taylor *et al.*, 2012). During leaf expansion and chloroplast division, many new initiations are taking place, but in mature leaves this is rare. This effect is also seen in dividing cells in the Arabidopsis root cap, with *ss4* mutants generally containing less starch. However, there was much more variability in starch content in the roots compared to the leaves, perhaps due to different granule initiation mechanisms in the two organs (Crumpton-Taylor *et al.*, 2013).

Starch synthesis can also be influenced by degradation processes. Starch synthesis is not completely abolished in *ss4* mutant leaves but is reduced to almost no starch in *ss3ss4* mutants, indicating that at least SS3 or SS4 is needed to provide substrates for downstream SSs. However, when an alpha-amylase 3 (AMY3) knockout is introduced, the *ss3ss4amy3* and *ss4amy3* mutants accumulate small granules, indicating that SS3 and SS4 are important for generating a glucan product that evades premature degradation by AMY3 (Seung *et al.*, 2016), allowing further processing into a crystalline granule.

All these observations highlight the sensitivity of starch granule initiation to different environmental conditions, either to the external environment, or to the internal metabolic and developmental environment. Mutants with several knockout genes are also likely to have pleiotropic effects, and many double or triple mutants of starch synthesis genes display phenotypes that are difficult to explain when considering the single mutants (Abt *et al.*, 2020). All processes must be considered in the context of carbon flux, as well as interdependencies and

overlapping roles in protein functions (such as SS3 and SS4), and the interdependency of many metabolic pathways.

1.4.3 The role of SS4 in generating a pre-amylopectin primer and *de novo* primer hypotheses

The role of SS4 as the first starch synthase to act during granule initiation is intriguing, as it suggests that SS4 could be the main enzyme to use the glucan primers generated *de novo* in new chloroplasts, to make a pre-amylopectin primer (Figure 1.3, first panel: primary granule initiation). However, the exact *in vivo* substrate of SS4 is not known, and there are several possible sources. Once a plant has begun making starch, there will always be an MOS pool with short MOSs of different lengths which can act as glucan substrates for starch synthases, resulting from starch degradation and trimming during starch synthesis (Figure 1.3, second panel: secondary granule initiation). However, it is unknown how the synthesis of MOSs begins in a chloroplast that has not yet initiated any starch. In bacteria, yeast, and mammals, glycogen synthesis is initiated via a self-glucosylating protein, in mammals this is a specialised enzyme called glycogenin (Lomako *et al.*, 2004). As there is no evidence for self-glucosylation of SS4 (Szydlowski *et al.*, 2009) or any other protein in plants, one possibility is that maltose could act as the *de novo* primer.

Previous studies have suggested that maltose is made *de novo* from photosynthesis and not solely as a degradation product from starch breakdown. This maltose could act as a primer for starch granule initiation. In ¹⁴C radiolabelling experiments, maltose was labelled asymmetrically on its non-reducing end (Linden *et al.*, 1975), and it is labelled faster than sucrose (Szecowka *et al.*, 2013). ¹⁴C pulse-chase experiments in *Arabidopsis* have shown that maltose and short MOSs are labelled before long MOSs (Niittyta, 2003). The pulse-chase experiments showed that *pgm* mutants do not accumulate label in MOSs, indicating glucose-1-phosphate is needed for *de novo* maltose synthesis, which could happen through a phosphorylase. Maltose synthesis could also start downstream of PGM with ADP-glucose as a substrate, requiring a nucleotide sugar transferase (Figure 1.2). Intriguingly, plastidial alpha-glucan phosphorylase (PHS1) interacts with SS4 in *Arabidopsis* (Malinova *et al.*, 2018) (see below), and barley PHS1 was able to use G1P alone as a substrate (Cuesta-Seijo *et al.*, 2016). While it is tempting to think of a role for PHS1 in generating the *de novo* maltose for SS4, *phs1* single mutants do not have a severe starch synthesis phenotype (Zeeman *et al.*, 2004), suggesting it is unlikely to be primarily responsible for *de novo* maltose synthesis.

In support of the maltose hypothesis, Cuesta-Seijo *et al.* (2016) investigated the substrate specificity of different barley starch synthases, and the shortest MOS that any barley SS could use *in vitro* was maltose, although the activity was low compared to other MOSs. The same was seen for Arabidopsis SS4, which cannot use glucose but can use maltose as a substrate (Brust *et al.*, 2013; Nielsen *et al.*, 2018). The barley SS4 had very low activity with long polysaccharides (amylopectin, glycogen) as substrates, but high specific activity on MOSs (defined as up to DP=8), suggesting SS4 is mostly involved in MOS elongation rather than amylopectin elongation. Small amounts of SS4 would be enough to create a large pool of MOSs to act as substrates for SBE processing (Cuesta-Seijo *et al.*, 2016). This would be compatible with the low SS4 activity seen in Arabidopsis compared with other SSs (Roldan *et al.*, 2007; Szydlowski *et al.*, 2009). One could imagine that not a lot of SS4 activity is needed if it is efficient at creating pre-amylopectin primers and is mainly involved at this initial step, after which these primers are processed by other enzymes.

Other starch synthases are able to elongate glucan chains of different lengths, including maltose, albeit at different levels of activity (Cuesta-Seijo *et al.*, 2016; Pfister *et al.*, 2016). Therefore, they would be able to elongate *de novo* maltose primers without SS4 to produce starch granules, but perhaps not in an efficient manner that ensures the correct timing of priming. Crumpton Taylor *et al.* (2013) also emphasize that maltose elongation cannot be the only function of SS4 for this reason and seeing as bacterial glycogen synthase can only partially complement *ss4*. SS4 might be needed to form the primer in the specific context of creating a nucleation centre, perhaps as part of a protein complex (Crumpton-Taylor *et al.*, 2013). Further investigation is needed to distinguish between starch synthase activity on the *de novo* primer directly from the Calvin cycle versus on longer SS4-primed glucan chains.

In summary, SS4 could have four connected roles: first, elongating glucan primers generated *de novo* (i.e. maltose), second, elongating longer glucan chains (products of maltose elongation and other MOS substrates resulting from glucan degradation), third, specifically coordinating the timing and location of starch granule initiation, and fourth, influencing granule morphology. SS4 could simultaneously perform these roles by concentrating glucan substrates within the nucleation centre created by a protein complex, enhancing the activities of enzymes within this complex (Burgy *et al.*, 2021; Crumpton-Taylor *et al.*, 2013; Seung & Smith, 2019; Szydlowski *et al.*, 2009).

1.4.4 SS4 domains are involved in protein interactions and localisation

The biochemistry of SS4 provides some clues into the specific roles of its protein domains. The SS4 C-terminal catalytic domain is highly conserved among SSs and prokaryotic glycogen synthases (Lu *et al.*, 2018). This conservation is confirmed by the crystal structure of the *AtSS4* catalytic domain, which shows a double Rossmann fold typical of GT-B group glycosyltransferases, with the active site in a cleft between the GT1 and GT5 domains (Nielsen *et al.*, 2018) (Figure 1.4B). Conversely, the N-terminus is unique to *AtSS4*, conserved among SS4 orthologs (Lu *et al.*, 2018) and predicted to contain coiled-coils (Letierrier *et al.*, 2008; Pfister & Zeeman, 2016) which often facilitate protein interactions (Mason & Arndt, 2004). The SS4 N-terminus is necessary for determining the anisotropic growth of *Arabidopsis* leaf starch granules and therefore their lenticular shape, and for SS4 localisation into distinct puncta in the chloroplast (Burgy *et al.*, 2021; Lu *et al.*, 2018).

1.4.4.1 Biochemical interactions of *AtSS4* with granule initiation proteins

AtSS4 has been found to interact with a number of proteins (summarised in Figure 1.6). *AtSS4* can form dimers *in vitro* and in bimolecular fluorescence complementation (BiFC) assays in *Nicotiana benthamiana* (Raynaud *et al.*, 2016). This was also seen in immunoprecipitation experiments (Seung *et al.*, 2017). *AtSS4* dimerization could play a role in its association into a higher-order structure, and this may allow MOS products to be directly passed on to other *AtSS4* proteins. *AtSS4* was also pulled down as an interactor of PROTEIN TARGETING TO STARCH 2 (*AtPTST2*) in an immunoprecipitation mass-spectrometry (IP-MS) experiment (Seung *et al.*, 2018). It has not been confirmed whether this interaction is direct or indirect, and it could be transient, as not many peptides were identified. However, it is worth noting that *AtPTST2* was pulled down with a truncated *AtSS4* protein without its coiled coil domain in a co-immunoprecipitation in *N. benthamiana*, where other *Arabidopsis* proteins were not present (Seung *et al.*, 2017).

AtPTST2 is also involved in granule initiation, as *Arabidopsis ptst2* mutants lacking the protein have reduced starch turnover, and zero or one large granule per chloroplast (Figure 1.4A). This is similar to *ss4*, but the granule shape remains lenticular rather than round. *AtPTST2* is hypothesised to bring long MOS substrates to SS4 (Seung *et al.*, 2017). In combination with the *AtSS4* dimers, the close association of long MOS substrates could promote crystallisation of glucan chains (Seung & Smith, 2019). A related protein in the same family, *AtPTST3*, was also identified in the *AtPTST2* IP-MS (Seung *et al.*, 2018). In co-immunoprecipitations in *N. benthamiana*, *AtPTST2* and *AtPTST3* interacted, but truncated *AtSS4* without its coiled coil only

interacted with *AtPTST2* (Seung *et al.*, 2017). *AtPTST3* function is still unclear: *ptst3* mutants have a less severe reduction in starch granule number compared to *ptst2*, but *ptst2ptst3* double mutants have a more extreme phenotype than either single mutant, indicating both are important for correct starch granule initiation.

Two more protein interactors were identified in the *AtPTST2* IP-MS (Seung *et al.*, 2018) – MAR-BINDING FILAMENT PROTEIN 1 (*AtMFP1*) and MYOSIN RESEMBLING CHLOROPLAST PROTEIN (*AtMRC*, also named PII1, PROTEIN INVOLVED IN STARCH INITIATION (Vandromme *et al.*, 2019), see section 1.5). Compared to *AtSS4*, these proteins might have a more stable association with *AtPTST2*, judging by the number of peptides identified in the IP-MS. *AtMFP1* is hypothesised to partially target *AtPTST2* to the stromal side of thylakoid membranes. It is not known what the function of thylakoid associated protein localisation would be, but Seung *et al.* (2018) hypothesised that *AtMFP1* limits the localisation of *AtPTST2* to distinct patches, directing the localisation of granule initiation. Interestingly, *AtSS4* has also been associated with the thylakoid membrane in its interaction with fibrillin proteins (Gamez-Arjona *et al.*, 2014; Raynaud *et al.*, 2016) (see next section).

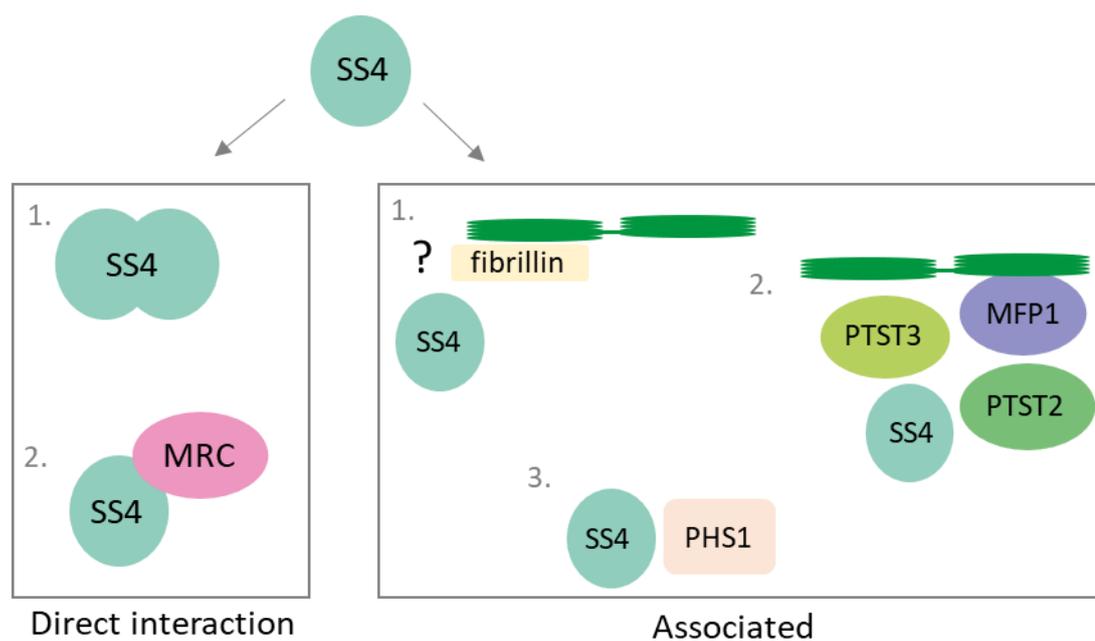


Figure. 1.6 Summary of known *AtSS4* interactions. The interaction partners are separated into confirmed direct interactions and associated proteins where direct interactions have not been determined.

1.4.4.2 Localisation of AtSS4 and associated granule initiation proteins

The punctate localisation of AtSS4 in the stroma has been observed several times (Lu *et al.*, 2018; Szydlowski *et al.*, 2009; Vandromme *et al.*, 2019) (Figure 1.7B). This distinct localisation is not likely directly on the surface of starch granules, as AtSS4 was found in the soluble fraction of leaf extracts (Szydlowski *et al.*, 2009), but AtSS4 may be associated with the area around the granules. Barley *HvSS4* also did not bind to any starches *in vitro* (Cuesta-Seijo *et al.*, 2016). This stromal localisation contrasts with the interaction of AtSS4 with fibrillin proteins and its fractionation with the thylakoid membrane rather than the stroma (Gamez-Arjona *et al.*, 2014; Raynaud *et al.*, 2016). However, AtSS4 did not co-fractionate with fibrillins in a Blue Native-PAGE assay from Arabidopsis leaf extracts (Lundquist *et al.*, 2017). It is possible that AtSS4 can associate with both the stroma and thylakoid membrane and change localisation in a dynamic process.

AtPTST2, AtMFP1, and AtMRC all display distinct localisation to puncta within chloroplasts, similar to AtSS4. However, AtPTST2 and AtMFP1 have more numerous puncta than AtSS4 and AtMRC (Figure 1.7). AtPTST2 and AtMFP1 colocalise (Seung *et al.*, 2018), but it is not known whether they also colocalise with AtSS4 and AtMRC, and how the more numerous puncta relate to the localising function of AtMFP1. Seung and Smith (2019) hypothesised that AtSS4, AtPTST2, AtPTST3, AtMFP1, and AtMRC form a 'granule initial' complex, a scaffold of proteins that concentrate starch granule initiation to specific locations in the chloroplast, to ensure correct priming of a crystalline granule. Nothing is known about the exact composition or stoichiometry of this potential complex, or the dynamics in which the complex could come together and dissociate. Perhaps the granule initial is such a large, concentrated protein complex that it would appear as spots in the chloroplast stroma, or it could be an inherent property of the individual proteins or interactions. Determining the mechanism by which these proteins are concentrated into distinct patches could shed light on how protein localisation influences granule initiation patterns. The granule initial proteins could also be tethered to additional structures inside the chloroplast, such as nucleoids. AtMFP1 has a DNA-binding domain (Jeong *et al.*, 2003) and could play a role in this association. It is worth noting that AtSS4 by itself has a tendency to promote crystalline glucan formation in a minimal yeast system without these other proteins of the granule initial, so these proteins are not strictly necessary for crystallisation itself (Pfister *et al.*, 2016), but could have roles in regulating the timing and location of granule initiation in Arabidopsis leaves.

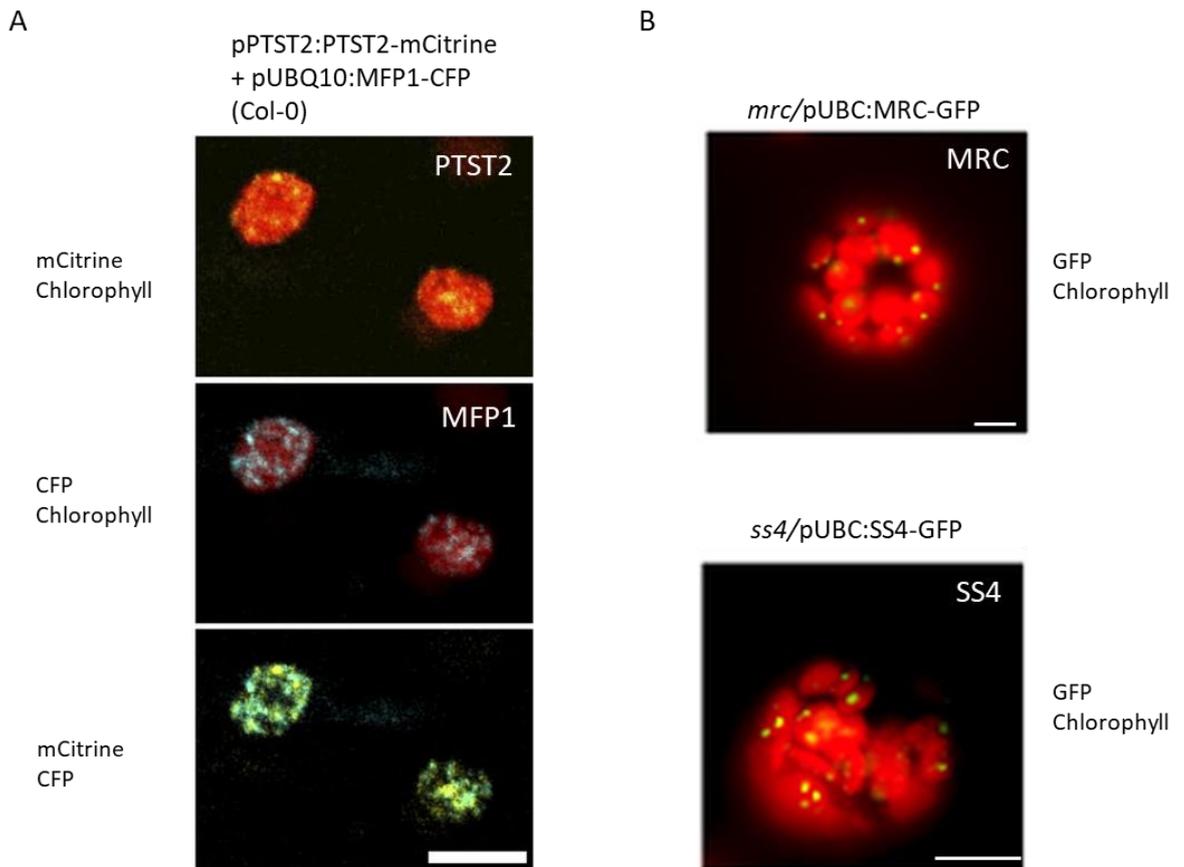


Figure 1.7. The granule initial proteins localise to distinct puncta in the chloroplast. A) Confocal microscopy images showing the punctate colocalisation of Arabidopsis PTST2 and MFP1 in a Col-0 background, in lower epidermal cells. Bar = 5 μ m. This image was adapted from (Seung *et al*, 2018), which is under a CC-BY license. **B)** Confocal microscopy images showing punctate localisation of MRC and SS4 in Arabidopsis protoplasts. Bar = 10 μ m. This image was adapted from (Vandromme *et al*, 2019), with permission from the publisher.

1.4.4.3 *AtSS4* interacts with *AtPHS1*

Finally, other than the associations in the granule initial and with fibrillins, *AtSS4* also interacts with *AtPHS1*, seen in an immunoblot of affinity-purified *AtPHS1* incubated with Arabidopsis protein crude extract (Malinova *et al.*, 2018). This enzyme catalyses a reversible reaction that either elongates glucan chains using glucose-1-phosphate as a substrate, or degrades glucan chains through phosphorolysis. Its exact role in starch metabolism is still uncertain (Zeeman *et al.*, 2010), but it could be involved in starch synthesis by elongating MOSs. In rice, PHS1 interacted with DPE1 in a complex, and this complex had enhanced synthesis of long MOSs *in vitro* compared to either protein alone (Hwang *et al.*, 2016). In Arabidopsis *ss4* mutants, PHS1 activity was upregulated. This was thought to be a measure against abiotic stress (Roldan *et al.*, 2007), as PHS1 is needed in Arabidopsis leaves for resilience against water deficit (Zeeman *et al.*, 2004). It is possible that PHS1 generates longer MOS substrates and its interaction with SS4 facilitates the usage of these substrates by SS4. However, there is no evidence for this yet, and *phs1* mutants do not differ from wild type in their starch content or number of granules per chloroplast (Malinova *et al.*, 2017; Zeeman *et al.*, 2004). It is not known if SS4 and PHS1 interact directly, and how the SS4-PHS1 interaction relates to other SS4 interactions.

In summary, SS4 has a specific role in granule initiation, consisting of its basic enzymatic role of a glycosyltransferase and its role in forming a pre-amylopectin primer, its protein interactions, and its specific punctate localisation, facilitated by its N-terminus.

1.5 MRC is a direct interactor of SS4, important for starch granule initiation

Arabidopsis MRC was independently discovered in two different studies – firstly in the AtPTST2 IP-MS mentioned above (Seung *et al.*, 2018), and also in a yeast two-hybrid (Y2H) screen as an interactor of AtSS4 (Vandromme *et al.*, 2019). The fact that it was identified in Y2H makes it the only known direct interactor of AtSS4 among the granule initiation proteins, together with AtSS4 itself. The function of AtMRC is unknown - it does not have any predicted enzymatic domains, and it has a predicted coiled-coil region spanning most of the protein (Seung *et al.*, 2018). Knockout mutants also have one large granule per chloroplast instead of the regular 5 – 7, similar to *ss4* (Figure 1.5), but the phenotype is not as severe as *ss4*. The *mrc* mutants do not have a growth phenotype, in contrast to the small and pale rosettes of *ss4*, and the granule shape remains lenticular rather than spherical (Seung *et al.*, 2018). This resembles the *ptst2* phenotype but is less severe as *mrc* chloroplasts still contain a starch granule in most chloroplasts, whereas *ptst2* mutants have many empty chloroplasts (Figure 1.5).

In one study, the total amount of starch in *mrc* leaves was even a little increased compared to WT at the end of the day (Seung *et al.*, 2018), but this was not found in another study (Vandromme *et al.*, 2019). There is more total starch at the end of the night in *mrc* compared to WT, but the rate of starch turnover was not measured so it is uncertain if this reflects a difference in turnover rate. Overall, the starch granule phenotype indicates that MRC is important for starch initiation, in relation to the number of granules per chloroplast and the size of the granule. There was no defect in the storage starch of root caps in Arabidopsis *mrc* mutants (Vandromme *et al.*, 2019). No *ss4mrc* double mutant has been studied in detail yet, and it would be interesting to see whether the granule phenotype in these double mutants resembles *ss4* mutants.

The biochemical mechanism whereby AtSS4 and AtMRC act together is not yet understood. Vandromme *et al.* (2019) proposed that AtMRC could aid in providing adequate substrates to AtSS4 or could be required for correct folding of AtSS4 or its interaction with other proteins. It is also unclear how the AtMRC-AtSS4 interaction relates to the AtSS4-AtPTST2 association and the AtPTST2-AtMFP1 association on the thylakoid membrane. The more numerous puncta of AtPTST2 and AtMFP1 compared to AtMRC and AtSS4 suggest a dynamic complex, with these four proteins associating in different ways. In a protein extraction from Arabidopsis leaves, Seung *et al.* (2018) found AtMRC in the stromal fraction, AtMFP1 in the thylakoid fraction, and AtPTST2 in both. In an *mfp1* mutant background, AtPTST2 was exclusively stromal. The question remains how the interactions between these proteins fit into the larger context of the granule initial. If AtPTST2 is directly delivering substrates to AtSS4, it could be interacting with AtSS4 alone, or with an AtSS4-AtMRC complex.

Other than *AtSS4*, the only other known interactor of *AtMRC* in Arabidopsis is the recently discovered *AtSS5* (Abt *et al.*, 2020). *AtSS5* is closely related to *AtSS4* and has predicted coiled coils at its N-terminus, but is not catalytically active. It does however influence starch granule number and size, with *ss5* mutants having a similar phenotype to *mrc* but less severe, with large lenticular granules but often more than the one granule per chloroplast as seen in *mrc* (Abt *et al.*, 2020). *mrcss5* mutants had a similar phenotype to *mrc* single mutants, demonstrating that *AtMRC* is epistatic over *AtSS5* (Abt *et al.*, 2020). *AtSS4* was not retrieved as an *AtSS5* interactor in IP-MS pull-downs, so Abt *et al.* (2020) hypothesised that *AtMRC* may interact with either *AtSS4* or *AtSS5*, at different times. The function of *AtSS5* and its relationship to the granule initial complex is unknown, but *AtSS5* might function in aspects of granule initiation that do not require *AtSS4*.

1.6 Protein complexes as a general feature in starch formation

In support of the granule initial hypothesis, Arabidopsis SS4 has been identified as part of large protein complexes in the chloroplast. In the Plant Proteome database, *AtSS4* (mature monomer: 113 kDa) was found in a >600 kDa fraction from size exclusion chromatography (SEC) separation of chloroplast stroma extract (Olinares *et al.*, 2010). Blue Native PAGE fractionation of chloroplast proteins identified *AtSS4* in the largest molecular weight fraction, above the 440 kDa marker (Lundquist *et al.*, 2017).

Protein complex formation could be a general feature in starch synthesis – protein complexes involving SBEs in wheat (Tetlow *et al.*, 2008; Tetlow *et al.*, 2004), maize (Hennen-Bierwagen *et al.*, 2008; Liu *et al.*, 2012), rice (Crofts *et al.*, 2015; Ida *et al.*, 2022) and barley (Ahmed *et al.*, 2015) demonstrate ways in which biosynthetic complexes can form during endosperm starch synthesis. Many rice endosperm starch synthesis enzymes co-eluted at high molecular weights of >700kDa, including rice SS4 which associated with a BE and other SSs (Crofts *et al.*, 2015; Ida *et al.*, 2022). Several of these complexes in cereal endosperms consist of different isoforms of SSs and BEs, such as the maize SS1-SS2a-SBE2b complex which is bound to starch granules during granule growth (Liu *et al.*, 2012), or the wheat SS1-SS2a-SBE2a/SBE2b complex (Tetlow *et al.*, 2008). There are likely many different starch biosynthesis protein complexes, with cases where the same protein is involved in different complex configurations (Crofts *et al.*, 2015; Ida *et al.*, 2022). The heterogeneity in protein complexes could be the case for the granule initiation mechanism as well.

Interestingly, SBE complexes found in wheat amyloplasts were not found in wheat chloroplasts (Tetlow *et al.*, 2004), suggesting different mechanisms for transient and storage starch synthesis.

When certain isoforms of SSs or SBEs are missing, other isoforms might also compensate by forming alternative complexes, such as maize *ZmSBE1* and substituting *ZmSBE2* complexes (Ahmed *et al.*, 2015) and rice *OsSS4* eluting at different high molecular weight fractions in wild type versus *ss2sbe2* mutants (Ida *et al.*, 2022). In addition to SS/SBE complexes, the rice *OsPHS1-OsDPE1* complex mentioned above (section 1.4.4.2) also influences substrate processing in starch synthesis (Hwang *et al.*, 2016). The close association of starch synthases and starch branching enzymes may help the coordination of different steps of starch granule formation. Many of these complexes are also regulated by phosphorylation (Tetlow *et al.*, 2008) (reviewed in (Crofts *et al.*, 2017)).

It is possible that Arabidopsis *SS4* is also under post-translational control, although this has not been investigated. Detection of the protein using an *AtSS4*-specific antibody (Roldan *et al.*, 2007) revealed a double band by immunoblotting, and it is still unclear why this double band appears. This band could be due to post-translational modifications. Phosphorylation prediction using the PhosPhAt database revealed two adjacent serines (positions 740, 741) as a predicted phosphosite, but with a low Mascot score of 23. The ATHENA database (Mergner *et al.*, 2020) identified a different phosphorylation site at serine 82, which would be interesting to investigate. The extra band on the immunoblot is unlikely to result from alternative splicing, as no splice variants are predicted from the *AtSS4* sequence, and proteins expressed from transgenes without introns (i.e: from *SS4* complementary DNA (cDNA)) still show the double band pattern.

1.7 Starch granule initiation in crops

Although most of our knowledge on the molecular mechanism of starch granule initiation comes from *Arabidopsis*, the pioneering work on starch synthesis was done in storage starches of crops. Now, we can combine this knowledge and start to compare the functions of recently discovered proteins from *Arabidopsis* with their orthologs in crops (Table 1.2). *Arabidopsis* is a useful model system to study SS4 in leaves and root caps (Crumpton-Taylor *et al.*, 2013), but it does not have starch in mature seeds, and is therefore not an ideal system to study both transitory and storage starch. Wheat is a suitable model to study starch in both chloroplasts and amyloplasts, and some work has been done on SS4 and PTST2 orthologs in wheat.

1.7.1 Granule initiation orthologs in crops

Orthologs of starch synthesis proteins have been studied in the starch crops wheat, barley, rice, potato, sweet potato and cassava (Table 1.2). From the granule initiation proteins, only SS4 and PTST2 have been studied in detail in various crops. The phenotypes of the mutants vary depending on the species, likely due to their different storage starch morphologies (Table 1.1, 1.2).

1.7.2.1 SS4 orthologs

Other than in *Arabidopsis*, SS4 has been studied in rice, barley, potato and wheat. In rice, there are two isoforms of SS4. Mutants deficient in *OsSS4a* have growth phenotypes and reduced leaf starch (Jung *et al.*, 2018). Lack of *OsSS4b* has a minor effect on compound granule formation in the endosperm, but affects the shape of rice endosperm granules in combination with *OsSS3*, with *ss3ss4* double mutants having round rather than polygonal starch granules (Toyosawa *et al.*, 2016). A double mutant of both *OsSS4* isoforms in rice still needs to be investigated. In barley, *HvSS4* function has been studied *in vitro*, as mentioned above, but its function has not been studied *in vivo* using mutants (Cuesta-Seijo *et al.*, 2016). *AtSS4* overexpression in potato tubers increased their starch content (Gamez-Arjona *et al.*, 2011), but the morphology of the granules was not investigated. This overexpression was also accompanied by pleiotropic changes in metabolism, with increased AGPase and sucrose synthase activity, and there were fewer but

larger tubers produced in the *AtSS4* overexpression line compared to WT (Gamez-Arjona *et al.*, 2011). The function of SS4 in potato is still unclear.

Wheat has one isoform of SS4 and it is preferentially expressed in leaves and embryo, with lower expression in the endosperm (Leterrier *et al.*, 2008). Within the endosperm, it is expressed higher at early stages of grain development (approx. 10 days post anthesis) than later stages (Hawkins *et al.*, 2021). In hexaploid wheat, loss of the D-genome homeolog of *TaSS4* results in fewer granules per chloroplast in the leaf (Guo *et al.*, 2017) and in tetraploid durum wheat, full *ss4* knockouts have fewer granules per chloroplast and reduced starch content in the leaves (Hawkins *et al.*, 2021), suggesting a similar role as in *Arabidopsis* leaves. However, endosperm starch from the durum knockout mutants mostly had an aberrant polyhedral granule shape, compared to the flattened, round shape of A-type granules in wild type, and these polyhedral granules sometimes also formed compound granules. These mutants had an irregular size distribution of granules, but no difference in the starch content of the grains (Hawkins *et al.*, 2021). This points to potential differences in granule initiation mechanisms in wheat leaves and endosperm.

1.7.2.2 *PTST2* orthologs

PTST2 has been studied in rice, barley, wheat and *Brachypodium distachyon*. In rice and barley, it is called FLOURY ENDOSPERM 6 (*FLO6*), and in wheat it is called B-GRANULE CONTENT 1 (*BGC1*). In the leaves, loss of *PTST2* consistently resulted in fewer granules per chloroplast in rice, barley, tetraploid wheat and *Brachypodium*, but to varying degrees (Watson-Lazowski *et al.*, 2022; Zhang *et al.*, 2022). In rice, wheat and *Brachypodium*, the granules were larger in the mutant compared to wild-type leaves, but not in barley (Watson-Lazowski *et al.*, 2022; Zhang *et al.*, 2022). In barley, the total starch content was lower than wild type, but in wheat this stayed the same. In *Brachypodium*, there was almost no starch in the leaves (Watson-Lazowski *et al.*, 2022). Curiously, in rice there was increased total starch in the mutant leaves compared to wild type (Zhang *et al.*, 2022).

In the endosperm, barley and wheat both normally have a bimodal starch granule size distribution, and the *Hvflo6* and *Tabgc1* knockout mutants (in both tetraploid and hexaploid wheat) produced irregularly shaped, compound-like granules instead (Chia *et al.*, 2020; Saito *et al.*, 2017; Suh *et al.*, 2004). Loss of two out of three homeologs of *BGC1* in hexaploid wheat specifically abolished the production of B-type granules, with the A-type granules normally shaped, suggesting *BGC1* gene dosage is also important in controlling granule initiation (Chia *et al.*, 2020). In *Brachypodium*, a species that forms simple granules in the endosperm, loss of *PTST2* also resulted in compound granule formation in the endosperm (Watson-Lazowski *et al.*, 2022). In rice,

flo6 mutants formed loosely packed small granules of irregular shapes instead of compound granules (Peng *et al.*, 2014; Zhang *et al.*, 2022).

These results demonstrate the roles of SS4 and PTST2 in starch granule initiation in several crop species. The presence of compound granules rather than simple and bimodal granules in *Brachypodium*, barley and wheat in *ptst2* and *ss4* mutants suggests an increase in the number of initiations per amyloplast. Therefore, PTST2 and SS4 might have a role in restricting granule initiation in the endosperm amyloplast, in contrast to their role in promoting granule initiation in leaves. There could also be both tissue-specific and species-specific effects of granule initiation proteins. Rice *OsFLO6* can interact with *OsSS4b* and *OsGBSS* in yeast two-hybrid and BiFC (Zhang *et al.*, 2022). Rice *OsFLO6* also physically interacts with *OsISA1*, but its functional relationship to *OsISA1* is not clear (Peng *et al.*, 2014). In contrast, Arabidopsis *AtPTST2* and *AtISA1* do not interact in the leaves (Seung *et al.*, 2017). Also, rice *OsFLO6* localises diffusely around the starch granules in rice leaves and endosperm (Zhang *et al.*, 2022), as opposed to the punctate localisation seen in Arabidopsis chloroplasts (Seung *et al.*, 2017).

1.7.2.3 Further discovery of determinants of granule morphology

The effect of gene knockouts must be studied in multiple tissues of the same species and the same tissue in multiple species, to understand both species-/tissue-specific granule initiation mechanisms as well as conserved and universal features. In addition to interspecies variation, we are only beginning to understand the possible variation in starch morphology within accessions/varieties of one species. Natural variation among accessions is a valuable resource for discovering novel factors influencing granule initiation, as seen through the identification of BGC1 by screening the ancestral grass *Aegilops* with varying B-type granule contents (Chia *et al.*, 2020; Howard *et al.*, 2011).

In addition to the reverse genetics approaches to look at Arabidopsis orthologs in crops, forward genetics screens in rice for defective endosperm and starch granule phenotypes have identified additional genes controlling amyloplast and compound granule development (Matsushima *et al.*, 2010; Matsushima *et al.*, 2014; Matsushima *et al.*, 2016; Zhang *et al.*, 2016). In addition to FLO6, several other floury endosperm mutants have been investigated. For instance, FLO7 reduces starch granule size in chloroplasts and affects amyloplast development (Zhang *et al.*, 2016). The SUBSTANDARD STARCH GRAIN genes SSG4 (Matsushima *et al.*, 2014) and SSG6 (Matsushima *et al.*, 2016) in rice also affect granule morphology in the endosperm and pollen, with mutants that have larger compound granules in the endosperm and larger simple granules in pollen. Rice *OsFLO6* (Zhang *et al.*, 2022) and wheat *TaSS4* (Hawkins *et al.*, 2021) also affect pollen starch,

where reduced starch reduces fertility. Starch granule initiation has been less well studied in tissues other than leaves and endosperm, and it is possible that granule initiation proteins have distinct and specific roles in other organs like pollen or tubers.

1.7.2 Wheat as a model to study granule initiation in leaves and endosperm

Wheat is an ideal system to study leaf and endosperm starch in one system. Starch properties of wheat grains are an important agricultural trait, with 85% of the grain weight consisting of carbohydrates, and 80% of that is starch (Shewry & Hey, 2015). There is still a lack of information on the differences between starch formation in chloroplasts and amyloplasts in one species, with studies often focusing on storage starches in crops.

The hexaploid *Triticum aestivum* is used for bread and the tetraploid *Triticum turgidum* spp. *durum* for pasta. For both these species, many genetic resources have been developed in recent years, including annotated genomes (International Wheat Genome Sequencing, 2018; Maccaferri *et al*, 2019) and transcriptomes (Borrill *et al*, 2016; Maccaferri *et al.*, 2019; Ramirez-Gonzalez *et al*, 2018). New ways of working with mutagenesis in these polyploid crops have been developed in recent years, such as Targeting Induced Local Lesions In Genomes (TILLING) mutant lines (Krasileva *et al*, 2017). The high tolerance for mutations in polyploid species increases the amount of mutant alleles that can be discovered compared to diploid species like *Arabidopsis* (Uauy *et al*, 2017). Effects of gene dosage can also be studied through knocking out individual homeologs of a gene.

Wheat leaf starch consists of several similar sized, lenticular granules in chloroplasts. However, like the other Triticeae, the endosperm starch granules have a bimodal size distribution composed of a population of large, flattened A-type granules (20-30 μm in diameter) and small, spherical B-type granules (2-7 μm in diameter) (Howard *et al.*, 2011). One A-type granule per amyloplast forms from 4 days post anthesis (dpa), and several B-type granules initiate 10 – 15 days after the A-type granules (at around 15 – 20 dpa), at least partially within protrusions from the amyloplast called stromules (Bechtel, 1990; Howard *et al.*, 2011; Langeveld *et al.*, 2000; Parker, 1985). Not much is known about initiation of wheat starch in either the leaves or grain, but the mechanisms might be quite different from *Arabidopsis* as grasses tend to rely more on sucrose as a carbon source in the leaves (Müller *et al*, 2018; Nie, 1995; Watson-Lazowski *et al.*, 2022). The temporal control of granule initiation is also unique to these bimodal starch granule species, so the underlying mechanisms will likely differ compared to simple and compound granule species.

MRC has not been studied in any other species than Arabidopsis yet. It is now very timely to investigate the biochemical mechanisms of Arabidopsis granule initial proteins in parallel with elucidating the role of their orthologs in wheat. This will shed light on similarities and differences in the granule initiation process in chloroplasts and amyloplasts and open the door to mechanistic discoveries.

Table 1.2. Examples of gene targets used for altering starch structure and physicochemical properties in crops. This table was made together with Erica Hawkins and David Seung and appears in (Chen *et al.*, 2021) under a CC-BY license. This table lists selected examples of studies that examined starch structure and physicochemical properties in mutants of starch biosynthesis genes in crop species. Primary effects on starch structure (listed first for each study) were used to illustrate links in Figure 1.8, although it should be noted that elimination of some gene targets results in multiple structural changes. The relative contribution these changes to the observed alterations in physicochemical properties is not known, and can only be dissected with alternative strategies that allow each structural property to be manipulated in isolation. Studies with promising alterations in granule size for which physicochemical properties should be determined are also listed. ? = no information.

Starch structure	Gene target	Species	Structural Property change(s)	Associated Physicochemical changes	Reference
Granule morphology	<i>SS4</i> <i>STARCH SYNTHASE 4</i>	wheat (<i>SS4</i> mutation)	1. compound granules with small polygonal granulae in place of most A-type granules	?	(Hawkins <i>et al.</i> , 2021)
	<i>BGC1/ FLO6</i> <i>B-GRANULE CONTENT 1/ FLOURY ENDOSPERM 6</i>	wheat (<i>BGC1</i> mutation)	1. compound granules with small polygonal granulae in place of most A-type granules 2. reduced gene dosage results in fewer B-type granules	?	(Chia <i>et al.</i> , 2020)
		barley (<i>FLO6</i> mutation)	1. mixture of simple, compound and semi-compound granules of heterogenous size 2. fractured granules	1. broadened range of starch gelatinisation temperature 2. altered pasting profile 3. faster starch hydrolysis rate 4. improved grain pearling quality	(Saito <i>et al.</i> , 2017; Suh <i>et al.</i> , 2004)
		rice (<i>FLO6</i> mutation)	1. defective compound granule formation - smaller irregular granules.	1. starch requires higher urea concentrations to gelatinise 2. altered pasting profile with lower viscosity	(Peng <i>et al.</i> , 2014)
	<i>SS3a</i> <i>SS4b</i> <i>STARCH SYNTHASE 3a/4b</i>	rice (<i>OsSS3a</i> <i>OsSS4b</i> double mutation)	1. polyhedral to spherical granules	?	(Toyosawa <i>et al.</i> , 2016)

	SSG4 or SSG6 <i>SUBSTANDARD STARCH GRAIN 4/6</i>	rice (mutation)	1. increased compound granule size	?	(Matsushima <i>et al.</i> , 2014; Matsushima <i>et al.</i> , 2016)
Amylopectin structure	SS2 <i>STARCH SYNTHASE 2</i>	rice	1. increase in amylopectin short chains 2. increased amylose content	1. lower pasting temperature	(Miura <i>et al.</i> , 2018)
		sweet potato	1. increase in amylopectin short chains	1. lower pasting temperature	(Takahata <i>et al.</i> , 2010)
		potato (<i>SBEII over-expression</i>)	1. increase in amylopectin short chains, more branches	1. reduced gelatinisation temperature 2. reduced viscosity 3. increased swelling power	(Brummell <i>et al.</i> , 2015)
Amylose content	GBSS <i>GRANULE-BOUND STARCH SYNTHASE</i>	Rice (GBSS gene editing)	1. reduced amylose content	1. reduced viscosity	(Zeng <i>et al.</i> , 2020)
		cassava	1. reduced amylose content	1. lower pasting temperature 2. higher final viscosity	(Bull <i>et al.</i> , 2018)
		sweet potato	1. reduced amylose content 2. absence of long amylopectin chains	1. higher gelatinisation temperature 2. increased digestibility	(Kitahara <i>et al.</i> , 2007; Noda <i>et al.</i> , 2002)
	SBE <i>STARCH BRANCHING ENZYME</i>	rice (SBEIIb gene editing)	1. increased amylose content 2. increase in amylopectin long chains 3. altered granule morphology	1. decreased digestibility 2. increased gelatinisation temperature	(Sun <i>et al.</i> , 2017)
		maize (<i>ae mutation</i>)	1. increased amylose content 2. increase in amylopectin long chains 3. altered granule morphology	1. decreased digestibility 2. increased gelatinisation temperature	(Li <i>et al.</i> , 2008; Liu <i>et al.</i> , 2012)
		potato (<i>SBEI/SBEII suppression</i>)	1. increased amylose content 2. long amylopectin chains, fewer branches 3. altered granule morphology	1. decreased digestibility	(Zhao <i>et al.</i> , 2018)
		sweet potato (SBEII suppression)	1. increased amylose content 2. increased phosphate content 3. altered amylopectin structure	1. higher viscosity 2. decreased digestibility	(Kitahara <i>et al.</i> , 2007)

Phosphate Content	<i>GWD</i> <i>GLUCAN, WATER</i> <i>DIKINASE</i>	barley (<i>StGWD over-expression</i>)	1. increased phosphate content 2. altered granule morphology	1. reduced gelatinisation temperature	(Carciofi <i>et al</i> , 2011)
		cassava (<i>StGWD over-expression</i>)	1. increased phosphate content	1. higher swelling power 2. clearer paste	(Wang <i>et al</i> , 2018)
		potato (<i>GWD suppression</i>)	1. decrease in phosphate content 2. increased amylose content	1. increased gelatinisation enthalpy	(Kozlov <i>et al</i> , 2007)
	<i>SEX4/LSF2</i> <i>STARCH EXCESS</i> <i>4/LIKE SEX FOUR 2</i>	potato (<i>SEX4</i> or <i>LSF2</i> suppression)	1. increase in phosphate content 2. altered amylopectin structure 3. reduced granule size	1. increased viscosity 2. increased swelling power	(Samodien <i>et al</i> , 2018)

1.8 Biotechnology prospects for modifying starch granules

Starch is not only used as an important source of nutrition, but its physicochemical properties also make it a valuable resource for non-food applications such as adhesives or coating, or in biodegradable plastics. Chemical, physical and enzymatic modifications have been used to tailor starches for their specific end-uses (Kaur *et al*, 2012), but our increased understanding of the biology behind starch granule morphology provides an opportunity to use genetics to modify these traits in a more environmentally friendly manner. This was previously not possible partly due to the limited range of modifications that could be made with previously known genes, and limited genes that could be modified without detriment to yield or use of transgenics.

Variation in granule morphology, amylose and amylopectin structure, amylose content and phosphate content are all structural starch traits which can influence starch physicochemical properties. The four main physicochemical properties evaluated in starch applications are gelatinisation/pasting temperature, viscosity, swelling power and digestibility; which are to some degree inter-dependent (Figure 1.8; Table 1.2) (Lindeboom *et al*, 2004). Digestibility is also sometimes considered a downstream property resulting from the other physicochemical properties, and the most studied aspect of this is the less digestible resistant starch, which is starch that resists digestion in the lower intestine and can be used by the gut microbiome (Li *et al*, 2019). One form of resistant starch is high amylose starch that can result from lack of SBEs (Li *et al*, 2019). For other structural properties, there is less of a direct correlation towards certain physicochemical properties, as this is also heavily dependent on the species and genetic background in question. However, certain trends can be seen in studies where starch physicochemical properties have been analysed in mutants of species with the same genetic background, summarised in Figure 1.8, with the details listed in Table 1.2.

Starch granule size can greatly influence the physicochemical properties of starch, but various effects have been seen depending on the source of the starch and the specific context in which it is measured (Li *et al*, 2021). No consistent physicochemical properties inherent to A-type or B-type granules have been determined either (Saccomanno *et al*, 2022). Small granules like the B-type granules in the Triticeae are undesirable in brewing as they are less easily digested during the brewing process (Bathgate & Palmer, 1973) and can be difficult to remove. However, there are also reports of higher *in vitro* digestibility of small granules compared to large ones for potato and maize starches, due to their larger surface area to volume ratio (Dhital *et al*, 2010). Small granules are desirable in pasta making due to their higher water absorption, creating firmer pasta (Soh *et al*, 2006).

As a promising development in tailoring wheat starch size, wheat *bgc1* mutants lacking B-type granules were developed using the TILLING population (named BlesT) and their developmental properties and starch physicochemical properties were measured after being grown in the field and the glasshouse (Saccomanno *et al.*, 2022). The final effects of the lack of *bgc1* still remain to be determined due to background mutations in the BlesT lines, but initial results are promising since the BlesT grains were softer and had higher water absorption and swelling power compared to wild type. However, it is difficult to say whether this means that the swelling power of B-type granules is less than A-type granules, as the BlesT lines also have slightly larger A-type granules than the mutant, demonstrating again that the effect of starch structural properties on physicochemical properties depends on the context and needs to be evaluated on a case-by-case basis.

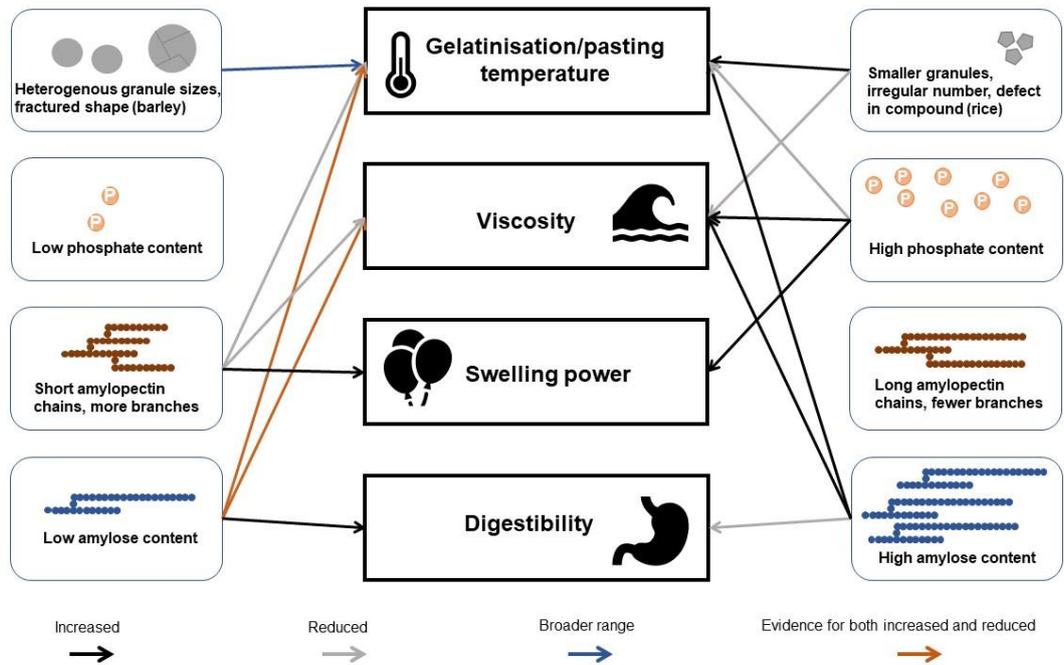


Figure.1.8 Influence of starch structure on starch physicochemical properties. Studies using mutants of various crop species have established potential associations between alterations in starch structure (blue boxes) and four main physicochemical properties (black boxes) (Table 1.2). Links between major structural alterations observed in the mutants and physicochemical properties are illustrated. Structural changes associated with an increase in measured physicochemical properties are shown with black arrows, while reductions are shown in grey. A change in some structural properties, such as a reduction in amylose content, may yield different physicochemical effects depending on species (orange arrows; Table 1.2). Heterogenous granule sizes have been observed to broaden the range of gelatinisation temperatures (blue arrow). This figure appears in (Chen *et al*, 2021) and is under a CC-BY license.

1.9 Aims of the thesis

It has been hypothesised that SS4 controls the seeding of starch granules, and thereby controls the timing of granule initiation by determining a centre of nucleation from where individual granules arise. This influences the size, shape and number of granules formed inside chloroplasts. Its physiological and genetic function has been extensively studied, but nothing is known about the biochemical mechanism that distinguishes it from other starch synthases, allowing it to perform its unique role in granule initiation. In Arabidopsis, several proteins have been identified in close association with SS4, including the coiled-coil protein MRC, together forming a potential granule initial protein complex from where starch initiation begins.

AtSS4 and *AtMRC* are both crucial components of starch granule initiation, but the mechanism by which they act together is not understood. The *AtSS4* N-terminal coiled coils could perhaps be forming an interaction with the MRC coiled coils. Considering that *AtSS4* might exist in a large complex, its direct interaction with *AtMRC*, and the lack of enzymatic domains in *AtMRC*, *AtMRC* could have a mainly structural role, forming a scaffold for the compartmentalisation of the granule initial complex. *AtMRC* could aid the concentration of glucan substrates and products to favour crystallisation into an ordered matrix. *AtMRC* interaction with *AtSS4* could also influence the activity of the *AtSS4* catalytic domain. In addition to protein interaction studies, structural information on both *AtSS4* and *AtMRC* could pave the way to understanding their function.

The overall aim of my thesis is to understand the biochemical mechanism of SS4 function, how this relates to its association with MRC, and how the function of these two proteins compares in Arabidopsis and wheat. Chapter 3 explores whether MRC also plays a role in granule initiation in wheat leaves and endosperm, by characterising several *T.turgidum mrc* mutant lines. This addresses the question of how much starch granule initiation mechanisms are conserved in different tissues in the same species as well as the same tissue in different species. In Chapter 4, I aim to understand the interaction between SS4 and MRC in more detail, through pairwise protein-protein interaction studies of both the Arabidopsis and wheat orthologs using co-immunoprecipitation, yeast two-hybrid, and localisation studies using fluorescently tagged proteins. Finally, in Chapter 5, I investigate the nature of the possible large SS4 complex, using size exclusion chromatography in plant tissue as well as *in vitro*. I demonstrate the successful purification of recombinant *AtSS4* and the potential for structural characterisation of an *AtSS4* dimer using transmission electron microscopy.

Chapter 2 – Materials and methods

Some sections of this chapter appear in (Chen *et al.*, 2022a) and (Chen *et al.*, 2022b). For standard media components not listed in the main text, see Appendix 3.

2.1 Plant material

2.1.1 *Triticum turgidum* for *mrc* phenotyping

The *mrc-1* and *mrc-2* mutant lines used in this thesis were previously generated in the Seung lab. Ethyl methanesulfonate (EMS) mutants of tetraploid wheat (*Triticum turgidum* cv. Kronos) carrying mutations in *TtMRC-A1* and the chromosome 6B pseudogene were identified from the wheat *in silico* TILLING database (<http://www.wheat-tilling.com>; (Krasileva *et al.*, 2017)) and obtained from the John Innes Centre (JIC) Germplasm Resource Unit. The selected mutants for *TtMRC-A1* were Kronos3272 (K3272), Kronos598 (K598) and Kronos4681 (K4681); while Kronos4305 (K4305) and Kronos3078 (K3078) were selected for the 6B pseudogene. From these mutants, we generated three different sets of lines. The *mrc-1* lines descend from a cross between K3272 and K3078, while the *mrc-2* lines descend from a cross between K4681 and K4305. For both crosses, *aaBB*, *AAbb* and *aabb* genotypes were obtained in the F2 generation. The *mrc-3* lines are the original K598 mutants. The KASP markers used to genotype the mutations are provided in Appendix 1B.

For experiments on grains and leaves, plants were grown in soil in a controlled environment room with fluorescent lamps supplemented with LED panels. The chambers were set to provide a 16-h light at 300 - 400 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ and 20°C, and 8-h dark at 16°C, with relative humidity of 60%. Grains from the first three tillers were harvested from mature, dry spikes (approximately 4 months after sowing). Leaves were harvested 10 days after germination, when the plants were at a two-leaf stage. The two leaves were pooled for starch quantification, and a section from the middle of the older leaf was used for light microscopy.

2.1.2 *Triticum turgidum* for RNA sequencing

Wheat plants were grown in controlled environment rooms (CERs) as above. Plants were marked for anthesis when an anther was visible from the spike, and this was taken as day 0. Checking for anthesis and harvesting were performed at 1 – 2 hours after midday. Four spikes from each individual plant were used, each for a different developmental time point. Three biological replicates per time point were collected from independent plants. Therefore, some of the biological replicates for separate time points were spread over several of the same plants. For sample collection, the whole spike was cut at the base, frozen in liquid nitrogen and stored at -80°C. Then, grains were threshed from the spikes on dry ice, avoiding the middle grain of each spikelet, and stored at -80°C again. Endosperms were subsequently dissected on dry ice, frozen in liquid nitrogen and stored at -80°C until RNA extraction.

2.1.3 *Arabidopsis thaliana*

Arabidopsis thaliana plants were grown in controlled environment rooms under 12-h light/12-h dark cycles, 20°C, 60 – 75% relative humidity and light levels of 150 $\mu\text{mol photons m}^{-2}\text{s}^{-1}$. These include Col-0 wild type, *ss4*, *mrc*, and transgenic lines expressing tagged Arabidopsis proteins (*ss4/pAtSS4:AtSS4-HA*, *ss4/pAtSS4:AtSS4-eGFP-HA*, *mrc/pAtMRC:AtMRC-HA*, *mrc/pAtMRC:AtMRC-RFP-HA*). Mutants are T-DNA insertion lines in the Columbia 0 (Col-0) background which have been used previously (Roldan *et al.*, 2007; Seung *et al.*, 2018). The *ss4* line is GABI_290D11 and *mrc* line is SAIL_1151_E06.

Arabidopsis ss4/pUBC:TaSS4-YFP and *mrc/p35S:TaMRC-HA* lines were previously generated in the Seung lab and phosphinothricin-resistant (for *ss4/pUBC:TaSS4-YFP*) or kanamycin-resistant (*mrc/p35S:TaMRC-HA*) T2 generation plants were selected. These lines were segregating for the transgenes and had not been selected for transgene single insertions, but were checked for protein expression by immunoblotting with an anti-GFP (TP401) or anti-HA (Abcam ab9110) antibody (Table 2.7).

Experiments were carried out on young leaves harvested from 5 - 7-week-old rosettes.

2.1.4 *Nicotiana benthamiana*

Nicotiana benthamiana plants for transient expression were grown in a glasshouse set to provide a minimum of 16 h light at 22°C, and a dark period of 20°C. These were used for *Agrobacterium tumefaciens* infiltration (strain GV3101, for details see section 2.9.4) at 4 – 6 weeks old, when there were two fully developed leaves. After infiltration, they were placed in a shaded area overnight, then placed on a windowsill until analysis, 72 hours after infiltration.

2.2 RNA-seq analysis of Kronos wheat developing endosperm

2.2.1 RNA extraction and sequencing

Total RNA was extracted (n=3) from frozen endosperm tissue using the RNeasy PowerPlant kit (Qiagen). An on-column DNase digest was incorporated during the extraction. Either 6 (for 6, 8 and 10 dpa), 4 (for 13, 15, and 18 dpa) or 3 (for 20 and 30 dpa) endosperms were pooled from each spike for each RNA extraction. Quality control, poly-A selection library preparation and RNA sequencing (RNA-seq) using a NovaSeq 6000 machine (Illumina) was carried out at Novogene (Cambridge, UK). Each sample returned between 78,762,223 and 169,888,168 million clean 150 base pair paired end reads. These clean RNA-seq reads can be downloaded from the NCBI Gene Expression Omnibus repository (Accession: GSE216253).

2.2.2 TPM value calculation and differential expression

Initial processing of RNA-seq data was carried out by Martin Vickers and Alexander Watson-Lazowski. Clean reads from above were pseudoaligned with *kallisto* to the Svevo v1 durum wheat transcriptome downloaded from Ensembl Plants, using canonical transcripts (Bray *et al*, 2016). The default settings were used for mapping. A range of 66.356 -77.040% of reads per sample aligned to the transcriptome. Both estimated count values and normalised expression as transcripts per million (TPM) for each gene were retrieved from the pseudoalignment. *DESeq2* (Love *et al*, 2014), run through the *DEapp* shiny (Li & Andrade, 2017), was then used to identify significantly differentially expressed genes (DEGs; significance cut-offs used - FDR \leq 0.05, \log_2 fold change $>$ 1.5 and a minimum count cut-off of 10 counts per million in at least three samples).

For data visualisation, average TPM values of the three biological replicates for each timepoint with their standard error of the mean were plotted using R.

2.3 Wheat bioinformatics

2.3.1 Publicly available endosperm expression data

Publicly available microarray expression data for rice was extracted from:

<https://ricexpro.dna.affrc.go.jp/> (Sato *et al.*, 2013). RNA-sequencing expression data for maize was extracted from (Qu *et al.*, 2016). For this paper, only one FPKM value per timepoint was listed, and was assumed to be the average value of the three biological replicates reported. The expression values were visualised using R.

2.3.2 *TtMRC* sequence analysis

The accession numbers corresponding to the genes investigated in Chapter 3 are: *TtMRC-A1* (TRITD6Av1G081580), *TaMRC-A1* (TraesCS6A02G180500), *TaMRC-D1* (TraesCS6D02G164600).

The analysis of the B-homeolog of wheat MRC was done in collaboration with Andy Chen. To characterize *TtMRC-B1* in tetraploid wheat, the whole genome-sequencing reads of *Triticum dicoccoides* (n = 10) and *Triticum turgidum ssp. durum* (n = 10) from (Zhou *et al.*, 2020) were aligned against the 'tetraploid' version of the Ref-Seqv1.0 Chinese Spring assembly (International Wheat Genome Sequencing, 2018). HiSat2-v-2.1.0 (Kim *et al.*, 2019) with the default settings was used for sequence alignment and the read alignments on the genetic signatures of retrotransposon insertion on *MRC-B1* (Figure 1B) were visualized using Integrated Genomics Viewer (Robinson *et al.*, 2011). Phylogenetic analyses were performed by David Seung as described in (Seung *et al.*, 2018).

2.4 Wheat phenotyping

2.4.1 Starch purification from mature and developing wheat endosperm

Starch was purified from mature grains using 3-6 grains per extraction. Dry grains were soaked overnight at 4°C in 5 mL of sterile water. The softened grains were homogenised in 10 mL sterile water using a mortar and pestle, and the homogenate was filtered through a 100 µm mesh. The starch was pelleted by centrifugation at 3000 × *g* for 5 minutes and resuspended in 2 mL of water. The resuspended starch was loaded on top of a 5 mL 90% Percoll (Sigma) cushion buffered with 50 mM Tris-HCl, pH 8, and was spun at 2,500 × *g* for 15 minutes. After the spin, it was verified that no intact granules were left in the Percoll interface. The starch pellet was washed twice with wash buffer (50 mM Tris-HCl, pH 6.8; 10 mM Ethylenediaminetetraacetic acid (EDTA); 4% SDS; and 10 mM dithiothreitol (DTT)), then three times with water, followed by a final wash in absolute ethanol. The starch was then air dried overnight.

For starch extraction from developing endosperm, developing grains were harvested at the indicated timepoints and were snap frozen in liquid nitrogen and stored at -80°C until analysis. Each grain was thawed just prior to extraction and the endosperm was dissected and placed into a chilled tube and weighed. The tissue was then homogenised in sterile water with a pestle, then filtered through a 60 µm mesh. The pellet was washed three times in 90% Percoll (Sigma) buffered with 50 mM Tris-HCl, pH 8, then three times with wash buffer (as above), followed by three times with water.

2.4.2 Total starch quantification from wheat leaves and endosperm

Starch was quantified in leaf tissue as previously described (Smith & Zeeman, 2006). Briefly, frozen leaf tissue was ground into a powder with a ball mill and then extracted with perchloric acid. Starch in the insoluble fraction of the extraction was gelatinised at 95°C and digested to glucose with α-amylase (Megazyme) and amyloglucosidase (Roche). The glucose released was measured using the hexokinase/glucose-6-phosphate dehydrogenase method (Roche). Starch content (in glucose equivalents) was calculated relative to the fresh weight of the leaves.

Five individual experiments were done to quantify the total end of day leaf starch content (Table 3.7). One set of genotypes was measured in experiments 1 – 3: WT, *mrc-1*, *mrc-2*, *mrc-3*. Another set was measured in experiments 4 and 5: WT, *mrc-1*, the double backcrossed *mrc-1* BC2 *aabb* and the wild-type segregant from that backcross, *mrc-1* BC2 AABB. Therefore, in the statistical

analysis, experiments 1, 2, 3 were pooled together and experiments 4 and 5 were pooled together. An analysis of variance (ANOVA) using a fixed effects model showed no interaction between the experiment effect and genotype effect, and experiment and genotype were set as fixed effects in the linear model. Individual linear models were fitted for end of day starch in experiments 1 – 3, end of day starch in experiments 4 & 5 and end of night starch in experiments 1 & 2. Then ANOVA and Tukey post-hoc tests using these linear models were performed for pairwise comparisons of the genotypes. For the details of number of replicates, see Table 3.7. I used the `lm()` function from the 'stats' package in R for the linear models, and the `emmeans()` function from the 'emmeans' package for calculating adjusted means and pairwise comparisons.

A similar method to leaves was used for starch quantification in grains. Mature grains (5-6 grains) were soaked overnight at 4°C in 5 mL of sterile water and were homogenised using a mortar and pestle. Developing endosperm tissue was extracted in 1 mL of sterile water with the pestle. Insoluble material in an aliquot of the homogenate was collected by centrifugation at 5,000 × *g* for 5 minutes, then washed once in 0.7 M perchloric acid, once in sterile water, then three times in 80% ethanol. The pellet was then resuspended in water. Starch in the pellet was gelatinised by heating at 95°C for 15 minutes, then digested using α -amylase (Megazyme) and amyloglucosidase (Roche), with glucose quantified as above for the leaves.

2.4.3 Grain morphometrics

This was carried out by David Seung and I performed the analysis. The number of tillers per plant was counted manually. The number of grains harvested per plant, plus grain area and thousand grain weight were quantified using the MARViN seed analyser (Marvitech GmbH, Wittenburg). Multiple grains from each plant (15 - 88 individual grains per plant) were measured for grain area to calculate the average value for each plant. All measured values were used in the plots of Figure 3.4 and the analysis of tables 3.1 – 3.4.

2.4.4 Leaf granule image analysis

For quantification of the number of granules per chloroplast, light microscopy images from three individual plants of each of experiments 2 and 5 were used (Table 3.7). Using the cell counter plug-in in Fiji (Schindelin *et al*, 2012), the number of granules per chloroplast were counted manually, until 200 – 250 chloroplasts were reached. Cells were chosen across the section, where

most of the granules were stained well and clearly visible, and cells directly next to vasculature were avoided. It should be noted that these are counts of granule sections, rather than the actual numbers of granules per chloroplast, as the actual number cannot be determined from two-dimensional sections. As much as possible, all chloroplasts in a cell were counted to avoid bias, although occasionally a few chloroplasts where the granules were unclear had to be omitted. Both lighter and darker stained granules were counted, but occasionally light-staining patches that were difficult to discern as granules were observed in all samples, and these were not included in the count.

For quantification of the granule size, the same samples as for the granules per chloroplast were analysed, using the 'analyse particles' function in Fiji, and the thresholding in an image was manually adjusted until the dark granules in the images were correctly separated. This was done for several images across the section, until 100 – 400 granules were reached for each plant. As much as possible, instances where the thresholding had fused multiple granules in proximity were removed, by checking obvious outliers against the original images.

For the statistics of the quantification of the number of granules per chloroplast, Statistical Services Centre Ltd (Reading, UK) was consulted, who wrote the statistical analysis script in R for us, available at [https://github.com/Jiaawen/2022_MRC_wheat_Rscripts/tree/main/Fig10]. A negative binomial mixed effects model was used, since count data was being analysed and I wanted to account for both the random effect of biological replicate and the frequency distribution of number of granules per chloroplast. I used the `mixed_model()` function with 'family=negative.binomial' from the 'GLMMadaptive' package, and the `emmeans()` function from the 'emmeans' package for pairwise comparisons in R. The data had overdispersion when fitting a Poisson model but not when using a negative binomial model. Individual plants were used as the random effect (for both experiments: 3 plants for each of 4 genotypes – 12 plants total), and genotype was used as fixed effect.

For the statistics of the granule area, a linear mixed effects model was used, also with the individual plants as random effect and genotype as fixed effect. I used the `lmer()` function from the 'lmerTest' package, and the `emmeans()` function from the 'emmeans' package for pairwise comparisons.

2.4.5 Coulter counter analysis of wheat endosperm starch

For profiles of granule size distribution, purified starch was suspended in Isoton II diluent (Beckman Coulter) and analysed with a Multisizer 4e Coulter counter fitted with a 70 µm aperture (Beckman Coulter). Granules were counted either in volumetric mode, measuring 1 mL from a total 100 mL volume preparation containing 20 µL of purified starch, or set to count at least 100,000 granules. For calculations of granule counts in volumetric mode, the number of granules per mg grain weight was back calculated to the total starting grain weight. The granules were sized, with the Coulter counter collecting the data using logarithmic bins for the granule diameter (standard settings).

For each plant, to calculate the mean A- and B-type granule size, as well as relative B-type granule volume, a mixed bimodal gaussian curve was fitted to the distribution using R (https://github.com/JIC-CSB/coulter_counter_fitting). As the data collection on the Coulter counter is set to logarithmic bins on the x-axis, for these calculations and for the traces in figures 3.6A and 3.10A, the x-axis was transformed to even bins, by changing the y-axis to volume percentage density (volume percentage for each bin divided by bin width). For each of the extracted phenotypes of mean A- and B-type granule size and relative B-type granule volume, individual linear models were fitted and a one-way ANOVA and Tukey post-hoc tests were performed for pairwise comparisons of the genotypes, using the `lm()` and `emmeans()` functions in R, from the 'stats' and 'emmeans' packages. Nine individual plants per genotype were used for this experiment.

2.4.6 Analysis of amylopectin structure and amylose content

This section was carried out by Robin Denley Bowers. Amylopectin structure and amylose content were analysed using purified starch. Amylopectin structure in terms of chain length distribution was quantified using High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) (Blennow, 1998). For amylose content, granules were dispersed in dimethyl sulfoxide (DMSO) and quantified using an iodine-binding method (Warren *et al*, 2016).

2.5 Light microscopy

The embedding and sectioning of samples was done by Elaine Barclay from the Bioimaging platform at the John Innes Centre.

2.5.1 Transmission light microscopy

2.5.1.1 Leaf material

For light microscopy of starch granules in Arabidopsis and wheat leaf chloroplasts, leaf material (1.5 mm x 1.5 mm squares) was harvested and fixed in 2.5% glutaraldehyde in 0.05 M sodium cacodylate, pH 7.4, which was vacuum infiltrated into the leaf. For wheat, leaf material was taken from the middle of the older of the two wheat leaves in 10-day-old wheat plants (same batch as those used for starch quantification). For Arabidopsis, leaf material was taken from the middle of an average sized leaf from a 5 – 7-week-old plant. For both samples, the leaf material was cut while avoiding the centre vein of the leaf. The segments were post-fixed with osmium, then dehydrated in an ascending ethanol series, and embedded in LR white resin using an EM TP embedding machine (Leica). Semi-thin sections (0.5 µm thick) were produced from the embedded leaves using a glass knife and were dried onto polytetrafluoroethylene (PTFE)-coated slides.

Chloroplasts and cell walls were stained using toluidine blue stain (0.5% toluidine blue 'O', 0.5% sodium borate) for 1 min. Starch was stained using reagents from the Periodic Acid-Schiff staining kit (Abcam): using a 30 min. incubation with periodic acid solution, followed by 5 min. with Schiff's solution. Then, chloroplasts and cell walls were again stained using toluidine blue stain for 1 min. The sections were mounted with Histomount (National Diagnostics) and imaged on a DM6000 microscope with 63X oil immersion lens (Leica).

2.5.1.1 Grain material

For light microscopy of endosperm sections from mature grains, thin sections (1 µm thick) of mature grains were made using a microtome fitted with a glass knife. Sections were mounted onto a glass slide and stained with 3% Lugol's iodine solution (Sigma) prior to imaging on a DM6000 (Leica) or AxioObserver Z1 (Zeiss) microscope.

For light/electron microscopy of developing endosperm tissue, developing grains (15 dpa) were harvested into 4% paraformaldehyde, 2.5% glutaraldehyde in 0.05 M sodium cacodylate, pH 7.4. The osmium post-fixation, dehydration and embedding into LR white resin was done as described above for leaves. For light microscopy, semi-thin sections were stained with toluidine blue and imaged as described above, by Erica Hawkins.

2.5.2 Confocal microscopy

For co-localisation, transient expression in *Nicotiana benthamiana* was used. The following constructs were used: pUBC:AtSS4-YFP, pUBC:AtMRC-RFP, pUBC:CTP-RFP (chloroplast transit peptide from Arabidopsis Rubisco small subunit), and P19. For more details on constructs and cloning, see section 2.8 and Table 2.3.

After 72 hours from infiltration, a leaf slice was cut from one of the infiltrated leaves, mounted on a microscope slide with water and imaged using a Leica SPX8 confocal microscope using the 63X water immersion lens. For imaging, the white light laser was set to a 514 nm excitation wavelength for yellow fluorescent protein (YFP) and chlorophyll autofluorescence and emission windows were set to 520 – 560 nm (YFP) and 650 – 700 nm (chlorophyll). A sequential channel set to a 589 nm excitation wavelength was used for red fluorescent protein (RFP), with an emission window of 596 – 640 nm (RFP). For all images in Figure 4.4, the same settings were used.

2.6 Plant material electron microscopy

2.6.1 Transmission electron microscopy

Transmission electron microscopy (TEM) grids were prepared and imaged at JIC Bioimaging by Elaine Barclay. Ultra-thin sections (approximately 80 nm) were produced from the embedded grains or leaves by sectioning with a diamond knife using a Leica UC7 ultramicrotome (Leica, Milton Keynes). The sections were picked up on 200 mesh copper grids which had been formvar and carbon coated, then stained with 2% (w/v) uranyl acetate for 1 hour and 1% (w/v) lead citrate for 1 minute, washed in distilled water and air dried. The grids were viewed in a FEI Talos 200C transmission electron microscope (Thermo Fischer Scientific, Eindhoven, Netherlands) at 200 kV and imaged using a Gatan 4k OneView CMOS detector (Gatan United Kingdom, Abingdon, Oxfordshire, United Kingdom) to record DM4 files.

2.6.2 Scanning electron microscopy

Scanning electron microscopy (SEM) was done by Elaine Barclay and David Seung. For imaging starch granules, a drop of purified starch suspended in water (5 mg/mL) was air-dried onto a glass coverslip attached onto an SEM stub. For imaging sections through developing endosperm, harvested grains were fixed in 2.5% glutaraldehyde in 0.05 M sodium cacodylate, pH 7.4. The fixative was removed by washing with 0.05 M sodium cacodylate, pH 7.4, after which the grains were dehydrated in an ascending ethanol series, and then subjected to critical point drying in a CPD300 instrument (Leica) according to the manufacturer's instructions. Thick transverse sections were produced from the dried grains and were glued onto SEM stubs. All stubs were sputter coated with gold (4 – 6 nm) and observed using either a Supra 55 VPFEF (Zeiss) or Nova NanoSEM 450 (FEI) SEM instrument.

2.7 Transmission electron microscopy of proteins

2.7.1 Negative staining

Carbon-coated 400 mesh copper grids (Agar Scientific) were glow-discharged for 20 seconds at 10 mA (Leica EM ACE200). 3.5 μ L of purified protein (0.02 mg/mL for SEC fraction from 4 L purification) was applied on the grids for 60 seconds, then excess sample was wicked away with Whatman No.1 filter paper. 3.5 μ L of 2% (w/v) uranyl acetate in H₂O was applied for 30 seconds, excess liquid was removed with filter paper, and the grid was left to air-dry.

2.7.2 Imaging conditions

Negative stained grids were imaged using the Talos F200C TEM operated at 200 kV, equipped with a 4k OneView CMOS detector to record DM4 files. Grids were first screened manually, then suitable grids were chosen for automated data acquisition using EPU v 2.7.0.5806 (Thermo Fisher Scientific, Eindhoven, Netherlands), collected by Jake Richardson. Each image had a 1 s exposure with a sample dose of approximately 40 e⁻/Å², defocus of -0.7 μ m, autofocus every 10 μ m, nominal magnification of 45,000 \times and a resolution of 3.5 Å/pixel.

2.7.3 Data processing and NS-TEM image analysis

Micrographs from the automated image collection were used for particle picking using the Relion 3.1.1 pipeline (Scheres, 2012; Zivanov *et al*, 2018), analysed by David Lawson and Michael Webster. In the first sample from the small-scale purification, 151 micrographs were collected. It was clear that the images suffered from some GroEL contamination, although there were clearly other particles present. Initially, 2D classes were generated from particles automatically picked using the reference-free Laplacian-of-Gaussian option. The best of these classes, excluding any obvious GroEL classes, were used as references for further autopicking. The resultant particle stack was cleaned up via several iterations of 2D classification and subset selection, again excluding any GroEL classes, to yield a set of promising classes (box size approx. 450 Å, circular mask diameter 250 Å). However, at this stage, there were only 3973 particles remaining, making further analysis difficult.

Therefore, these 2D classes were used as templates for making new autopick templates from the second dataset obtained from the 4L protein purification, starting from 899 micrographs and 537,792 particles. Three rounds of 2D classification were done, each with 25 iterations of alignment. Subset selection was performed while excluding GroEL classes, producing final 2D class averages from 93 884 particles (box size approx. 350 Å, circular mask diameter 220 Å).

Using these final 2D classes, a 3D density map was generated in Relion, at a resolution of 22.5 Å.

2.8 Cloning

In this thesis I used Gateway cloning and the seamless cloning methods of In-Fusion cloning and Gibson Assembly. For transient expression in *N.benthamiana* and for yeast two-hybrid (Y2H), I used Gateway cloning because there were existing Gateway vectors suitable for these experiments. For creating transgenic Arabidopsis lines, I used In-Fusion cloning to create optimal constructs without additional *att* sequences resulting from Gateway cloning.

For most of the Gateway constructs, David Seung had previously generated donor plasmids. For cloning new constructs, inserts were amplified from plasmid templates, genomic DNA (gDNA), or synthesised DNA, using Q5[®] High-Fidelity DNA Polymerase (New England Biolabs) or CloneAmp HiFi PCR Premix from the In-Fusion kit (Takara Bio) with polymerase chain reaction (PCR) conditions according to the manufacturer's protocol. The PCR products were separated by electrophoresis on a 1% agarose gel and visualised using a UV transilluminator. The band of expected size was cut out and purified from the agarose gels using the Zymoclean Gel DNA Recovery Kit (Zymo Research) or Monarch[®] DNA Gel Extraction Kit (New England Biolabs). Unless stated otherwise, the assembled plasmids were transformed into chemically competent *Escherichia coli* DH5 α cells (Library Efficiency[™] DH5 α Competent Cells, Invitrogen) and the QIAprep Spin Miniprep Kit (Qiagen) was used for plasmid isolation from *E.coli*. All kits were used according to manufacturer's instructions, and samples were eluted with distilled water. The sequences were verified by Sanger sequencing, performed by Eurofins Genomics with the Mix2Seq kit service or by Source Bioscience. Other than custom sequencing primers listed in Appendix 1, standard sequencing primers for the relevant vectors were used. Chromatograph traces were analysed in CLC Main Workbench (Qiagen).

2.8.1 Gateway cloning

For all Gateway cloning experiments, the coding sequence (CDS) of *AtSS4*, *AtMRC*, *TaSS4* (full length and truncated versions), *TaMRC* or the chloroplast transit peptide (CTP) sequence of the Rubisco small subunit from Arabidopsis (Kim *et al*, 2010) were used, previously cloned into the pDONR221 Gateway donor vector by David Seung.

The Gateway cloning system (Invitrogen) relies on the exchange of *att*-sites between DNA sequences and is a two-step reaction. First, the desired gene fragment is inserted into a donor vector (here pDONR221) through recombination of *attB1* and *attB2* sites on either side on the fragment and *attP* sites on the donor vector, using the enzyme BP clonase, forming the *attL* sites

in the entry plasmid. In the second reaction, the *attL* sites recombine with the *attR* sites on the destination vector, catalysed by the enzyme LR clonase. The reaction components are listed in tables 2.1 and 2.2. The reactions were performed at room temperature for 1 hour, then transformed into *E.coli*. The destination vectors used in this thesis are pUBC-YFP, pUBC-RFP, pB7RWG2, pB7YWG2, pJCV52, pGBKT7, pGADT7. Primers used for these cloning reactions were designed with the correct *att* sites and are listed in Appendix 1.

Table 2.1 BP reaction components.

BP reaction component	Amount for total 5 μ L reaction (μ L)
DNA fragment (10 ng/ μ L)	1
Donor plasmid (approx. 150 ng/ μ L)	0.5
TE buffer (Tris-EDTA buffer)	2.5
BP clonase	1

Table 2.2 LR reaction components.

LR reaction component	Amount for total 5 μ L reaction (μ L)
Entry plasmid (10 – 500 ng/ μ L)	0.5
Destination plasmid (10 – 500 ng/ μ L)	0.5
TE buffer (Tris-EDTA buffer)	3
LR clonase	1

2.8.2.1 Transient expression in *N.benthamiana*

For transient expression in *Nicotiana benthamiana*, the Gateway vectors pUBC-YFP, pUBC-RFP and pJCV52 were used. These all have C-terminal tags. *AtSS4* in pUBC-YFP was previously cloned by David Seung.

2.8.2.2 Y2H

Stanislav Kurass and Brendan Fahy helped me with the cloning of the yeast two-hybrid (Y2H) constructs. The Gateway vectors pGBKT7 and pGADT7 (Clontech) were used, which have the Gal4 activation domain (AD) or binding domain (BD) tagged at the N-terminus of the protein of interest. pDONR221 constructs of *TaSS4* (full length and truncated) and *TaMRC* previously generated by David Seung were used as donor plasmids. These constructs were originally used for

plant expression, so they have the chloroplast transit peptide (CTP) included. The presence of the CTP does not make a difference to the behaviour of the constructs, and the CTP is likely not cleaved in yeast. I also did a Y2H experiment with full-length *TaSS4* and *TaMRC* sequences used for *E.coli* expression, without their CTPs, which gave the same result as using the plant constructs (not shown). The constructs used also do not have a stop codon directly at the end of the sequence, but there is a stop codon present in pGBKT7 and pGADT7 downstream from the end of the coding sequence, after 28 and 13 amino acids respectively. The murine p53 in the pGBKT7 vector and its interacting partner the SV40 large T-antigen in the pGADT7 vector were used as positive controls (Clontech PT3024-1).

2.8.2 In-Fusion cloning/Gibson Assembly

In-Fusion cloning (Takara Bio) and Gibson Assembly (New England Biolabs) are similar cloning methods which both use a seamless primer design without additional vector-specific fragments in the cloned constructs. I used In-Fusion cloning for most of the new constructs that I made for this project, which are the ones used for creating Arabidopsis transgenic lines, and the *AtMRC* construct used for *E.coli* expression. Gibson Assembly was used to clone *AtSS4* into a custom vector backbone, as In-Fusion cloning was not used in the lab yet at the time. Both cloning reactions and restriction digest reactions were done according to the enzyme manufacturer's protocols.

2.8.3.1 Arabidopsis transformation constructs

I cloned the tagged *AtSS4* and *AtMRC* constructs for generating Arabidopsis transgenic lines using the In-Fusion kit (Takara), with the Gateway plasmid pK7YWG2 as a backbone. This is a binary vector for stable plant transformation using *Agrobacterium tumefaciens* and has a kanamycin resistance marker for plant expression. I cut this vector using HindIII and XbaI restriction enzymes to remove all the sequences from promoter to terminator, including the tag in the original plasmid, and inserted my own sequences in between these two cuts. In In-Fusion cloning, the DNA fragments are amplified with primers that create 15 base pair (bp) overhangs which are homologous to the ends of a linearised vector backbone, and the two fragments are annealed using the In-Fusion enzyme mix. In this reaction, 3' exonuclease activity creates complementary overhangs of the insert and backbone, and no ligase or polymerase is needed. The constructs

were transformed into Stellar competent cells from the In-Fusion kit, according to the manufacturer's instructions.

The promoter fragments were cloned from *Arabidopsis* gDNA (provided by David Seung). The *AtSS4* and *AtMRC* CDS were amplified from existing plasmids (*AtSS4* in pUBC-YFP, *AtMRC* in pDONR221, both cloned previously by David Seung). The linker-GFP-HA and linker-RFP-HA sequences were synthesised (gBlocks, Integrated DNA Technologies, Appendix 2), and the linker-HA fragment was PCR-amplified using the linker-GFP-HA as backbone. The NOS terminator sequence was amplified from another position in the pK7YWG2 plasmid. For the HA-tag constructs, the four fragments were all PCR-amplified with In-Fusion primers for complementary overhangs, gel extracted, and cloned into one seamless sequence in the pK7YWG2 digested backbone. For primers, see Appendix 1. The fluorescent protein-tagged constructs were cloned using the HA-tagged constructs as template. The amplified p*AtMRC* promoter has a deletion of one adenine in a repeat adenine sequence compared to the sequence on Ensembl plants, but no obvious binding motifs were identified in this region and this construct was still taken forward.

2.8.3.2 Recombinant expression in *E.coli* constructs

For recombinant protein expression in *E.coli*, I cloned an *E.coli* codon-optimised *AtMRC* fragment (*AtMRC_cdEC*) into the pOPINE *E.coli* expression vector using In-Fusion cloning. I used a synthesised *E.coli* codon-optimised *AtMRC* DNA fragment as a template (gBlocks, Integrated DNA Technologies, Appendix 2), and used a pOPINE vector cut with PmlI and NcoI restriction enzymes as the backbone.

David Seung had previously generated an *E.coli* codon-optimised *TaSS4* construct (*TaSS4_cdEC*) in a modified version of the pPROEXHtb vector, where the N-terminal histidine tag (His-tag) was replaced with a Strep-II-tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys), and the ampicillin resistance was replaced with spectinomycin resistance, here named pPROEXHtb2. There is a missense mutation in the CDS of this construct, where a cytosine to thymine change caused a Ser286Leu change at the end of the *TaSS4* coiled-coil domain.

I used the above plasmid (*TaSS4_cdEC* in pProExHtb2) to amplify out the backbone to clone an *AtSS4* construct, using the Gibson assembly kit (New England Biolabs). I used an *E.coli* codon-optimised *AtSS4* CDS from a synthesised DNA fragment as a template (gBlocks, Integrated DNA Technologies, Appendix 2), amplified with Gibson assembly primers to create complimentary overhangs. In Gibson assembly, a 5' exonuclease chews back 5' sequences for annealing of insert and backbone, a polymerase fills in the gaps of the annealed regions, and a DNA ligase seals the

nick. The cloned plasmid (AtSS4_cdEC) was transformed into *E.coli* DH5α competent cells from the Gibson Assembly cloning kit (New England Biolabs).

2.8.3 Plasmids

The plasmids used in this thesis are listed in Table 2.3. The primers used for the cloning of these plasmids are listed in Appendix 1. The antibiotics used in the cloning reactions and their working concentrations are listed in Table 2.4.

Table 2.3 Plasmids used in this thesis. The rows are coloured by the purpose of the plasmid. D.S. = David Seung

Construct	Vector	Purpose	Cloning	Notes
pAtSS4-AtSS4-L-HA	pK7YWG2_HindIIIxbaI	Arabidopsis transformation	In-Fusion	
pAtSS4-AtSS4-L-eGFP-HA	pK7YWG2_HindIIIxbaI	Arabidopsis transformation	In-Fusion	
pAtMRC-AtMRC-L-HA	pK7YWG2_HindIIIxbaI	Arabidopsis transformation	In-Fusion	One adenine deletion in promoter
pAtMRC-AtMRC-RFP-HA	pK7YWG2_HindIIIxbaI	Arabidopsis transformation	In-Fusion	One adenine deletion in promoter
pUBC:AtSS4-YFP	pUBC-YFP	localisation in <i>N.benthamiana</i>	Gateway	cloned by D.S
pUBC:CTP-YFP	pUBC-YFP	localisation in <i>N.benthamiana</i>	Gateway	
pUBC:CTP-RFP	pUBC-RFP	localisation + co-IP in <i>N.benthamiana</i>	Gateway	
pUBC:AtMRC-RFP	pUBC-RFP	localisation + co-IP in <i>N.benthamiana</i>	Gateway	
p35S:CTP-RFP	pB7RWG2	localisation in <i>N.benthamiana</i>	Gateway	
p35S:AtSS4-HA	pJCV52	Co-IP in <i>N.benthamiana</i>	Gateway	
p35S:TaSS4-YFP	pB7YWG2	Co-IP in <i>N.benthamiana</i>	Gateway	cloned by D.S
p35S:TaSS4dCC-YFP	pB7YWG2	Co-IP in <i>N.benthamiana</i>	Gateway	cloned by D.S
p35S:TaSS4N-YFP	pB7YWG2	Co-IP in <i>N.benthamiana</i>	Gateway	cloned by D.S
pUBC:TaMRC-RFP	pUBC-RFP	Co-IP in <i>N.benthamiana</i>	Gateway	
AtSS4_cdEC	pPROEXHtb2	<i>E. coli</i> expression	Gibson assembly	
TaSS4_cdEC	pPROEXHtb2	<i>E. coli</i> expression		cloned by D.S., missense mutation
AtMRC_cdEC	pOPINE	<i>E. coli</i> expression	In Fusion	
TaSS4	pGBKT7	Y2H	Gateway	
TaSS4 dCC	pGBKT7	Y2H	Gateway	
TaSS4 N	pGBKT7	Y2H	Gateway	
TaMRC	pGBKT7	Y2H	Gateway	
TaSS4	pGADT7	Y2H	Gateway	
TaSS4 dCC	pGADT7	Y2H	Gateway	
TaSS4 N	pGADT7	Y2H	Gateway	
TaMRC	pGADT7	Y2H	Gateway	

Table 2.4 Antibiotics used in this study and their working concentrations. All stocks were 1000 × and filter sterilised with a 0.2 µm filter (Sartorius), and stored in -20 °C. All stocks were dissolved in water, except for rifampicin which was dissolved in MeOH.

Antibiotic	Stock components	Concentration (mg/L)
Ampicillin	Ampicillin sodium salt	50
Carbenicillin	Carbenicillin disodium	100
Gentamycin		20
Kanamycin	Kanamycin monosulfate	50
Rifampicin		50
Spectinomycin	Spectinomycin dihydrochloride pentahydrate	100

2.9 Transformation of organisms

2.9.1 *E.coli* transformation

Chemically competent *Escherichia coli* DH5α cells (Library Efficiency™ DH5α Competent Cells, Invitrogen) were used for plasmid production, and chemically competent *E.coli* BL21 (DE3) ΔglgAP cells (generated by David Seung, Szydlowski *et al.*, 2009) were used for expression of SS4 and MRC proteins for purification. 50 µL aliquots of cells were mixed with 1 – 2 µL of plasmid DNA at approx. 10 - 200 ng/µL, incubated on ice for 5 minutes, heat shocked in a 42 °C water bath for 45 seconds, then placed on ice for 1 – 2 minutes. 1 mL of SOC media was added to the cells and samples were incubated at 37 °C, shaking at 300 revolutions per minute (rpm) for 1 hour. Then they were plated on LB (Lysogeny Broth)-agar plates (1% w/v) with the appropriate antibiotics and were incubated overnight at 37°C. Colonies were picked and grown in 5 mL of LB media over night at 37 °C, 225 rpm. DNA was isolated, then sent for sequencing with appropriate primers. The correct plasmids were chosen based on the sequencing and kept for further use.

2.9.2 *Agrobacterium* transformation

Agrobacterium tumefaciens strain GV3101 electrocompetent cells (previously generated in the Seung lab) were used for transformation. 50 µL aliquots of cells were mixed with 1 – 2 µL of plasmid DNA at approx. 10 - 200 ng/µL, then placed in 1 mm electroporation cuvettes (Molecular BioProducts), and transformed by electroporation at 1.25 kV, 200 Ω, using an electroporator (Bio Rad pulse controller). 1 mL of LB media was added to the cells and samples were incubated at

28°C, 300 rpm for 2 hours. Then they were streaked onto LB-agar plates with the appropriate antibiotics and were incubated at 28°C for 2 days.

2.9.3 Bacterial glycerol stocks for storage

For both *E.coli* and *A.tumefaciens* cultures amplifying the correct plasmids, 500 µL of a 5 mL overnight culture was mixed with 500 µL of 40% glycerol, frozen and stored at -80 °C. For growth of the same cultures again, an aliquot of the glycerol stock was scooped with a pipette tip and placed into a 5 mL overnight culture with appropriate antibiotics.

2.9.4 *Nicotiana benthamiana* transient transformation

For co-localisation and co-immunoprecipitation experiments, *Agrobacterium* (GV3101) 5 mL cultures were grown for 48 hours from single colonies or from glycerol stocks. These cultures were spun down at 3023 × *g* for 10 minutes and resuspended in MMA buffer (10 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.6; 10 mM MgCl₂; 0.1 mM acetosyringone), then the OD₆₀₀ was adjusted to 0.2 for the P19 silencing suppressor plasmid (Win & Kamoun, 2004) and 0.3 for the constructs of interest, and desired combinations of constructs were mixed 1:1:1 with P19. These samples were infiltrated into two leaves of *Nicotiana* (4 – 6 weeks old) using a blunt syringe, using the first two fully unfurled leaves from the top of the plant, until most of the leaf was infiltrated. Plants were placed in a shaded area overnight, then placed on the windowsill the next day. 72 hours after infiltration, the infiltrated leaves were used for relevant experiments.

2.9.5 *Arabidopsis thaliana* stable transformation by floral dipping

Arabidopsis plants [*ss4* (GABI_290D11) and *mrc* (SAIL_1151_E06) background] were transformed with *Agrobacterium* (GV3101) using the floral dip method (Zhang *et al*, 2006). A single colony of *Agrobacterium* with the desired plasmid (or an aliquot of glycerol stock from the relevant strain) was grown in 5 mL cultures of LB media with relevant antibiotics for 2 days at 28 °C, 200 rpm. This pre-culture was used to inoculate a 200 mL culture, grown at 28 °C overnight (16 – 24 hours). Cells were centrifuged at 4000 × *g* for 10 minutes, then resuspended in 200 mL infiltration medium (5% sucrose; 0.05% Silwet in H₂O). *Arabidopsis* developing flowers were dipped in this suspension,

then covered in cling film and kept dark overnight. The plants were unwrapped the next day and grown until transgenic seeds were harvested.

2.10 Generation of Arabidopsis transgenic lines

2.10.1 Screening by antibiotic resistance

For transgenic lines obtained from floral dipping, seeds from independent T0 dipped plants were harvested (T1 seeds) and screened on Murashige-Skoog (MS)-agar plates (3% sucrose, 0.8% agar) containing kanamycin (50 mg/L) or the herbicide phosphinothricin (15 mg/L), dependent on the construct tested. The plates were stratified and grown in a CER with conditions as described in section 2.1.3. After about two weeks, susceptible plants with pale leaves could be discerned from resistant plants with green leaves and any resistant plants were transferred to soil. These plants were grown to maturity and the seeds harvested as T2 seeds. These were sown in selective media again and screened for single insertion of the transgene by looking for lines with a Mendelian segregation of 3:1 for resistant to susceptible plants. Resistant plants were transferred to soil again, and seeds collected as T3 seeds. These were sown on selective media for a T3 screen of homozygous lines which were segregating for 100% resistance, and these lines were kept as stable transgenic lines. This process is summarised in figure 2.1.

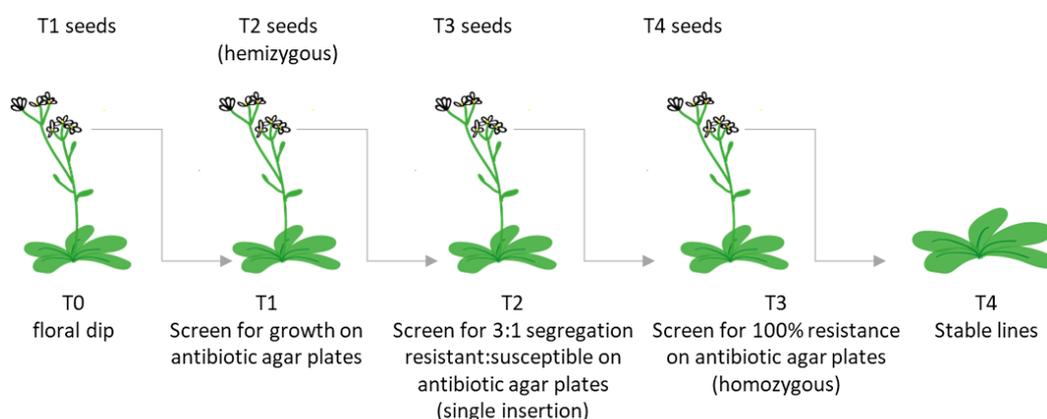


Figure 2.1 Schematic of the generation of Arabidopsis transgenic lines. This scheme summarises the selection process of single insertion, homozygous transgenic Arabidopsis plants, and the nomenclature used in this work.

2.10.2 Seed sterilisation

Prior to sowing on selection media, seeds were sterilised either using ethanol or bleach. The ethanol sterilisation was used for the T1 screen, when seeds were evenly sowed across the plates, and the bleach wash was used for T2 and T3 screens when a set amount of seeds were plated to count for segregation.

For ethanol sterilisation for the T1 screen, Arabidopsis seeds were placed in a miniprep column from the QIAprep Spin Miniprep kit (used columns washed in 80% ethanol), washed twice in 70% ethanol by incubating then centrifuging at $11\ 000 \times g$ for 30 seconds, then washed once more in 100% ethanol. The seeds were dried overnight in a sterile laminar flow cabinet, then sown on MS agar plates (3% sucrose, 0.8 % agar).

For bleach sterilisation in T2 and T3 screens, Arabidopsis seeds were placed in a 1.5 mL tube, incubated with 50% ethanol for 1 minute, and the ethanol was removed. Then the seeds were incubated with a bleach solution (25% household bleach in water; 0.01% Triton-X-100) for 10 – 15 minutes, the bleach was removed, and seeds were washed three times with sterile distilled water for 2 minutes. Finally, the water was removed and seeds were resuspended in 0.2% MS agar, and plated on MS agar plates (3% sucrose, 0.8 % agar) using a pipette.

2.11 Generation of Arabidopsis lines expressing *TaSS4* and *TaMRC*

2.11.1 Crossing

To test whether the presence of both *TaSS4* and *TaMRC* transgenes could complement the *mrc* mutant in Arabidopsis, the T2 parental lines *ss4/TaSS4-YFP* and *mrc/TaMRC-HA* previously generated by David Seung were crossed, which could be either homozygous or heterozygous for the transgene. I generated lines with an *mrc* mutant background and expressing both the *TaSS4-YFP* and *TaMRC-HA* transgenes. These plants may have an *ss4* background or a wild type (WT) background for the *AtSS4* gene, but this does not matter for the purposes of the question whether presence of *TaSS4* and *TaMRC* can complement an *mrc* mutant background, as the presence of *TaSS4* can at least partially complement the *ss4* mutant (Figure 4.2).

F1 plants that contained both transgenes were generated from this cross. These plants would be at least heterozygous for the transgenes. Then plants were generated from selfing of these F1 lines, named the F2. The selfed progeny from the F2 plants were named F3 lines, which were the

ones used for light microscopy. These plants were selected by a first round of antibiotic resistance selection on MS agar plates (3% sucrose, 0.8% agar, kanamycin for *TaMRC*, phosphinothricin for *TaSS4*), and were then genotyped for the presence of *TaSS4*, *TaMRC* and *mrc* mutant background. This process is summarised in figure 2.2.

2.11.2 Genotyping

Richard Goram from the DNA extraction and genotyping platform at the John Innes Centre performed the DNA extraction from *Arabidopsis* leaves. A young leaf was harvested into a plastic 96-well box, and frozen at -20°C until use. 37 mL (per 96-well box) of extraction buffer (0.1M Tris-HCl pH 7.5; 0.05 EDTA pH 8.0; 1.25% sodium dodecyl sulfate (SDS)) was heated to 65°C, RNaseA (10 mg/mL, 100 µL per box) was added, and 333 µL of the extraction buffer solution was added to each tube. The box was sealed and shaken in a Genogrinder for 2 minutes at 1750 rpm, pulse-spun until 3000 rpm was reached, then incubated at 65°C for 45 minutes. The box was incubated at 4°C for 15 minutes to cool down to room temperature, and pulse-spun to 5000 – 6000 rpm, then 167 µL of 67M ammonium acetate was added to each tube. The box was sealed and shaken for 15 seconds, incubated at -20°C for 15 – 20 minutes, and centrifuged for 8 minutes at 6000 rpm. Then 400 µL of the supernatant was transferred to cold 1.2 mL storage plates with 240 µL of ice-cold isopropanol in each well. The plate was sealed and shaken for 15 seconds, then pulse-spun, and the DNA was precipitated for 15 – 20 minutes at -20°C. The sample was centrifuged for 12 minutes at 5200 rpm to pellet the DNA, and the pellet was washed in 350 µL of 70% ethanol. The plate was centrifuged for 12 minutes at 5200 rpm and the pellet was dried overnight at room temperature or for 30 minutes at 65°C. The DNA pellet was resuspended in 200 µL water.

For genotyping of the *TaSS4*(-YFP) and *TaMRC*(-HA) transgenes in the *mrc-3* T-DNA insertion line background, GoTaq G2 GreenMaster (Promega) was used. For *TaSS4* genotyping, primers DSB112 and Z1385_eYFP_seq_rev were used. For *TaMRC* genotyping, primers DSB174 and Z1901_T35S were used. For *mrc-3* genotyping, the SAIL LB1 primer was used, and either the primer pairs DSB352+DSB353 or JC80+JC81. For details on the primers used, see Appendix 1A.

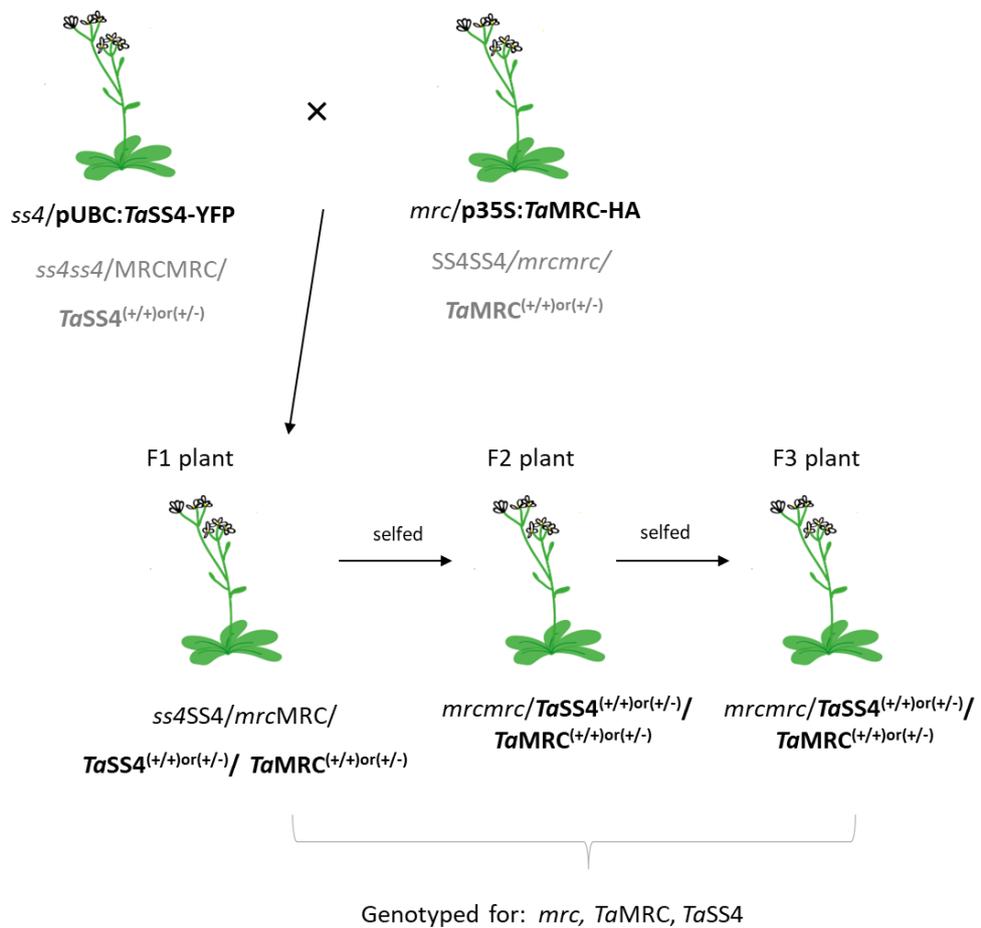


Figure 2.2 Schematic of the generation of Arabidopsis lines expressing *TaSS4* and *TaMRC*. This scheme summarises the cross of the *ss4/pUBC:TaSS4-YFP* and *mrc/p35S:TaMRC-HA* parent line and subsequent generation of selfed progeny. The plants were genotyped for the *mrc* mutant background and the *TaSS4* and *TaMRC* transgenes, to select plants that were at least heterozygous for the transgenes. It is not known whether the plants are homozygous (+/+) or heterozygous (+/-) for the *TaSS4* and *TaMRC* transgenes.

2.12 Biochemistry methods

2.12.1 Crude protein extraction from Arabidopsis leaves and wheat leaves

2.12.1.1 Verifying protein expression in transgenic lines using immunoblotting

Protein expression of individual transgenic Arabidopsis plants (*ss4/pUBC:TaSS4-YFP*, *mrc/p35S:TaMRC-HA*, *ss4/pAtSS4:AtSS4-HA*, *ss4/pAtSS4:AtSS4-eGFP-HA*, *mrc/pAtMRC:AtMRC-HA*, *mrc/pAtMRC:AtMRC-RFP-HA*) was confirmed by crude protein extraction from 7 mm leaf discs. These were ground in a ball mill with 3 – 4 glass beads (30/s, 1 min.), then mixed with 1 × Laemmli buffer and heated at 95°C for 2 minutes. The protein screen was done by running an immunoblot of these samples (see section 2.13.3) using the appropriate anti-HA or anti-GFP antibody, and secondary antibodies conjugated to an infra-red fluorescent dye (Table 2.7), visualised on the Odyssey Classic imaging system (LI-COR). Plants with strong protein expression were selected for further experiments.

2.12.1.2 For size exclusion chromatography

6-week-old Arabidopsis rosettes were harvested at midday by cutting at the base of the rosette. Wheat seedlings were harvested at midday, 10 days after germination, when they had 2 fully unfurled leaves and were cut at the base of the node of these two leaves. Both tissues were frozen in liquid nitrogen and kept at -80°C until use. The leaves were homogenised with glass homogenisers in extraction buffer (100 mM Tris-HCl; 5% v/v glycerol; 150 mM KCl; 5 mM DTT, 1× cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche)), adding 1 mL buffer for every 100 mg tissue. Homogenates were centrifuged at 18 000 × *g*, 4°C, for 10 minutes. 2 mL of the supernatant was filtered using 0.2 µm filters and concentrated using centrifugal filter devices with a 30 kDa molecular weight cut-off (Amicon) by centrifuging at 3160 × *g*, to a volume of approx. 250 µL.

2.12.2 Protein quantification

For protein quantification, the Bradford assay was done in a BMG Omega Plate Reader in 96-well plates, using 200 μL reactions measured at 595 nm. $4 \times 200 \mu\text{L}$ reactions were made for each sample in one tube and 3 technical replicates were measured from each of these. Each reaction contained 160 μL protein + water, and 40 μL Bradford reagent (Bio-Rad Protein Assay). The standard curve for the Bradford assay was made with bovine serum albumin (BSA) (Melford) (Table 2.5).

Table 2.5 Bradford assay standard curve preparation with BSA.

Final μg of BSA in 200 μL reaction	Amount of water to add (μL)	Amount of 100 $\mu\text{g}/\text{mL}$ BSA stock to add
0	640	0
0.5	620	20
1	600	40
1.5	580	60
2	560	80
2.5	540	100
3	520	120

Protein concentrations were quantified using Bradford assay and were confirmed to be in the same range before proceeding with size exclusion chromatography. For Arabidopsis, the three Col-0 replicates had 6 – 8 $\mu\text{g}/\mu\text{L}$ total protein, the *ss4* replicates had 8 – 12 $\mu\text{g}/\mu\text{L}$ total protein, and the *mrc* replicates had 11- 27 $\mu\text{g}/\mu\text{L}$ total protein. For wheat, the three Kronos wild type replicates had 6 – 9 $\mu\text{g}/\mu\text{L}$ total protein, the *ss4* replicates had 3 – 7 $\mu\text{g}/\mu\text{L}$ total protein.

2.12.3 SDS-PAGE and Western Blot

For sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), gels were either pre-made (10% or 7.5% Mini-PROTEAN TGX Precast Protein Gels, Bio-Rad) or self-made (Table 2.6). The gels were run using a Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell and gels were transferred onto nitrocellulose membranes using the Trans-Blot Turbo Transfer System (Bio Rad). The membranes were incubated with 5% milk in TBST (20 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) Tween 20) for 1 hour, then with the relevant primary antibody (see table 7.2) overnight at 4 °C with gentle shaking. The next day, the membrane was washed 3 \times with TBST, incubated with secondary antibody for 2 hours with gentle shaking, and washed 3 \times with TBST again. For chemiluminescence, SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fischer Scientific) was used and detected using an ImageQuant 500 or 800 (GE Healthcare Life Sciences).

For fluorescence detection, the membranes were additionally washed 2 × with TBS (20 mM Tris HCl pH 7.5, 150 mM NaCl), then detected using an Odyssey Classic imaging system (LI-COR).

Table 2.6 Self-made SDS-PAGE gel components.

	1 x 1 mm 10% Gel	1 x 1 mm 7.5 %Gel
RESOLVING GEL		
40% Glycerol	1.094 mL	1.094 mL
30% acrylamide (Severn or Bio Rad)	1.667 mL	1.25 mL
1.5 M Tris pH 8.8	1.25 mL	1.25 mL
Distilled water	0.99 mL	1.406 mL
20% SDS (Severn)	25 µL	25 µL
10% Ammonium Persulfate	40 µL	40 µL
Tetramethylethylenediamine (TEMED)	2.5 µL	2.5 µL
STACKING GEL		
30% acrylamide (Severn or Bio Rad))	0.25 mL	0.25 mL
1 M Tris pH 6	0.12 mL	0.12 mL
Distilled water	1.63 mL	1.63 mL
20% SDS (Severn)	10 µL	10 µL
10% Ammonium Persulfate	16 µL	16 µL
Tetramethylethylenediamine (TEMED)	1.5 µL	1.5 µL

Table 2.7 List of primary and secondary antibodies used. For use, antibodies were diluted in 5% milk TBST.

Antibody	Product name	Dilution	Host
Anti-HA	Abcam ab9110	1:5000	Rabbit
Anti-Strep	Abcam ab76949	1:2000	Rabbit
Anti-GFP	Torrey Pines TP401	1:5000	Rabbit
Anti-His	Abcam ab9110	1:5000	Mouse
Anti-Actin	Sigma A0480	1:10000	Mouse
Anti-rabbit (IR fluorescence)	Licor IRDye 800CW Donkey anti-Rabbit 926-32213	1:10000	Donkey
Anti-mouse (IR fluorescence)	Licor IRDye 680RD Donkey anti-Mouse 926-68072	1:10000	Donkey
Anti-rabbit	Sigma A0545 Anti-rabbit IgG (whole molecule)-Peroxidase	1:20000	Goat
Anti-mouse	Sigma A0168 Anti-mouse IgG (Fc Specific)-Peroxidase	1:10000	Goat

2.12.4 *In vitro* recombinant protein expression in *E. coli*

2.12.4.1 Solubility test

Plasmids encoding recombinant proteins (the constructs from section 2.8.3.2 – *AtSS4*, *TaSS4* and *AtMRC*) were transformed in BL21 (DE3) Δ glgAP cells (lacking the endogenous glycogen synthase, so as not to interfere with the starch synthesis proteins) using the heat shock method. 5 mL precultures were grown overnight, then 1 mL of the preculture was used to inoculate a 100 mL main culture. The growth of the main culture was monitored over 2 hours, until $OD_{600} = 0.5 - 0.7$.

1 mL of uninduced sample (U) was collected (pellet frozen at -20°C), then protein expression was induced by adding 1mM Isopropyl β - d-1-thiogalactopyranoside (IPTG). Cultures for each construct were incubated at both 18°C and 22°C overnight. The OD₆₀₀ was measured for each sample, and an induced sample (I) was taken for each, taking the same number of cells as the uninduced sample. Then the pellet was spun down at 5000 $\times g$, 10 min., 4°C. 10 mL lysis buffer (50 mM Tris-HCl, pH7.5; 300 mM NaCl; 40 mM Imidazole; 1 mg/mL lysozyme; 1 \times cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche); 2 mM dithiothreitol (DTT)) was added and samples were placed on ice for 30 min. The samples were sonicated, with the sonicator (Soniprep 150 Plus (MSE)) set at 30 sec. \times 2, amplitude 7.8 μ m with an exponential probe. 1 mL of crude extract sample (C) was taken. Then another 1 mL was taken and centrifuged at 20817 $\times g$, 10 min., 4°C. The supernatant was collected as the soluble fraction (S) and the pellet was resuspended in 1 mL lysis buffer, as the insoluble pellet fraction (P). Collected samples were analysed on an SDS-PAGE gel, first stained with Coomassie stain (InstantBlue, Abcam), and for specific detection of protein also immunoblotted with the relevant antibodies.

For SDS-PAGE gel loading for the Western blot, U and I fractions were resuspended in 150 μ L 1 \times Laemmli buffer, then further diluted 1 in 4 (with 1 \times Laemmli buffer), and 7.5 μ L was loaded. For the crude extract (C) and pellet (P), 10 \times Laemmli buffer was added, then samples were diluted 1 in 6 as they were very viscous and 7.5 μ L was loaded. For the soluble fraction (S), 10 \times Laemmli buffer was added, the sample was diluted 1:1, and then further diluted 1 in 4, with 7.5 μ L loaded.

2.12.4.2 Expression and purification

Glycerol stocks of the relevant bacterial strains were streaked out on LB agar plates with the correct antibiotic (for Strep-tagged proteins in BL21 (DE3) Δ glgAP: spectinomycin and kanamycin, for His-tagged proteins: ampicillin and kanamycin) and grown at 37°C overnight. A single colony from a plate or a swab from the glycerol stock itself was used to inoculate a 5 mL LB culture (for 500 mL cultures) or 100 mL culture (for 4 L cultures) with relevant antibiotics and grown at 37°C, at 200 rpm shaking overnight. The small culture was then used to inoculate a larger volume of LB culture with antibiotics - 5 mL into 500 mL or 20 mL into 1 L. For large scale purification, 4 \times 1 L cultures were grown for each construct. Cultures were grown at 37°C until OD₆₀₀ was approx. 0.6. Protein expression was induced by adding 1 mM of IPTG, and the cultures were grown at 18°C, shaking at 200 rpm overnight.

The following day, the cultures were centrifuged at 5000 $\times g$, 10 minutes at 4°C. The pellet was frozen at -80°C. For Strep-tag purifications, it was then resuspended in 30 mL (for 500 mL) or 50

mL (for 4 L) buffer W (100 mM Tris-HCl pH 8.0; 150 mM NaCl; 1 mM EDTA) with 1 mg/mL of lysozyme added. For His-tag purifications, the pellet was resuspended in 30 mL lysis buffer (50 mM Tris pH 7.5; 200 mM NaCl; 20 mM imidazole, 1 × cComplete™, EDTA-free Protease Inhibitor Cocktail (Roche), 2 mM DTT, 1 mg/mL lysozyme).

The sample was incubated on ice or at 4°C for 30 - 60 minutes, then sonicated on ice. For 500 mL cultures, the sonicator (Soniprep 150 Plus (MSE)) settings were: 3 × 30 s with 1 minute pause in between, amplitude 7.8 μm with an exponential probe. For 4 L cultures, a Vibra-Cell sonicator was used with settings: 40% amplitude, on 1 second off 3 seconds, on time 4 minutes. The samples were centrifuged at 20 000 × *g*, 10 - 15 minutes, at 4°C. The supernatant was kept on ice until further protein purification.

Strep-tagged proteins were purified using a Strep-Tactin Sepharose agarose resin (IBA Lifesciences) and His-tagged proteins were purified using a Ni-NTA agarose resin (Qiagen). All subsequent purification steps were performed at 4°C.

For the Strep-tag purifications, 1 mL Strep-Tactin resin bed volume was used for the 500 mL cultures and 3 mL Strep-Tactin resin bed volume was used for 4 L cultures. Strep-Tactin Sepharose 50 % suspension (IBA Lifesciences) was loaded on a Poly-Prep gravity flow column (Bio Rad) or self-made column using a syringe plugged with Miracloth, using 2 × the volume of the resin bed. If re-using a column, the storage buffer was washed out using 2 × 4 column volumes (CVs) of buffer W. If using for the first time, the resin was left to settle for 0.5 – 1 hour. The column was equilibrated with 2 CVs buffer W. The cleared lysates were centrifuged at 20 000 × *g*, 5 minutes, at 4°C. The supernatant was then loaded onto the column. The column was washed 5 × with 1 CV buffer W, and the eluate was collected in 1 CV fractions. Then 6 × 0.5 CVs of buffer E (supplied by IBA kit – 100 mM Tris-HCl pH 8.0; 150 mM NaCl; 1 mM EDTA; 2.4 mM desthiobiotin) were run through the column and the eluate was collected in 0.5 CV fractions. Alternatively, batch purification was used for the incubation of the lysate with resin and wash steps (see below for His-tag purification steps). For regeneration of the column, it was washed 3 × with 5 CVs of buffer R (supplied by IBA kit – 100 mM Tris-HCl pH 8.0; 150 mM NaCl; 1mM EDTA; 1 mM HABA) and overlaid with 2 mL buffer W or buffer R for storage.

For the His-tag purifications, a batch purification method was used, where all incubation and wash steps were done by resuspending the contents in a 50 mL tube, centrifuging at 500 × *g* for 4 minutes, and removing the supernatant. First, 2 mL of the Ni-NTA agarose resin 50 % suspension (Qiagen) was washed with 10 mL distilled water, then with 10 mL lysis buffer. The cleared lysates were incubated with the resin for 60 minutes at 4°C, with gentle inverting. This suspension was then centrifuged at 500 × *g* for 4 minutes, the supernatant was removed, and the resin was moved to a 15 mL tube, washed five times with 5 mL Triton buffer (50 mM Tris pH 7.5; 200 mM

NaCl; 20 mM imidazole; 2 mM DTT; 0.05% (w/v) Triton X-100) and then washed five times with 5 mL Wash buffer (50 mM Tris pH 7.5; 200 mM NaCl; 20 mM imidazole; 2 mM DTT). The resin (with some Wash buffer) was transferred to a self-made column, and proteins were eluted with 3 × 1 mL Elution buffer 1 (50 mM Tris pH 7.5; 300 mM NaCl; 100 mM imidazole), 3 × 1 mL Elution buffer 2 (50 mM Tris pH 7.5; 300 mM NaCl; 250 mM imidazole), and 2 × 1 mL Elution buffer 3 (50 mM Tris pH 7.5; 300 mM NaCl; 500 mM imidazole).

For both Strep-tag and His-tag purifications, the elution fractions were run on an SDS-PAGE gel and stained with Coomassie stain (InstantBlue, Abcam) or immunoblotted with anti-Strep or anti-His antibody (see Table 2.7). Then fractions with a significant amount of tagged protein were pooled and concentrated using a 30 kDa cut-off centrifugal filter (Amicon).

2.12.5 *In vitro* starch synthase activity assay

For the recombinant *AtSS4* and *TaSS4* proteins, an adenosine diphosphate (ADP) /nicotinamide adenine diphosphate hydride (NADH) assay was performed on a BMG Omega Plate Reader at 37°C. This reaction is based on the conversion of phosphoenolpyruvate (PEP) to pyruvate, by pyruvate kinase (PK), while converting ADP to adenosine triphosphate (ATP), then the conversion of pyruvate to lactate by lactate dehydrogenase (LDH), while converting NADH to nicotinamide adenine dinucleotide (NAD⁺). The ADP in the first reaction is provided by the activity of starch synthase, released when adding adenosine diphosphoglucose (ADP-glucose) onto a glucan substrate. NADH absorbs light at 340 nm, so that the signal goes down as NADH is converted to NAD⁺.

Reaction mixtures of 100 µL total volume were made, with three technical replicates of each reaction. The average kcat was taken from the three replicates. Each reaction contained: 100 mM bicine (pH 8), 25 mM potassium acetate, 2 mM magnesium chloride, 5 mM DTT, 0.05% BSA, 1 mM PEP, 1.6 mM NADH, 2 units/100 µL of PK in ammonium sulphate 3.2M pH 6, 2 units/ 100 µL of LDH in ammonium sulphate 3.2 M pH 6, 1 µg of purified SS4 enzyme, 20 mg/mL of glucan substrate (maltotriose – Sigma), and distilled water up to 100 µL. An initial absorbance measurement was taken at 340 nm, then 6 mM of ADP-glucose was added, and the absorbance at 340 nm was measured for 50 minutes.

The absorbance was plotted against the time, and the slope of the curve before the depletion of the substrate was used as the rate of absorbance per second. An NADH standard curve was made in water, and the slope of this curve was used to calculate the amount of NADH in µmol turned

over per second per amount of SS4 enzyme added to the reaction. The k_{cat} was calculated from this number divided by the μmol amount of SS4 enzyme, resulting in the turnovers per second.

2.12.6 Size Exclusion Chromatography (SEC)

2.12.6.1 Plant material

Crude protein extracts were spun at maximum speed for 1 minute in a tabletop centrifuge, then 100 μL of sample was loaded onto a Superose 6 Increase 10/300 column (Cytiva) connected to an ÄKTExpress system. The column was equilibrated in filtered (0.2 μm filter) and degassed running buffer (100 mM Tris-HCl; 5% v/v glycerol; 150 mM KCl; 5 mM DTT). Samples were run at 0.25 $\mu\text{L}/\text{min}$, at 4°C, and collected in 1 mL fractions. The column was calibrated using the HMW standards kit (GE Healthcare Life Sciences), see Appendix 5.

Because the SEC heavily dilutes the proteins, fractions within and surrounding the range of the molecular weight standards were concentrated using acetone precipitation (see below). For Arabidopsis samples, the full 1 mL fraction was acetone precipitated, but for the wheat samples, 330 μL was precipitated. The samples were resuspended in 50 μL 1 × Laemmli buffer, heated at 95°C, then loaded on a 10% or 7.5 % SDS-PAGE gel (Mini-PROTEAN TGX, Bio-Rad). For SDS-PAGE, 20 μL of the resuspended sample was loaded.

SS4 was detected by immunoblotting with an anti-AtSS4 antibody (Roldan *et al.*, 2007) or anti-TaSS4 antibody (Hawkins *et al.*, 2021) as primary antibody, and a secondary antibody conjugated to an infra-red fluorescent dye (IR Dye 800 CW, donkey anti-rabbit) or a peroxidase (Sigma A0545). Protein bands were visualised using an Odyssey Classic imaging system (LI-COR) or ImageQuant 500 or 800.

2.12.6.2 Recombinant protein

SEC was performed for the recombinant Strep-tagged AtSS4 and TaSS4 proteins as for the plant samples described above, but the samples were run in buffer W (100 mM Tris-HCl pH 8.0; 150 mM NaCl; 1 mM EDTA), which was used for the Strep-tag purification of recombinant proteins. To check for calibration in the new buffer, the thyroglobulin standard from the HMW standards kit (GE) was run with buffer W on the Superose 6 Increase 10/300 column. It eluted in the same

fraction as when calibrated in the buffer in section 2.11.6.1 (with the peak at 13.3 mL), so the column calibration from above was still used as a reference for these samples.

2.12.7 Acetone precipitation

Either the full 1 mL fraction or 330 μ L aliquots of protein sample from an SEC fraction were taken and 4 \times sample volume of cold acetone (cooled to -20°C) was added. Then the sample was vortexed and incubated for 60 minutes at -20°C , centrifuged for 10 minutes at 13 000 – 15 000 $\times g$, and the supernatant was decanted. 0.5 mL of cold acetone was added, the sample was vortexed, and centrifuged at 13 000 – 15 000 $\times g$ for 10 minutes. The acetone was removed with a pipette, and the sample was left to dry at room temperature for 30 minutes. Then the sample was resuspended in 50 μ L 1 \times Laemmli buffer, heated for 5 minutes at 95°C and stored at -20°C .

2.13 Protein-protein interaction analysis

2.13.1 Co-immunoprecipitation (co-IP) in *N. benthamiana*

Transiently transformed *N. benthamiana* plants (see section 2.9.4) were used for co-immunoprecipitation (co-IP) experiments. For *AtSS4* and *AtMRC* constructs, 8 \times 1 cm leaf discs (4 discs per leaf) were harvested from individual plants, and for *TaSS4* and *TaMRC* constructs, 4 \times 1 cm leaf discs (2 per leaf) were harvested. These were homogenised using a glass homogeniser with 2.6 mL extraction buffer (50 mM Tris-HCl pH 8; 150 mM NaCl; 1% Triton X-100; 1 \times cComplete™, EDTA-free Protease Inhibitor Cocktail (Roche); 1 mM DTT). The homogenate was spun at 20,000 $\times g$ for 5 minutes, 4°C , and the supernatant was transferred into a new tube and spin was repeated. For the input sample, 90 μ L of the supernatant was mixed with 10 μ L of 10 \times Laemmli buffer and heated at 95°C for 5 minutes. 1.2 mL of the supernatant was incubated with 50 μ L μ MACS HA magnetic beads (Miltenyi Biotec) at 4°C for 1 hour. The beads were captured on a μ Column (Miltenyi Biotec) on a magnetic stand, washed five times with wash buffer (50mM Tris-HCl pH 8; 300 mM NaCl; 1% Triton X-100; 1 \times cComplete™, EDTA-free Protease Inhibitor Cocktail (Roche), 1 mM DTT), then 20 μ L of heated elution buffer (μ MACS kit, 85°C) was added to the column, incubated for 5 minutes, another 80 μ L was added and eluate was collected. For the Arabidopsis constructs, 50 μ L of the prepared input sample and 10 μ L of the immunoprecipitated

sample (IP) was loaded on an SDS-PAGE gel and checked with an immunoblot. For the wheat constructs, 30 μ L of both the input and IP were loaded.

2.13.2 Yeast two-hybrid

The Gal4-based yeast two-hybrid assay (Fields & Song, 1989) was performed using the yeast (*Saccharomyces cerevisiae*) strain AH109. The lithium acetate method for transformation (Gietz & Woods, 2001) was used. AH109 yeast cells were grown on an YPAD agar plate at 30°C, then a single colony from this plate was used to inoculate a 5 mL YPAD culture and incubated at 30 °C overnight. 1 mL of this culture was used to inoculate each of 2 \times 100 mL YPAD cultures and the cultures were grown until the OD₆₀₀ was at 0.6 – 0.8. The cultures were centrifuged at 3214 \times g for 5 minutes, and resuspended in 25 mL total sterile water. This suspension was centrifuged at 1000 \times g for 5 minutes and the pellet was resuspended in 3 mL of sterile 100 mM lithium acetate (LiOAc). Approx. 200 ng of each DNA construct was added in the relevant combinations in separate tubes, then 281 μ L of transformation mixture (240 μ L 50 % PEG, 36 μ L 1 M LiOAc, 5 μ L carrier DNA (Takara Bio)) was added to each combination. 100 μ L of yeast cells was added to each transformation reaction, the mixture was vortexed for one minute, incubated at 30°C for 30 minutes, then incubated at 42°C for 30 minutes, and centrifuged for 5 minutes at 700 \times g. The pellet was resuspended in 100 μ L of 0.9% NaCl and each transformation was plated on double dropout (yeast medium without leucine and tryptophan (-LW), Y0750, Merck) yeast selection plates with SD+ agar, and incubated at 30 °C for 2 – 3 days.

A clump of yeast cells from successful transformations was used to inoculate 5 mL SD+ -LW cultures, and the cultures were incubated at 28 – 30°C at 200 rpm overnight. 2 mL of each culture was centrifuged at 10 621 \times g for 5 minutes, and the pellets were resuspended in 1 mL sterile water. Serial dilutions of these cultures were made: no dilution, 1 in 10, 1 in 100, and 1 in 500. 10 μ L of each dilution for each transformation was spotted on both a -LW SD+ agar plate and a quadruple dropout SD+ agar plate (-LWHA: medium without leucine, tryptophan, histidine, adenine, Y2021, Merck). These plates were left to dry in the laminar flow cabinet, and incubated at 30°C, checking for yeast growth up to 10 days from incubation.

2.14 Affinity purification Mass spectrometry

2.14.1 Affinity purification from Arabidopsis

Arabidopsis rosettes were harvested at midday and frozen in liquid nitrogen, then stored at -80°C. For three biological replicates of sample and control, three independent lines of *ss4/pAtSS4:AtSS4-HA* transgenic lines were used (with two rosettes per sample for lines 18-6 and 42-1 and one rosette for line 1-2) and three Col-0 plants. For each 200 mg of tissue, 1 mL of extraction buffer (50mM Tris-HCl pH 8; 150mM NaCl; 1% Triton X-100; 1 × protease inhibitor (Roche); 1mM DTT) was added, then leaves were crushed using a glass homogeniser. The purification was done as for the co-immunoprecipitation in *N.benthamiana* (section 2.13.1), but the columns were washed with extraction buffer rather than wash buffer. The enrichment of the target protein in the eluate was verified by immunoblotting as described above.

2.14.2 Preparation of samples for mass spectrometry

The remaining IP samples were run on a 10% resolving SDS-PAGE gel, just until all sample was run into the gel. Gel slices of the bands were cut out (approx. 5 × 10 mm) and cut into 3 – 4 slices. Gel slices were prepared according to standard procedures adapted from (Shevchenko *et al*, 2006). Briefly, the slices were washed with 50 mM TEAB buffer pH 8 (Sigma), incubated with 10 mM DTT for 30 min at 65°C followed by incubation with 30 mM iodoacetamide (IAA) at room temperature (both in 50 mM TEAB).

From here onwards, the samples were processed by Gerhard Saalbach and Carlo Martins from the JIC proteomics team. After washing and dehydration with acetonitrile, the gels were soaked with 50 mM TEAB containing 10 ng/μL Sequencing Grade Trypsin (Promega) and incubated at 40°C for 8 hours. The extracted peptide solution was dried down, and the peptides dissolved in 0.1% TFA/3% acetonitrile. Aliquots were analysed by nanoLC-MS/MS on an Orbitrap Eclipse™ Tribrid™ mass spectrometer coupled to an UltiMate® 3000 RSLCnano LC system (Thermo Fisher Scientific, Hemel Hempstead, UK). The samples were loaded and trapped using a pre-column with 0.1% TFA at 15 μL min⁻¹ for 4 min. The trap column was then switched in-line with the analytical column (nanoEase M/Z column, HSS C18 T3, 100 Å, 1.8 μm; Waters, Wilmslow, UK) for separation using the following gradient of solvents: A (water, 0.1% formic acid) and B (80% acetonitrile, 0.1% formic acid) at a flow rate of 0.2 μL min⁻¹ : 0-3 min. 3% B; 3-10 min. increase B to 7% (curve 4); 10-70 min. linear increase B to 37%; 70-90 min. linear increase B to 55%; followed by a ramp to 99%

B and re-equilibration to 3% B. Data were acquired with the following mass spectrometer settings in positive ion mode: MS1/OT: resolution 120K, profile mode, mass range m/z 300-1800, AGC 4e5, fill time 50 ms; MS2/IT: data dependent analysis with the following parameters: 2 s cycle time in IT turbo mode, centroid mode, isolation window 1 Da, charge states 2-5, threshold 1.9e4, HCD CE = 33, AGC target 1e4, max. inject time Auto, dynamic exclusion 1 count, 15 s exclusion, exclusion mass window ± 10 ppm.

2.14.3 Mass spectrometry data analysis

The raw data was processed in Proteome Discoverer 3.0 (Thermo Scientific, Waltham, USA), by Gerhard Saalbach and Carlo Martins. Spectra were recalibrated and filtered for top 20 peaks per 100 Da. Identification was performed using the Chimerys search node with the inferys_2.1_fragmentation prediction model (MSAID, Munich, Germany) on the TAIR10_pep_20101214 *Arabidopsis thaliana* protein sequence database (arabidopsis.org, 35,386 entries) plus the Maxquant contaminants database (245 entries). Parameters were enzyme trypsin, 2 missed cleavages, oxidation (M) as variable and carbamidomethylation (C) as fixed modification, 0.6 Da fragment tolerance. Evaluation was performed using Percolator. For peak detection and quantification, the Minora Feature Detector was used with a min. trace length of 7 and S/N 3. After normalisation to total peptide amount the quantification was based on the top 3 unique peptides per protein. Missing values were imputed by low abundance resampling. For hypothesis testing a background-based t-test was applied. Maximum fold change was set to 100. Results were exported to Microsoft Excel and filtered to remove contaminants and proteins with less than 2 unique peptides.

2.15 Statistics and data visualisation

All statistical analyses were done in R version 4. Overall, I used the emmeans() function from the 'emmeans' package throughout for pairwise comparisons. For most experiments, I used linear models using the lm() function from the 'stats' package and did one-way ANOVAs with Tukey post-hoc tests, unless specified otherwise. The 'ggplot2' package was used for data visualisation.

Chapter 3 – The distinct roles of MRC in wheat leaves and endosperm

This chapter contains elements of my two first-author papers (available as preprints at the time of writing), (Chen *et al.*, 2022a) and (Chen *et al.*, 2022b).

3.1 Introduction

We are only beginning to understand how starch granule formation is initiated, and the factors underpinning the vast diversity in granule initiation patterns observed between different organs and species (Chen *et al.*, 2021; Seung & Smith, 2019; Tetlow & Emes, 2017). A prime example of diverse granule initiation patterns between species can be observed in the seed endosperms of grasses (Matsushima *et al.*, 2013). Grass species of the Triticeae, including important cereal crops such as wheat, barley and rye, have a unique bimodal size distribution of starch granules in the grain endosperm – containing large, flattened A-type granules (20-30 µm in diameter) and small, round B-type granules (2-7 µm in diameter) (Howard *et al.*, 2011). The initiation of these two different types of granules is both spatially and temporally separated, with A-type granules initiating as early as 4 days post anthesis (dpa) in the main amyloplast compartment, and B-type granules initiating 10 – 15 days after that in stromules (Bechtel, 1990; Howard *et al.*, 2011; Langeveld *et al.*, 2000; Parker, 1985).

Recent work has identified proteins important for determining the initiation of bimodal starch granules in wheat. These include BGC1 in wheat/*Aegilops* - which is orthologous to FLO6 in barley and rice and PTST2 in *Arabidopsis* and *Brachypodium* (Chia *et al.*, 2020; Peng *et al.*, 2014; Saito *et al.*, 2017; Seung *et al.*, 2017; Watson-Lazowski *et al.*, 2022). BGC1 in wheat has a dose-dependent effect on granule initiation, where partial reductions in gene dosage can almost eliminate B-type granules without affecting A-type granule formation, whereas complete loss of function also causes defective A-type granule formation, including the formation of some compound/semi-compound granules that arise from multiple initiations (Chia *et al.*, 2020; Howard *et al.*, 2011; Saccomanno *et al.*, 2022).

SS4 is also required for normal A-type granule formation. In its absence, compound granules form in place of most A-type granules (Hawkins *et al.*, 2021). The increased number of initiations per amyloplast that led to compound granule formation in these mutants was unexpected since both SS4 and PTST2 promote granule initiation in *Arabidopsis* leaves, and mutants lacking either

protein have reduced numbers of starch granules per chloroplast (Roldan *et al.*, 2007; Seung *et al.*, 2017). These observations suggest that the proteins involved in granule initiation are to some extent conserved between species and organs, but they can act differently depending on the patterns of granule initiation in the species/tissue.

In this chapter, I explored the function of MRC (also known as PII1) in wheat, in both chloroplasts of leaves and amyloplasts of the endosperm. MRC was identified as a non-enzymatic protein that promotes starch granule initiation in Arabidopsis leaves (Seung *et al.*, 2018; Vandromme *et al.*, 2019). It is predicted to have coiled coils and is a conserved protein in land plants (Seung *et al.*, 2018). Consistent with previous work in Arabidopsis, I found that MRC promotes starch granule initiation in wheat leaves, but MRC had an unexpected, distinct role in the temporal control of B-type granule initiation in the wheat endosperm. MRC is expressed at the early stages of grain development, and wheat *mrc* mutants have severe alterations in the starch granule size distribution relative to the wild type, with smaller A-type granules and a higher relative volume of B-type granules. This phenotype arises from the early initiation of B-type granules in the mutant, suggesting MRC represses B-type granule initiation during early grain development. This distinct role of MRC in the wheat endosperm demonstrates how the function of granule initiation proteins can be adapted to mediate specific patterns of granule initiation among different species/tissues.

3.2 Results

3.2.1 The wheat orthologs of MRC are encoded on chromosomes 6A and 6D

The starch granule initiation protein MRC is highly conserved among land plants (Seung *et al.*, 2018). To determine the role of MRC in wheat, David Seung searched the wheat genome for genes encoding MRC orthologs. He ran a BLASTp search using the amino acid sequence of Arabidopsis MRC (*AtMRC*, At4g32190) against the protein sequences from both tetraploid durum wheat (Svevo v1.1) (Maccaferri *et al.*, 2019) and hexaploid bread wheat (IWGSG Chinese Spring) (International Wheat Genome Sequencing, 2018) genomes on Ensembl Plants. For hexaploid wheat (*Triticum aestivum*), the two top protein hits were TraesCS6A02G180500.1 (encoded on chromosome 6A) and TraesCS6D02G164600.1 (encoded on chromosome 6D), which shared 95% sequence identity with each other and were predicted as homeologs on Ensembl. Both genes had a two-exon structure like the Arabidopsis gene (Seung *et al.*, 2018), and were in syntenic positions on the A and D genomes (Figure 3.1).

For tetraploid durum wheat (*Triticum turgidum*), the top protein hit was TRITD6Av1G081580.1 (encoded on chromosome 6A), which was identical in nucleotide and amino acid sequence to TraesCS6A02G180500.1. To determine whether these proteins were true orthologs of *AtMRC*, David Seung repeated the phylogenetic analyses of MRC homologs from a previous study (Seung *et al.*, 2018) with the wheat protein sequences included. The 6A and 6D proteins clustered together on the tree, distinctly within the grass clade containing the rice and maize sequences (Figure 3.2A). This confirms that the proteins are the wheat orthologs of MRC. They will hereafter be referred to as *TaMRC-A1* (TraesCS6A02G180500) or *TtMRC-A1* (TRITD6Av1G081580), and *TaMRC-D1* (TraesCS6D02G164600). The *AtMRC* and *TaMRC-A1* sequences shared 31.6% sequence identity and 55.8% sequence similarity (pairwise amino acid sequence alignment in EMBOSS-NEEDLE), and the two MRC orthologs shared coiled-coil predictions across the entire sequence (as predicted by PCOILS), which is discussed in more detail in Chapter 4. Notably, there was no full gene model for MRC on chromosome 6B, or anywhere else on the B genome, either in the durum or bread wheat genome.

I then worked collaboratively with Andy Chen to investigate the absence of *MRC* on the B genome. In the syntenic region of chromosome 6B in Chinese Spring, there was a stretch of sequence that had homology to exon 2 and the beginning of the 3' UTR (untranslated region) (Figure 3.1A). Interestingly, around 14 kb downstream of that, there was a region highly similar to the end of the 3' UTR of *TaMRC-A1*. There was a complete *gypsy* retrotransposon annotated between the two 3' UTR fragments in the transposable element annotation of the wheat

reference genome around the exon 2 fragment (Daron *et al*, 2014) (Figure 3.1B). Further, we identified the 5 bp target site duplication (GAGAT, which is part of the 3'UTR) and the inverted terminal repeat (TGTA and TTACA at the start and end of the retrotransposon, respectively) characteristic of retrotransposon insertions. The distance between the 5' end of the exon 2 fragment and its upstream neighbouring gene (TraesCS6B02G205000, a respiratory burst oxidase homolog) was much larger than the distance between *TaMRC* and the homeologs of the same neighbouring gene on 6A and 6D, indicating a large insertion in this 6B region (Figure. 3.1A). Indeed, a fragment of another *gypsy* retrotransposon was found ca. 16 kb upstream of the exon 2 fragment.

Additionally, there was a sequence with homology to the 5' UTR of *TaMRC-A1* (85% identity over 243 bp) just 3 bp upstream of the exon 2 fragment, which suggested that a ~1.3 kbp deletion (based on A-genome distances) removed some of the 5' UTR, all of exon 1, intron 1 and the start of exon 2 of *TaMRC-B1*. Similar to Chinese Spring, there were identical disruptions in *TaMRC-B1* sequences with retrotransposon insertions and deletions in ten additional wheat genome assemblies (Walkowiak *et al*, 2020). Overall, it appears that a deletion and a series of transposon insertions severely disrupted *MRC* on chromosome 6B in bread wheat.

Since a B-genome copy was also absent from durum wheat, it is likely that the disruption of *MRC* on chromosome 6B preceded the second hybridisation that resulted in hexaploid wheat. To further investigate when the disruption of *MRC-B1* occurred, we looked for *MRC-B1* in more tetraploid wheat accessions by aligning genome sequencing reads from *Triticum dicoccoides* (wild emmer) (n=10) and *Triticum turgidum* ssp. *durum* (pasta wheat) (n=12) against the A and B genomes of Chinese Spring (Zhou *et al.*, 2020).

The exon 1 deletion and the retrotransposon insertion at the 3' end were detected in all accessions (except for a few lines that had poor sequencing depth in the region), suggesting that *TaMRC* on 6B was disrupted before or immediately after the hybridization of diploid ancestors carrying the A and B genomes. To distinguish these possibilities, we examined *MRC* in *Aegilops speltoides*, the diploid species thought to be most closely related to the progenitor species of the wheat B genome. We ran a BLASTn search on the genome assembly of the *Aegilops speltoides* accession TS01 using the coding sequence of *TaMRC-A1* (Li *et al*, 2022) (Appendix 4). The top hit was on chromosome 6S (homologous to chromosome 6B in wheat) with intact exons one and two (97.5% sequence identity), and the translated protein sequence had 96% identity to *TaMRC-A1* with BLASTp. Thus, it is likely that *MRC* is intact in *Aegilops speltoides*, suggesting that the loss of the B-homeolog occurred shortly after the hybridisation that gave rise to tetraploid wheat (Figure 3.1C). It is therefore expected that all tetraploid wheats have one *MRC* homeolog (on chromosome 6A), and all hexaploids have two (on chromosomes 6A and 6D).

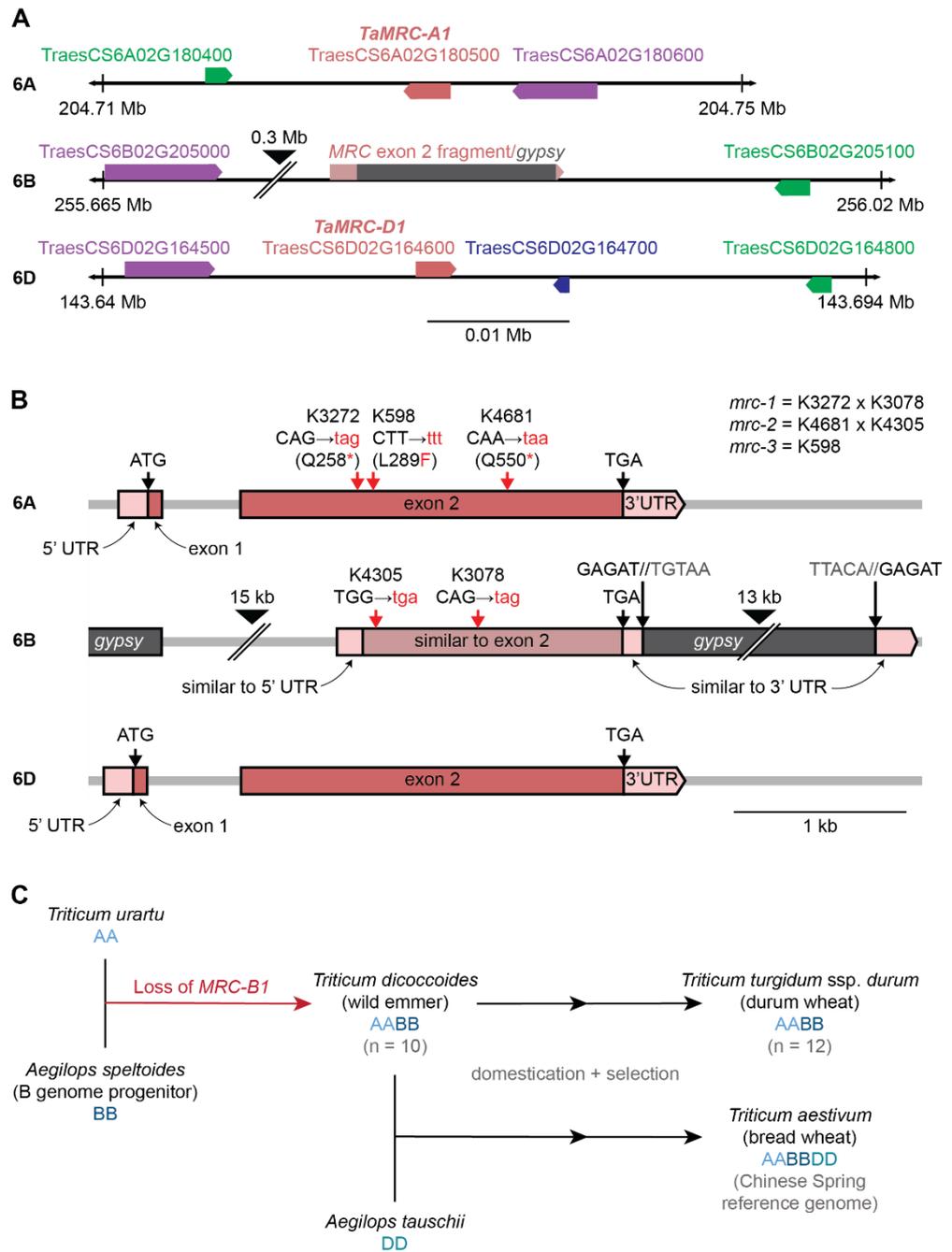


Figure 3.1. *MRC* homeologs in wheat are encoded on chromosomes 6A and 6D, with a disruption of the 6B homeolog. A) Location of *TaMRC* homeologs on chromosome 6A and 6D. The pink boxes represent *TaMRC* homeologs, while homeologs of the adjacent genes are shown in green (cytochrome P450 family protein), purple (respiratory burst oxidase homolog) and blue (uncharacterised protein). Arrowheads on the boxes indicate direction of transcription. The syntenic region on chromosome 6B has a large insertion, depicted with a black arrowhead. The diagram is drawn to scale, and chromosome coordinates of the region are indicated. **B)** Gene models of the *TaMRC-A1* and *-D1* homeologs and 6B pseudogene. Exons are represented with pink boxes, while light pink boxes represent the 5' and 3' UTRs. On the 6B region, areas with sequence similarity to exon 2 and UTRs of *TaMRC-A1* are indicated, as well as the location of *gypsy* retrotransposons. The locations of the mutations in the *mrc* mutants are depicted with red arrows, and the mutated codons/amino acids are shown in red letters. Large black arrowheads show where the sequence has been truncated for illustration - the length of truncated sequence is indicated above. **C)** Summary of species analysed for the loss of *MRC-B1* during wheat hybridisation. This figure appears in (Chen *et al.*, 2022a). Panels A and B were made by David Seung.

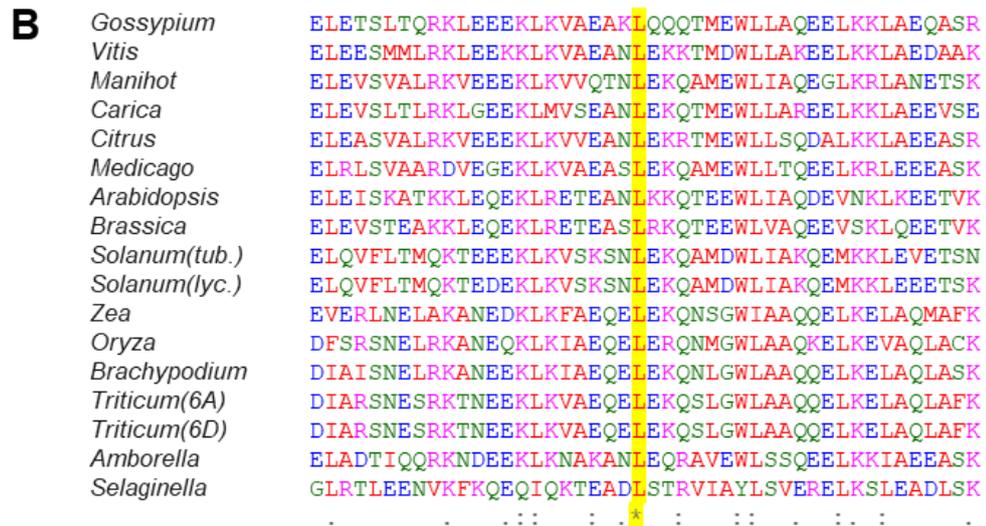
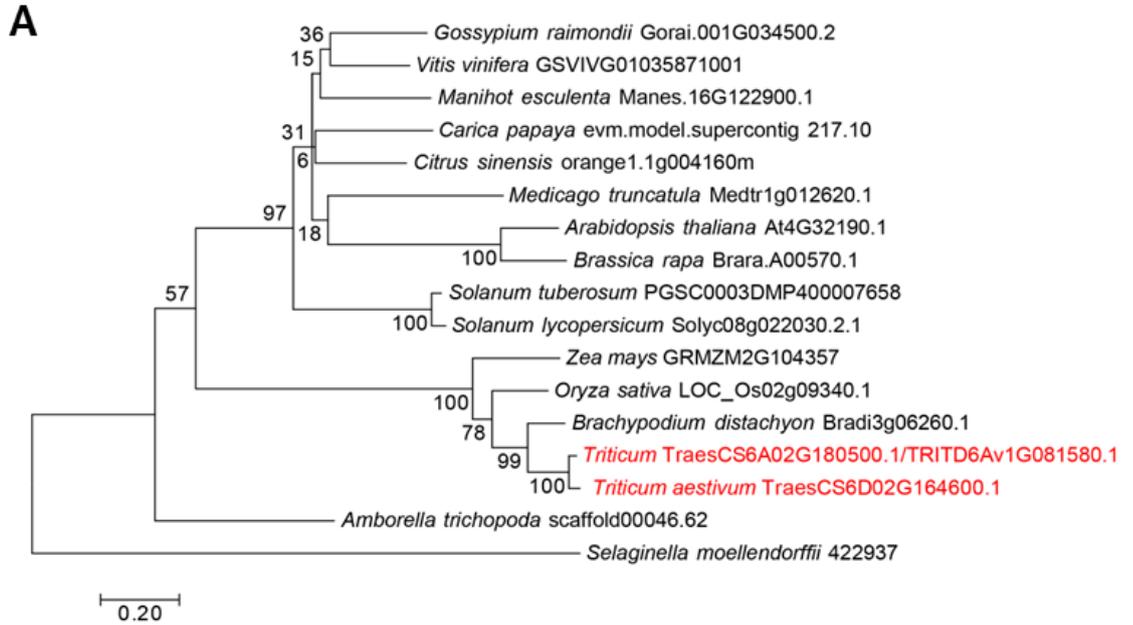


Figure 3.2. Phylogenetic analysis of wheat MRC sequences. **A)** Phylogenetic tree of MRC orthologs. The amino acid sequences of the orthologs identified in Seung *et al.* 2018 were aligned with the wheat sequences. A maximum likelihood tree was assembled with 1000 bootstraps. Bootstrap values are shown next to each node. Branch lengths indicate the number of substitutions per site. **B)** Multiple sequence alignment of MRC orthologs generated with Clustal Omega. The region surrounding Leu289 (highlighted in yellow) is shown. Symbols under the alignment indicate conserved residues with: complete identity (*), highly similar properties (:), or weakly similar properties (.). Colours of residues represent side chain properties: hydrophobic (red), acidic (blue), basic (magenta) and other (green). This figure appears in (Chen *et al.*, 2022a) and was made by David Seung.

3.2.2 Expression analysis during early grain development reveals distinct patterns of starch granule initiation genes

I investigated the expression patterns of *TtMRC* in durum wheat using RNA sequencing (RNA-seq) data. The wheat expression browser (Borrill *et al.*, 2016; Ramirez-Gonzalez *et al.*, 2018) contained expression data for hexaploid bread wheat, where transcripts of the 6A and 6D homeologs were present in both leaves and grains, suggesting that MRC plays a role in both these tissues (Figure 3.3A). However, the expression browser did not contain detailed information regarding gene expression in the starchy endosperm during grain development. I therefore created my own RNA-seq dataset on dissected endosperms of durum wheat (*Triticum turgidum* cv. Kronos) throughout grain development. The Kronos variety was chosen as it is the genetic background of the tetraploid wheat TILLING mutants characterised in this chapter.

I harvested developing grains from plants grown in controlled environment chambers - starting from the earliest timepoint in which the starchy endosperm could be reliably dissected from the pericarp (6 dpa) and until the grain turned yellow, indicating maturation (30 dpa) (Figure 3.3B). RNA was extracted from endosperm tissue dissected from the grains, with three biological replicates per timepoint, with each replicate consisting of multiple grains (3-6 depending on the timepoint) harvested from an independent plant. The RNA-seq using Illumina technology was outsourced to Novogene (Cambridge), and initial processing and analyses of raw data were done with the support of Martin Vickers from the JIC Informatics Platform and Alexander Watson-Lazowski. Their analysis showed that the RNA-seq yielded between 78,762,223 and 169,888,168 reads per sample. Pseudoalignment of reads using *kallisto* (Bray *et al.*, 2016) to the annotated transcripts from the Svevo v1 durum wheat transcriptome (using Ensembl canonical transcripts only) (Maccaferri *et al.*, 2019) resulted in 66.356-77.040% of reads per sample being successfully pseudoaligned. Transcripts per million (TPM) values for each detected transcript were calculated, and differentially expressed genes (DEGs) for all pairwise comparisons of the sampled timepoints were determined, using DESeq2 (FDR \leq 0.05 and \log_2 fold change $>$ 1.5).

Most of the granule initiation genes, including *TtMRC*, were identified as DEGs. *TtMRC-A1* (TRITD6Av1G081580.1) showed a peak of expression during early grain development (8 dpa), but strongly decreased in expression between 10 – 20 dpa (Figure 3.3C). It therefore appears that MRC is expressed in the wheat endosperm almost exclusively during early grain development. In contrast, BGC1 expression increased during grain development and is highly expressed at 15-20 dpa when B-type granules initiate (Figure 3.3C), which is consistent with its role in promoting B-type granule initiation during later grain development (Chia *et al.*, 2020). SS4 was primarily

expressed during early grain development (Figure 3.3C), consistent with a role in establishing correct granule number per amyloplast in early grain development (Hawkins *et al.*, 2021).

To assess how specific these expression patterns are to species that have A- and B-type granules, I examined the expression patterns of these genes in the endosperm of rice and maize using publicly available data. Rice produces compound granules, where granules are initiated during early grain development and there is no second wave of granule initiation in later grain development that leads to B-type granules; and maize produces simple granules without distinct populations of different granule sizes (Chen *et al.*, 2021). Among the granule initiation genes, BGC1 had the most different expression patterns when comparing durum wheat with rice and maize. Rather than peaking during mid-late grain development, expression of *OsFLO6* in rice (the rice ortholog of wheat BGC1) was highest at the beginning of endosperm development and then decreased, increasing again only slightly in late grain development (after 20 dpa) (Figure 3.3D, (Sato *et al.*, 2013)). Maize has two orthologs of BGC1, called PTST2a and PTST2b (named as the *Arabidopsis* genes). PTST2a expression does not change much during maize endosperm development, but PTST2b was also high during early endosperm development, dipping and then plateauing in later endosperm development (Figure 3.3E, (Qu *et al.*, 2016)). It is therefore tempting to speculate that change in the BGC1 expression pattern is one of the key factors that facilitates the formation of A- and B-type granules.

Overall, the RNA-seq analyses showed that *TtMRC* is only expressed during early endosperm development, in an expression pattern that is opposite to that of *TtBGC1*. Durum wheat has a distinct BGC1 expression pattern compared to other cereals that have compound and simple starch granules rather than bimodal granules.

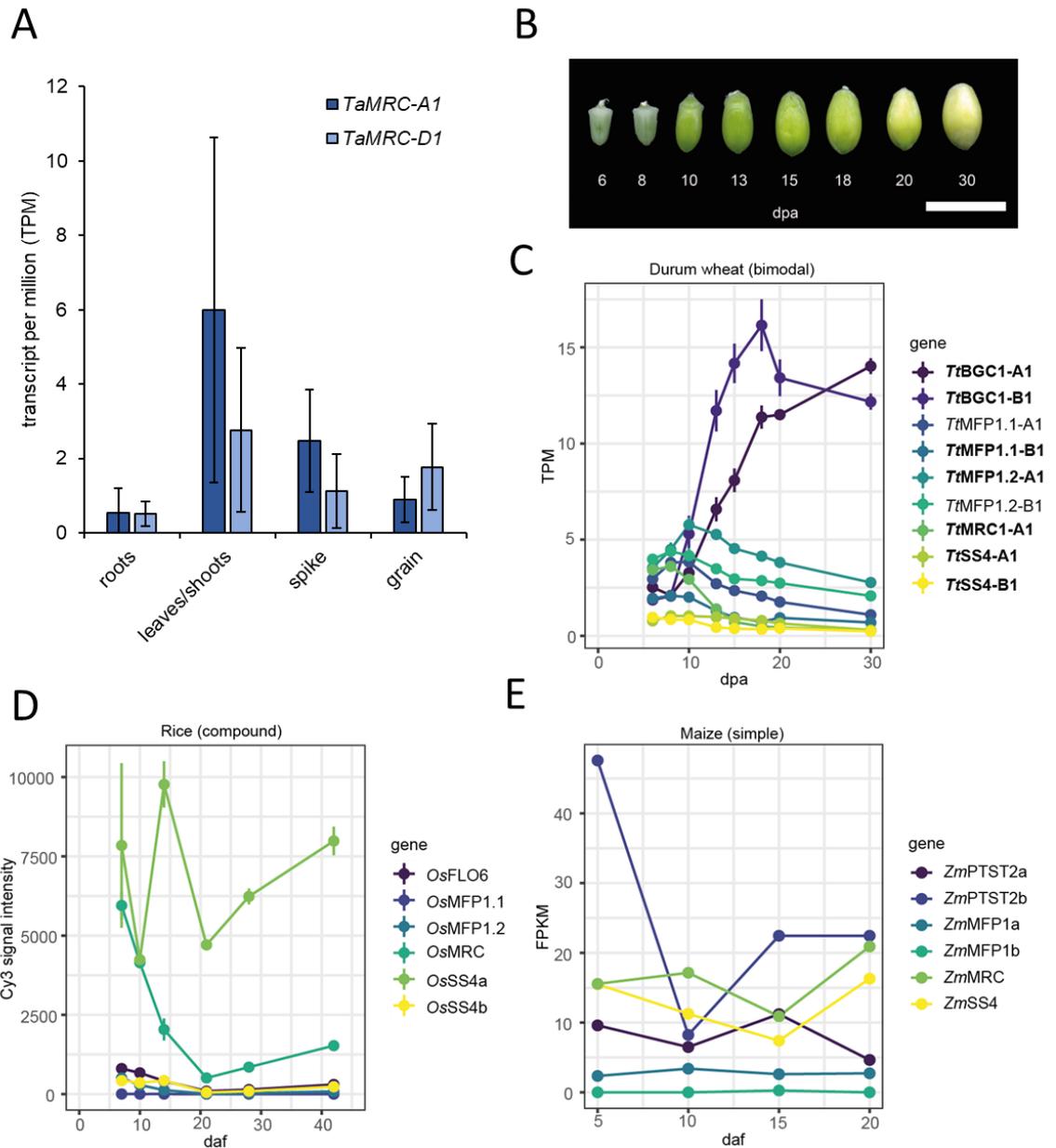


Figure 3.3. MRC is expressed in wheat leaves and endosperm during early grain development. A) Expression of *TaMRC* homeologs between different tissue types in *Triticum aestivum*. Data were obtained from the wheat expression browser, and values represent transcript per million (TPM) \pm standard error of the mean (SEM) from $n = 89$ (roots), 48 (leaves/shoots), 280 (spike), 166 (grain) samples from different experiments. **B)** Morphology of *Triticum turgidum* grains harvested at each timepoint for RNA-seq analysis. Bar = 1 cm, dpa = days post anthesis. **C)** Average TPM values ($n = 3$) of starch granule initiation genes in durum wheat endosperm. Genes in bold were identified as differentially expressed genes (DEGs). **D)** Average expression values ($n=3$) from microarray analysis of starch granule initiation genes from rice developing endosperm, extracted from a public database (<https://ricexpro.dna.affrc.go.jp/>) (Sato *et al*, 2013). FLO6/BGC1 = LOC_Os03g48170, MFP1.1 = LOC_Os05g08790, MFP1.2 = LOC_Os01g08510, MRC = LOC_Os02g09340, SS4a = LOC_Os01g52260, SS4b = LOC_Os05g45720 **E)** Expression values ($n=3$) from RNA sequencing analysis of starch granule initiation genes from maize developing endosperm, using data from (Qu *et al*, 2016). PTST2a = GRMZM2G104501, PTST2b = GRMZM2G122274, MFP1a = GRMZM2G106233, MFP1b = GRMZM2G142413, MRC = GRMZM2G104357, SS4 = GRMZM2G044744. For E, no data was available on the error values. dpa = days post anthesis. daf = days after flowering. Panels from this figure appear in (Chen *et al.*, 2022a) and (Chen *et al.*, 2022b).

3.2.3 Loss of MRC does not affect the growth of wheat plants, or grain development

To study the function of MRC in wheat, I characterised mutants of durum wheat (*Triticum turgidum* cv. Kronos) defective in MRC. These mutants were previously generated by Brendan Fahy, Cristobal Uauy and David Seung prior to the start of my project. They used the wheat *in silico* TILLING mutant resource, which has an EMS-mutagenised population of Kronos with exome-capture sequencing data for identification of lines with mutations of interest (Krasileva *et al.*, 2017). They obtained three mutants that were likely to cause a loss of function in *TtMRC-6A* (Figure 3.1B). The K3272 and K4681 lines contained premature stop codons in place of codons for the 258th and 550th amino acids respectively. In addition, there was a third line (K598) that contained a missense Leu289Phe mutation, which was predicted to be deleterious to protein function by SIFT scoring (Ng & Henikoff, 2006). The Leu289 residue is highly conserved in all MRC orthologs (Figure 3.2B).

Since the 6B homeolog of MRC has likely become a pseudogene, *TtMRC-A1* would likely be the only functional MRC homeolog in tetraploid wheat. However, to rule out the possibility that the fragment of exon 2 on chromosome 6B affects MRC function, we also obtained the K4305 and K3078 lines which contain two different premature stop codon mutations in the putative reading frame of the exon. The *mrc-1* lines were generated by crossing K3272 and K3078, and the *mrc-2* lines by crossing K4681 and K4305, where lines homozygous for either the 6A (F_2 aaBB) or 6B (F_2 AAbb) mutation were isolated, or both (F_2 aabb). The *mrc-1* double mutant line was backcrossed twice to wild type (WT), and the wild-type segregant (BC_2F_2 AABB) and the homozygous double mutant (BC_2F_2 aabb) were selected. These will be hereafter referred to as BC_2 AABB and BC_2 aabb. No backcrossing was done for the *mrc-2* line. The *mrc-3* line was the TILLING line K598 with the L289F missense mutation, and no crossing was done.

There was no consistent effect of the independent *mrc* mutations on plant growth or grain development under our growth conditions. None of the *mrc* mutant plants appeared different from WT with respect to growth or the number of tillers per plant (Figure 3.4A, B, Table 3.1). The number of grains per plant also did not differ in the mutants, except for a slight decrease in *mrc-3* and the wild-type segregant (*mrc-1* BC_2 AABB) compared to the WT (Figure 3.4C, Table 3.2). The morphology of the mature grains of the mutants was indistinguishable from the WT (Figure 3.4D), and there were no differences in thousand grain weight (TGW) and grain size between the WT and any of the three mutants *mrc-1*, *mrc-2*, *mrc-3*, or between the WT and the backcrossed *mrc-1* (*mrc-1* BC_2 aabb); but the wild-type segregant (*mrc-1* BC_2 AABB) had a slightly higher TGW and grain size compared to WT (Figure 3.4E, F, Table 3.3, 3.4). This suggests that some of the background mutations in the wild-type segregant may have affected grain development, but

these effects are small. The measurements in this section were done by David Seung, and I performed the data analysis.

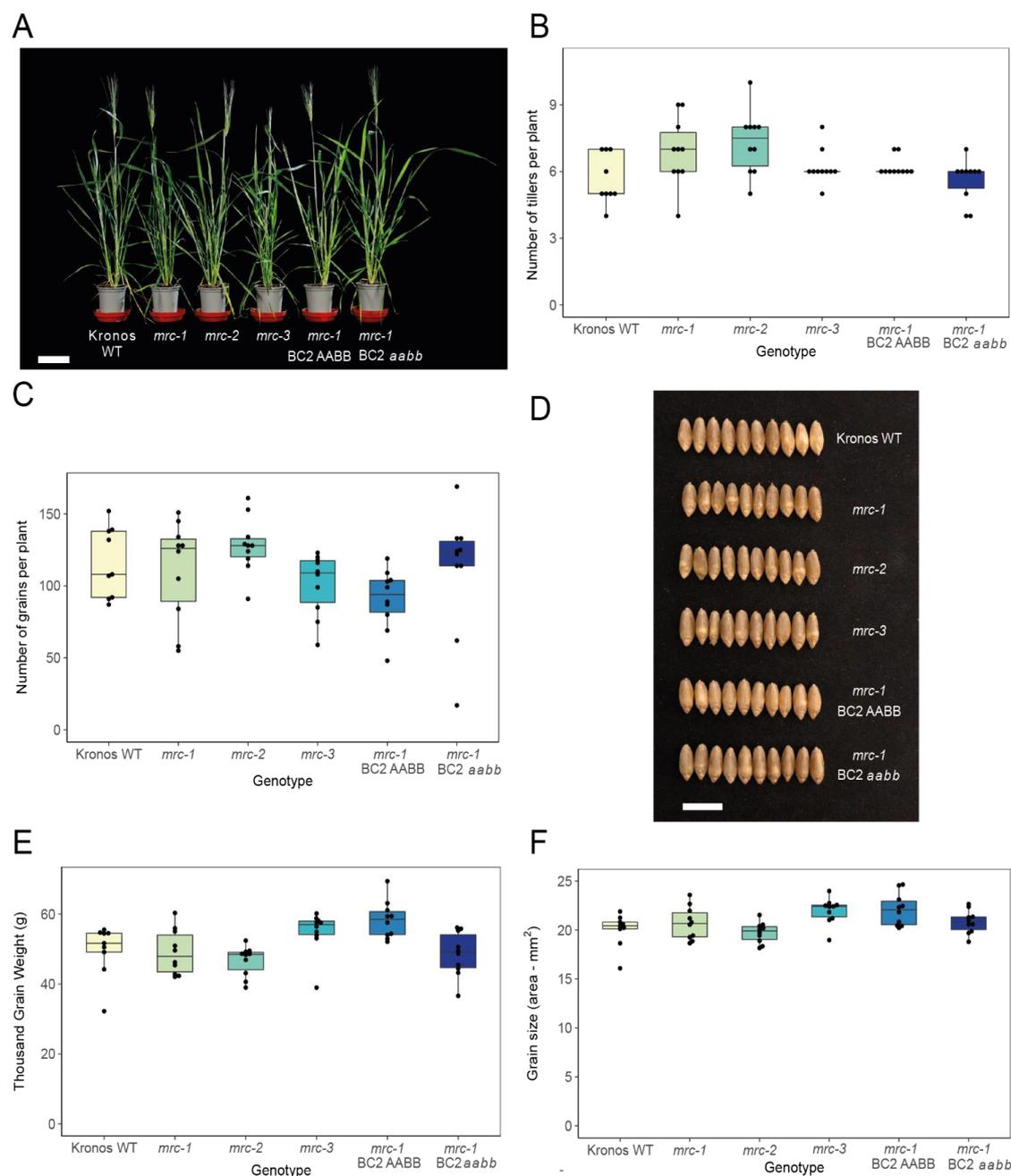


Figure 3.4 Mutations in *MRC* do not affect plant growth or the number and size of grains. A) Photographs of 7-week-old wild type (WT) and *mrc* mutant plants. Bar = 9 cm. **B)** Number of tillers per plant. **C)** Number of grains per plant. **D)** Photographs of mature grains. Bar = 1 cm. **E)** Thousand Grain Weight (TGW) of mature grains. **F)** Grain size, measured as the average area of individual mature grains (15 – 88 grains) from each plant. For statistical analysis of B, C, E, F, see Table 3.1 – 3.4. For all boxplots, each box encloses the middle 50% of the distribution, the middle line is the median and the whiskers are the minimum and maximum values within $1.5 \times$ the interquartile range. Dots are measurements from individual plants, with the same group of plants measured for all phenotypes and $n = 9 - 10$ individual plants. This figure appears in (Chen *et al.*, 2022a).

Table 3.1. Pairwise comparisons of number of tillers per plant. The table shows ratios of pairwise comparisons between genotypes, based on a Poisson regression model with post-hoc tests performed on the log scale and intervals back-transformed from the log scale. For the post-hoc tests, p-value and 95% confidence intervals were adjusted using Bonferroni method. Statistics were done in R, using the glm() function from the 'GLMMadaptive' package. All pairwise comparisons were done using the emmeans() function from the 'emmeans' package. SE = standard error, 95% CI = 95% confidence interval, df = degrees of freedom.

Tiller number per plant						
contrast	ratio	SE	z.ratio	p.value	- 95% CI	+ 95 CI
WT / <i>mrc-1</i>	0.82	0.152	-1.07	1	0.48	1.41
WT / <i>mrc-2</i>	0.78	0.142	-1.39	1	0.45	1.32
WT / <i>mrc-3</i>	0.91	0.173	-0.48	1	0.52	1.59
WT / <i>mrc-1</i> BC2 AAB <i>B</i>	0.91	0.173	-0.48	1	0.52	1.59
WT / <i>mrc-1</i> BC2 aab <i>b</i>	1.01	0.196	0.06	1	0.57	1.79
<i>mrc-1</i> / <i>mrc-2</i>	0.95	0.159	-0.34	1	0.58	1.55
<i>mrc-1</i> / <i>mrc-3</i>	1.11	0.195	0.61	1	0.67	1.86
<i>mrc-1</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1.11	0.195	0.61	1	0.67	1.86
<i>mrc-1</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.23	0.222	1.16	1	0.73	2.09
<i>mrc-2</i> / <i>mrc-3</i>	1.18	0.203	0.95	1	0.71	1.95
<i>mrc-2</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1.18	0.203	0.95	1	0.71	1.95
<i>mrc-2</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.30	0.232	1.49	1	0.77	2.20
<i>mrc-3</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1	0.180	1.36E-15	1	0.59	1.69
<i>mrc-3</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.11	0.204	0.55	1	0.64	1.90
<i>mrc-1</i> BC2 AAB <i>B</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.11	0.204	0.55	1	0.64	1.90

Table 3.2. Pairwise comparisons of number of grains per plant. The table shows ratios of pairwise comparisons between genotypes, based on a Poisson regression model with post-hoc tests performed on the log scale and intervals back-transformed from the log scale. For the post-hoc tests, p-value and 95% confidence intervals were adjusted using Bonferroni method. Statistics were done in R, using the glm() function from the 'GLMMadaptive' package. All pairwise comparisons were done using the emmeans() function from the 'emmeans' package. Comparisons where p<0.05 are highlighted in grey. SE = standard error, 95% CI = 95% confidence interval, df = degrees of freedom.

Grain number per plant						
contrast	ratio	SE	z.ratio	p.value	- 95% CI	+ 95 CI
WT / <i>mrc-1</i>	1.05	0.045	1.026	1	0.92	1.19
WT / <i>mrc-2</i>	0.91	0.038	-2.335	0.293151	0.80	1.03
WT / <i>mrc-3</i>	1.15	0.051	3.117	0.027373	1.01	1.31
WT / <i>mrc-1</i> BC2 AAB <i>B</i>	1.28	0.058	5.464	6.95E-07	1.12	1.46
WT / <i>mrc-1</i> BC2 aab <i>b</i>	1.04	0.045	1.005	1	0.92	1.18
<i>mrc-1</i> / <i>mrc-2</i>	0.87	0.036	-3.452	0.008351	0.77	0.98
<i>mrc-1</i> / <i>mrc-3</i>	1.10	0.048	2.147	0.477097	0.97	1.24
<i>mrc-1</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1.23	0.055	4.554	7.88E-05	1.08	1.40
<i>mrc-1</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.00	0.042	-0.021	1	0.88	1.13
<i>mrc-2</i> / <i>mrc-3</i>	1.26	0.053	5.583	3.55E-07	1.12	1.43
<i>mrc-2</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1.41	0.061	7.956	2.67E-14	1.24	1.60
<i>mrc-2</i> / <i>mrc-1</i> BC2 aab <i>b</i>	1.15	0.047	3.431	0.009028	1.02	1.30
<i>mrc-3</i> / <i>mrc-1</i> BC2 AAB <i>B</i>	1.12	0.051	2.418	0.23417	0.98	1.28
<i>mrc-3</i> / <i>mrc-1</i> BC2 aab <i>b</i>	0.91	0.040	-2.168	0.452384	0.80	1.03
<i>mrc-1</i> BC2 AAB <i>B</i> / <i>mrc-1</i> BC2 aab <i>b</i>	0.81	0.036	-4.575	7.13E-05	0.71	0.93

Table 3.3. Pairwise comparisons of Thousand Grain Weight (TGW). The TGW was measured using a MARVIN seed analyser. The table shows differences of pairwise comparisons between genotypes, based on a linear regression model and one-way ANOVA with Tukey post-hoc test. Statistics were done in R using the `lm()` function from the 'stats' package. All pairwise comparisons were done using the `emmeans()` function from the 'emmeans' package. Comparisons where $p < 0.05$ are highlighted in grey. SE = standard error, 95% CI = 95% confidence interval, df = degrees of freedom.

Thousand Grain Weight (g)							
contrast	difference	SE	df	t.ratio	p.value	- 95% CI	+ 95 CI
WT - <i>mrc-1</i>	0.597	2.768	53	0.215819	0.999932	-7.588	8.782
WT - <i>mrc-2</i>	3.031	2.768	53	1.094812	0.881287	-5.154	11.216
WT - <i>mrc-3</i>	-5.275	2.768	53	-1.90551	0.410261	-13.460	2.910
WT - <i>mrc-1</i> BC2 AABB	-8.684	2.768	53	-3.13658	0.031545	-16.869	-0.499
WT - <i>mrc-1</i> BC2 aabb	1.104	2.768	53	0.3986	0.998628	-7.081	9.289
<i>mrc-1</i> - <i>mrc-2</i>	2.433	2.695	53	0.903079	0.944029	-5.533	10.400
<i>mrc-1</i> - <i>mrc-3</i>	-5.873	2.695	53	-2.17945	0.264477	-13.840	2.094
<i>mrc-1</i> - <i>mrc-1</i> BC2 AABB	-9.281	2.695	53	-3.44427	0.013618	-17.248	-1.314
<i>mrc-1</i> - <i>mrc-1</i> BC2 aabb	0.506	2.695	53	0.187789	0.999966	-7.461	8.473
<i>mrc-2</i> - <i>mrc-3</i>	-8.306	2.695	53	-3.08253	0.036314	-16.273	-0.340
<i>mrc-2</i> - <i>mrc-1</i> BC2 AABB	-11.715	2.695	53	-4.34735	0.000851	-19.681	-3.75
<i>mrc-2</i> - <i>mrc-1</i> BC2 aabb	-1.927	2.695	53	-0.71529	0.979257	-9.894	6.039
<i>mrc-3</i> - <i>mrc-1</i> BC2 AABB	-3.408	2.695	53	-1.26481	0.802435	-11.375	4.558
<i>mrc-3</i> - <i>mrc-1</i> BC2 aabb	6.379	2.695	53	2.367244	0.186556	-1.588	14.346
<i>mrc-1</i> BC2 AABB - <i>mrc-1</i> BC2 aabb	9.787	2.695	53	3.632056	0.007919	1.820	17.754

Table 3.4. Pairwise comparisons of grain size. The grain size was measured as the average area of individual mature grains (15 – 88 grains) from each plant, using a MARViN seed analyser. The table shows differences of pairwise comparisons between genotypes, based on a linear regression model and one-way ANOVA with Tukey post-hoc test. Statistics were done in R using the `lm()` function from the ‘stats’ package. All pairwise comparisons were done using the `emmeans()` function from the ‘emmeans’ package. Comparisons where $p < 0.05$ are highlighted in grey. SE = standard error, 95% CI = 95% confidence interval, df = degrees of freedom.

Grain size (area mm ²)							
contrast	difference	SE	df	t.ratio	p.value	- 95% CI	+ 95 CI
WT - <i>mrc-1</i>	-0.684	0.668	53	-1.02467	0.907566	-2.658	1.290
WT - <i>mrc-2</i>	0.279	0.668	53	0.418244	0.998272	-1.695	2.253
WT - <i>mrc-3</i>	-1.951	0.668	53	-2.92194	0.054466	-3.925	0.0231
WT - <i>mrc-1</i> BC2 AABB	-2.077	0.668	53	-3.11028	0.033791	-4.051	-0.103
WT - <i>mrc-1</i> BC2 aabb	-0.809	0.668	53	-1.2119	0.82914	-2.783	1.165
<i>mrc-1</i> - <i>mrc-2</i>	0.964	0.650	53	1.482452	0.676692	-0.958	2.885
<i>mrc-1</i> - <i>mrc-3</i>	-1.267	0.650	53	-1.94926	0.384721	-3.188	0.655
<i>mrc-1</i> - <i>mrc-1</i> BC2 AABB	-1.393	0.650	53	-2.14276	0.281888	-3.314	0.529
<i>mrc-1</i> - <i>mrc-1</i> BC2 aabb	-0.125	0.650	53	-0.19236	0.999961	-2.046	1.797
<i>mrc-2</i> - <i>mrc-3</i>	-2.230	0.650	53	-3.43171	0.01411	-4.152	-0.309
<i>mrc-2</i> - <i>mrc-1</i> BC2 AABB	-2.356	0.650	53	-3.62521	0.00808	-4.278	-0.435
<i>mrc-2</i> - <i>mrc-1</i> BC2 aabb	-1.089	0.650	53	-1.67481	0.55407	-3.010	0.833
<i>mrc-3</i> - <i>mrc-1</i> BC2 AABB	-0.126	0.650	53	-0.1935	0.99996	-2.047	1.796
<i>mrc-3</i> - <i>mrc-1</i> BC2 aabb	1.142	0.650	53	1.756897	0.501633	-0.780	3.063
<i>mrc-1</i> BC2 AABB - <i>mrc-1</i> BC2 aabb	1.268	0.650	53	1.950401	0.384065	-0.654	3.189

3.2.4 Loss of MRC greatly alters starch granule size distributions in the mature endosperm

3.2.4.1 Loss of MRC has variable effects on total endosperm starch

To determine the effect of the *mrc* mutations on starch synthesis in grains, I first looked at total starch content of mature grains. The quantification of starch per gram flour was carried out by Erica Hawkins, and I performed the subsequent data analysis. Starch content was largely similar between WT and the *mrc* mutants. Although some pairwise comparisons showed $p < 0.05$ (WT with *mrc-2* and *mrc-3*), the confidence intervals of the difference in means were still close to zero, suggesting that any effect of the mutations was small (Figure 3.5A). To analyse starch granule number and granule size, I performed starch purification and Coulter counter analyses, with technical assistance from David Seung. I performed the data analyses on these samples. I found that there were some minor differences between genotypes in the total number of granules relative to grain weight (granules/mg grain) (Figure 3.5B) but overall there was not a strong effect of loss of MRC. The backcrossed *mrc-1* mutant had more starch granules per unit grain weight than both the WT and the wild-type segregant, but the non-backcrossed *mrc-1* mutant was not significantly different to the WT. The *mrc-2* mutant also had relatively more granules than the WT, but *mrc-3* did not.

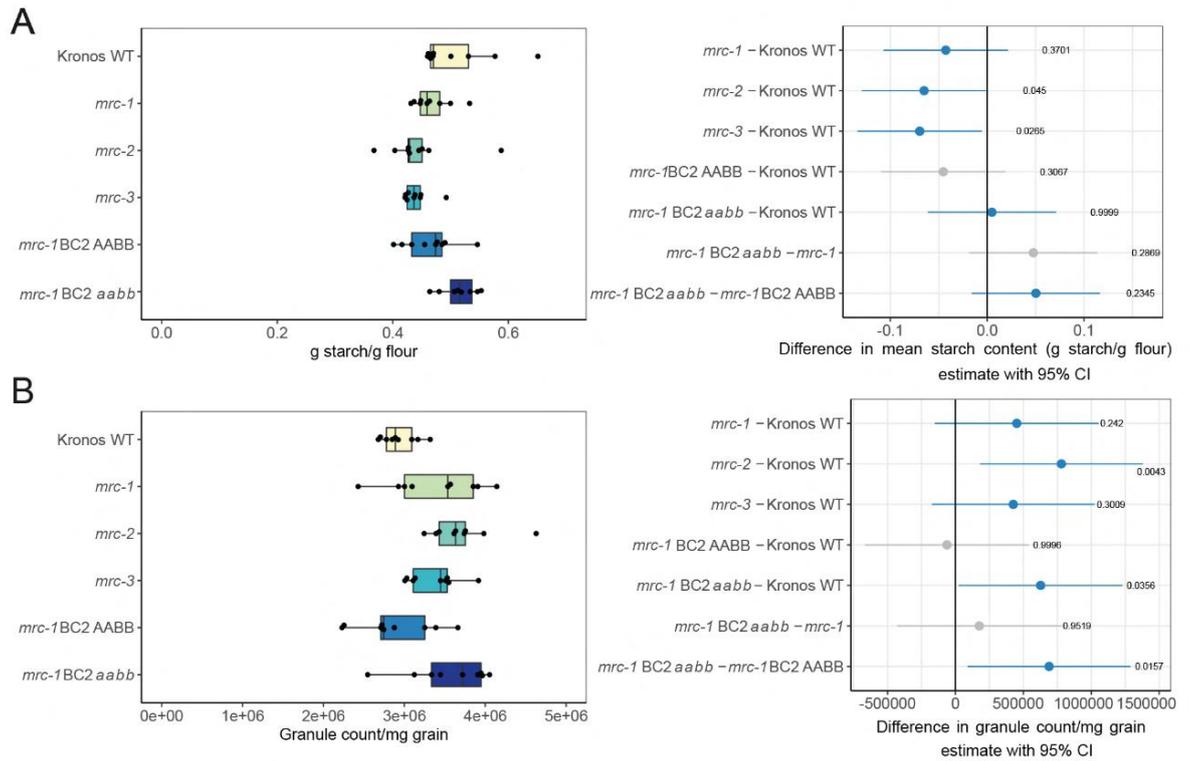


Figure 3.5 Mutations in *MRC* have variable effects on total starch in mature grains. A) Total endosperm starch content in mature grains in WT and *mrc* mutants. **B)** Number of starch granules per mg of grain, counted with a Coulter counter. For A and B, dots on the boxplots indicate individual plants, with $n = 8 - 9$. Each box encloses the middle 50% of the distribution, the middle line is the median and the whiskers are the minimum and maximum values within $1.5 \times$ the interquartile range. All statistical analyses were performed using a linear model, with a one-way ANOVA and Tukey post-hoc test. Panels on the right indicate differences in means between genotypes from pairwise comparisons based on these models. The difference in means is indicated by a dot, with whiskers showing the 95% confidence interval (CI) of this difference, with the corresponding p-value. Grey indicates the WT or *mrc-1* mutant compared to the backcrossed line with the equivalent genotype at the *MRC* loci, and blue indicates all other pairwise comparisons. This figure appears in (Chen *et al.*, 2022a).

3.2.4.2 Mature endosperm of *mrc* mutants has a shifted starch granule size distribution compared to wild type

Granule size distributions were determined from the Coulter counter data by plotting the percentage of starch volume in each size bin relative to the total volume of starch measured. I observed clear bimodal distributions for all genotypes, with a peak corresponding to A-type granules (~ 18 – 25 μm) and a peak corresponding to B-type granules (~3 – 10 μm). However, the distribution profiles in the mutants were very different from the WT, with the peak corresponding to B-type granules being more prominent in the mutants (Figure 3.6A). The profiles of the backcrossed *mrc-1* lines were similar to their non-backcrossed equivalents.

I fitted a bimodal log-normal distribution curve to the granule size distributions of each sample to estimate the total volume percentage of B-type granules and the mean sizes of A- and B-type granules. Comparing the means of these extracted values between genotypes showed a higher B-type granule percentage (by volume) for all three mutants (*mrc-1*, *mrc-2* and *mrc-3*) compared to the WT (Figure 3.6B) The strongest increase was seen for *mrc-1*, and this increase was consistent when comparing the double backcrossed *mrc-1* (*mrc-1* BC2 *aabb*) with WT and the wild-type segregant (*mrc-1* BC2 AABB). There was a small increase in the volume percentage of B-type granules in the wild-type segregant compared to the WT, but the difference was much smaller than between the other genotypes.

The higher B-type granule volume percentage in *mrc* mutants could be due to an increase in both B-type granule size and number. The mean B-type granule size was larger than WT for *mrc-1* (and *mrc-1* BC2 *aabb*) and *mrc-2*, but not *mrc-3* (Figure 3.6D). By contrast, the mean A-type granule diameter was smaller for all three mutants than for WT (Figure 3.6C). Thus, the increased proportion of B-type granule volume may be due to a combination of smaller A-type granules and larger and perhaps more numerous B-type granules.

To see whether the altered granule size distributions in the *mrc* mutants were accompanied by changes in starch granule morphology, Erica Hawkins took light microscopy images of iodine-stained thin sections of mature grains. Like the WT, all mutants had flattened A-type and round B-type granules (Figure 3.7A). Similarly, no defects in A- or B-type granule shape were observed in the mutants using scanning electron microscopy (SEM) (Figure 3.7B).

To see whether the change in granule size distribution was influenced by changes in starch polymer biosynthesis, Robin Denley Bowers measured the starch polymer structure using High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD), and amylose content using iodine. Both of these traits were measured using purified starch,

and neither was affected by loss of functional MRC. The *mrc-1* mutant, with the strongest alteration in granule size distribution, had normal amylopectin structure and amylose content (Figure 3.8).

Overall, these data suggest that MRC is required for the normal size distribution of starch granules in wheat endosperm. In tetraploid wheat, mutants lacking in the 6A copy of MRC consistently had a higher relative volume of B-type granules in the endosperm than the WT, and smaller A-type granules. This change in granule size distribution occurred without accompanying changes in total starch content, starch granule shape, amylose content or amylopectin structure.

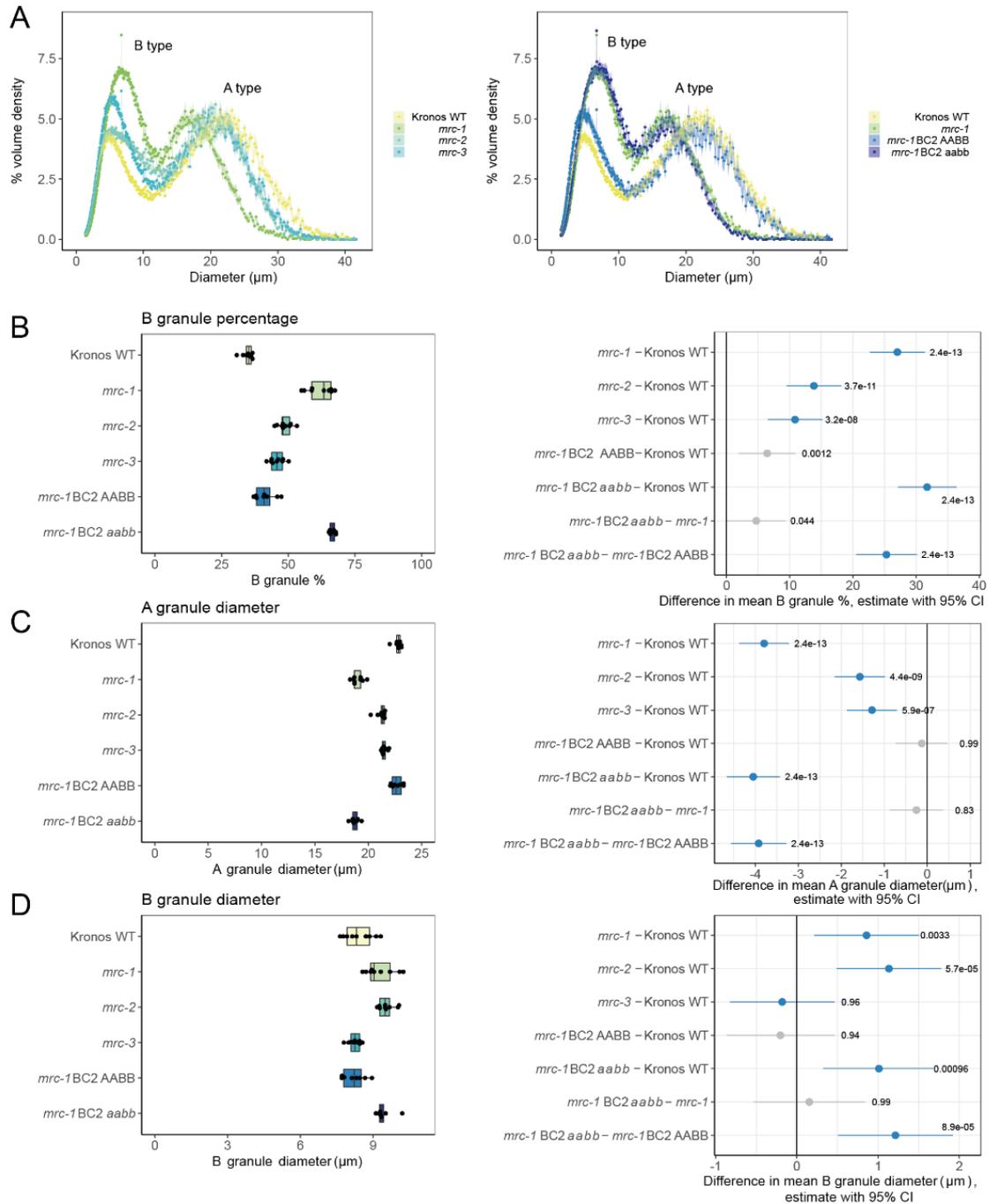


Figure 3.6. Endosperm starch in mature grains of *mrc* mutants have altered granule size distribution. **A)** Coulter counter traces with evenly binned x-axes show a bimodal distribution of granule sizes from purified wheat endosperm starch. Data points are mean values from 9 individual plants of each genotype (3 grains from each plant), with the standard error of the mean shown as a shaded ribbon. WT and *mrc-1* in the left and right panel are the same data. **B, C, D)** Mean B-type granule percentage volume (B), A-type granule diameter (C) and B-type granule diameter (D) values extracted from bimodal log-normal distribution curves fitted to Coulter counter traces of individual plants. Extracted values and boxplots are shown on the left, where dots indicate mean values from individual plants ($n = 9$). Each box encloses the middle 50% of the distribution, the middle line is the median and the whiskers are the minimum and maximum values within $1.5 \times$ the interquartile range. All statistical analyses were performed using linear models for each panel, with a one-way ANOVA and Tukey post-hoc test. Panels on the right indicate differences in means between genotypes from pairwise comparisons based on these models. The difference in means is indicated by a dot, with whiskers showing the 95% confidence interval (CI) of this difference, with the corresponding p-value. Grey indicates the WT or *mrc-1* mutant compared to the backcrossed line with the equivalent genotype at the *MRC* loci, and blue indicates all other pairwise comparisons. This figure appears in (Chen *et al.*, 2022a).

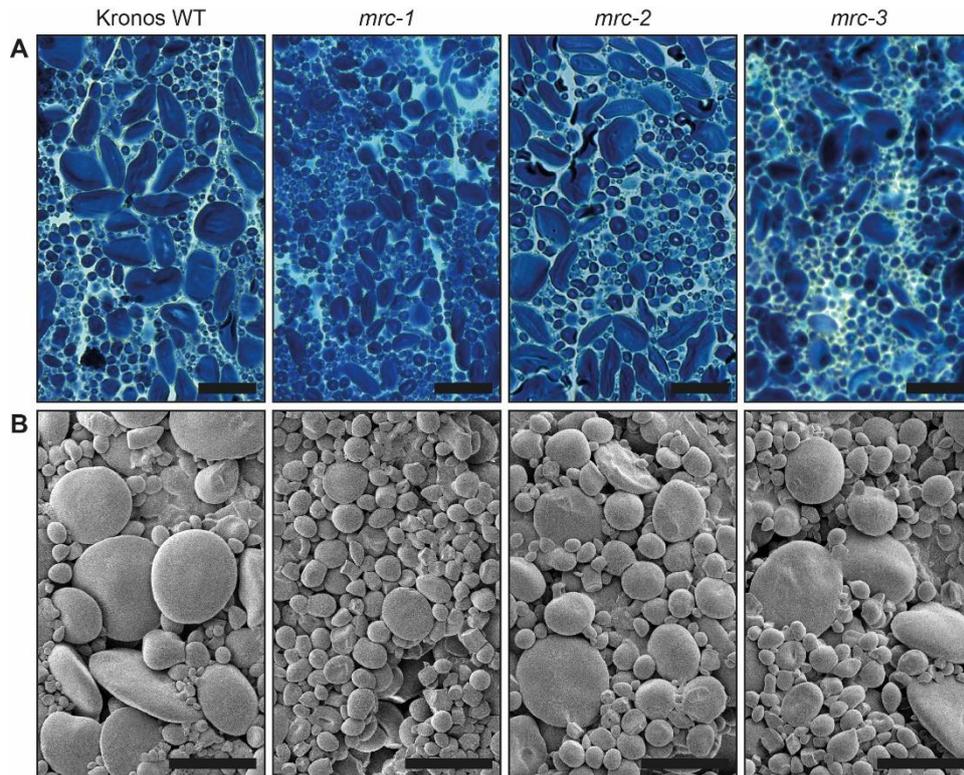


Figure 3.7. Endosperm starch granules of mature grains in three wheat *mrc* mutants are similar in shape compared to wild type. A) Thin sections of mature endosperm tissue were stained with Lugol's solution and imaged using light microscopy. Bar = 40 μ m. **B)** Purified endosperm starch granules were observed using scanning electron microscopy (SEM). Bar = 20 μ m. This figure appears in (Chen *et al.*, 2022a).

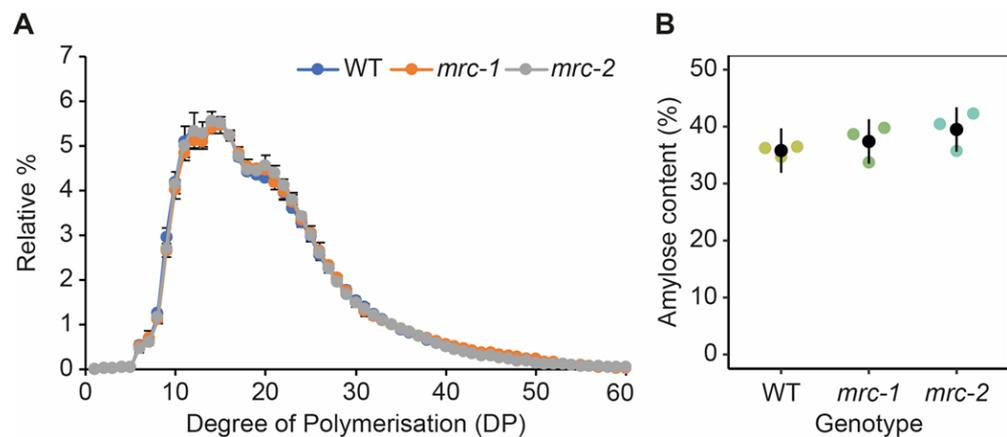


Figure 3.8 Chain length distribution and amylose content of the *mrc* mutants. A) Chain length distribution of *mrc-1* and *mrc-2* starch. Purified starch was debranched and analysed with High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). The area of peaks corresponding to chains of each degree of polymerisation (DP) was expressed as a percentage of the summed peak area for DP 1–60. Values are the mean \pm SEM from three replicate measurements. **B)** Amylose content of of *mrc-1* and *mrc-2* starch quantified using iodine colourimetry. Values represent mean \pm 95% CI from three replicate measurements. This figure appears in (Chen *et al.*, 2022a) and the experiments were performed by Robin Denley Bowers.

3.2.5 Loss of MRC results in the early initiation of B-type granules

To understand how MRC affects the size distribution of endosperm starch granules, and its specific effects on A-type or B-type granules, I investigated granule initiation during grain development in the *mrc* mutant with the strongest phenotype, *mrc-1*. All experimental work in this section was performed by David Seung, and I performed all the data analysis and interpretation.

The total starch content and number of starch granules were measured in dissected endosperms of developing grains harvested 8, 14, 20 and 30 days post anthesis (dpa). The total starch content of the endosperm increased between each time point, and there was no significant difference between the mutant and the WT at any time point (Figure 3.9A). At the 8 dpa timepoint, the mutant and the WT contained similar numbers of starch granules. Interestingly, for the two subsequent time points (14 dpa and 20 dpa), the mutant had almost twice as many starch granules as the WT, despite having similar starch contents (Figure 3.8B). The largest increase in granule number during grain filling was observed between the 20 and 30 dpa timepoints in the WT, and between the 14 and 20 dpa timepoints in the mutant. At the 30 dpa timepoint, the difference in granule number between the mutant and WT decreased. In both the WT and mutant, the number of starch granules decreased between the 8 and 14 dpa timepoints. The reason for this is unknown, but it has also been observed in *Aegilops* species – which are close relatives of wheat (Howard *et al.*, 2011).

In WT endosperm, there was a unimodal distribution of starch granule sizes at the 8 and 14 dpa timepoints, and only A-type granules with their characteristic flattened morphology were observed using SEM (Figure 3.10). The A-type granules grew substantially in size between the two timepoints, seen as a shift in the granule size distribution peak. B-type granules only became prominent at the 20 dpa timepoint. In the *mrc-1* mutant A-type granules were initially the same size as those of WT (at 8 dpa), but subsequently grew more slowly than wild-type granules. By contrast, B-type granules were already present at 14 dpa in the *mrc-1* mutant (seen as a distinct shoulder that appeared in the granule size distribution), considerably earlier than in the WT. Taken together, these data suggest that the larger number of granules observed between 14-20 dpa in the *mrc-1* endosperm compared to the WT (observed in Figure 3.9B) is due to the early initiation of B-type granules in the mutant.

In conclusion, MRC is required for the temporal control of B-type granule initiation during wheat grain development. It is expressed during early grain development, and its loss leads to the early initiation of B-type granules. I therefore propose that MRC acts as a repressor of B-type granule formation in the developing wheat endosperm during early grain development.

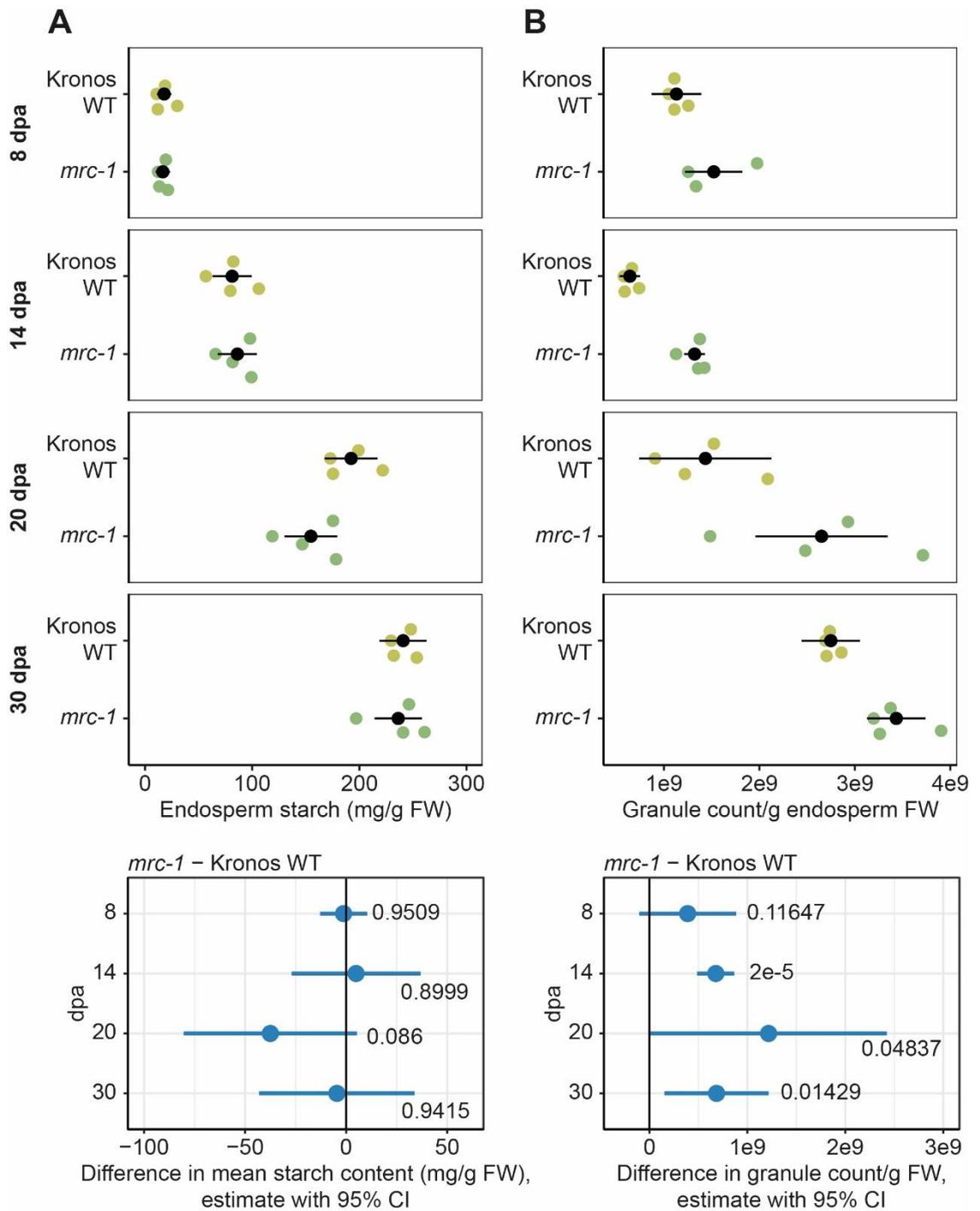


Figure 3.9. In developing endosperm, granule number increases in *mrc-1* compared to WT despite having similar starch content. The endosperm was dissected from developing grains of WT and *mrc-1*, harvested at 8, 14, 20 and 30 dpa, with $n = 3 - 4$ individual plants for each genotype per time point, indicated by coloured dots in the upper panels, with the mean \pm 95% confidence interval (CI) in black dots and whiskers. **A)** Starch content of the endosperm. Values are expressed relative to the fresh weight of the dissected endosperm. **B)** Starch granule number in the endosperm. Starch was purified from dissected endosperm and the number of granules was determined using a Coulter counter. Values are expressed relative to the fresh weight of the dissected endosperm. For A and B, individual linear models were fitted to the data of each time point, with a one-way ANOVA and Tukey post-hoc test to compare the means of WT and *mrc-1*. Panels on the bottom summarise the differences in means from these linear models, indicated by a dot, with whiskers showing the 95% CI of this difference, with the corresponding p-value. This figure appears in (Chen *et al.*, 2022a).

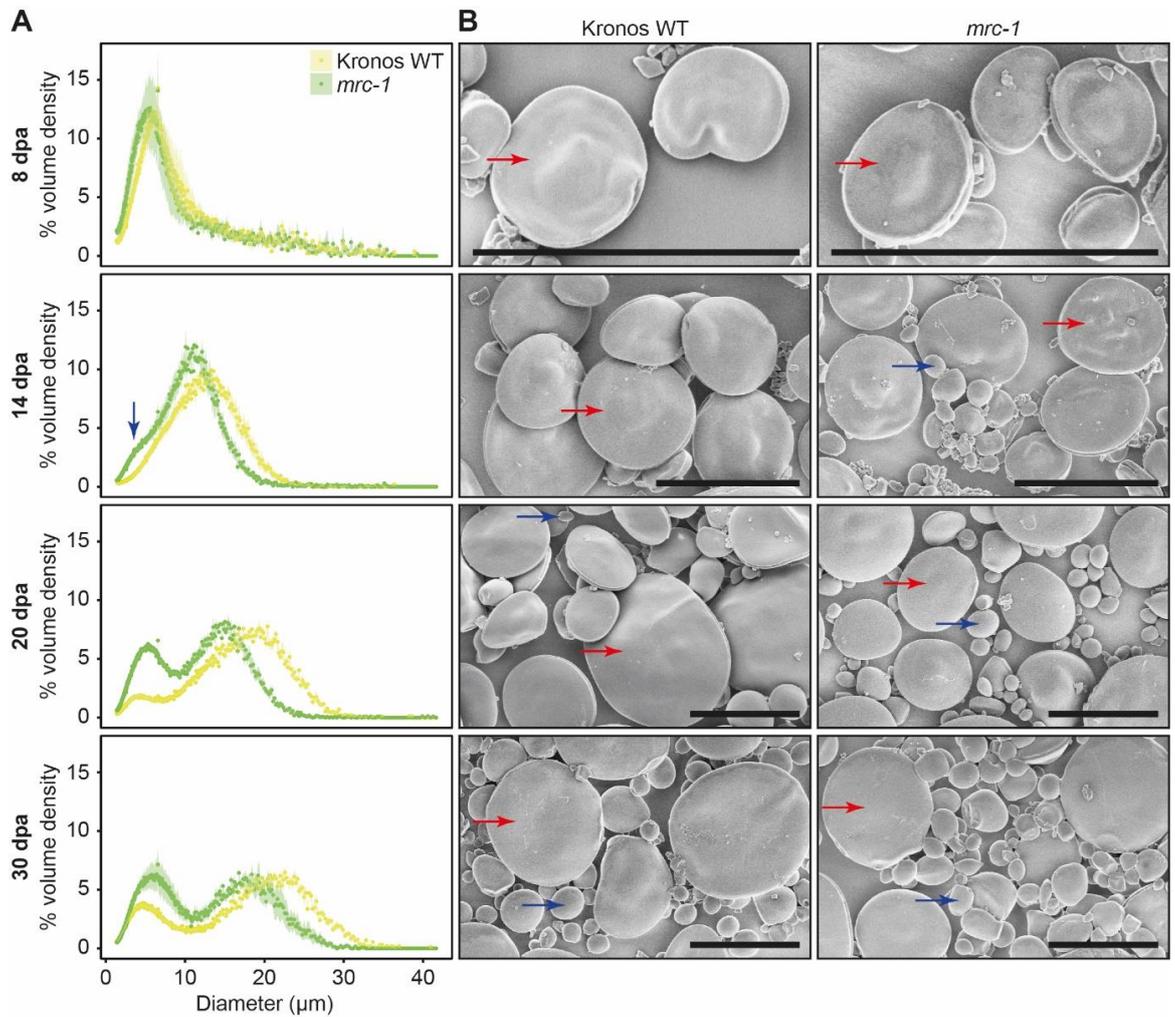


Figure 3.10. The *mrc-1* mutant initiates B-type granules earlier in grain development than the WT. The endosperm was dissected from developing grains of WT and *mrc-1* harvested at 8, 14, 20 and 30 dpa. **A)** Coulter counter traces with evenly binned x-axes show the starch granule size distribution of endosperm starch. Distributions are the average of $n = 4$ measurements, each carried out on grains harvested from a different plant (three grains per measurement). Data points are mean values from these 4 measurements, with the standard error of the mean shown as a shaded ribbon. The B-type granule peak in the *mrc-1* mutant at 14 dpa is indicated with a blue arrow. **B)** Endosperm starch granules were observed using scanning electron microscopy (SEM). Bars = 20 μm . Examples of A-type granules and B-type granules are marked with red and blue arrows respectively. This figure appears in (Chen *et al.*, 2022a).

3.2.6 *TaMRC* promotes granule initiation in wheat leaves

Considering the role of MRC in promoting granule initiation in *Arabidopsis* leaves (Seung *et al.*, 2018; Vandromme *et al.*, 2019), it was surprising that MRC repressed B-type granule formation in wheat endosperm. I therefore investigated whether this repressive role also applies to granule initiation in wheat leaves, which would suggest a divergence in MRC function between wheat and *Arabidopsis*; or whether the role of MRC in wheat leaves is the same as in *Arabidopsis* but it has a distinct function in granule initiation in the endosperm.

3.2.6.1 *Loss of MRC results in fewer granules per chloroplast in young wheat leaves*

To investigate whether MRC affected the starch granule number per chloroplast in wheat leaves, I analysed light microscopy (LM) images of young wheat leaf sections, stained for starch. I found that *mrc-1* and *mrc-2* had fewer granules per chloroplast than the WT (Figure 3.11 A, C, Table 3.5). As in the endosperm, the *mrc-1* mutant showed the strongest effect, having almost 50% fewer granules per chloroplast than the WT (Table 3.5). Interestingly, there was no effect of *mrc-3*, suggesting that the L289F mutation does not affect MRC function in leaves (Figure 3.11C, Table 3.5).

To assess granule number, shape and size in leaf chloroplasts of the wheat mutants, leaf tissue was harvested from the middle of the older of two leaves of 10-day-old wheat seedlings and sections were imaged with light microscopy (Figure 3.11 A, B). The number of granules per chloroplast was quantified from these images. I measured three individual plants per genotype for each experiment and compared the mean granules per chloroplast between genotypes using a negative binomial mixed effects model with individual biological replicates as random effect. I first compared all three *mrc* mutants with the WT (Figure 3.11C, Table 3.5), then I compared the backcrossed *mrc-1* lines with the WT (Figure 3.11 D, Table 3.5).

The backcrossed *mrc-1* (*mrc-1* BC2 *aabb*) had fewer granules per chloroplast than WT, similar to the non-backcrossed *mrc-1*. However, I was not able to make meaningful comparisons with the wild-type segregant (*mrc-1* BC2 *AABB*) because – as for comparisons of TGW and grain size above – values for the wild-type segregant differed from those of the WT, with significantly fewer granules per chloroplast than WT. Overall, the smaller number of granules in *mrc* chloroplasts than in WT chloroplasts suggests that MRC promotes granule initiation in wheat leaves, pointing to distinct roles of MRC in different organs.

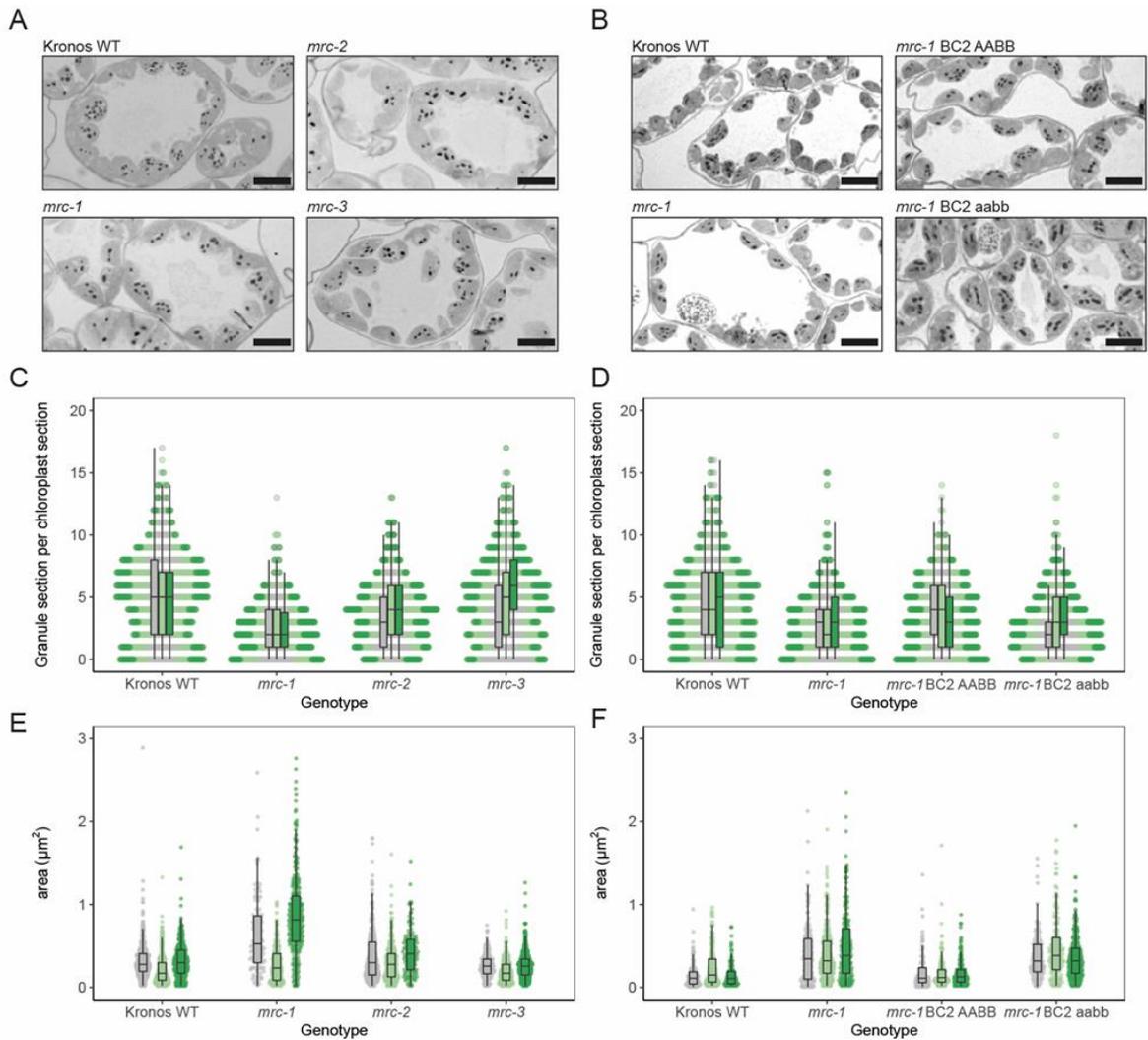


Figure 3.11. Chloroplasts of *mrc-1* mutants have fewer and larger starch granules than the WT. A, B) Representative light microscopy images of thin sections (500 nm) from the middle of the older of two leaves in 10-day old wheat seedlings, collected at the end of day. Sections were stained for starch using periodic acid and Schiff staining. Scale bar = 10 µm. **C, D)** Distributions of counts of granule sections per chloroplast section, counted from light microscopy images as in A and B. The three different colours correspond to three individual plants of each genotype, per experiment. For each sample, 200 – 250 chloroplasts were counted, with absolute values varying between samples. Dots represent individual chloroplasts. **E, F)** Distributions of starch granule sizes measured as the area in the light microscopy images as in A and B. The three different colours correspond to three individual plants of each genotype, per experiment. 100 – 400 granules were measured in each sample. Dots represent individual granules. For all boxplots, each box encloses the middle 50% of the distribution, the middle line is the median and the whiskers are the minimum and maximum values within 1.5 × the interquartile range. Measurements were made on the same batch of plants as Figure 3.12, experiment 2 (panels A, C, E here) and experiment 5 (panels B, D, F here). This figure appears in (Chen *et al.*, 2022a).

Table 3.5. Pairwise comparisons of leaf granule per chloroplast in young wheat leaves. Negative binomial mixed effect models were used, with biological replicate (3 of each genotype) as random effect, using Tukey post-hoc tests. Individual models for each experiment were used. SE = standard error, 95% CI = 95% confidence interval.

contrast	Experiment	ratio	SE	- 95% CI	+ 95 CI	z.ratio	p.value
WT / <i>mrc-1</i>	2	1.997	0.190	1.564	2.550	7.271	2.17E-12
WT / <i>mrc-2</i>	2	1.390	0.131	1.091	1.769	3.497	0.00265
WT / <i>mrc-3</i>	2	1.091	0.102	0.858	1.387	0.929	0.789273
<i>mrc-1</i> / <i>mrc-2</i>	2	0.696	0.067	0.544	0.890	-3.783	0.000892
<i>mrc-1</i> / <i>mrc-3</i>	2	0.546	0.052	0.427	0.698	-6.336	1.41E-09
<i>mrc-2</i> / <i>mrc-3</i>	2	0.785	0.074	0.616	1.000	-2.564	0.05063
<i>mrc-1</i> BC2 <i>aabb</i> / <i>mrc-1</i>	5	1.030	0.097	0.810	1.311	0.318	0.988897
<i>mrc-1</i> BC2 AABB / <i>mrc-1</i>	5	1.234	0.115	0.971	1.569	2.256	0.108551
<i>mrc-1</i> BC2 AABB / <i>mrc-1</i> BC2 <i>aabb</i>	5	1.198	0.111	0.943	1.522	1.943	0.210022
WT / <i>mrc-1</i>	5	1.635	0.152	1.287	2.077	5.285	7.52E-07
WT / <i>mrc-1</i> BC2 <i>aabb</i>	5	1.587	0.147	1.251	2.014	4.982	3.76E-06
WT / <i>mrc-1</i> BC2 AABB	5	1.325	0.122	1.045	1.679	3.048	0.012347

Table 3.6. Pairwise comparisons of granule size in young wheat leaves. Linear mixed effect models were used, with biological replicate (3 of each genotype) as random effect, using Tukey post-hoc tests and Satterthwaite degrees of freedom calculation. Individual models for each experiment were used. SE = standard error, df = degrees of freedom, 95% CI = 95% confidence interval.

contrast	Experiment	difference	SE	- 95% CI	+ 95 CI	df	t.ratio	p.value
WT - <i>mrc-1</i>	2	-0.304	0.125	-0.706	0.0975	7.998	-2.42484	0.149207
WT - <i>mrc-2</i>	2	-0.087	0.125	-0.488	0.3150	7.979	-0.69054	0.89788
WT - <i>mrc-3</i>	2	0.045	0.125	-0.357	0.4461	7.941	0.356537	0.983352
<i>mrc-2</i> - <i>mrc-1</i>	2	-0.218	0.126	-0.619	0.1841	8.055	-1.73159	0.368589
<i>mrc-3</i> - <i>mrc-1</i>	2	-0.349	0.125	-0.750	0.0529	8.017	-2.77906	0.090736
<i>mrc-3</i> - <i>mrc-2</i>	2	-0.131	0.125	-0.533	0.2704	7.998	-1.04602	0.729078
<i>mrc-1</i> BC2 <i>aabb</i> - <i>mrc-1</i>	5	-0.034	0.0394	-0.160	0.0924	7.990	-0.85905	0.825337
<i>mrc-1</i> BC2 AABB - <i>mrc-1</i>	5	-0.255	0.0398	-0.381	-	8.317	-6.41375	0.000799
<i>mrc-1</i> BC2 AABB - <i>mrc-1</i> BC2 <i>aabb</i>	5	-0.221	0.0400	-0.348	0.0949	8.470	-5.5362	0.002041
WT - <i>mrc-1</i>	5	-0.260	0.0396	-0.386	0.1335	8.122	-6.56687	0.000748
WT - <i>mrc-1</i> BC2 <i>aabb</i>	5	-0.226	0.0397	-0.352	0.0996	8.274	-5.68343	0.001851
WT - <i>mrc-1</i> BC2 AABB	5	-0.005	0.0401	-0.131	0.1218	8.607	-0.11473	0.999416

3.2.6.2 Loss of MRC has variable effects on granule size and total starch content in young leaves

As loss of MRC reduced the number of granules per chloroplast in wheat leaves, I investigated whether this was accompanied by an effect on the granule size and total starch content. While the effect of the *mrc-1* mutation on starch granule number per chloroplast was consistent, granule size and total starch content appeared to be interdependent and varied between the experiments. I quantified the size of granules from the LM images by measuring the area of the granules in the sections. The distributions showed a trend towards larger granule sizes in the *mrc-1* mutant (Figure 3.11 E, F). However, after fitting a linear mixed effects model with individual biological replicates as random effect (Table 3.6), pairwise comparisons showed no differences between any of the non-backcrossed *mrc* mutants and WT. However, in the experiment comparing the backcrossed *mrc-1* mutants to WT and wild-type segregant, the granules were larger in both backcrossed and non-backcrossed *mrc-1* mutants. Aside from the differences in size, there was no visible difference in starch granule shape in any of the mutants compared to WT.

Similar variability between experiments was seen in total starch content, which was measured in both leaves of 10-day-old wheat seedlings at the end of the day, when maximum starch is expected. The first three experiments compared the genotypes WT, *mrc-1*, *mrc-2* and *mrc-3* (Experiments 1, 2, 3; Table 3.7), and the pooled data showed a lower starch content in all *mrc* mutants compared to WT (Figure 3.12A). The next two experiments (Experiments 4 and 5, Table 3.7) compared end of day starch content in WT, *mrc-1*, *mrc-1* BC2 *aabb* and *mrc-1* BC2 AABB. By contrast, these experiments showed no difference in starch content between *mrc-1* and WT (Figure 3.12B). In fact, when comparing *mrc-1* BC2 *aabb* with *mrc-1* BC2 AABB or with the WT, there was a small increase in starch content in the backcrossed mutant. The reason for the discrepancy between the results from experiments 1 – 3 compared to experiments 4 and 5 is unknown, but it may indicate that total starch content in wheat leaves is particularly variable and dependent on many environmental and physiological factors.

Most interestingly, lower starch content in *mrc* mutants compared to the WT was observed in the same batch of plants where we did not observe a significant difference in granule size between *mrc* mutants (Experiment 2, Figure 3.11E), and there was identical (or greater) starch content observed for *mrc-1* mutants in the same batch of plants where we observed larger granules for these mutants (Experiment 5, Figure 3.11F). Since the reduction in granule number described above was consistently observed in all experiments, it is plausible that the effect on granule size depends on starch content, such that the *mrc-1* mutant only produces larger granules under conditions where total starch content is equal. Pooled data from plants harvested at the end of

the night (quantified in experiments 1 and 2) showed very low starch content in all three *mrc* mutants, indicating substantial nocturnal starch turnover in all genotypes, and no differences between WT and mutants (Figure 3.12C).

Table 3.7. Summary of leaf starch quantification experiments. ED = End of Day, EN = End of night. Non-valid sample = sample lost during experiment or omitted because technical replicates were too divergent. All experiments were done in the same CER, on a 16-h day, 8-h night cycle.

Experiment	Date	Genotypes	Time	Number of biological reps per genotype planned	Total non-valid samples deleted	Total outliers deleted from final dataset	Final n
1	Oct-19	WT, <i>mrc-1</i> , <i>mrc-2</i> , <i>mrc-3</i>	ED+EN	6	3	0	ED: 6,6,5,6; EN:6,6,4,6
2	Dec-19	WT, <i>mrc-1</i> , <i>mrc-2</i> , <i>mrc-3</i>	ED+EN	6	5	0	ED: 6,6,6,5; EN: 4,6,5,5
3	Jun-20	WT, <i>mrc-1</i> , <i>mrc-2</i> , <i>mrc-3</i>	ED	20	8	0	18,16,19,19
4	Jan-21	WT, <i>mrc-1</i> , <i>mrc-1</i> BC2 AABB, <i>mrc-1</i> BC2 <i>aabb</i>	ED	15	10	1	15,12,10,12
5	Feb-21	WT, <i>mrc-1</i> , <i>mrc-1</i> BC2 AABB, <i>mrc-1</i> BC2 <i>aabb</i>	ED	12	3	0	10,12,12,11

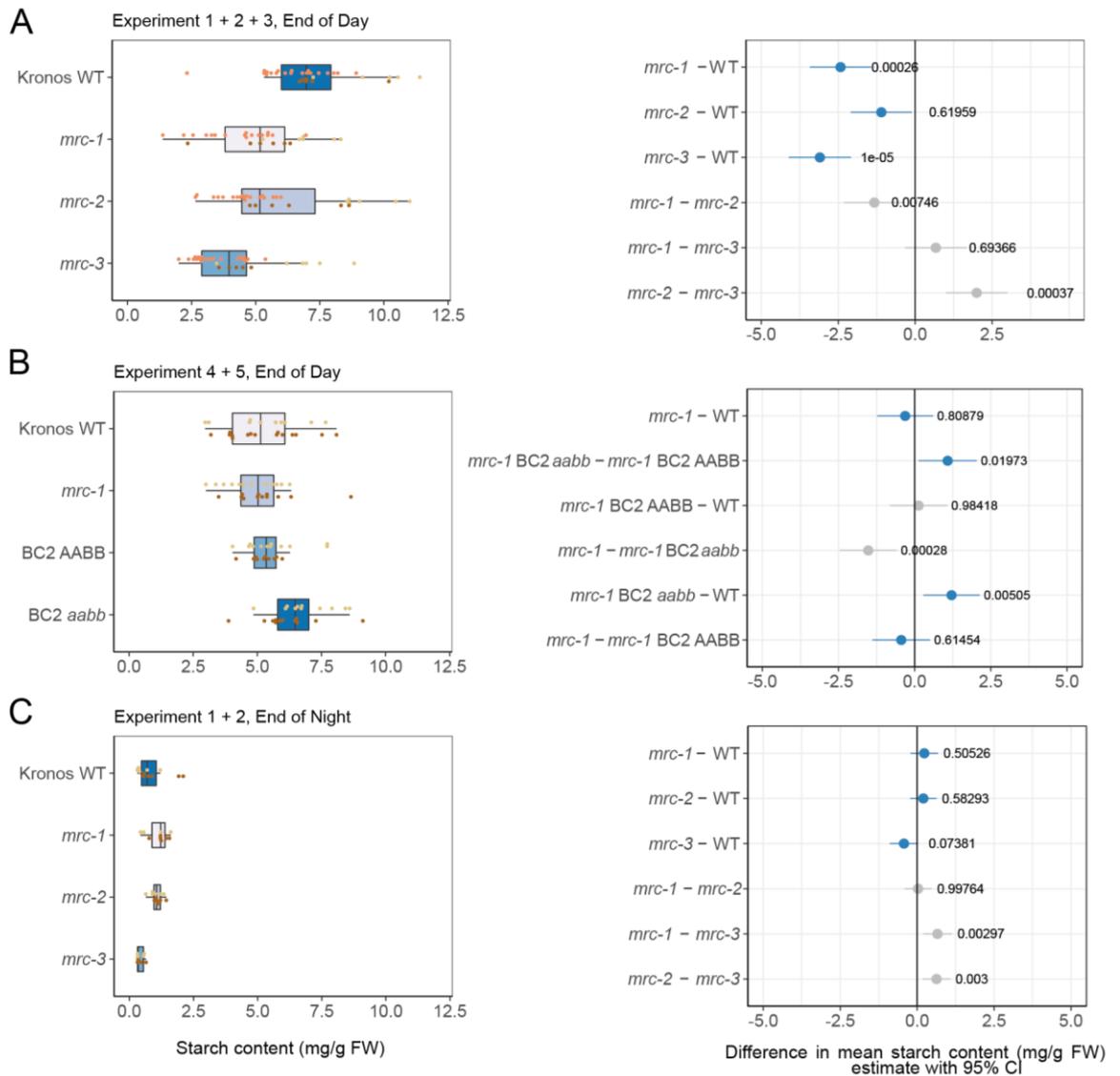


Figure 3.12. Loss of MRC has variable effects on the total end of day (ED) leaf starch content in wheat leaves. A) Pooled starch content data of wheat leaves harvested at the end of day (ED) in experiments 1 (dark brown, $n = 5 - 6$ per genotype), 2 (light brown, $n = 5 - 6$ per genotype) and 3 (orange, $n = 16 - 19$ per genotype). **B)** Pooled starch content data from wheat leaves in experiments 4 (dark brown, $n = 10 - 15$ per genotype) and 5 (light brown, $n = 10 - 12$ per genotype) harvested at ED. **C)** Pooled starch content data of wheat leaves harvested at the end of night (EN) in experiments 1 (dark brown, $n = 4 - 6$ per genotype) and 2 (light brown $n = 4 - 6$ per genotype) For all panels, raw data and boxplots are shown on the left, where dots indicate values from individual wheat plants, with colours indicating each experiment. Each box encloses the middle 50% of the distribution, the middle line is the median and the whiskers are the minimum and maximum values within $1.5 \times$ the interquartile range. All statistical analyses were performed using a linear model with genotype and experiment as fixed effects for each panel, ANOVA and Tukey post-hoc tests. Panels on the right indicate differences in adjusted means of total starch content based on these models and pairwise comparisons of the genotypes. The difference in means is indicated by a dot, with whiskers showing the 95% confidence interval (CI) of this difference, with the corresponding p-value. For A and C, blue indicates comparisons between mutant and wild type, while grey indicates comparisons between mutants. For B, grey indicates the WT or *mrc-1* mutant compared to the backcrossed line with the equivalent genotype at the *MRC* loci, and blue indicates all other pairwise comparisons. This figure appears in (Chen *et al.*, 2022a).

3.3 Discussion

3.3.1 A novel role for MRC in repressing B-type granule initiation during endosperm starch synthesis

Starch granule initiation remains the most enigmatic part of the starch synthesis process, where we understand little about how the diverse numbers and morphologies of starch granules are determined in our most important crops (Abt & Zeeman, 2020; Chen *et al.*, 2021; Seung & Smith, 2019). Here, I shed new light on the temporal regulation of granule initiation in wheat endosperm, by demonstrating the unique role of the protein MRC in the repression of B-type granule initiation during early grain development.

The expression pattern of MRC during grain development is consistent with the change in onset of B-type granule initiation observed in the mutant. MRC is expressed in the endosperm between 6–10 dpa but decreases rapidly in expression between 10–15 dpa, remaining low after 15 dpa (Figure 3.3). This decline in expression coincides with when B-type granules start to form in the WT (between 14–20 dpa; Figures 3.9, 3.10). In the *mrc-1* mutant, B-type granules were initiated earlier than in the WT (Figure 3.10), which could be due to the loss of B-type granule repression by MRC in early grain development. There was an increased number of starch granules in the *mrc-1* mutant compared to WT from 14 to 30 dpa (Figure 3.9), and this increase remained apparent in the mature grains of the backcrossed *mrc-1* mutant (Figure 3.5).

I propose that the early initiation of B-type granules results in drastically altered starch granule size distributions in the endosperm (Figure 3.13). The early appearance of B-type granules in the *mrc-1* mutant could present competition with A-type granules for substrates for granule growth (i.e., ADP-glucose) from an earlier stage of grain development than in the WT, resulting in a higher volume of B-type granules and lower volume of A-type granules in the mature mutant grains (Figure 3.6). At 8 dpa, before the appearance of B-type granules, the size distribution curves of A-type granules were almost identical between mutant and WT (Figure 3.10), and it was only at the later stages of grain development following B-type granule initiation that the A-type granules became smaller in the mutant compared to WT. The lack of difference in A-type granule number and size at 8 dpa suggests that the *mrc-1* mutation did not affect A-type granule initiation.

The data show that the early B-type granule initiation alone could result in the increased proportion of B-type granule volume in mature grains, but it is difficult to make conclusions about the effect of MRC on B-type granule number specifically. The higher B-type granule volume could be caused by a combination of smaller A-type granules and larger and perhaps relatively more

numerous B-type granules (Figure 3.6). I measured a decrease in A-type granule size and an increase in B-type granule size in the *mrc-1* and *mrc-2* mutants (Figure 3.6), and in some lines there was also a slight increase in total granule number (Figure 3.5B). However, it should be noted that there is currently no method for specifically quantifying the number of B-type granules, considering our definition of B-type granules comes from curve fitting volume/size distribution graphs from the Coulter counter, and small differences in B-type granule number may be difficult to detect as the B-type granules already make up a much larger proportion of the total granule number in WT. For *mrc-3*, a higher B-type granule volume percentage was observed (Figure 3.6B) without a measurable increase in B-type granule size or in total granule number (Figure 3.5B), but it is still possible that there is a relative increase in B-type compared to A-type granule number.

Despite all three *mrc* mutants having an increased proportion of B-type granule volume, they differed in severity (Figure 3.6). The strongest effect on B-type granule percentage was seen in *mrc-1*, and since the phenotypes of the *mrc-1* backcrossed and non-backcrossed lines were similar, the severity of this line is not due to background mutations. The *mrc-1* line may have the strongest phenotype because the premature stop codon occurs earlier in the coding sequence than in *mrc-2*, and it is possible that the truncated protein in *mrc-2* is partially functional. The *mrc-3* mutant had the weakest phenotypes, suggesting the Leu289Phe mutation might also produce a partially functional protein. Future experiments should further explore the biochemistry of the truncated *mrc-2* version of the MRC protein and the MRC with the Leu289Phe mutation, for instance by looking at whether these mutated versions of MRC affect interactions with SS4 (see Chapter 4 for more details on protein interaction between *TaMRC* and *TaSS4*).

The *mrc-1* wild-type segregant likely has some background mutations, as several of its phenotypes differed from the WT. The presence of background mutations is not unusual in these EMS-mutagenised wheat TILLING lines (Uauy *et al.*, 2017), and importantly, the effects observed in the wild-type segregant were minor compared to the effects observed in the mutants.

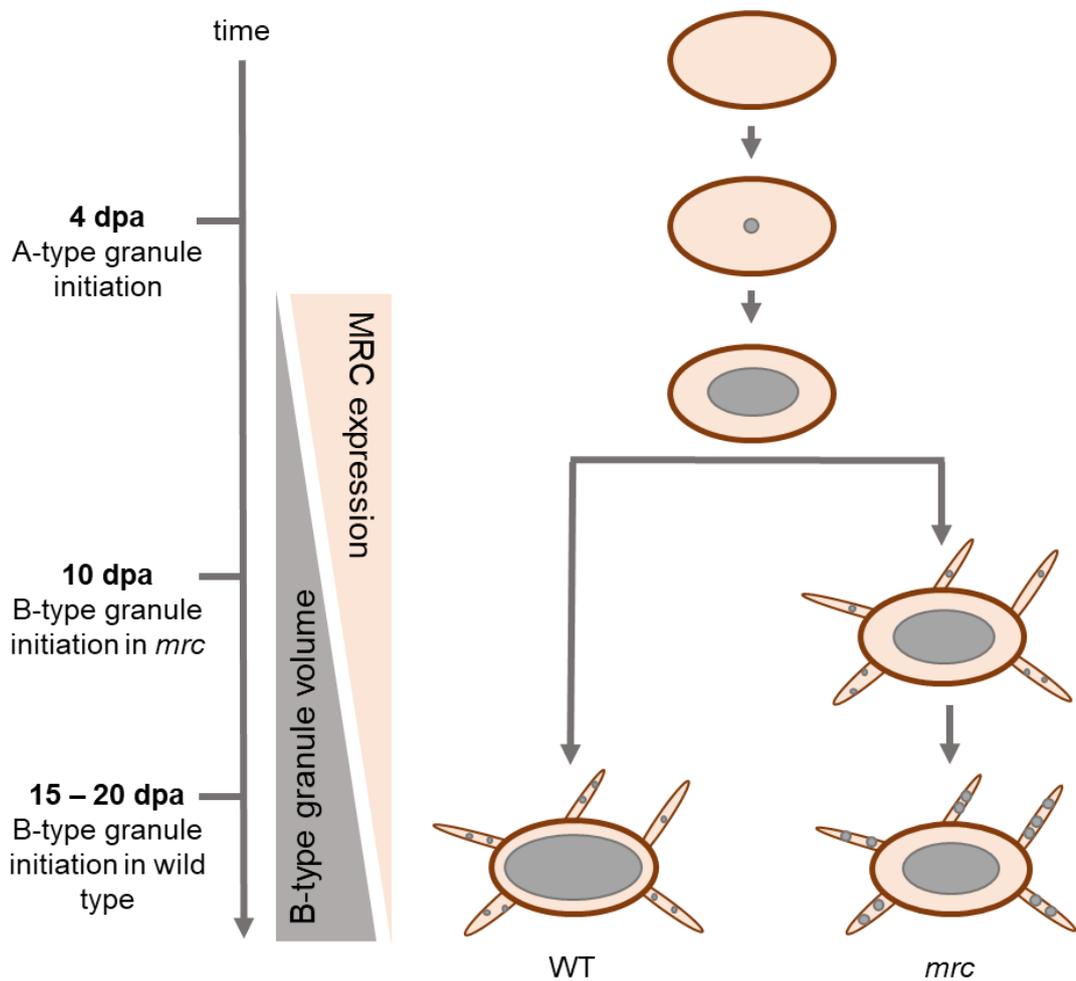


Figure 3.13. Model of MRC function in wheat developing endosperm. MRC is required for the control of the timing of B-type granule initiation during early grain development. In wild type, A-type granules initiate around 4 days post anthesis (dpa), and B-type granules initiate around 15 – 20 dpa. The expression of MRC during early endosperm development could prevent the B-type granule formation, and B-type granule volume increases as MRC expression decreases after approx. 10 dpa. This process is disrupted in mutants lacking a functional MRC protein, and B-type granules initiate earlier in endosperm development. The early initiation of B-type granules provides them with more time and substrate to grow, leading to a higher volume of B-type granules in the mutant at grain maturity compared to the wild type, and a concomitant decrease in A-type granule size in the mutant. This figure appears in (Chen *et al.*, 2022a).

3.3.2 MRC has tissue-specific roles in promoting or repressing granule initiation

Since wheat produces starch in both leaves and endosperm, the effects of *mrc* mutations could be explored in both organs. MRC has contrasting functions in the endosperm compared to the leaves. Rather than repressing granule initiation, MRC promotes granule initiation in the leaves, as in Arabidopsis leaves, although the mutant phenotype in wheat appears to be less severe than in Arabidopsis. In Arabidopsis *mrc* mutants, there is one large granule per chloroplast rather than the multiple smaller granules in wild-type chloroplasts (Seung *et al.*, 2018; Vandromme *et al.*, 2019). Wheat *mrc* chloroplasts have fewer granules per chloroplast, but there are still multiple granules in each chloroplast (Figure 3.11). This difference in phenotype severity is similar to those observed for the orthologs of other granule initiation proteins of Arabidopsis (BGC1 and SS4) in wheat (Hawkins *et al.*, 2021; Watson-Lazowski *et al.*, 2022). The *mrc-1* and *mrc-2* mutants both have fewer granules per chloroplast compared to WT, with *mrc-1* having the strongest effect, consistent with the endosperm phenotype.

The primary effect of the *mrc* mutations in leaves was the reduction in granule number per chloroplast, but we also saw varying effects on granule size. In one set of experiments, we detected larger granules in *mrc-1* compared to WT, although in another set we did not (Figure 3.11). This appears to be linked to whether a change in total leaf starch content was observed or not, since in both experiments, there was a reduced number of starch granules in the mutants. In the experiment where there was no difference in total starch at the end of the day, the individual (fewer) granules were larger in the mutant, as would be expected. However, in the experiment where the total starch content was lower in the *mrc-1* mutant compared to WT, the increase in granule size was not observed.

In Arabidopsis, lack of MRC does not reduce the total end of day leaf starch content (Seung *et al.*, 2018; Vandromme *et al.*, 2019). Overall, in wheat leaves the effect of *mrc* mutations on the total end of day starch content was variable in our experiments (Figure 3.12). This could reflect differences in the regulation of starch turnover in wheat vs. Arabidopsis leaves, since soluble sugars make up the majority of the wheat leaf carbohydrate content and are thought to be more important as a storage reserve than starch (Müller *et al.*, 2018; Nie, 1995; Watson-Lazowski *et al.*, 2022). These differences may also contribute to the less severe effect on granule number in wheat *mrc* mutants compared to Arabidopsis leaves. The effect of MRC on leaf starch granule initiation is likely to differ depending on the developmental stage, so future studies should focus on the biochemistry of starch granule initiation across leaf and plant development to reveal a complete picture of the role of MRC and other granule initiation proteins in the wheat

chloroplast. In *Arabidopsis*, reduced *AtSS4* abundance has a much stronger effect on granule formation in younger leaves compared to mature leaves (Crumpton-Taylor *et al.*, 2013).

3.3.3 Exploring the biochemical basis of the contrasting roles of MRC between organs

MRC is conserved in land plants and is predicted to be a long coiled-coil protein with no known enzymatic domains ((Seung *et al.*, 2018), see also Chapter 4). It is possible that MRC can exert opposite effects on granule initiation by interacting with different interaction partners. The *Arabidopsis AtMRC* can interact directly with *AtSS4* in a yeast two-hybrid experiment (Vandromme *et al.*, 2019), and in *Arabidopsis* leaves *AtMRC* was also pulled down in association with *AtPTST2*, the ortholog of wheat *BGC1* (Seung *et al.*, 2018). Therefore, it is possible that in wheat, protein-protein interactions also play an important role in the function of MRC, although the nature of the interactions may be quite different. Nothing is currently known about the biochemical interactions of starch granule initiation proteins in wheat, but mutant studies have begun to uncover the distinct roles of *Arabidopsis* starch granule initiation orthologs in wheat, and as for MRC, they are all quite different from their role in *Arabidopsis*. In the endosperm, *SS4* restricts granule initiation during early grain development to ensure proper A-type granule formation and mutants defective in *SS4* form compound granules (Hawkins *et al.*, 2021). Similar, but less severe, is the *bgc1* mutant in wheat which also forms irregular compound-like granules but has a higher proportion of normal-looking A-type granules than *ss4* mutants (Chia *et al.*, 2020; Hawkins *et al.*, 2021). *BGC1* not only represses granule initiation during A-type granule formation, but also promotes B-type granule formation, as mutants with reduced gene dosage of *BGC1* have fewer B-type granules than WT, with no apparent impact on A-type granules (Chia *et al.*, 2017; Chia *et al.*, 2020).

However, a B-type granule suppressing mechanism during early grain development, like reported here for MRC, has not previously been described, and suggests that MRC and *BGC1* might have opposing roles in the wheat endosperm, despite their similar initiation promoting roles seen in leaves of *Arabidopsis* (Seung *et al.*, 2017; Seung *et al.*, 2018) and wheat (Watson-Lazowski *et al.*, 2022). In support of this, the RNA-sequencing analysis of the developing wheat endosperm shows opposite expression patterns for MRC and *BGC1* throughout grain development (Figure 3.3). The pattern of *BGC1* transcript accumulation is consistent with the pattern of protein accumulation observed through immunoblotting (Hawkins *et al.*, 2021).

Wheat *SS4* also promotes leaf granule initiation, as mutants had mostly starchless chloroplasts in the leaves (Hawkins *et al.*, 2021). Together with our finding for MRC, this suggests that there may

be more distinct differences for granule initiation proteins between different organs of the same species in comparison to the same organs from different species. The Arabidopsis granule initiation proteins localise to distinct puncta in the chloroplast, but we do not yet know anything about the localisation of these wheat orthologs in the chloroplast or amyloplast. In the future, exploring the protein localisation, protein-protein interactions, and the connection between the two in both tissues could reveal what underpins these distinct functions of MRC.

The specific effect of the L289F mutation in the *mrc-3* mutant on endosperm starch may provide an important clue in these studies, since *mrc-3* had an increased B-type granule percentage in the endosperm, but no differences in granule number in leaves. Determining whether the L289F mutation affects protein conformation, interactions and localisation might help dissect the differences in MRC function in leaves versus endosperm.

3.3.4 MRC as a gene target for biotechnological modification of starch granule size

These results establish MRC as a promising gene target for modifying starch granule size distribution in wheat, specifically to achieve smaller starch granules and a narrower granule size distribution range than conventional cultivars. This adds to the repertoire of lines with variable starch granule size distributions in wheat, as was done with the BlessT lines (Saccomanno *et al.*, 2022). Since the repression of B-type granule initiation is likely to be a role specific to Triticeae species carrying a bimodal size distribution of endosperm starch, it remains to be determined what the role of MRC is in cereal species that do not have a bimodal distribution of starch granules, such as those that have compound granules (e.g. rice) or simple granules (e.g. maize). Also, oats have a bimodal distribution of starch granules, with large compound granules and smaller simple granules. However, in oat, the smaller granules initiate at the same time as the larger compound granules (Saccomanno *et al.*, 2017), and it would therefore be interesting to determine if differences in MRC function play a role in timing the initiation of the small granules during oat endosperm development. Exploring the role of MRC in multiple crop species would therefore not only reveal the molecular differences that result in distinct spatiotemporal patterns of granule initiation between species, but also potentially increase its biotechnological potential.

Chapter 4 – Biochemical interactions between SS4 and MRC

4.1 Introduction

The proteins SS4 and MRC promote granule initiation in both Arabidopsis and wheat leaves, but the mechanisms by which they achieve this may differ between the two species, possibly facilitated by variations in protein-protein interactions. In this chapter, I will explore whether SS4 and MRC functions are conserved in Arabidopsis and wheat, investigate whether *AtSS4* and *AtMRC* could act together by colocalising, and compare the pairwise protein interactions between *AtSS4+AtMRC* and *TaSS4+TaMRC*.

4.1.1 SS4 and MRC are conserved in land plants

Starch synthases are generally conserved enzymes and fall into two distinct clades. One of these contains SS1, SS2 and GBSS, and the other contains SS3, SS4, SS5, SS6 (Abt *et al.*, 2020; Ball & Morell, 2003; Helle *et al.*, 2018; Leterrier *et al.*, 2008). SS4 is conserved in the green lineage, and there are SS4 orthologs identified in the green alga *Chlamydomonas reinhardtii* (Ball & Morell, 2003; Leterrier *et al.*, 2008). MRC is conserved in land plants, with an ortholog in the lycophyte *Selaginella moellendorffii* but none in green algae (Seung *et al.*, 2018; Vandromme *et al.*, 2019). Merida and Fettke (Merida & Fettke, 2021) proposed that there could be a core set of granule initiation proteins that includes SS4 that is already present in green algae, with other proteins like MRC only playing a role since the evolution of land plants.

In support of the conservation of SS4 function, the *Brachypodium distachyon* SS4 could complement the reduced growth and starch content phenotypes of the Arabidopsis *ss4* mutant (Raynaud *et al.*, 2016), but the role of SS4 in *Brachypodium* itself has not yet been characterised. The *ss4* mutant of durum wheat had a severe reduction in starch granule number in leaves, with 80% of chloroplasts not containing any starch, and the ones that did contain starch had multiple small round granules rather than the flattened ones in wild type (Hawkins *et al.*, 2021).

Arabidopsis *ss4* mutants also have many empty chloroplasts but the ones that do contain starch have a single, large, round granule compared to the multiple flattened ones in wild type (Roldan *et al.*, 2007). A similar comparison can be made for MRC, as seen in Chapter 3. MRC promotes granule initiation in wheat leaves, as it does in Arabidopsis, but the mutant phenotype is much less severe in wheat leaves.

Although the granule initiation function of SS4 and MRC is conserved in wheat compared to Arabidopsis leaves, the exact mechanism will likely differ in the two species. This could be due to both physiological and biochemical factors. Starch is a more important source of carbohydrates in Arabidopsis leaves than in wheat leaves, so its metabolism is likely not as tightly regulated in wheat compared to Arabidopsis leaves. Also, wheat does not contain the same set of granule initiation proteins as Arabidopsis. The proteins SS5 and PTST3 were identified as part of the Arabidopsis granule initiation machinery and are crucial for proper granule initiation as demonstrated by the fewer and larger granules in the *ptst3ss5* double mutant (more severe than *ptst3* and *ss5* single mutants) (Abt *et al.*, 2020). However, PTST3 and SS5 are not present in wheat (Abt *et al.*, 2020; Seung *et al.*, 2017). As PTST3 and SS5 interact with other granule initiation proteins in Arabidopsis (PTST2 and MRC respectively), there are known differences in the biochemical interactions of granule initiation proteins in the two species, and perhaps there are also biochemical differences between the common granule initiation orthologs in Arabidopsis and wheat.

4.1.2 The direct interaction between SS4 and MRC is of particular interest among SS4-associated proteins

Currently, the protein-protein interactions of SS4 and MRC have been best studied in Arabidopsis, but nothing is yet known about the interactions of the wheat orthologs. Both Arabidopsis and *Brachypodium* SS4 interacted with themselves in a bimolecular fluorescence complementation (BiFC) assay using transiently expressed proteins in *Nicotiana benthamiana* (Raynaud *et al.*, 2016). *AtSS4* also formed dimers in an *in vitro* assay using truncated recombinant proteins expressed in *E.coli*. A conserved region in between the N-terminal coiled-coils and the C-terminal catalytic domain (amino acids 471 – 515) was necessary for the *AtSS4-AtSS4* interaction in BiFC (Raynaud *et al.*, 2016), highlighted in Figure 4.1B in magenta. Two separate yeast two-hybrid (Y2H) screens to identify direct interactors of Arabidopsis SS4 have been published: one screen identified the fibrillin 1a (FBN1a) and fibrillin 1b (FBN1b) proteins (Gamez-Arjona *et al.*, 2014) and the other identified MRC (Vandromme *et al.*, 2019). The *fbn1afb1b* double mutant did not have a granule initiation phenotype, whereas the *mrc* mutant had a granule initiation phenotype where only one large granule was found in each chloroplast. This makes the MRC-SS4 interaction especially interesting to study.

Curiously, SS4 was found exclusively in the thylakoid fraction in one study, where it was also associated with the FBN proteins (Gamez-Arjona *et al.*, 2014), but in the stromal fraction in another study (Seung *et al.*, 2018). This unresolved localisation suggests it might be present in

both, in a dynamic localisation. AtSS4 and the FBN proteins were found in separate molecular weight fractions in a large-scale chloroplast proteomic study, not supporting the interaction of these proteins (Lundquist *et al.*, 2017). In the Plant Proteome Database (<http://ppdb.tc.cornell.edu/>), most of the experiments listed AtSS4 in the chloroplast stroma, but a few have also noted nucleoid or thylakoid associations.

The relationship between AtSS4 and AtMRC interaction and punctate localisation in the chloroplast still remains to be explored. AtSS4 (and *Brachypodium* SS4 expressed in *N. benthamiana*) and AtMRC both localised to distinct puncta in the chloroplast (Raynaud *et al.*, 2016; Seung *et al.*, 2018; Vandromme *et al.*, 2019). The localisation pattern of AtSS4 and AtMRC looked very similar and was different from the punctate localisation pattern of the other granule initiation proteins AtMFP1 and AtPTST2, in line with a specific interaction between AtSS4 and AtMRC (see Introduction). Their co-localisation has not been shown specifically, but in Arabidopsis protoplasts, AtSS4 and AtMRC were not necessary for the punctate localisation of one another (Vandromme *et al.*, 2019).

In this chapter, I examined Arabidopsis *ss4* and *mrc* lines complemented with their wheat orthologs and showed that the functions of the two proteins are in fact conserved to different extents. I studied the co-localisation of AtSS4 and AtMRC using transient expression in *N. benthamiana* and made the surprising discovery that AtMRC localisation can influence AtSS4 localisation in this system. I demonstrated using co-immunoprecipitation in *N. benthamiana* that the wheat orthologs of SS4 and MRC, like the Arabidopsis orthologs, interact *in planta*. The interaction between the wheat orthologs was confirmed using yeast two-hybrid, with which I additionally demonstrated that both the N and C terminus of *TaSS4* are sufficient for its interaction with *TaMRC* and with itself. Overall, this chapter expands on previous evidence of the pairwise interaction between SS4 and MRC, laying the groundwork for further comparisons between the biochemical mechanisms of Arabidopsis and wheat orthologs.

4.2 Results

4.2.1 *TaSS4* can partially complement *ss4* in Arabidopsis, but *TaMRC* cannot complement *mrc*

In previous studies of the wheat SS4 and MRC (Hawkins *et al.*, 2021) (Chapter 3 of this thesis), we focused on tetraploid durum wheat, *Triticum turgidum*, which has A- and B-genome homeologs of SS4 and an A-genome homeolog of MRC. The *T. turgidum* SS4 sequences were 99-100% identical to the bread wheat (*Triticum aestivum*) Chinese Spring reference genome, and the two *T. turgidum* homeologs were 97% identical to each other (Hawkins *et al.*, 2021). The *T. turgidum* *MRC-A1* sequence was identical to the *T. aestivum* one. For the remainder of the analyses on wheat SS4, the sequence of the B-genome copy of *T. aestivum* SS4 was used (*TaSS4-B1*), as the Chinese Spring reference genome has a premature stop codon in *TaSS4-A1* (Hawkins *et al.*, 2021). The wheat MRC will be referred to as *TaMRC*.

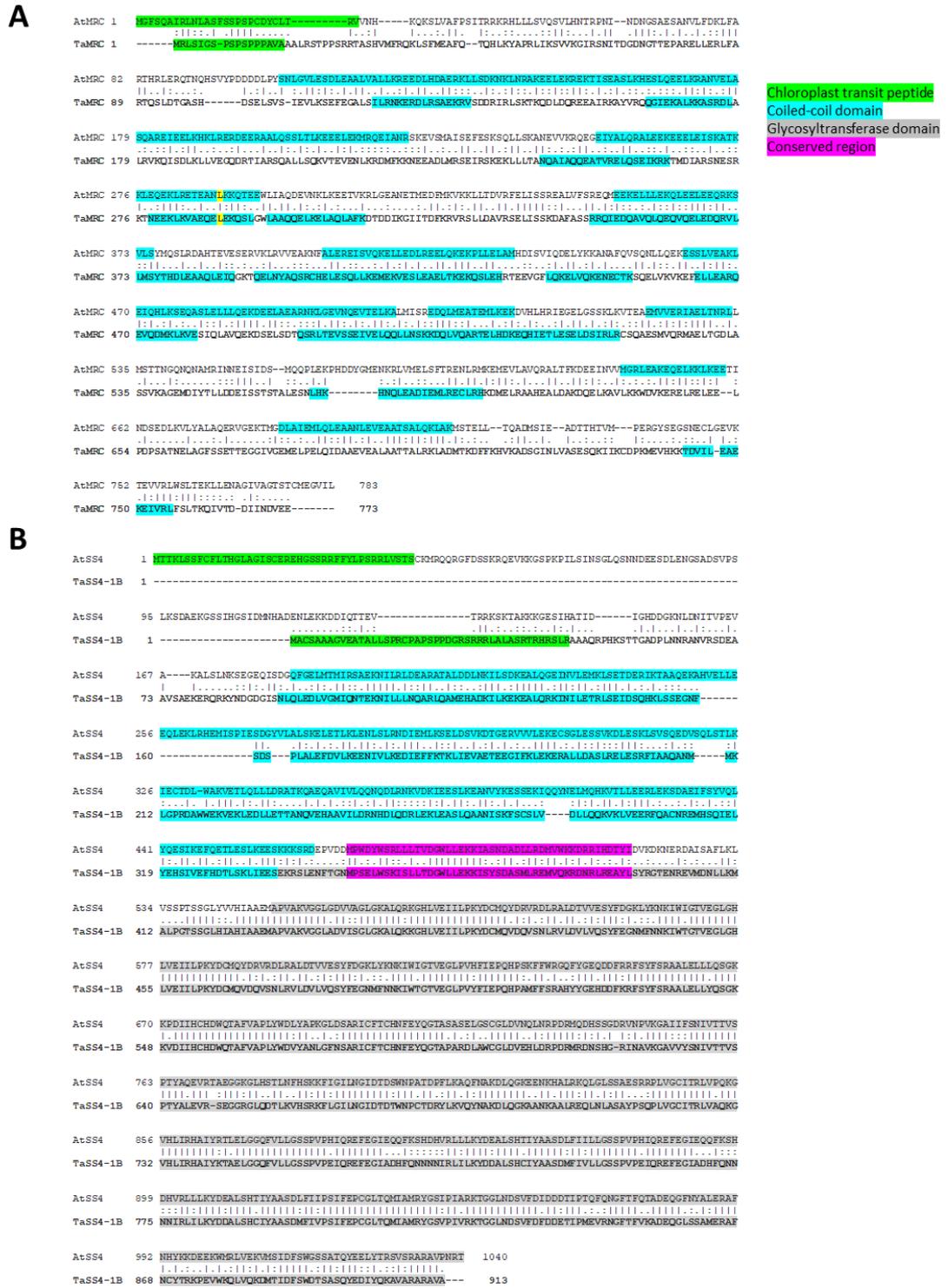
Pairwise alignments (using EMBOSS NEEDLE) of *AtMRC* and *TaMRC* amino acid sequences showed 31.6% sequence identity and 55.8% sequence similarity (Figure 4.1A). The *AtSS4* and *TaSS4-B1* sequences shared 48.2% sequence identity and 62.9% sequence similarity (Figure 4.1B). The two MRC orthologs shared coiled-coil predictions across the entire sequence (as predicted by PCOILS, Figure 4.1A) and the two SS4 orthologs shared coiled-coil predictions across their N-termini (Figure 4.1B). The sequence similarity, conservation of domains and predicted coiled coils suggest that some of the SS4 and MRC functions could be conserved in Arabidopsis and wheat.

To test for functional conservation of SS4 and MRC between Arabidopsis and wheat, I analysed transgenic Arabidopsis complementation lines expressing the wheat proteins. These lines were generated by David Seung: *TaSS4-B1-YFP* was expressed under the Ubiquitin 10 promoter (pUBQ, here named pUBC for the vector used with C-terminal tag) in an *ss4* mutant background and *TaMRC-HA* was expressed under the cauliflower mosaic virus 35S promoter (p35S) in an *mrc* mutant background (Figure 4.2). Two independent T1 lines were isolated for each of these lines using the phosphinothricin or kanamycin resistance selection markers and protein immunoblots to screen for lines expressing the tagged proteins. Then T2 plants from each of those lines were checked again by immunoblotting to select individuals with good protein expression for sectioning and light microscopy. These plants could be either homozygous or heterozygous and may contain multiple insertions.

Expression of the *TaSS4-B1-YFP* construct could partially complement the *ss4* mutant phenotype. The *ss4* plants were small and pale due to build-up of ADP-glucose (Crumpton-Taylor *et al.*, 2013; Ragel *et al.*, 2013), but the complemented lines were not pale and resembled the wild-type

(Figure 4.2A). As previously reported, the *ss4* mutants had many empty chloroplasts, with the chloroplasts containing starch having one large round granule. Light microscopy (LM) showed that the complemented lines contained multiple flattened granules in the chloroplasts, as the wild type (Figure 4.2A). Upon closer examination using transmission electron microscopy (TEM), some of the starch granules in the complemented lines resembled the flattened morphology of the wild type, but others were more irregularly shaped (Figure 4.2A). I contributed these results to the publication (Hawkins *et al.*, 2021).

Expression of the p35S:*TaMRC*-HA construct could not complement the *mrc* mutant phenotype, as the granule number and morphology was not visibly different from *mrc*, with the chloroplasts mostly containing one large flattened granule (Figure 4.2B). This suggests that the function of *TaSS4* in determining starch granule number and morphology is partially conserved in Arabidopsis, but that of *TaMRC* is not. As I hypothesised that *TaMRC* and *TaSS4* act together, I also generated an Arabidopsis line expressing both the p35S:*TaMRC*-HA construct and the *TaSS4*-1B-YFP construct, in an *mrc* mutant background. I crossed the two transgenic lines mentioned above, and selected F3 plants that had an *mrc* mutant background and were either expressing only the p35S:*TaMRC*-HA construct or both p35S:*TaMRC*-HA and pUBC:*TaSS4*-1B-YFP (Figure 4.2C). Both lines still had mostly one large, flattened granule per chloroplast, so the presence of both *TaSS4* and *TaMRC* is also not sufficient to complement the *mrc* granule initiation phenotype in Arabidopsis.



Chloroplast transit peptide
Coiled-coil domain
Glycosyltransferase domain
Conserved region

Figure 4.1 Pairwise amino acid sequence alignment between wheat and Arabidopsis MRC and SS4 shows high sequence similarity between Arabidopsis and wheat orthologs. **A)** Alignment between *AtMRC* and *TaMRC*-A1 orthologs. The residue highlighted in yellow is the L289 residue mutated in the *mrc-3* line from Chapter 3. Coiled coils are predicted by PCOILS with a score of >0.5 and a 14 amino acid window. **B)** Alignment between *AtSS4* and *TaSS4*-B1 orthologs (this is the sequence used for cloning in the rest of this thesis). The conserved domain was identified in (Raynaud *et al*, 2016). The predicted chloroplast transit peptides (CTPs) and domains for the proteins are highlighted. The MRC CTPs were predicted using WolfPsort and those for SS4 were predicted using ChloroP. Pairwise alignments were done using EMBOSS NEEDLE (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). (|) = identical residue, (:) = residues with strongly similar properties, (.) = residues of weakly similar properties.

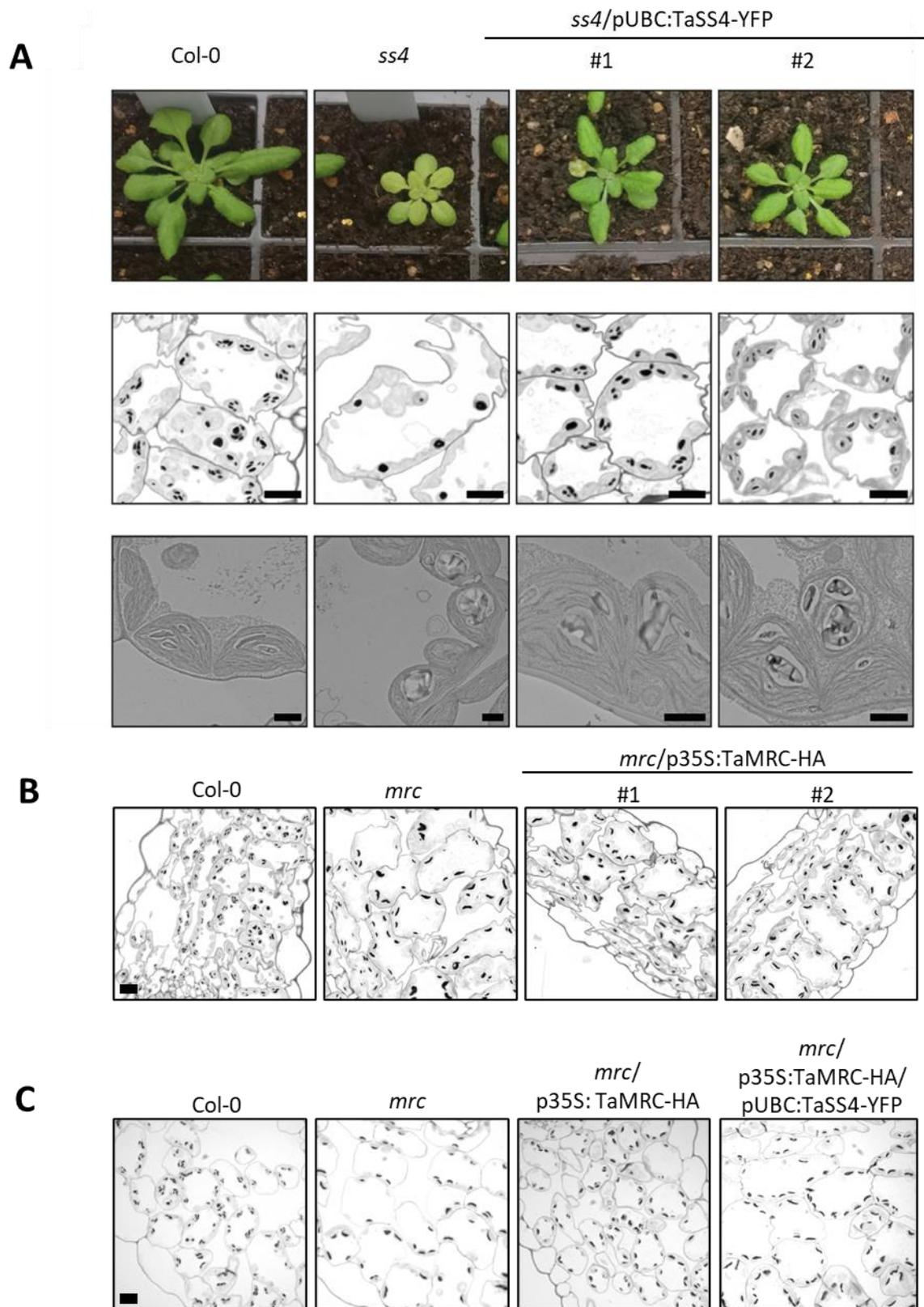


Figure 4.2 *TaSS4* can partially complement *Arabidopsis ss4*, but *TaMRC* cannot complement *mrc*. **A)** Growth phenotypes, light microscopy (LM) leaf sections and transmission electron microscopy (TEM) leaf sections of *Arabidopsis* lines expressing *TaSS4* in an *ss4* background. This figure appears in (Hawkins *et al*, 2021), and is under a CC-BY license. **B)** LM of leaf sections of *Arabidopsis* lines expressing *TaMRC* in an *mrc* mutant background. **C)** LM of leaf sections of *Arabidopsis* lines expressing *TaMRC* or *TaMRC* and *TaSS4* in an *mrc* mutant background. The plants were generated from a cross of the *mrc/p35S:TaMRC-HA* and *ss4/pUBC:TaSS4* lines, and were not genotyped for their *ss4* background.

4.2.2 AtSS4 and AtMRC colocalise to puncta in the chloroplast

AtSS4 and *AtMRC* both show distinct punctate localisation in the chloroplast (Lu *et al.*, 2018; Seung *et al.*, 2018; Szydlowski *et al.*, 2009; Vandromme *et al.*, 2019). Although their abundance looks similar, it is not known whether they are the same puncta. Therefore, I performed a co-localisation experiment of the two proteins. I cloned fluorescent protein-tagged (C-terminal tag) *AtSS4* and *AtMRC* under a Ubiquitin 10 (UBQ) promoter and expressed them transiently in *Nicotiana benthamiana* using Agrobacterium-mediated infiltration of leaves. For controls, I used fluorescent proteins fused to the Arabidopsis Rubisco small subunit chloroplast transit peptide (CTP) (Kim *et al.*, 2010), to facilitate chloroplast localisation. Confocal microscopy was used to study the localisation of these proteins in fresh *N. benthamiana* epidermal tissue.

Figure 4.3 shows the localisation pattern of the four constructs when expressed on their own. The CTP-YFP and CTP-RFP controls both localised diffusely in the chloroplast. *AtSS4*-YFP had a diffuse pattern in most chloroplasts as well, whereas *AtMRC*-RFP localised to distinct puncta in the chloroplast, as reported previously (Seung *et al.*, 2018; Vandromme *et al.*, 2019). The mostly diffuse localisation pattern of *AtSS4*-YFP was reproducible in multiple experiments, with only some chloroplasts containing puncta within the diffuse signal (Figure 4.4A). This localisation pattern for *AtSS4*-YFP was surprising, as previously a punctate localisation signal was reported for *AtSS4* when transiently expressed in *N. benthamiana* (Gamez-Arjona *et al.*, 2014) and after stable transformation in Arabidopsis (Lu *et al.*, 2018).

To test whether *AtSS4* and *AtMRC* co-localise, I co-expressed both *AtSS4*-YFP and *AtMRC*-RFP. After co-expression, very distinct co-localised puncta were seen, with no detectable diffuse *AtSS4*-YFP signal (Figure 4.4C). The change in *AtSS4*-YFP localisation, from diffuse signal to discrete puncta, was not observed when *AtSS4*-YFP was co-expressed with CTP-RFP – indicating that the effect on *AtSS4* localisation is specific to the co-expression with *AtMRC* (Figure 4.4B). This suggests that in the *N. benthamiana* transient expression system, presence of *AtMRC* promotes the specific localisation of *AtSS4* to puncta.

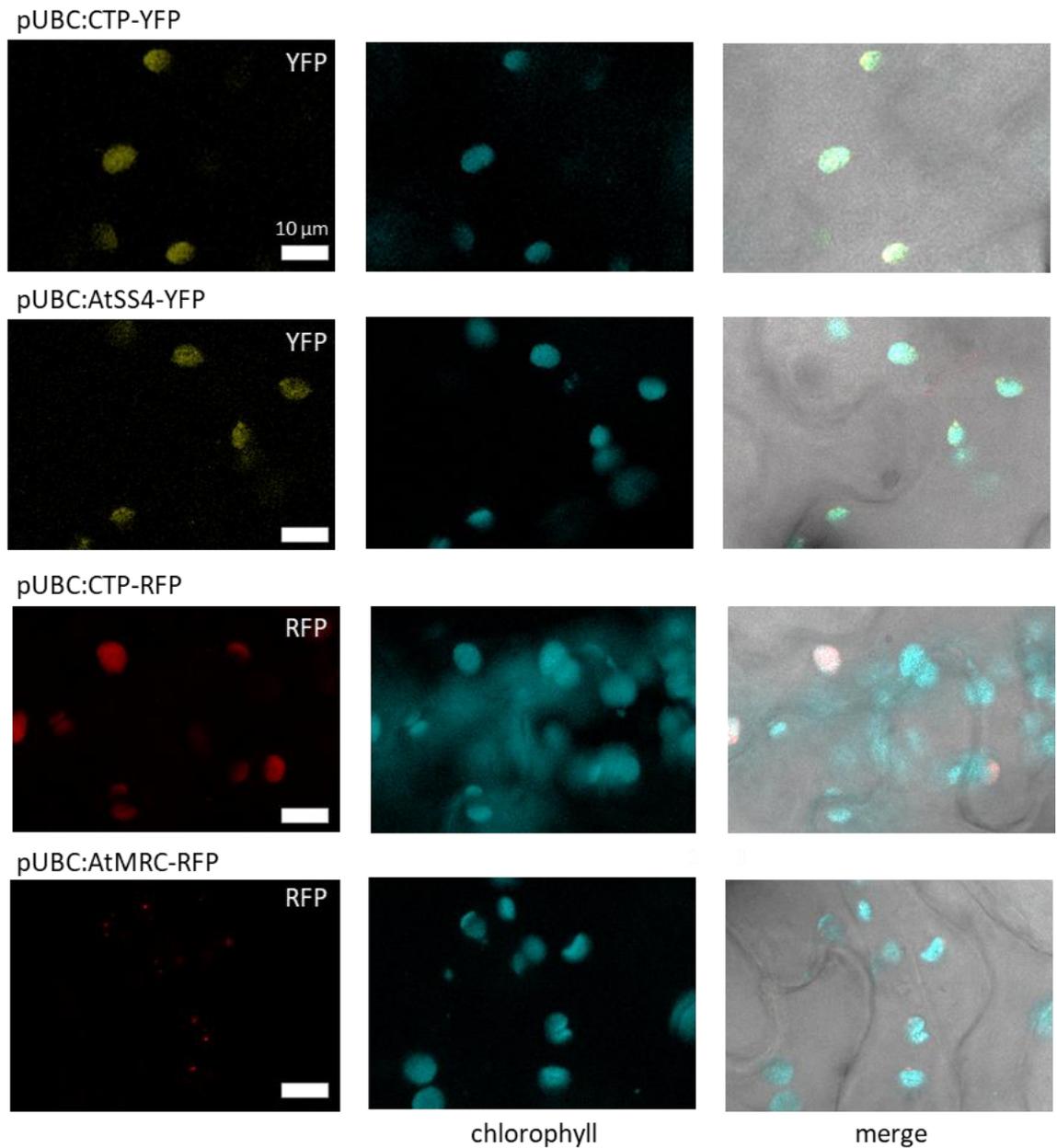


Figure 4.3 Localisation patterns of expression constructs in *Nicotiana benthamiana*. All constructs were under the Ubiquitin 10 (UBQ) promoter. YFP and RFP with a chloroplast transit peptide (CTP) from the Arabidopsis Rubisco small subunit were used as controls. Scale bar = 10 µm

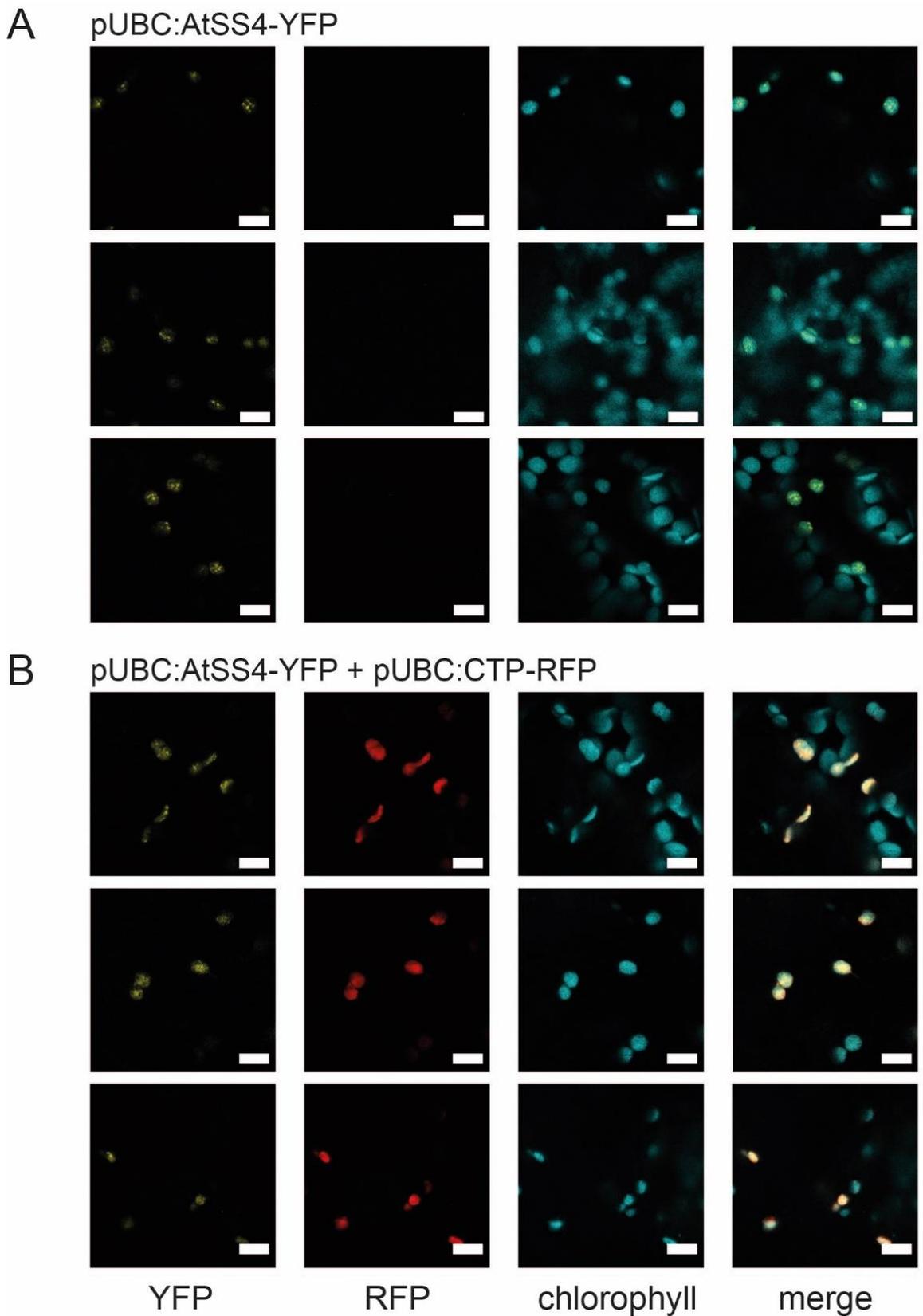


Figure 4.4 Co-localisation of AtSS4 and AtMRC using transient expression in *Nicotiana benthamiana*. Three independent plants for each combination of constructs are shown. The same microscope settings were used for all images. **A)** Localisation of AtSS4-YFP alone is mainly diffuse, with some puncta visible occasionally. **B)** Control sample showing diffuse localisation of CTP-RFP and diffuse localisation of AtSS4-YFP. Scale bar = 10 μ m. All plants were also infiltrated with the gene silencing suppressor P19.

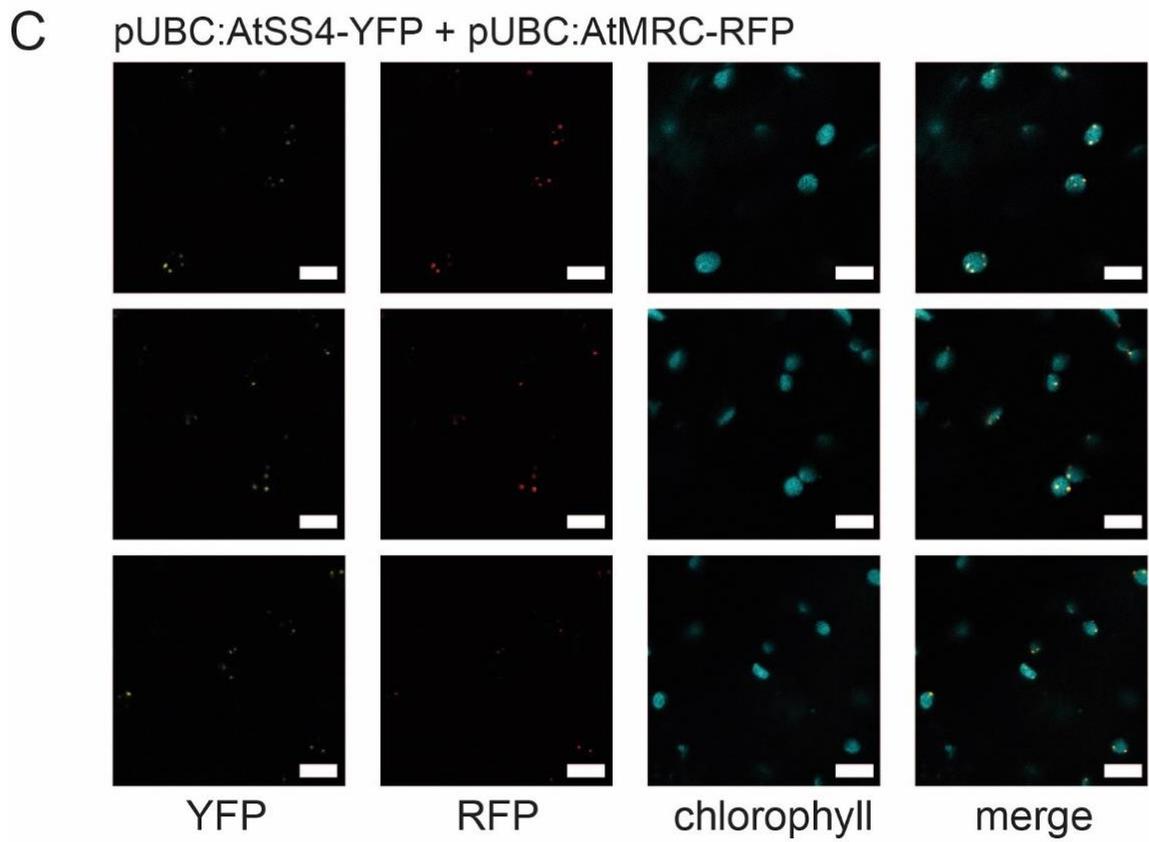


Figure 4.4 continued. C) Co-localisation of AtSS4-YFP and AtMRC-RFP in distinct puncta in the chloroplast. Scale bar = 10 μ m. All plants were also infiltrated with the gene silencing suppressor P19.

4.2.3 SS4 and MRC interact in co-immunoprecipitation

The co-localisation of *AtSS4* and *AtMRC* strongly supports the interaction of these two proteins observed using yeast two-hybrid (Vandromme *et al.*, 2019). To confirm whether the co-localisation in *N. benthamiana* was indeed a biochemical interaction *in planta*, I performed a co-immunoprecipitation (co-IP) using the same transient expression system as for the co-localisation. In this experiment, I used an *AtSS4*-HA construct (C-terminal tag) under the p35S promoter, along with the *AtMRC*-RFP construct from the localisation experiments. After co-expression in *N. benthamiana*, I made protein extracts and used magnetic beads to pull down *AtSS4*-HA. Immunoblotting the IP fraction showed the presence of *AtSS4*-HA, demonstrating the successful pulldown of *AtSS4*-HA, despite the weak bands in the input sample suggesting that the protein was not strongly expressed (Figure 4.5).

AtMRC-RFP was also present in the IP fraction when co-expressed with *AtSS4*-HA, suggesting it was pulled down together with *AtSS4*-HA. No RFP signal was detected for either of the negative controls, where *AtSS4*-HA was co-expressed with CTP-RFP or where *AtMRC*-RFP was expressed alone. Therefore, the presence of *AtMRC*-RFP in the IP was due to a specific interaction with *AtSS4*-HA, and not due to unspecific interaction of *AtSS4*-HA with the RFP tag or of *AtMRC*-RFP with the column. In the input, the signal for *AtMRC*-RFP looked stronger in the co-infiltration with *AtSS4*-HA compared to *AtMRC*-RFP alone, and both had a much weaker signal compared to the CTP-RFP control, although this was not quantified (Figure 4.5).

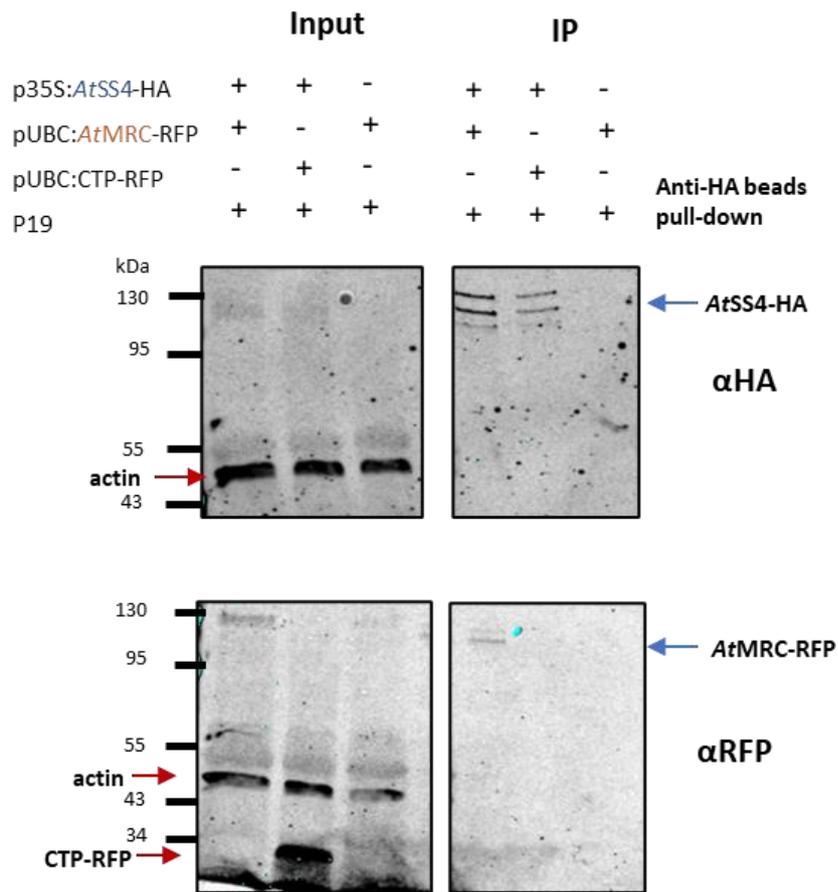


Figure 4.5 Co-immunoprecipitation (co-IP) of *AtSS4*-HA and *AtMRC*-RFP using transient expression in *N.benthamiana*. Co-IP was performed using anti-HA magnetic beads. 50 μ L of input and 10 μ L of IP were loaded. Anti-HA immunoblots are shown on top, and anti-RFP immunoblots are shown below.

I then explored whether the wheat orthologs of SS4 and MRC could also interact. I supervised a summer student, Stanislav Kurass, to perform a co-IP of *TaSS4* and *TaMRC* expressed in *N. benthamiana*, including truncations of *TaSS4*, previously cloned by David Seung (Figure 4.6A). *TaSS4* N is the N-terminal region of *TaSS4*, including the coiled coils but not including the conserved region identified in (Raynaud *et al.*, 2016) (see Figure 4.1B). *TaSS4* dCC is the protein lacking the coiled coil region, including a short N-terminal sequence, the conserved region and the C-terminal glycosyltransferase domain. All *TaSS4* constructs were expressed under the 35S promoter and had a C-terminal YFP tag. *TaMRC* was expressed under the pUBQ promoter, with a C-terminal RFP tag. *TaMRC*-RFP was pulled down together with the full length *TaSS4*. The truncated *TaSS4* constructs were also able to pull down *TaMRC*-RFP, but the *TaMRC*-RFP signal on the immunoblot was much weaker for these samples compared to the pull-down with the full-length *TaSS4*-YFP, when equal volumes were loaded. Therefore, both the *TaSS4* coiled coil region and *TaSS4* without the coiled coil were sufficient for interaction with *TaMRC*, but the interaction was likely strongest with the full-length *TaSS4*.

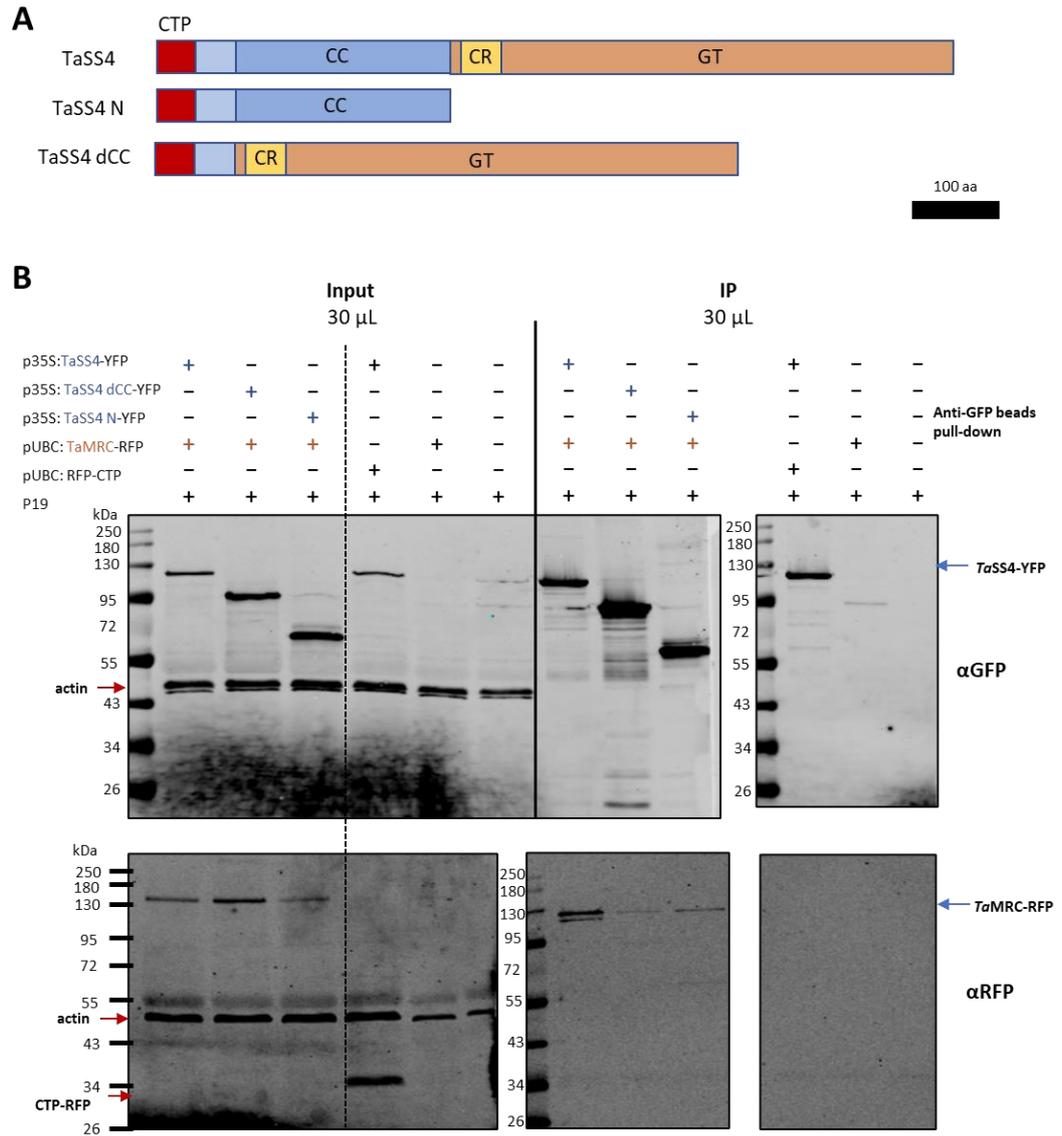


Figure 4.6 Co-immunoprecipitation (co-IP) of *TaSS4* and *TaMRC* using transient expression in *N. benthamiana*. **A)** Schematic of full length and truncated *TaSS4* proteins used in this experiment. The *TaSS4* dCC construct contained the CTP and a short N-terminal region before the coiled coil. CTP = chloroplast transit peptide, CC = coiled coil, CR = conserved region, GT = glycosyltransferase domain. **B)** Co-IP of full length and truncated *TaSS4* with full length *TaMRC*, performed using transient expression in *N.benthamiana* leaves. Anti-GFP immunoblots are shown on top, and anti-RFP immunoblots are shown below. 30 μ L of input and 30 μ L of IP were loaded.

4.2.4 *TaSS4* and *TaMRC* interact in yeast two-hybrid

To test whether the interaction between *TaSS4* and *TaMRC* observed in the co-IP was likely to be a direct interaction, I used a yeast two-hybrid assay (Y2H). I also tested interactions with the truncated versions of *TaSS4*, as used in the IP experiment in Figure 4.6A. These experiments were performed under my supervision by Stanislav Kurass and Brendan Fahy. *TaSS4* and *TaMRC* were cloned into pGBKT7 (with the GAL4 binding domain, BD) and pGADT7 (with the GAL4 activation domain, AD), which are N-terminally tagged Y2H vectors. In Y2H, the full-length *TaSS4*, *TaSS4* N and *TaSS4* dCC were all sufficient to interact directly with *TaMRC* (Figure 4.7), in accordance with the results of the co-IP. The Y2H assay is not technically quantitative, therefore it is difficult to draw conclusions on the strength of these interactions, but when looking at the most diluted samples (1/500), the yeast cells expressing full length *TaSS4* + *TaMRC* had more growth than the truncated *TaSS4* combinations, also consistent with the co-IP.

These constructs had the CTP of both *TaSS4* and *TaMRC* included. In yeast, the CTP is unlikely to be cleaved as it is in plants, so the proteins will be expressed with their N-terminal tags attached. The presence of the CTP also likely does not affect protein folding and function, as I previously also saw interaction between *TaMRC* and *TaSS4* using constructs without the CTP. Some pink colonies can be seen in the assay, especially on the -LW plates, which is due to pigment accumulation under low levels of adenine (Smirnov *et al.*, 1967).

We also included combinations to test for oligomerisation of either protein, since *AtSS4* was previously reported to dimerise, and nothing is known about whether MRC can oligomerise. *TaSS4* was able to interact with itself (Figure 4.7, middle panel). The full length *TaSS4*, *TaSS4* N and *TaSS4* dCC were all sufficient to interact with any of these *TaSS4* fragments. The conserved region (Raynaud *et al.*, 2016) was not necessary for the interaction of *TaSS4* fragments, as *TaSS4* N without the conserved region was sufficient for *TaSS4*-*TaSS4* interaction. This is different to what was reported for *AtSS4*, where the conserved region was necessary for *AtSS4*-*AtSS4* interaction in bimolecular fluorescence complementation (BiFC) (Raynaud *et al.*, 2016).

The *TaMRC* + *TaMRC* assay did not show any yeast growth, suggesting *TaMRC* does not interact with itself under these conditions. However, it should be noted that yeast carrying *TaMRC* with the AD did not grow as well as yeast carrying *TaMRC* with the BD, as demonstrated by decreased yeast growth when looking at both orientations of *TaMRC* + *TaSS4* (Figure 4.7, bottom panel). *TaMRC* + *TaSS4* dCC grew with *TaMRC* with the BD but did not grow with *TaMRC* with the AD. The assay with *TaMRC* with the AD is therefore not as sensitive as with the BD. It is possible that a very weak interaction of *TaMRC* with itself could be missed due to this lower sensitivity.

Overall, Y2H provides strong evidence for an interaction between *TaSS4* and *TaMRC* and oligomerisation of *TaSS4*, but there is no evidence to support oligomerisation of *TaMRC*.

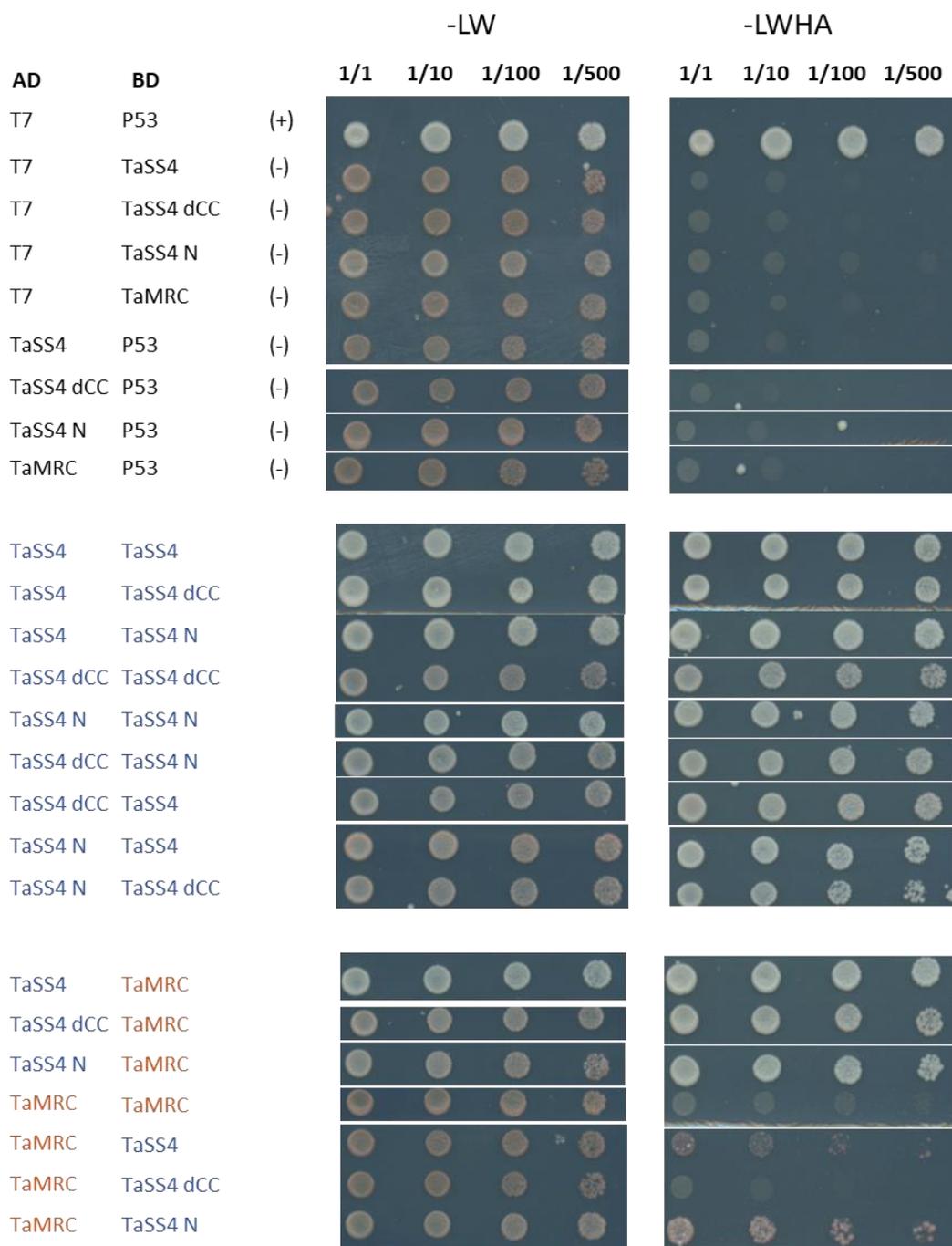


Figure 4.7 Yeast two-hybrid of pairwise interactions between *TaSS4* and *TaMRC*. Images show all pairwise comparisons between truncated and full length *TaSS4* with full length *TaMRC*, in both the binding domain and activation domain vector. Controls are on top, with *TaSS4* interactions in the middle and *TaSS4* + *TaMRC* interactions on the bottom. AD = GAL4 activation domain (pGADT7 vector), BD = GAL4 binding domain (pGBKT7 vector). -LW = double dropout media without leucine and tryptophan, -LWHA = quadruple dropout media without leucine, tryptophan, histidine, and adenine. Dilutions of yeast cultures are indicated as 1/1, 1/10, 1/100, 1/500. (+) = positive control, (-) = negative control.

4.3 Discussion

In this chapter, I have shown that both *AtSS4+AtMRC* and *TaSS4+TaMRC* can interact directly with each other, in a heterologous yeast system and in the plant system of *N. benthamiana*. There is partial conservation of SS4 function in wheat and Arabidopsis leaves, as *TaSS4* could complement the Arabidopsis *ss4* phenotype. However, *TaMRC* could not complement the Arabidopsis *mrc* phenotype, even in the presence of *TaSS4*. My protein interaction studies suggest that this divergence does not come from a difference in the ability of the SS4 and MRC orthologs to interact with each other. However, there may still be ortholog-specific properties of the domains involved in protein interaction, species-specific availability of other granule initiation proteins, and different cellular and physiological contexts that may influence how the proteins act between the two species.

4.3.1 Both the N-terminus and C-terminus of *TaSS4* are sufficient for protein interactions

In the co-IP and Y2H assays of *TaSS4* and *TaMRC* (Figure 4.6, 4.7), full length *TaSS4*, *TaSS4* N (without the conserved region) and *TaSS4* dCC were all sufficient for interaction of *TaSS4* with itself and with *TaMRC*, therefore neither the coiled coil nor the glycosyltransferase domain was specifically necessary, suggesting the entire *TaSS4* protein is involved in protein interactions. However, the full length *TaSS4* might still be the most effective interactor, as in the co-IP the interaction of *TaMRC* with full length *TaSS4* seemed stronger than with both truncated versions. In the Y2H this difference was not as obvious, although there were some small decreases in yeast growth for the truncated *TaSS4*.

The co-IP may be a more sensitive assay than the Y2H. The nature of the readouts in co-IP and Y2H are very different, and the fact that I detected interactions between SS4 and MRC via these two different methodologies lends robustness to the results. In Y2H, the readout is transcriptional activation of a gene that results in yeast growth, whereas in co-IP, the immunoblot detection is a more direct readout of the presence of a protein partner. The yeast system is also a more heterologous one, where the proteins are localised to the nucleus in an entirely different environment, whereas the co-IPs performed here are still in a plant system, albeit not a native one. This makes the yeast less reflective of the *in planta* behaviour of the proteins, but also more likely to provide an indication of direct protein-protein interactions, rather than through other potential protein interactors in the plant chloroplast environment. It is possible that the protein interactions in the chloroplast are more transient, possibly due to presence of other granule

initiation proteins. Although both systems are overexpressing the proteins, the relative amounts of protein expression may differ between the two systems, which could affect the sensitivity of the assays.

Raynaud *et al.* (2016) saw that the conserved region between the coiled-coils and the glycosyltransferase domain in SS4 was necessary for the dimerization of AtSS4 in BiFC. Using truncated AtSS4 constructs, they also showed that the coiled coil was not necessary for this dimerization. However, they did not test whether the coiled-coil domain without the conserved region would be sufficient for dimerization, as they used truncations where the conserved region was included in the N-terminal AtSS4 construct. In this chapter, *TaSS4 N* did not include the conserved region, and *TaSS4 dCC* did. To make any clear comparisons between the Arabidopsis and wheat orthologs, I would need to test the impact of equivalent truncations in AtSS4 and *TaSS4* on their pairwise protein interactions, but the current results suggest that the conserved region is not specifically important in the studied *TaSS4* protein interactions.

4.3.2 The granule initiation protein interaction networks are different in Arabidopsis and wheat

Although both the Arabidopsis and wheat SS4 and MRC interact with each other, the biochemical context in which this interaction happens is different. Wheat does not have all the granule initiation proteins that are important in Arabidopsis (PTST3 and SS5), so different interactions with other granule initiation proteins might contribute to the distinct mechanisms by which the proteins act in these two species. Perhaps this is part of the reason why *TaMRC* cannot complement the Arabidopsis *mrc* phenotype.

It is possible that different SS4 domains are important for different interaction partners. AtSS4 was previously pulled down with AtPTST2 in an affinity purification mass spectrometry experiment, and a co-IP showed that the AtSS4 coiled coil was not necessary for interaction with AtPTST2 (Seung *et al.*, 2017). The AtSS4 coiled coil was necessary for its interaction with AtFBN1b in BiFC, and the conserved region was not (Raynaud *et al.*, 2016). However as mentioned above, Raynaud *et al.* (2016) did not test whether that same AtSS4 coiled coil fragment that interacted with AtFBN1b could dimerise with other AtSS4 fragments. We also do not know if fibrillin proteins are involved in SS4 interactions in wheat, and if this would be relevant in the amyloplast in the endosperm. Fibrillin proteins are associated with plastoglobules, lipoprotein subcompartments which are attached to thylakoid membranes (Austin *et al.*, 2006), which are not present in amyloplasts.

It is uncertain whether *Ta*MRC interacts with itself in Y2H, but any interaction is likely to be quite weak. I would have to confirm any lack of interaction using additional methods such as co-IP. The only other interaction that has been identified for MRC is between *At*MRC and the catalytically inactive *At*SS5 using co-IP (Abt *et al.*, 2020). Plants lacking *At*SS5 have fewer starch granules per chloroplast, but the degree of reduction is not as severe as *ss4* or *mrc* mutants. *At*SS5 can interact with itself and with *At*MRC, and its N-terminal coiled coil domain is necessary for both these interactions. *At*SS5 was not identified in a pull-down with *At*SS4, so it is uncertain whether *At*SS4, *At*SS5 and *At*MRC interact in one complex and Abt *et al.* (2020) hypothesised that *At*MRC might interact with *At*SS4 and *At*SS5 at different times. If we can determine the exact interaction domains between *At*SS4 and *At*MRC, then it would be interesting to compare these with the interaction domains of *At*SS5 and *At*MRC, as well as those of *At*SS4 and the *At*FBN proteins. However, there is no SS5 ortholog in wheat. If interactions with SS5 and FBN are lacking for the wheat MRC and SS4 orthologs, the different cellular context in which *Ta*SS4 would interact with itself and with *Ta*MRC could contribute to a different mechanism of interaction.

4.3.3 *At*MRC is not necessary but sufficient for directing the punctate localisation of *At*SS4

Curiously, in the current study *At*SS4-YFP localised both diffusely and in puncta under transient expression in *N. benthamiana*, different from the solely punctate localisation seen in Arabidopsis in multiple studies. When Raynaud *et al.* (2016) looked at transient expression of *At*SS4 in *N. benthamiana*, they saw distinct puncta as in Arabidopsis. The simultaneously diffuse and punctate pattern that I saw looks more like the previously observed localisation pattern of the truncated *At*SS4 fragment with some of the coiled coil region removed (Gamez-Arjona *et al.*, 2014; Raynaud *et al.*, 2016). However, these studies did use the p35S promoter, rather than the UBQ10 promoter used in the current study. It is possible that the p35S promoter enhances punctate localisation due to overexpression artefacts. When I expressed the CTP-RFP control construct under a p35S promoter, this also displayed punctate localisation, demonstrating the importance of the promoter when looking at protein localisation. Previous studies in plants have highlighted that the p35S promoter may cause localisation artefacts (Zhu *et al.*, 2018). A survey of human proteins determined a negative correlation between gene expression and protein aggregation rates, possibly demonstrating that in general, the aggregation rates of proteins are finely tuned and overexpression could cause aberrant aggregation of proteins that assemble into puncta (O'Connell *et al.*, 2012; Tartaglia *et al.*, 2007). I used the UBQ10 promoter in this study as this did not induce punctate localisation of CTP-RFP, although the expression was still much higher than *At*MRC-RFP (Figure 4.5) or *Ta*MRC-RFP (Figure 4.6).

It is possible that expression of Arabidopsis proteins in the non-native *N. benthamiana* system influences their localisation patterns. The localisation pattern of AtSS5 was distinctly punctate when observed in Arabidopsis, but it was both diffuse and punctate under transient expression in *N. benthamiana*, even under the p35S promoter (Abt *et al.*, 2020). In the current study, AtSS4 was simultaneously diffuse and punctate when expressed on its own but was exclusively punctate when co-expressed with AtMRC. Using double mutant studies, Vandromme *et al.* (2019) had shown that AtMRC and AtSS4 do not need one another for their punctate localisation patterns in Arabidopsis protoplasts, but perhaps other Arabidopsis starch initiation proteins could be accounting for this punctate localisation. In the absence of the Arabidopsis orthologs in a *N. benthamiana* system, AtSS4 alone might be unable to specifically localise to puncta without any of its native partners. I propose that AtMRC is not necessary but sufficient for directing AtSS4 punctate localisation. It would also be interesting to see if any other combinations of granule initiation proteins produce similar results, such as AtSS4 + AtPTST2 localisation. The punctate AtPTST2 localisation pattern is not the same as that of AtSS4, but AtPTST2 localisation may still influence that of AtSS4, and AtPTST2 is hypothesised to bring glucan substrates to AtSS4 (Seung *et al.*, 2017).

The AtSS4 N-terminus (including conserved region) is crucial for ensuring correct starch granule morphology, anisotropic growth of Arabidopsis starch granules and for correct punctate localisation of the protein in the chloroplast (Burgy *et al.*, 2021; Lu *et al.*, 2018; Raynaud *et al.*, 2016). Future studies that focus on the interaction domains of AtSS4 and AtMRC might reveal whether any role of the AtSS4 N-terminus in protein interactions is connected to its importance for punctate localisation. The biological function of this punctate localisation is not known, and it is tempting to speculate that the puncta could correspond to starch granule initiation points. Punctate protein localisation has been associated with concentrating protein complexes into dynamic biomolecular condensates (Banani *et al.*, 2017; O'Connell *et al.*, 2012), and studying the dynamics of granule initiation proteins could help in understanding whether these processes are also involved in the biochemistry of starch granule initiation.

In this chapter, I have shown similarities and differences between Arabidopsis and wheat granule initiation protein orthologs. I have shown the pairwise interaction of AtSS4 and AtMRC proteins, confirming their direct interaction and co-localisation, but this does not reveal anything about the native cellular context in which these interactions happen. For this, I looked at the native multimeric state of AtSS4 in leaf tissue in Chapter 5.

Chapter 5 – SS4 dimers could be part of a large protein complex

5.1 Introduction

Most of our knowledge on the molecular functions of SS4 and MRC come from *Arabidopsis*, therefore I decided to initially explore the biochemistry of the *Arabidopsis* orthologs in more detail. Over the years, increasing evidence has supported the crucial role of AtSS4 in the process of starch granule initiation. We have learned a lot about its role in the control of granule morphology, from the initial identification of the reduced numbers of starch granules and reduced growth phenotype of *Arabidopsis ss4* mutants (Roldan *et al.*, 2007), to further characterisation of its protein interactions, activity and localisation (Gamez-Arjona *et al.*, 2014; Lu *et al.*, 2018; Raynaud *et al.*, 2016; Szydlowski *et al.*, 2009), its substrate priming role in starch biosynthesis (Crumpton-Taylor *et al.*, 2013; Ragel *et al.*, 2013) and most recently its specific role in the anisotropic growth of *Arabidopsis* leaf starch granules (Burgy *et al.*, 2021). However, we still do not understand the basic biochemical mechanisms that allow AtSS4 to perform these roles, and how they are related to its starch synthase activity.

Elucidating the structure of SS4 and understanding exactly how it interacts with other proteins will greatly advance our understanding on these biochemical mechanisms. The crystal structure of the conserved catalytic C-terminal domain of AtSS4 was determined in 2018 (Nielsen *et al.*, 2018), but the structure of the full-length protein including its unique N-terminus remains unknown, other than that the N-terminus is predicted to have coiled coils (Leterrier *et al.*, 2008; Szydlowski *et al.*, 2009). The next big aim would be to uncover the structural mechanism of how the N-terminus affects AtSS4 interaction and activity, considering that it is precisely this domain that seems to be necessary for the punctate localisation of AtSS4, and its function in determining correct granule shape and growth (Burgy *et al.*, 2021; Lu *et al.*, 2018). As coiled-coil domains are known for their protein-protein interactions (Mason & Arndt, 2004), it is reasonable to speculate that the SS4 N-terminus has a specific role in protein-protein interactions. However, as discussed in Chapter 4, the wheat SS4 N-terminal coiled-coil domain does not seem to be strictly necessary for its dimerization or interaction with MRC, but it can still be sufficient for mediating these interactions, and the interactions are stronger with than without it. It may also be involved in other protein interactions. AtSS4 also associates with AtPHS1 in immunoprecipitation (Malinova *et al.*, 2018), and as discussed, interacts with FBN1a/FBN1b at the thylakoid membrane (Gamez-Arjona *et al.*, 2014; Raynaud *et al.*, 2016), and was pulled down with PTST2 in the stroma (Seung *et al.*, 2018) to make up the hypothesised 'granule initial' (Seung & Smith, 2019).

There is some indication from large-scale proteomic fractionation experiments that AtSS4 may exist in Arabidopsis leaves as a large protein complex. Forming large protein assemblies can be a way to concentrate enzyme substrates. In plant cellulose biosynthesis, cellulose synthase proteins likely organise into a hexamer of trimers (Purushotham *et al.*, 2020), synthesising several glucan polymers in proximity, possibly facilitating the crystallisation of these individual chains into a microfibril that forms the ordered cellulose structure (Haigler & Roberts, 2019; Purushotham *et al.*, 2020). Some mutants in cellulose synthase proteins have altered cellulose crystallinity (Haigler & Roberts, 2019). For starch, it is still not known exactly how the nucleation of a single granule takes place, or how the elongation of simple glucan chains transitions to a semi-crystalline granule and polymer elongation on the granule surface. In the Plant Proteome Database, SS4 (113 kDa) was found in a >600 kDa fraction from size exclusion chromatography (SEC) separation of chloroplast stroma extract (Olinares *et al.*, 2010). Blue Native PAGE fractionation of chloroplast proteins identified SS4 in the largest molecular weight fraction, above the 440 kDa marker (Lundquist *et al.*, 2017). How the formation of these large complexes relates to the observed interactions of AtSS4 with itself and other proteins is not known.

In this chapter, I examined protein complex formation of the Arabidopsis and wheat SS4 proteins (AtSS4 and TaSS4-B1) in leaves. I saw that both Arabidopsis and wheat SS4 natively exist at high molecular weight (HMW) fractions when leaf protein extracts were analysed by size exclusion chromatography (SEC). I generated transgenic Arabidopsis lines with tagged AtSS4 and AtMRC, providing a useful resource for further protein characterisation. Using one of these lines, I performed affinity purification mass spectrometry (AP-MS) with AtSS4-HA as bait and I only pulled down AtSS4 itself as a significant hit, suggesting that perhaps AtSS4 mainly exists as a homocomplex. Recombinant AtSS4 and TaSS4 in expressed in *E. coli* also existed at HMW fractions in SEC, suggesting that the proteins can form higher order structures by themselves. However, negative stain transmission electron microscopy (NS-TEM) of recombinant AtSS4 identified particles that were most consistent with AtSS4 dimers. The discrepancy between the particles observed with NS-TEM and the HMW elution pattern observed in SEC could suggest that there are heterogenous assemblies of SS4 oligomers *in vitro*, with smaller dimeric assemblies being more stable. Alternatively, AtSS4 could function as a dimer but have an extended shape that causes the SEC elution pattern to run higher than the expected size of a dimer formed by a globular protein.

5.2 Results

5.2.1 SS4 is part of a high molecular weight assembly in Arabidopsis leaves and wheat leaves

5.2.1.1 SS4 exists at high molecular weight in Arabidopsis leaves

Considering the presence of AtSS4 in high molecular weight fractions in large scale proteomic experiments, I decided to confirm this by probing for AtSS4 in size-exclusion chromatography fractions. I extracted total soluble protein from 6-week-old whole Arabidopsis rosettes, concentrated this to between 6 – 27 µg/µL total protein (Col-0: 6 – 8 µg/µL total protein, *ss4*: 8 – 12 µg/µL total protein, *mrc*: 11- 27 µg/µL total protein) and ran 100 µL of the extract on a Superose 6 Increase 10/300 size exclusion chromatography (SEC) column, with a fractionation range of 5 – 5000 kDa for globular proteins.

I ran an immunoblot of the fractions with an anti-AtSS4 antibody (Roldan *et al.*, 2007). In the Col-0 wild-type plants, I detected AtSS4 mostly in three fractions, with two of them corresponding to the elution fractions of a 669 kDa standard (Figure 5.1A, for elution of standards see Appendix 5). As the mature monomer of AtSS4 is only 113 kDa (without chloroplast transit peptide), this confirms the presence of AtSS4 in a high molecular weight species. I did not detect AtSS4 in the monomer or dimer range, which would be around fractions 17 or 18. In the *ss4* negative control plants, there was no AtSS4 signal. In the *mrc* plants, the elution pattern was similar to the wild-type plants, indicating there were no major changes in the assembly of AtSS4 when lacking a functional AtMRC. These elution patterns were reproducible, as they were seen for three replicate protein extracts for each genotype (Appendix 6).

5.2.1.2 SS4 exists at high molecular weight in wheat leaves

As SS4 is important in both Arabidopsis and wheat leaves (Chapter 3, (Hawkins *et al.*, 2021)) and wheat does not have all the proteins involved in granule initiation in Arabidopsis (SS5, PTST3) (Abt *et al.*, 2020; Seung *et al.*, 2017), I investigated whether the high molecular weight complex seen in Arabidopsis leaves is also present in wheat leaves. I performed SEC on total soluble leaf protein from 10-day-old wheat leaves. Here too, I saw the presence of TaSS4 at high molecular weight

fractions, similar to Arabidopsis. *TaSS4* is present in fractions 14 and 15, where the two highest markers at 669 and 440 kDa elute (Figure 5.1B, red asterisks). The protein band corresponding to *TaSS4* was absent in the *ss4* control, but I also observed strong unspecific bands in the *ss4* control that ran very close to the *TaSS4* band (Figure 5.1B, blue circles). This is likely due to the anti-*TaSS4* antibody being less specific than the anti-*AtSS4* antibody (Hawkins *et al.*, 2021) and the prominence of this band in blots with fractions from the *ss4* control may be due to the absence of the *TaSS4* protein, which could favour unspecific binding. The unspecific band was seen in almost all of the fractions of the wheat *ss4* control samples. This experiment still requires some optimisation to resolve whether the exact fractions where *TaSS4* elutes are the same as *AtSS4*, but the results clearly demonstrate that SS4 is also present at high molecular weight in wheat leaves.

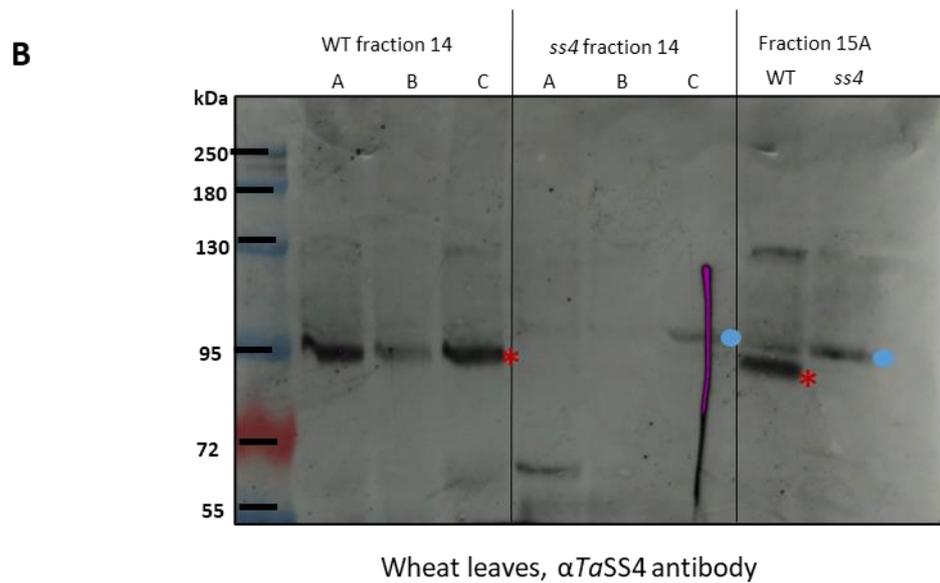
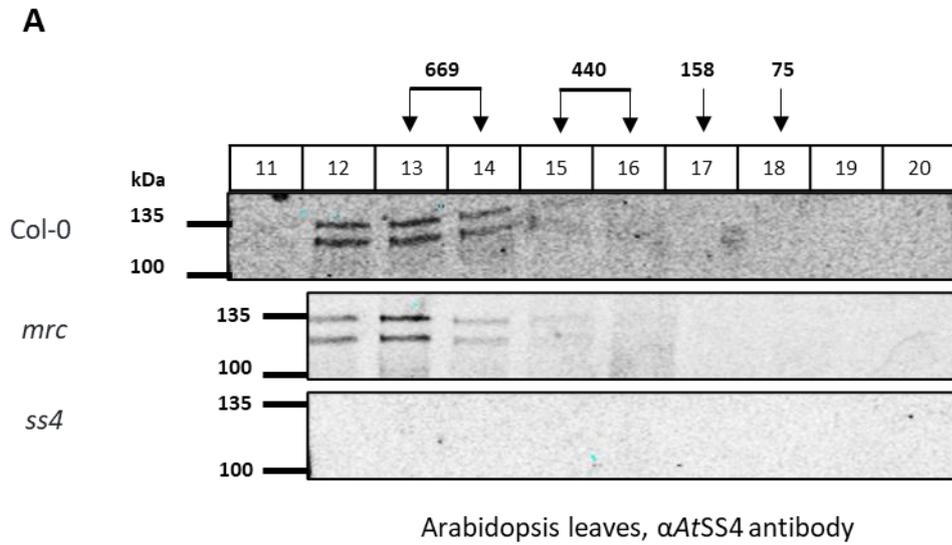


Figure 5.1. SS4 exists at high molecular weight in size exclusion chromatography (SEC) fractions of Arabidopsis and wheat leaf total soluble protein extracts. A) Immunoblot of SEC fractions of extracts from whole rosettes of 6-week-old Arabidopsis plants, for Col-0, *mrc*, and *ss4* as a negative control. All three samples were blotted with anti-AtSS4 antibody (Roldan *et al.*, 2007). Equal volumes were loaded in each lane. These are representative blots from 3 biological replicates, see Appendix 6. **B)** Immunoblot of SEC fractions 14 and 15 (~440 – 669 kDa) from 10-day-old wheat leaf extracts, for WT and *ss4*. A, B and C are biological replicates. An anti-TaSS4 antibody was used (Hawkins *et al.*, 2021). Red asterisks indicate the TaSS4 band in WT, whereas blue circles indicate the non-specific background band just above, in *ss4*. A) was detected by infrared fluorescence, B) was detected by chemiluminescence.

5.2.2 Development of transgenic Arabidopsis lines

To study the biochemical behaviour of *AtSS4* and *AtMRC* in their native environment, I produced a new set of Arabidopsis transgenic lines expressing tagged versions of the proteins under their native promoters. I designed expression constructs of tagged *AtSS4* and *AtMRC* using seamless In-Fusion cloning. An example of this is given in Figure 5.2A, for a construct of pAtSS4:*AtSS4*-HA.

Expression of these constructs under their native promoters was achieved by cloning upstream of the transcription start site of the genes in the genomic sequence (1500 bp for pAtSS4, 1674 bp for pAtMRC). This was followed by the coding sequence (CDS) of *AtSS4* or *AtMRC*, and then a glycine linker sequence (GGGS) \times 3, followed by a tag and an NOS terminator sequence. I made three versions of tags: HA (YPYDVPDYA), eGFP-HA and RFP-HA. I cloned HA constructs for both *AtSS4* and *AtMRC*, eGFP-HA for *AtSS4*, and RFP-HA for *AtMRC*. The HA-tag was chosen as it is a small tag with a small chance of disrupting the native protein structure, and it could be used for affinity purification of the protein from plants. The fluorescently tagged constructs could be used in the future to study the localisation and behaviour of the proteins using confocal microscopy. The addition of HA-tags to these fluorescent protein-tagged constructs can allow easy biochemical comparison with the HA-tagged constructs. A summary of the constructs is shown in Figure 5.2B.

These expression modules were then assembled into the backbone of the pK7YWG2 binary vector used for plant transformation, through In-Fusion cloning. It should be noted that the cloned pAtMRC promoter has a deletion of one adenine in a repeat adenine sequence compared to the sequence on Ensembl Plants, 1013 bp upstream of the start codon, but no obvious binding motifs were identified in this region and this construct was still taken forward.

The *AtSS4* constructs were transformed into Arabidopsis *ss4* plants and the *AtMRC* constructs were transformed into *mrc* plants, using Agrobacterium-mediated floral dipping (Zhang *et al.*, 2006). The T1-generation transformants were screened on MS (Murashige & Skoog) agar plates (3% sucrose, 0.8% agar) with kanamycin (for the pK7YWG2 vector) and were also screened for protein expression by immunoblotting with an anti-HA antibody. Successful transformants were identified for all four constructs in the T1 generation (examples shown in Figure 5.2C). Seeds from these T1 plants were collected. The *ss4*/pAtSS4:*AtSS4*-HA lines were prioritised for further experiments below, and three independent stable lines were generated after screening the T2 generation for lines with single insertions, and homozygous lines were obtained in the T3 generation. For *mrc*/pAtMRC:*AtMRC*-HA, only one successful T1 plant was identified and was also shown as a single insertion line in the T2 screen, therefore the T3 seeds (from 14 plants) have been collected for this line, but an additional independent line still has to be isolated. For *mrc*/pAtMRC:*AtMRC*-RFP-HA, the T2 seeds have been collected (from 7 plants including 3

independent T0 parents), ready for a T2 screen. For *ss4/pAtSS4:AtSS4-eGFP-HA*, T3 seeds (from 13 plants) have been collected for one line, but an additional independent line still needs to be isolated from a T2 screen.

I first characterised the phenotype of the homozygous *ss4/pAtSS4:AtSS4-HA* lines (Figure 5.3). The transgenic plants had wild-type rosette morphology rather than the pale, dwarfed *ss4* mutant phenotype (this was also seen for the *ss4/pUBC:TaSS4-YFP* lines in Figure 4.2A), suggesting that the expression of *pAtSS4:AtSS4-HA* had complemented the *ss4* phenotype. LM sections from young leaves also showed complementation of the *ss4* starch granule phenotype, where the complemented lines had multiple granules per chloroplast, as in wild type. However, the degree of complementation seemed to vary among the three different lines, with the 1-2 line most resembling the wild type (Figure 5.3). Leaves of lines 42-1 and 18-6 still had some chloroplasts containing no starch granules, or some abnormally shaped starch granules.

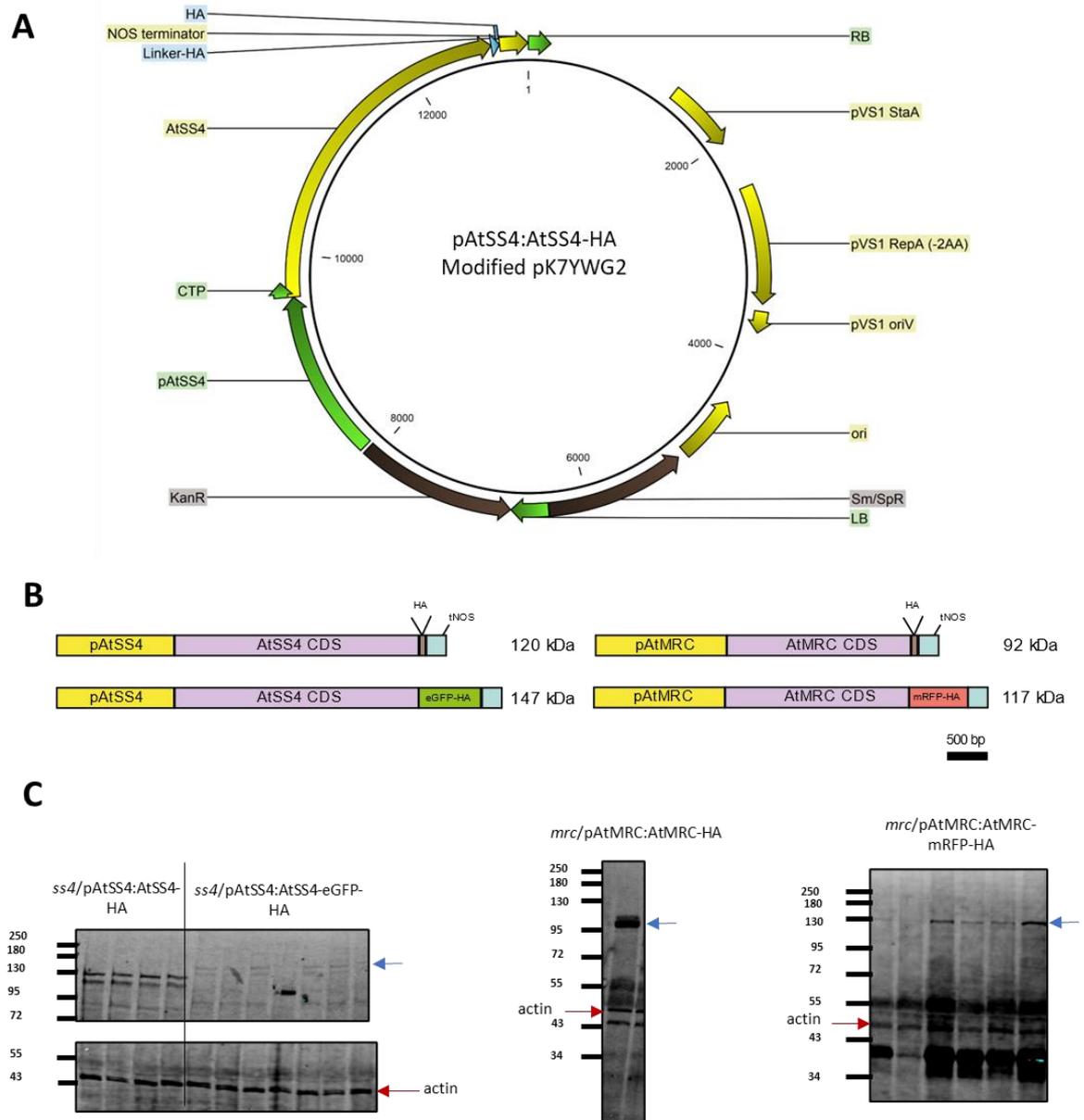


Figure 5.2. The development of transgenic Arabidopsis lines expressing tagged *AtSS4* and *AtMRC*. **A)** A modified pK7YWG2 plant expression vector was used to clone the *AtSS4* and *AtMRC* constructs using In-Fusion cloning. This is the plasmid of the pAtSS4:*AtSS4*-HA construct as an example. LB = T-DNA left border, RB = T-DNA right border, KanR = kanamycin resistance gene, Sm/SpR = streptomycin/spectinomycin resistance gene, ori = *E. coli* origin of replication **B)** Schematic of constructs for Arabidopsis transformation, cloned using seamless In Fusion cloning. **C)** Anti-HA immunoblots showing detection of transgenes in the T1 generation plants transformed with the constructs in B. Blue arrows indicate the bands corresponding to the proteins of interest: *AtSS4*-eGFP-HA, *AtMRC*-HA, and *AtMRC*-mRFP-HA.

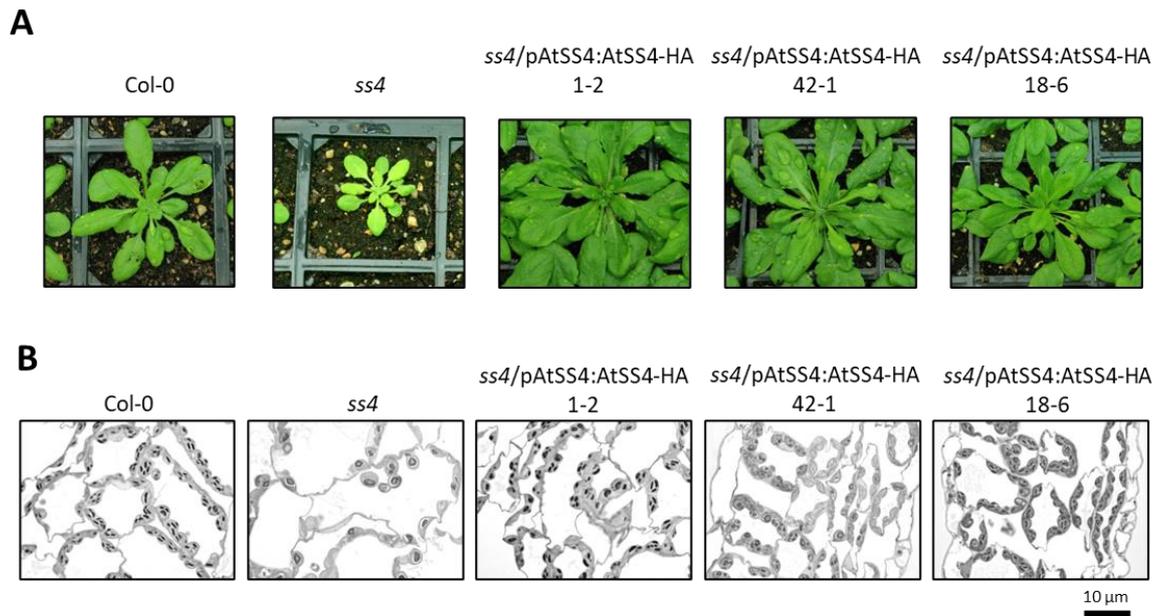


Figure 5.3. Three independent stable lines of the *ss4/pAtSS4:AtSS4-HA* lines show complementation of the *ss4* phenotype to varying degrees. A) Examples of the complemented growth of the transgenic lines compared to the *ss4* background. The images are all on the same scale. **B)** Examples of light microscopy leaf sections from the middle of an approx. 6-week-old *Arabidopsis* leaf, showing complemented starch granule phenotypes in the transgenic lines compared to the *ss4* background.

5.2.3 Affinity purification mass spectrometry of *AtSS4-HA* in *Arabidopsis* leaves

The SEC elution patterns suggest that *AtSS4* is present in a large protein complex, but it does not provide information on the composition of this complex. I therefore used the stable *ss4/pAtSS4:AtSS4-HA* lines for affinity purification mass spectrometry (AP-MS) of the *AtSS4-HA* protein, to see whether *AtSS4* is associated with other proteins like those in the granule initial such as *PTST2* (with which *SS4* was pulled down in an earlier experiment (Seung *et al.*, 2018)). I performed the affinity purification and prepared the samples for trypsin digestion, after which the mass spectrometry and data analysis were outsourced to Gerhard Saalbach and Carlo Martins in the Proteomics platform at the John Innes Centre.

The affinity purification was performed using anti-HA beads, using equal amounts of plant tissue per volume of extraction buffer. After the pull-down, I ran 5 µL of each sample on an immunoblot to confirm the presence of *AtSS4-HA* (Figure 5.4). The three lines were not even in their expression, which is consistent with their uneven complementation of the *ss4* phenotype (section 5.2.2). Line 1-2 had the highest *AtSS4-HA* expression (Figure 5.4) and also the best complementation of *ss4* phenotype (Figure 5.3).

AtSS4 was pulled down in the bait samples (Table 5.1), and no *AtSS4* peptides were pulled down in any of the Col-0 controls, confirming that the pull-down worked. The most *AtSS4* peptides were detected in the 1-2 line, in line with it having the highest expression of *AtSS4*-HA among the three lines. After statistical analysis, *AtSS4* was the main protein that was significantly enriched in the bait samples (*ss4/pAtSS4:AtSS4*-HA) versus the wild type controls, and the only one with an adjusted p-value <0.05 after filtering. The starch initiation related proteins were notably absent from the list of other enriched proteins with a non-adjusted p-value <0.05. The only protein from this list that has been characterised in detail is JACALIN-LECTIN LIKE1 (*AtJAC1*) (AT3G16470), which functions in flowering time control through the regulation of RNA processing (Xiao *et al.*, 2015), thus is unlikely to influence starch granule initiation in the chloroplast. These proteins were also all identified in large scale omics studies, so are likely not specific interactors of *AtSS4* (Table 5.1).

In this experiment we were unable to differentiate between the bait itself attached to the column or additional *AtSS4*-HA proteins copurifying with the bait *AtSS4*-HA protein. However, considering that *AtSS4* can dimerise (Raynaud *et al.*, 2016) and native SS4 was previously pulled down together with SS4-RFP in a transgenic *Arabidopsis* line (Seung *et al.*, 2017), it is plausible that there are multiple *AtSS4* monomers that make up the hypothesised large protein complex. As *AtSS4* is found in HMW fractions in SEC and no stable interactions with other proteins are detected in the AP-MS, the oligomerisation of *AtSS4* may be a key component of the HMW complex.

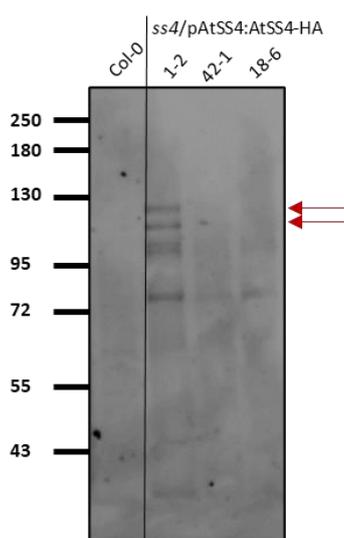


Figure 5.4. *AtSS4*-HA is pulled down at various abundances in transgenic lines. Immunoblots of 5 μ L from the affinity-purification of *ss4/pAtSS4:AtSS4*-HA lines, using anti-HA magnetic beads, blotted with an anti-HA antibody. Col-0 was used as a control. The rest of the samples was used for mass spectrometry. Red arrows indicate the double band characteristic for full length *AtSS4*.

Table 5.1 Interaction partners of AtSS4-HA identified in an AP-MS of ss4/pAtSS4:AtSS4-HA lines. Three biological replicates were used for both the bait ss4/pAtSS4:AtSS4-HA lines and the Col-0 controls. Statistical analyses were performed by Gerhard Saalbach and Carlo Martins as described in the materials and methods.

Accession	Description	Coverage [%]	# Peptides	# Unique Peptides	Score CHIMERYS Identification	Abundance Ratio: (AtSS4-HA) / (WT)	Abundance Ratio P-Value: (AtSS4-HA) / (WT)	Abundance Ratio Adj. P-Value: (AtSS4-HA) / (WT)
AT4G18240.1	starch synthase 4	25	24	24	81.84	9.226	0.0001579	0.0292205
AT1G54020.2	GDSL-like Lipase/Acylhydrolase superfamily protein	11	4	4	22.38	6.94	0.00113795	0.1403475
AT3G28220.1	TRAF-like family protein	26	10	10	60.02	3.803	0.0318287	0.5607920
AT3G16470.1	Mannose-binding lectin superfamily protein	45	14	13	53.98	3.621	0.0397305	0.6124862
AT2G39330.1	jacalin-related lectin 23	41	19	18	75.9	3.461	0.048447	0.6659976

5.2.4 Optimisation of recombinant SS4 and MRC purification from *E. coli*

5.2.4.1 *In vitro* expression of SS4 and MRC

As I hypothesised that SS4 could form higher-order assemblies on its own, I wanted to test this in isolation using recombinant SS4 protein expressed in *Escherichia coli*. Because of the interaction between SS4 and MRC, I also expressed recombinant MRC, with a view to looking at *in vitro* protein interactions.

When Raynaud *et al.* (2016) expressed recombinant AtSS4 in *E. coli*, they saw a significant amount of degradation products, and did not obtain enough full length protein to perform robust SEC experiments. Instead, they used an N-terminally truncated fragment, which ran as a dimer. They used the pET45b vector with an N-terminal His-tag. I used different expression vectors and a codon optimised CDS (Appendix 2) to try a new expression and purification approach.

David Seung had previously made a construct for *E. coli* codon-optimised TaSS4 expressed in a modified pPROEXHtb vector, where the N-terminal His-tag was replaced with a StrepII-tag, and the ampicillin resistance was replaced with spectinomycin resistance, here named pPROEXHtb2. Using Gibson assembly, I used this vector as a backbone to clone a new *E. coli* codon-optimised AtSS4 construct, using the AtSS4 CDS (Figure. 5.5). Both of these constructs were used for expression and purification from *E. coli*. However, one caveat is that later I discovered a missense mutation in the TaSS4 construct, where a cytosine to thymine change caused a Ser286Leu change at the end of the coiled-coil domain. The results from this construct are still included in the remainder of this thesis for reference. I cloned a codon-optimised AtMRC CDS (Appendix 2) into the pOPINE vector with a C-terminal His-tag.

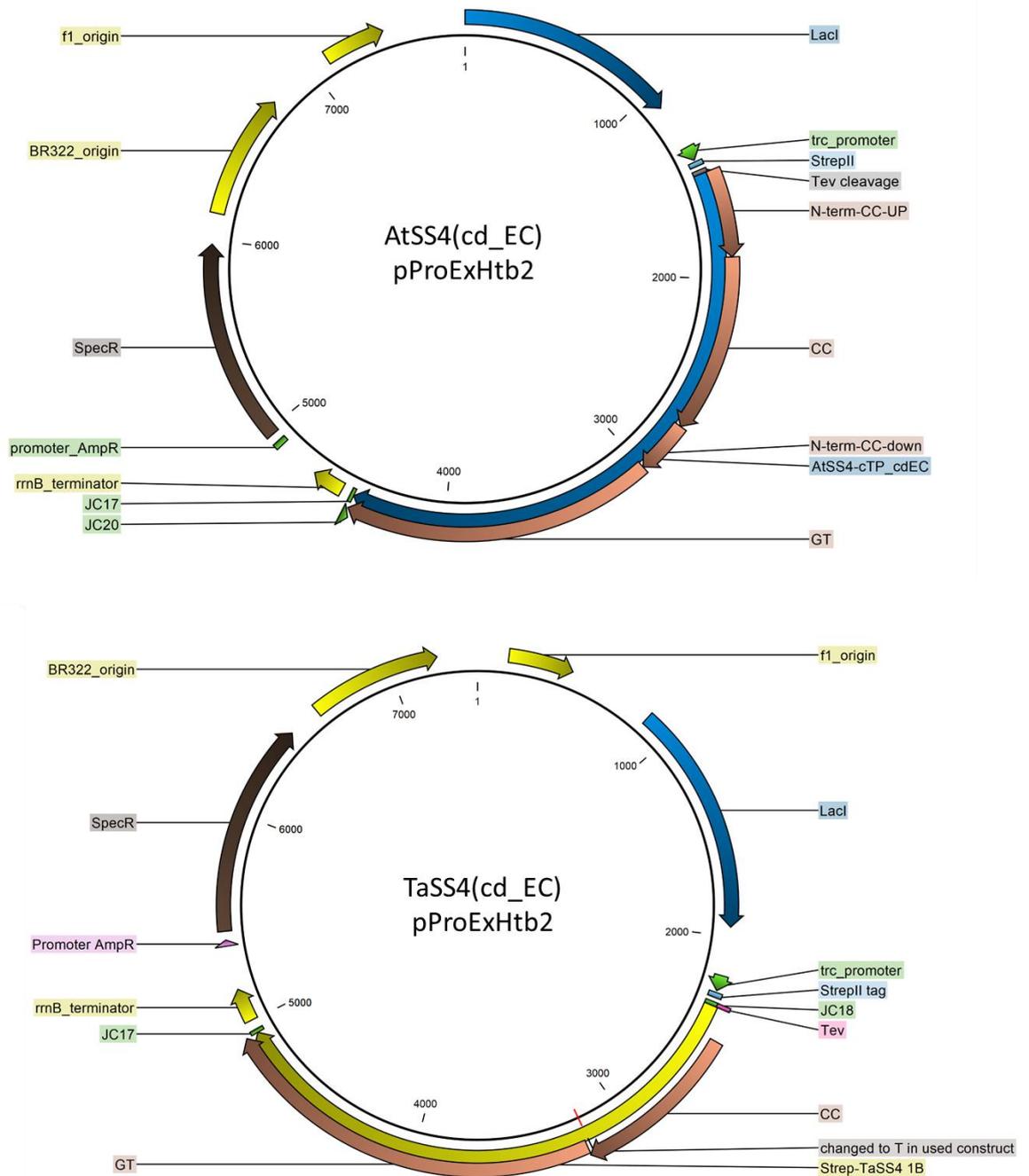


Figure 5.5. Plasmid maps of SS4 recombinant expression constructs. A) *AtSS4* coding sequence (CDS) codon optimised for *E.coli* in an edited pProExHtb vector with spectinomycin resistance and an N-terminal StrepII-tag. **B)** Same as for A, but the *TaSS4* CDS. A missense C → T mutation was present in this construct, indicated by a red line. This resulted in a serine → leucine change, and was only detected after using this construct for expression and purification. Annotations: CC = coiled coil, GT = glycosyltransferase domain. The 'cd_EC' means codon optimised for *E.coli*, which includes removal of the chloroplast transit peptides.

I obtained soluble protein for recombinant *AtSS4*, *TaSS4*, and *AtMRC* in a small scale (100ss mL) solubility test. I expressed all constructs in BL21 (DE3) Δ glgAP *E. coli* cells lacking endogenous glycogen synthase (GS). The Strep-*AtSS4* and Strep-*TaSS4* proteins were soluble (Figure. 5.6A) but there were many lower molecular weight bands in the anti-Strep immunoblots as well, which could correspond to degradation products like those observed in (Raynaud *et al.*, 2016). The bands for *AtSS4* and *TaSS4* are enriched in all the induced fractions – the induced sample before lysis of cells (I), the crude extract after lysis (C) and the pellet after spinning down the lysed cells (P). For the *AtMRC*-His solubility test, the protein was already seen in the uninduced sample, indicating leaky expression (Figure 5.7A). There were also degradation products seen in the anti-His immunoblot, although these largely disappeared in the soluble fraction, where the full-length protein was the main band. This band ran slightly higher than expected for an 88 kDa protein.

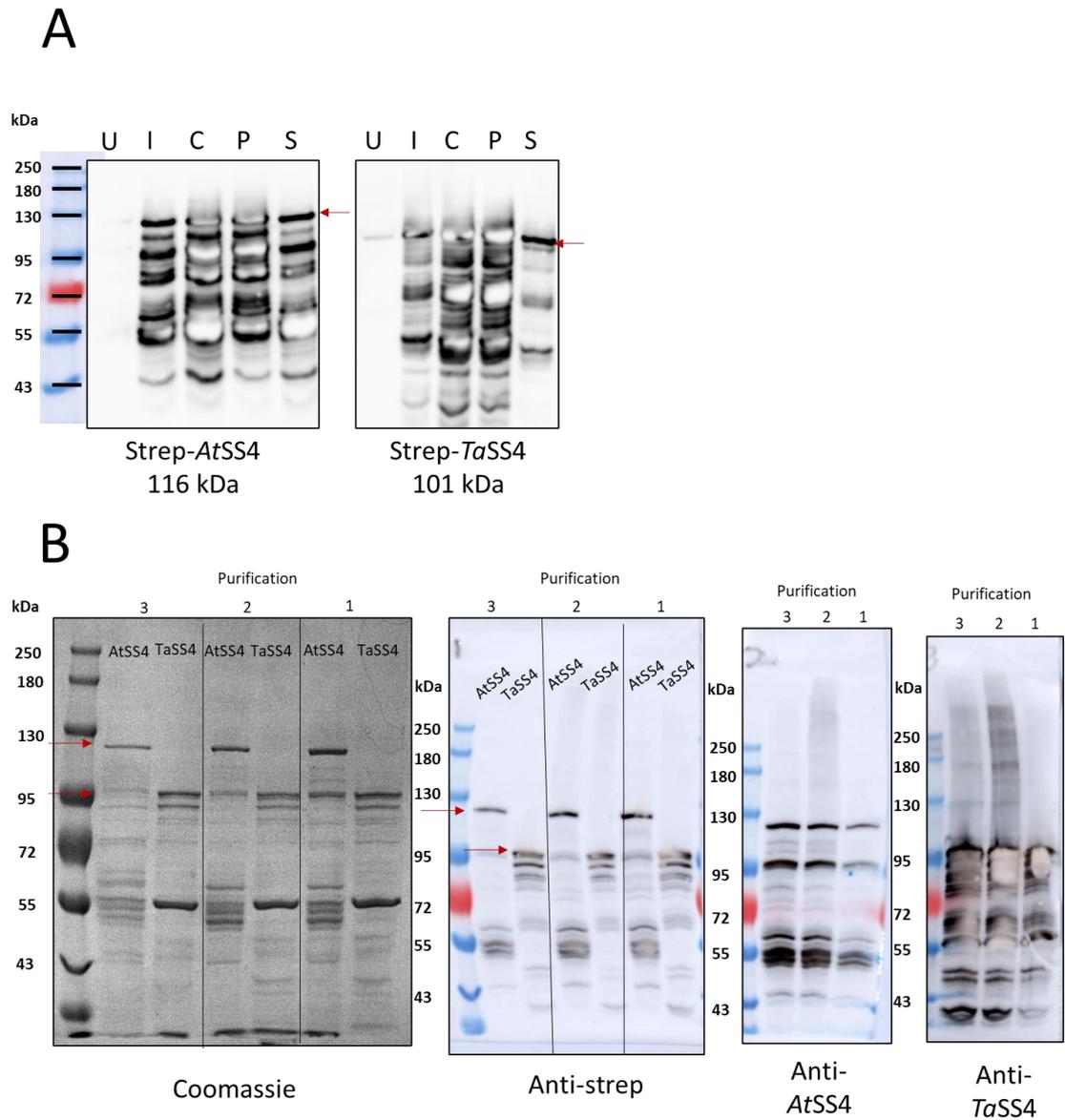


Figure 5.6. *In vitro* purification of SS4 proteins produces full length SS4 proteins but also degradation products and contamination. A) Immunoblots blotted with anti-Strep antibody show solubility tests of small scale (100 mL) expression cultures of Strep-tagged SS4 proteins. SS4 proteins are present in the soluble fraction. U = uninduced, I = induced, C = crude, P = pellet, S = soluble fraction. **B)** Coomassie-stained SDS-PAGE gel and immunoblots (with relevant antibodies used to blot indicated beneath) of three *in vitro* purifications of AtSS4 and TaSS4 proteins using half or all of 500 mL cultures. 1 μ g of protein was loaded on the Coomassie gel, and 0.2 μ g of protein was loaded on the immunoblots. Red arrows indicate relevant full length SS4 proteins.

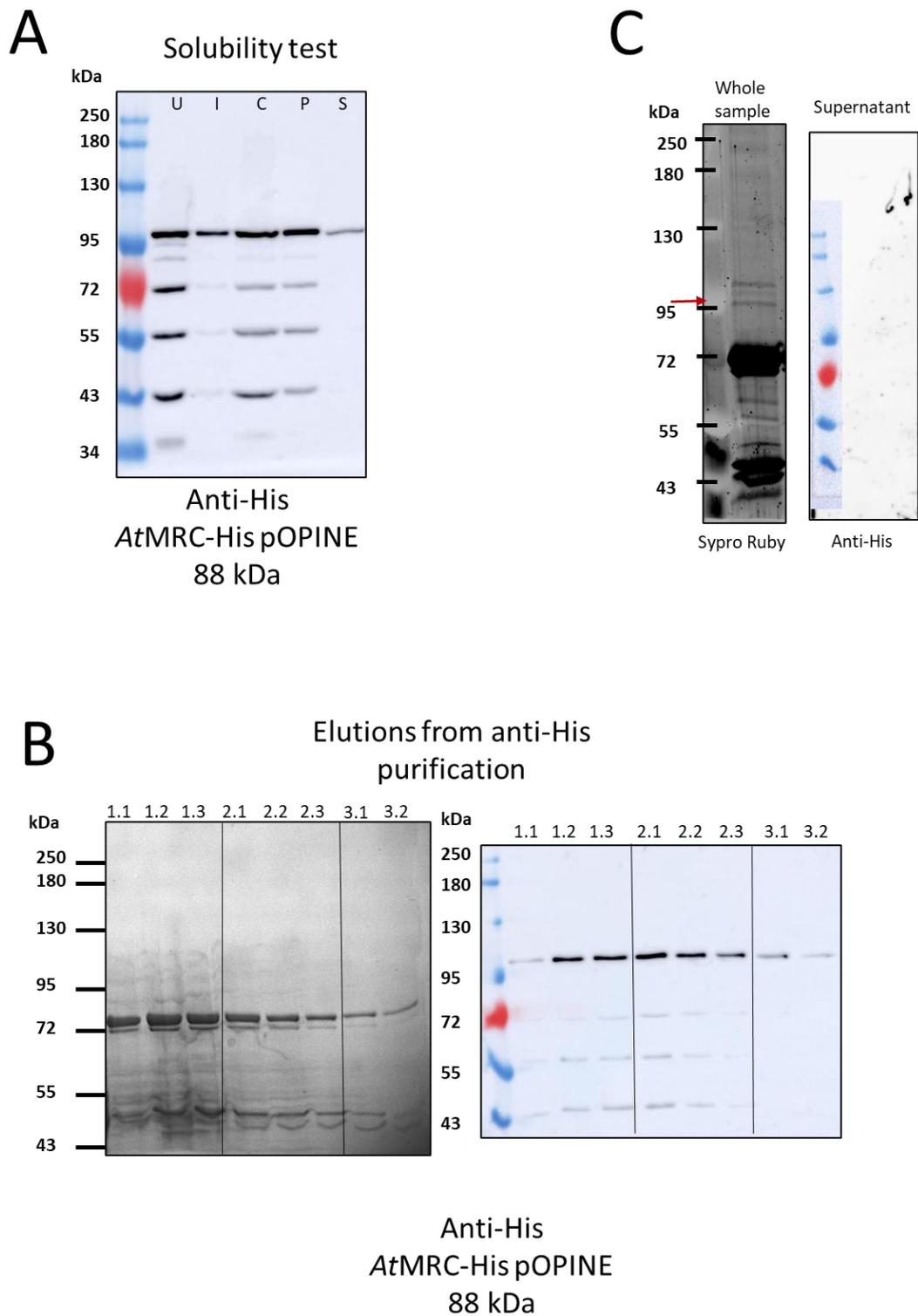


Figure 5.7. Recombinant *AtMRC* is soluble but can precipitate. **A)** Immunoblot blotted with anti-His antibody shows the solubility test of a small scale (100 mL) expression culture of His-tagged *AtMRC* protein. U = uninduced, I = induced, C = crude, P = pellet, S = soluble fraction. **B)** Coomassie-stained SDS-PAGE gel (left) and immunoblot blotted with anti-His antibody (right) of the eluted fractions from a 500 mL anti-His purification of recombinant *AtMRC*. **C)** Pooled and concentrated recombinant *AtMRC* from the purification in B, after precipitation after storage. Sypro-Ruby stained SDS-PAGE gel of total (cloudy) sample (left) and an immunoblot of just the supernatant blotted with anti-His antibody (right). The red arrow shows the position of *AtMRC* in the total protein stain, but this is not present in the supernatant immunoblot.

5.2.4.2 *In vitro* purification of SS4 and MRC

As the SS4 proteins were soluble, I continued with protein purification using larger cultures of 500 mL (using either half or all of it in one purification). The purification was done using a Strep-Tactin Sepharose agarose resin (IBA Lifesciences) and relevant elution fractions were pooled and concentrated. The purifications were successful, but not entirely pure (Figure. 5.6B). The full-length protein was the main band in the *At*SS4 purification, but in the *Ta*SS4 purification there was another prominent band at 55 kDa visible in the total protein stain but not in the immunoblot, indicating contamination. For *At*SS4, there were some contaminants and degradation products around the 55 kDa marker, similar to those observed in (Raynaud *et al.*, 2016). However, because the full-length protein was still the main band in the purifications, I continued with further analysis using these proteins.

To test whether the purified SS4 proteins were functional and could perform their glucan elongating activity, I used an *in vitro* starch synthase activity assay (an ADP/NADH-based assay) to measure their activity. Both SS4 proteins were active (Table 5.2, Figure 5.8), using maltotriose efficiently as a substrate with an average *k_{cat}* of 3.7 turnovers/second for *At*SS4 and 0.5 turnovers/second for *Ta*SS4.

Table 5.2 Calculations of average kcat values for *in vitro* starch synthase activity assays of recombinant AtSS4 (A) and TaSS4 (B). abs = absorbance

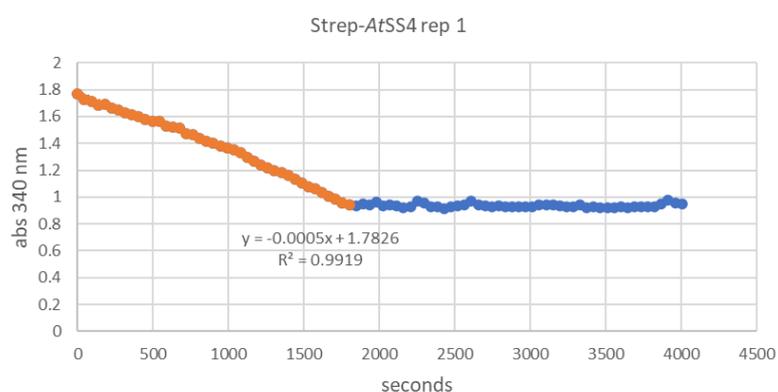
A

	abs per s	μmol of NADH per second per 0.0086 nmol AtSS4	kcat (turnovers per second)
AtSS4 rep1	0.0005	3.25E-05	3.78E+00
AtSS4 rep2	0.0005	3.25E-05	3.78E+00
AtSS4 rep 3	0.0005	3.25E-05	3.78E+00
average			3.78E+00

B

	abs per s	μmol of NADH per second per 0.0099 nmol TaSS4	kcat (turnovers per second)
TaSS4 rep1	9.00E-05	5.84E-06	5.90E-01
TaSS4 rep2	8.00E-05	5.19E-06	5.25E-01
TaSS4 rep 3	8.00E-05	5.19E-06	5.25E-01
average			5.47E-01

A



B

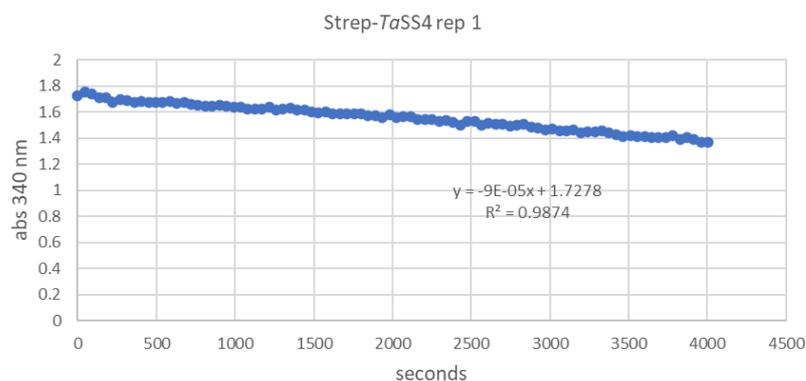


Figure 5.8. *In vitro* starch synthase activity assays of purified recombinant SS4 proteins using maltotriose as a substrate. A) AtSS4 *in vitro* activity measurement, using protein from purification 3 from Figure 5.6 B) Same as A, but for TaSS4. Graphs are each one of three total technical replicates, with calculations of kcat for each technical replicate and average kcat listed in Table 5.2.

For AtMRC, I performed a 500 mL purification using an anti-His agarose resin (Qiagen Ni-NTA resin). This produced quite a low yield, as no band of correct size was detected in the total protein stain, but when immunoblotted with an anti-His antibody, the correct band was visible (Figure 5.7 B). In the purification, the degradation products were also far less abundant than in the non-soluble fractions (Figure 5.7A, B). However, a prominent band at 72 kDa in the total protein stain indicated high levels of contamination. Elution fractions with prominent bands in the immunoblot were pooled and concentrated.

However, after concentrating there was some protein precipitation. I ran a small aliquot of the whole extract on an SDS-PAGE gel and stained it with Sypro Ruby stain, spun down the sample and ran an immunoblot of the supernatant. In the total protein stain, the AtMRC band was visible, although at much lower abundance than the contamination at 72 kDa (Figure 5.7C). However, in the immunoblot of the supernatant, the AtMRC protein was not detectable (Figure 5.7C), suggesting it had indeed precipitated. Therefore, this process requires further optimisation to determine under which conditions the AtMRC protein is soluble and when it starts to precipitate.

Overall, the expression of AtMRC is promising, but it was not a stable soluble protein *in vitro*. The AtSS4 purification was cleaner than the TaSS4 purification, so I focused on AtSS4 for further *in vitro* characterisation.

5.2.5 The SS4 protein alone forms high molecular weight assemblies *in vitro*

Considering the importance of SS4 for my central question regarding the biochemical mechanism of starch granule initiation, and the fact that its purification was the most promising, I focused on SS4 for further *in vitro* biochemical characterisation. I increased the culture volume to a 4 L purification (Figure 5.9A), to have enough sample to run with SEC. I obtained purifications of approx. 8 µg/µL of Strep-*At*SS4 and 5 µg/µL of Strep-*Ta*SS4, loading 100 µL on the SEC. As with the small-scale purifications, the *At*SS4 purification looked purer than the *Ta*SS4 one, but both still contained some contamination and degradation products.

I ran these samples on the same Superose 6 Increase 10/300 column as the plant extracts in section 5.2.1.1. The UV traces for both *At*SS4 and *Ta*SS4 showed a sharp peak for the void in fraction 8 (Figure 5.9B, C), which likely contains large aggregates above 2 MDa, the molecular weight of the void marker blue dextran. Both proteins also showed a broad peak around fractions 17 and 18 and two smaller peaks at fractions 20 – 22. These last two peaks are at the end of the column volume (24 mL), so are perhaps caused by sample build-up at the end of the column. Since the UV signal cannot discriminate between full length SS4 and any contaminants, I ran the fractions on SDS-PAGE gels and stained for total proteins using Coomassie Blue (Figure 5.9 D, E). I found that full length SS4 was in fractions 8 – 14, in fairly equal abundance across these fractions. Surprisingly, fractions 17 and 18 did not contain detectable amounts of the full length SS4 proteins, despite the strong UV signal in these fractions.

It is uncertain where the strong UV signal in fractions 17 and 18 is coming from, and why fractions 9 – 14 did not display a strong UV signal despite the abundance of the SS4 protein in these fractions. It could be due to aggregation of smaller degradation or contamination products masking the UV trace signals of the full-length protein, but in that case one would expect stronger additional bands on the gel, with increasing abundance across the fractions and peaking in fractions 17 and 18. It is also unlikely that the UV trace signal is coming from smaller products not visible on the gel, as in the full purification in Figure 5.6 there were not a lot of additional products seen below the 55 kDa marker.

Notably, the elution pattern of the full-length SS4 proteins observed on the SDS-PAGE gels was broader and at higher molecular weight fractions (Figure 5.9 D, E) than what was seen in the plant extracts (Figure 5.1). The recombinant SS4 proteins are already seen in the void fraction (fraction 8), suggesting it might be aggregating to very large assemblies *in vitro*, or perhaps it is aggregating in a way that does not elute according to the expected size for globular proteins, which an SEC column is calibrated for.

The SDS-PAGE gels also showed that contaminants became more prominent in the later fractions, from fraction 13 (for *AtSS4*, Figure 5.9D) or fraction 11 (for *TaSS4*, Figure 5.9E) onwards. From fraction 15 onwards, for both *AtSS4* and *TaSS4*, fractions mostly contained contaminants. One contaminant around the 55 kDa marker was in both SS4 purifications but especially prominent in the *TaSS4* purification (Figure 5.9 D, E indicated by a red asterisk). After negative staining TEM (see section 5.2.7), this was identified as the bacterial chaperonin GroEL (Braig *et al*, 1994), an oligomeric complex consisting of fourteen subunits of 57 kDa, to make an approx. 800 kDa complex. This is consistent with its elution at/before the highest fraction marker at 669 kDa. Curiously, in the *AtSS4* sample GroEL elutes in lower molecular weight fractions than in the *TaSS4* sample. In summary, the SEC was effective in separating out fractions enriched in the full-length SS4 protein, containing less GroEL and other contaminants/degradation products. Thus, it has been both an analytical method demonstrating the association of SS4 proteins *in vitro* and the presence of contaminating products in the purification, and also a purification method allowing separation of contaminants from the SS4 protein in individual fractions.

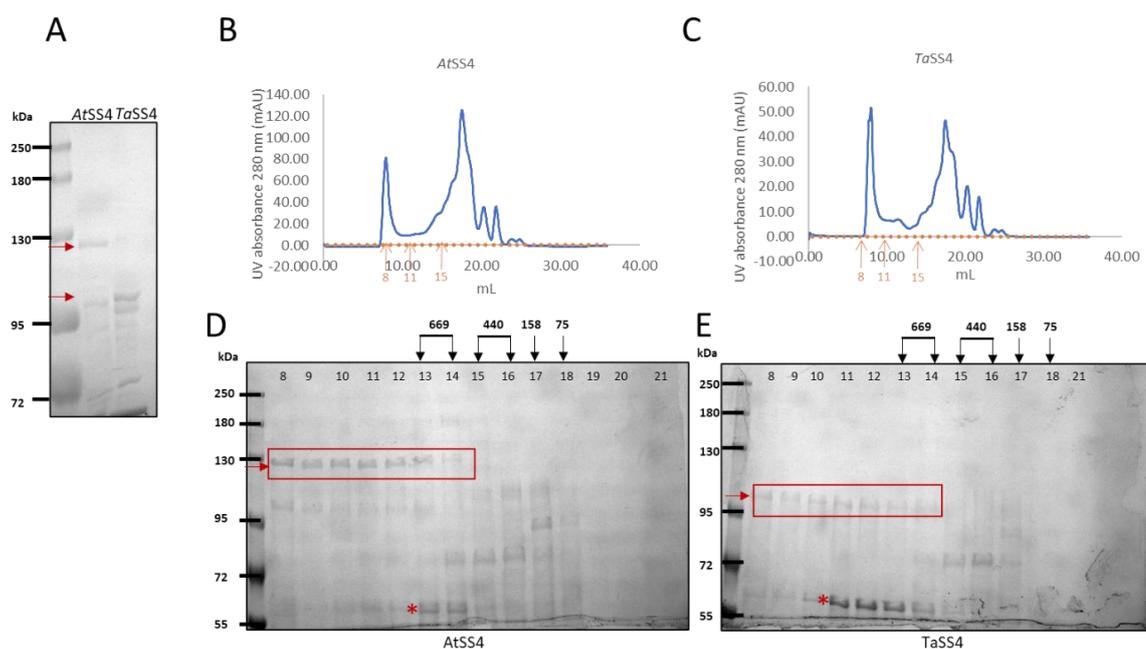


Figure 5.9. The SS4 protein alone forms high molecular weight assemblies *in vitro*. **A)** Coomassie-stained SDS-PAGE gel from a 4 L purification of SS4 proteins, with approx. 1 μ g of protein loaded. Red arrows indicate the full-length SS4 proteins **B), C)** Superose 6 Increase 10/300 size exclusion chromatography (SEC) UV traces of purified SS4 proteins. 1 mL fractions used to collect the eluate are marked in between orange dots, with fractions 8, 11, and 15 highlighted with arrows. **D), E)** Coomassie-stained SDS-PAGE gels of the fractions collected from B and C. 13.5 μ L of protein from each fraction was loaded. Red boxes and arrows indicate the bands for the full length SS4 proteins. Red asterisks indicate contamination from bacterial GroEL.

5.2.6 Structural prediction of SS4 proteins suggests association of the N-termini

To hypothesise how the SS4 proteins might be interacting with each other, I looked at structural predictions of the proteins using AlphaFold (Jumper *et al*, 2021; Varadi *et al*, 2022). During the final year of this project, the development of AlphaFold has been a great technical advancement, with many people using it as a tool to aid in structural biochemistry work. David Lawson and Michael Webster helped me run predictions that were not already available on the public AlphaFold database (<https://alphafold.ebi.ac.uk/>), to run and interpret protein-protein interactions.

Structural predictions for *AtMRC* and *AtSS4* were available on the AlphaFold database (Figure 5.10A). *AtMRC* is predicted to be a long coil, similar to what has been predicted in the literature. However, the relative arrangement of these coils (for instance into coiled coils) is not accurately predicted, shown in the Predicted Aligned Error (PAE) plot. The C-terminal catalytic domain of *AtSS4* is predicted with high local accuracy (pLDDT score) as well as high inter-domain accuracy, shown in the PAE plot. This prediction is also very similar to the solved crystal structure (Nielsen 2018), demonstrating the high accuracy that AlphaFold is capable of. The *AtSS4* N-terminus is predicted to have coiled regions as well as unresolved/unstructured regions, which have relatively low confidence local accuracy, and low inter-domain accuracy in the PAE plot. A similar pattern was seen for the prediction of *TaSS4* (Figure 5.10B), which was done using ColabFold (Mirdita *et al*, 2022). The N-terminus had a slightly different PAE plot compared to *AtSS4*, but the relative arrangement of the coils was still difficult to interpret here.

In both the *AtSS4* and *TaSS4* predictions, the conserved region (Raynaud *et al.*, 2016) (see Figure 4.1) corresponded to a short coiled region in between the extended coils and the catalytic domain, highlighted with a green outline (Figure 5.10 A,B). This region was also predicted with higher local accuracy in the *AtSS4* model compared to the rest of the coiled region. This conserved region might have a unique role which is structurally determined.

I then used ColabFold to predict protein-protein interactions. No interactions were predicted with *AtMRC* and *AtSS4*, or with two full-length *AtSS4* proteins. However, when using just part of the N-terminus of *AtSS4* (residues 151 – 540), AlphaFold predicted the dimerization of two N-termini into a coiled coil with high accuracy (Figure 5.10C). When running ColabFold, it predicts five models, and the highest-ranking model is shown in Figure 5.10C, with the PAE plots of all five models shown beneath. All of these show some confidence in the relative arrangement of the two coils, indicated by regions of low error in the off-diagonal boxes. The coils are arranged in parallel, so that the two catalytic domains would appear on the same end. This could be a possible arrangement of the dimerization of two *AtSS4* proteins, and perhaps higher order structures

could form with multimers of these dimers. The conserved regions on the C-terminal end form independent coiled regions which do not seem to interact with each other. The SS4 N-terminus was previously predicted to have a coiled coil on its own, but the AlphaFold prediction shows that it could in fact form one coiled coil using two N-termini. A similar prediction was seen for *TaSS4* (residues 46 – 394, without CTP). Both *AtSS4* and *TaSS4* N-terminus dimer predictions also have an N-terminal coiled section folding back on itself (Figure 5.10C, D).

As AlphaFold predicted some regions of MRC and SS4 could be unstructured, I also ran the MRC and SS4 proteins in two disorder prediction softwares, PONDR (Romero *et al*, 2001) and IUPred3 (Erdos *et al*, 2021). PONDR predicted some regions of protein disorder across the MRC proteins, and at the N-termini of the SS4 proteins (Figure 5.11A). IUPred3 mainly predicted disordered regions at the beginning of the SS4 proteins (Figure 5.11B). For *AtSS4*, the first PONDR region and the IUPred3 region overlapped with the first disordered region in the AlphaFold model, the two other PONDR regions spanned two coiled regions of the AlphaFold model. For *TaSS4*, the first PONDR region and the IUPred3 region also corresponded to the first disordered region in the AlphaFold model, and the second PONDR region spanned two coils in the AlphaFold prediction. It is striking that the *AtSS4* IUPred3 region started from right after the CTP up to residue 165, ending around where the dimerization of the coils is predicted to start. For *AtMRC*, the first two PONDR regions span two coils in the AlphaFold model, and the third region spans two coils and a disordered region. However, according to the IUPred3 prediction and the high pLDDT scores in the AlphaFold model, *AtMRC* might not be so disordered, but the relative arrangement of the coils could still be flexible.

These structural predictions provide us with more hypotheses on the structural basis of *AtSS4* oligomerisation. Chapter 4 indicated that for *TaSS4*, both the N and C terminus were likely involved in protein interaction, and here we show that the interaction of two N-termini could be through forming a coiled coil. Flexible regions could be involved in the interaction between MRC and SS4.

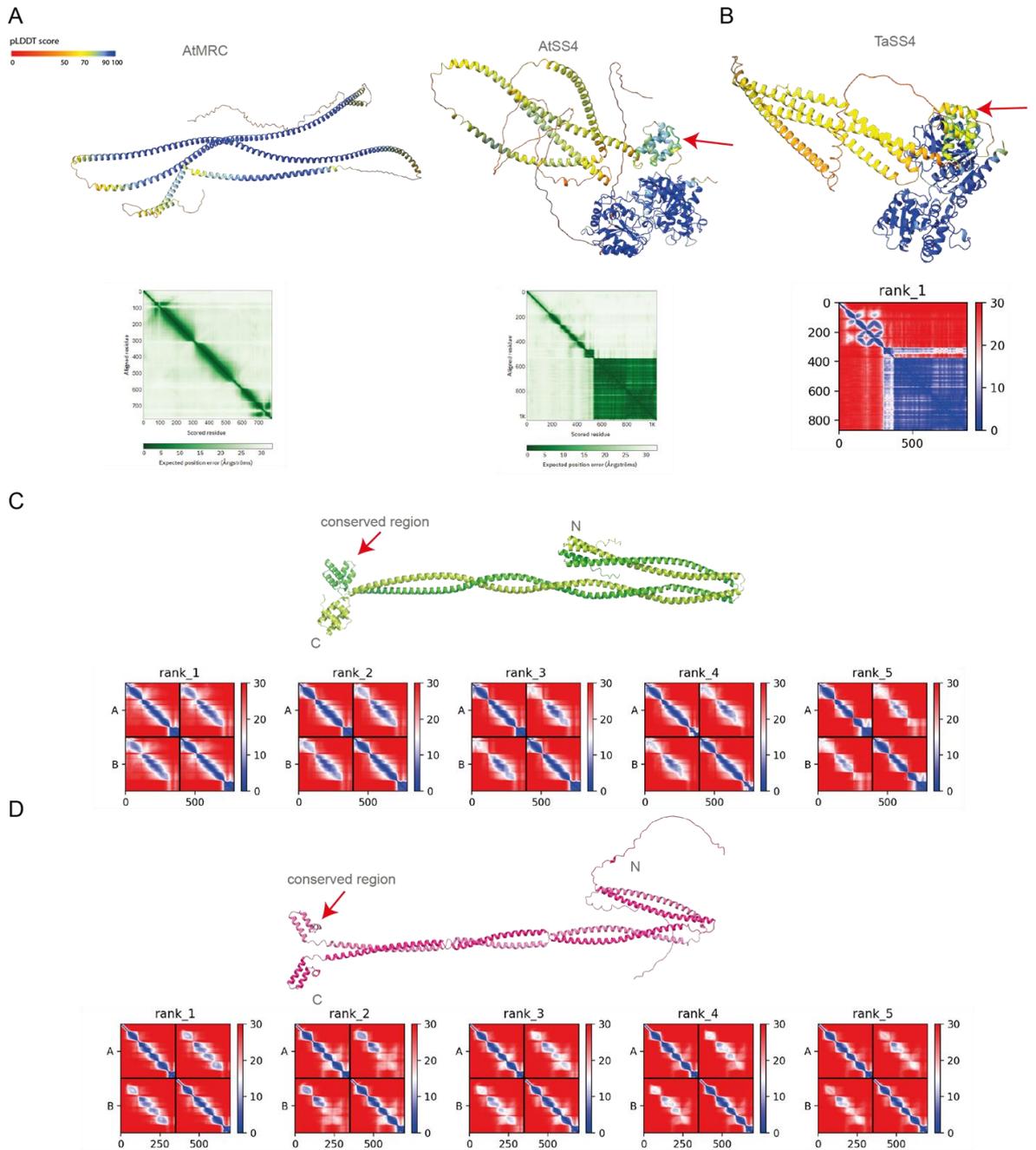


Figure 5.10 Structural predictions of SS4, MRC and SS4 dimerization. A) AlphaFold models of AtSS4 and AtMRC extracted from the <https://alphafold.ebi.ac.uk/> database. These models still have the chloroplast transit peptide (CTP) included. **B)** AlphaFold model of TaSS4 (without CTP) run in Colabfold (the highest ranking model out of 5 is shown). For A and B: Residues are coloured according to AlphaFold's per-residue confidence score (pLDDT, local Distance Difference Test) from 0 to 100, indicating local accuracy. 0 – 50: should not be interpreted, may be disordered, 50 – 70: low confidence, caution, 70 – 90: backbone expected to be modelled well, 90 – 100: high accuracy expected. Plots below show Predicted Aligned Error (PAE), indicating low inter-domain accuracy for coiled regions. Red arrows highlight a region of the structure outlined in green, which is the conserved region as defined in (Raynaud *et al.*, 2016) **C)** AlphaFold model of the dimerization of two N-termini of AtSS4 (residues 151 – 540), run in ColabFold (the highest ranking model out of 5 is shown). The PAE plots of all 5 models are shown below. **D)** Same as C, but for the TaSS4 N-termini, from right after the CTP (residues 46 – 394).

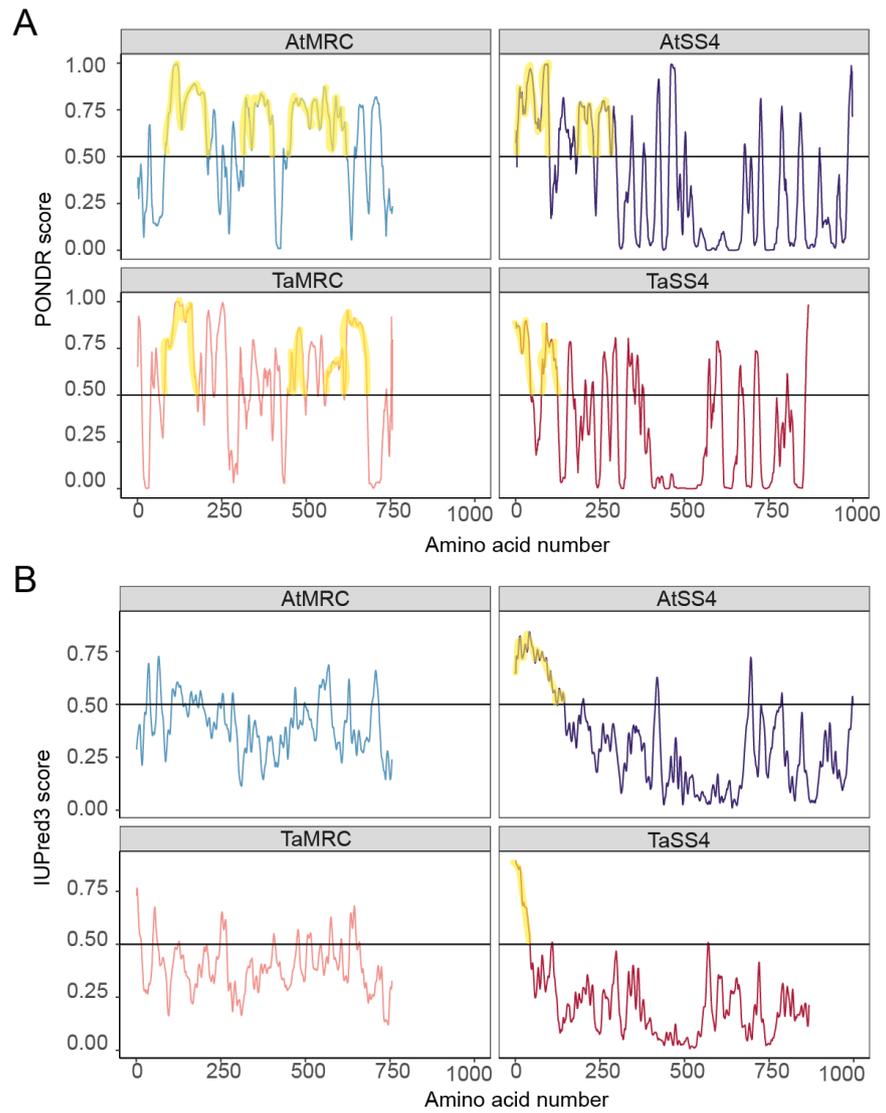


Figure 5.11 Protein disorder prediction of AtSS4, TaSS4, AtMRC and TaMRC. A) Intrinsic disorder prediction by POND, using the VL-XT algorithm. **B)** Intrinsic disorder prediction by IUPred3, using the long disorder analysis. For both A and B, yellow highlighted regions have a relatively high confidence of disorder prediction according to the length of sequence (>30 amino acids). Sequences are without CTP.

5.2.7 Negative staining TEM reveals potential dimeric assemblies of *AtSS4 in vitro*

Since *AtSS4* might assemble into high molecular weight homooligomeric complexes in the plant (section 5.2.3) as well as *in vitro* (section 5.2.5) and recombinant *AtSS4* was the best purified protein with the least contamination, I continued with *AtSS4* for negative stain transmission electron microscopy (NS-TEM) on the SEC-purified protein. I was assisted in these analyses by Jake Richardson from the Bioimaging team at the John Innes Centre. First, I prepared negative stain (NS) grids from several of the fractions from the SEC of the 4 L purification of recombinant Strep-*AtSS4* (Figure 5.9D). From these, I selected fraction 11 to continue with for further analysis. Fraction 11 is from a similar size range as where *AtSS4* was seen in the SEC of plant extracts (Figure 5.1) and it had minimal contamination of GroEL observed under the TEM compared to neighbouring fractions. For *TaSS4*, the contamination of GroEL was very high in comparison to the protein of interest (consistent with the protein gel in Figure 5.9E), so this sample could still benefit from further purification and I did not continue with this sample for the following analyses.

Grids prepared from independent purifications looked similar to each other (Figure 5.12 A, B). The 4 L purification (Figure 5.12B) had a higher yield of protein than the smaller purification (Figure 5.12A), and this was also visible on the NS grids as a higher density of particles. Fraction 11 from the SEC after the 4L purification had a concentration of 0.02 $\mu\text{g}/\mu\text{L}$, whereas the fraction from the smaller scale purification (half of a 500 mL purification) had too low a concentration to be measured by Bradford assay. The fraction from the smaller scale purification was further concentrated using a 30 kDa centrifugal filter concentrator, but the fraction from the 4 L purification was used directly on the grid.

Most visible were the large GroEL particles, with examples circled (Figure 5.12A, B). Apart from these, there was no evidence of large and abundant particles. This suggested that there were no homooligomeric assemblies of the megadalton size range present, as the HMW SEC elution patterns might have suggested (Figure 5.9D). The micrographs indicated a prevalence of smaller particles, of which the shape was indiscernible by eye. To get a better resolution of these small particles, Jake Richardson set up an automated image collection of these grids. Together with David Lawson and Michael Webster, we extracted particle images from the micrographs and performed two-dimensional classification to select particles that were not contaminants (selecting out GroEL).

Initial processing of a small micrograph subset was performed to generate low-resolution classes (Figure 5.12C). This was done using a first set of 151 micrographs (from the sample in Figure 5.12A), containing 3973 particles. These initial 2D classes were then used as a template to pick particles from the second set of 899 micrographs (the sample in Figure 5.12B). After multiple

rounds of automated particle picking, 2D classification and subset selection, a final set of 2D class averages was produced from 93,884 particles (Figure. 5.12D).

The particles picked were indeed the small particles abundant in the micrographs. They are not in the large size ranges suggested by the SEC and are more consistent with the size of an *AtSS4* dimer. In some of the class averages, two distinct round densities can be seen, which could represent two catalytic domains. The distance between the two densities as depicted by the arrows in Figure 5.12D is approx. 160 Å. There were multiple class averages that showed this morphology, likely presenting different views of the same shape (Figure. 5.12D, red boxes). There are also a few class averages with less distinct densities, which could resemble a different shape containing four densities (orange boxes). Flexible domains like that predicted for the N-termini would not usually be visible in a negative-stained micrograph and we cannot achieve higher resolution using negative staining.

To better assess the size of the 2D class averages relative to the *AtSS4* monomer, we generated a 3D density map using all the 2D classes from figure 5.12D (Figure 5.13). If there are two different conformations represented in the 2D classes as indicated in Figure 5.12D, these will be averaged together in the 3D density map. Figure 5.12 displays this map next to two monomers of the crystal structure of the *AtSS4* catalytic domain (PDB: 6GNE), showing that the size of the 2D class averages would be consistent with a particle size of an *AtSS4* dimer. The arrows in the second panel of figure 5.12 indicate a distance of approx. 160 Å and could represent the corresponding distance measured in Figure 5.12D.

It is puzzling that we did not see the *AtSS4* protein in dimeric size fractions in the SEC SDS-PAGE gels (Figure 5.9D), considering smaller molecular weight particles are the major species in the NS-TEM micrographs. However, SEC separates proteins by both size and shape, so if the full-length *AtSS4* is a relatively flexible protein with a potentially extended N-terminus, it could follow an SEC elution pattern that is at higher molecular weight than expected for a globular protein of the same size. Alternatively, it is also possible that we are only observing the most stable conformations in the NS-TEM, and the larger assemblies are disrupted during earlier processing, such as the grid preparation.

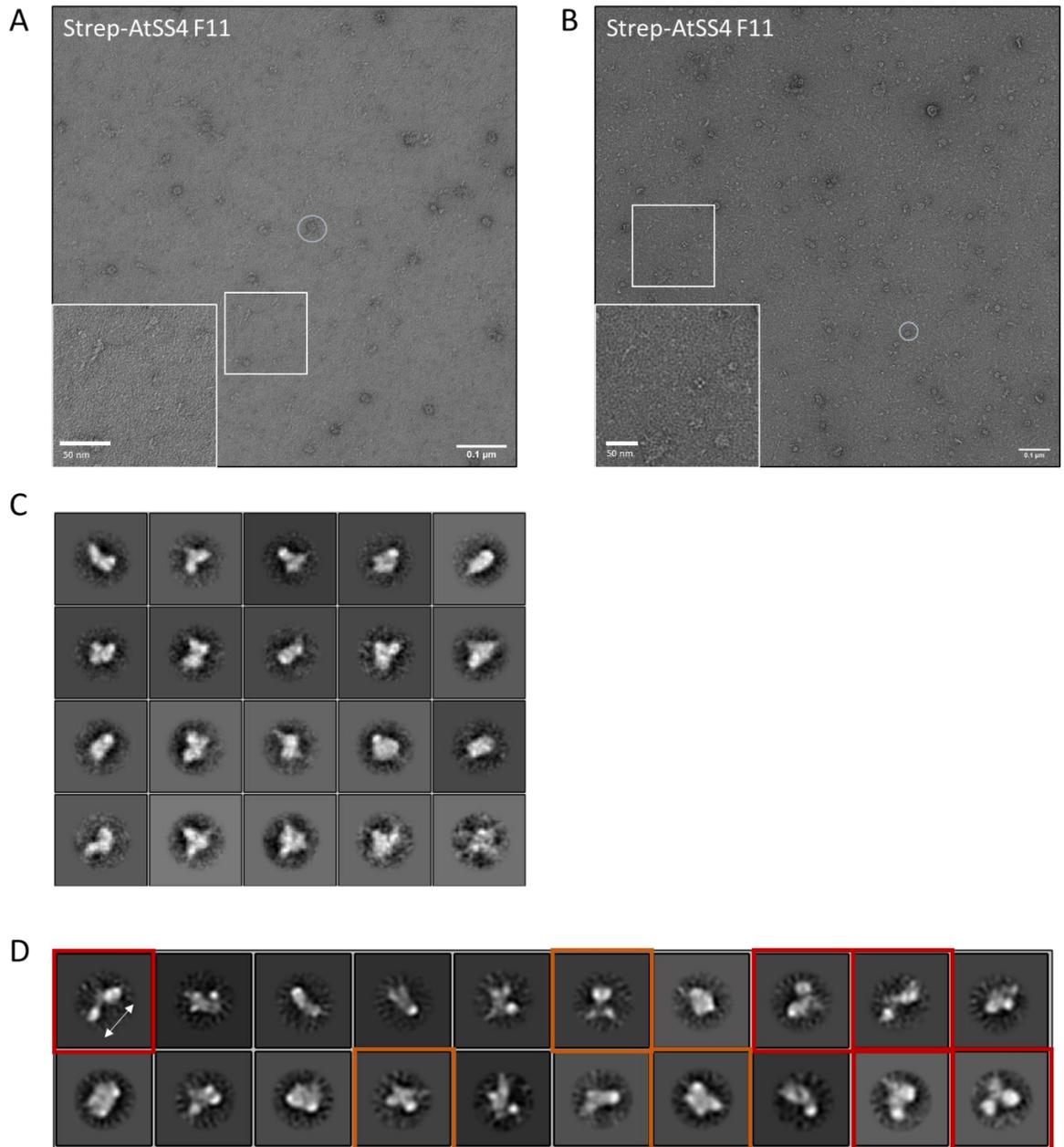


Figure 5.12 Negative stain TEM reveals potential dimeric assemblies of AtSS4. A) Negative stain TEM micrograph of recombinant Strep-AtSS4 fraction 11 after half of a 500 mL purification and size exclusion chromatography (SEC). **B)** Same as A, but for the large scale 4 L purification of AtSS4. The insets in A and B show enlarged images of the areas indicated in the white boxes on the main image. **C)** 2D classes based on 3973 particles, from the grid in A, showing a range of different views, as well as some compositional and/or conformational heterogeneity (box size approx. 450 Å, circular mask diameter 250 Å) **D)** 2D classes based on 93,884 particles, from the grid in B, using the 2D classes from C as autopicking templates (box size approx. 350 Å, circular mask diameter 220 Å). Boxes of the same colour enclose averages that resemble the same shapes. The distance between the arrows is approx. 160Å.

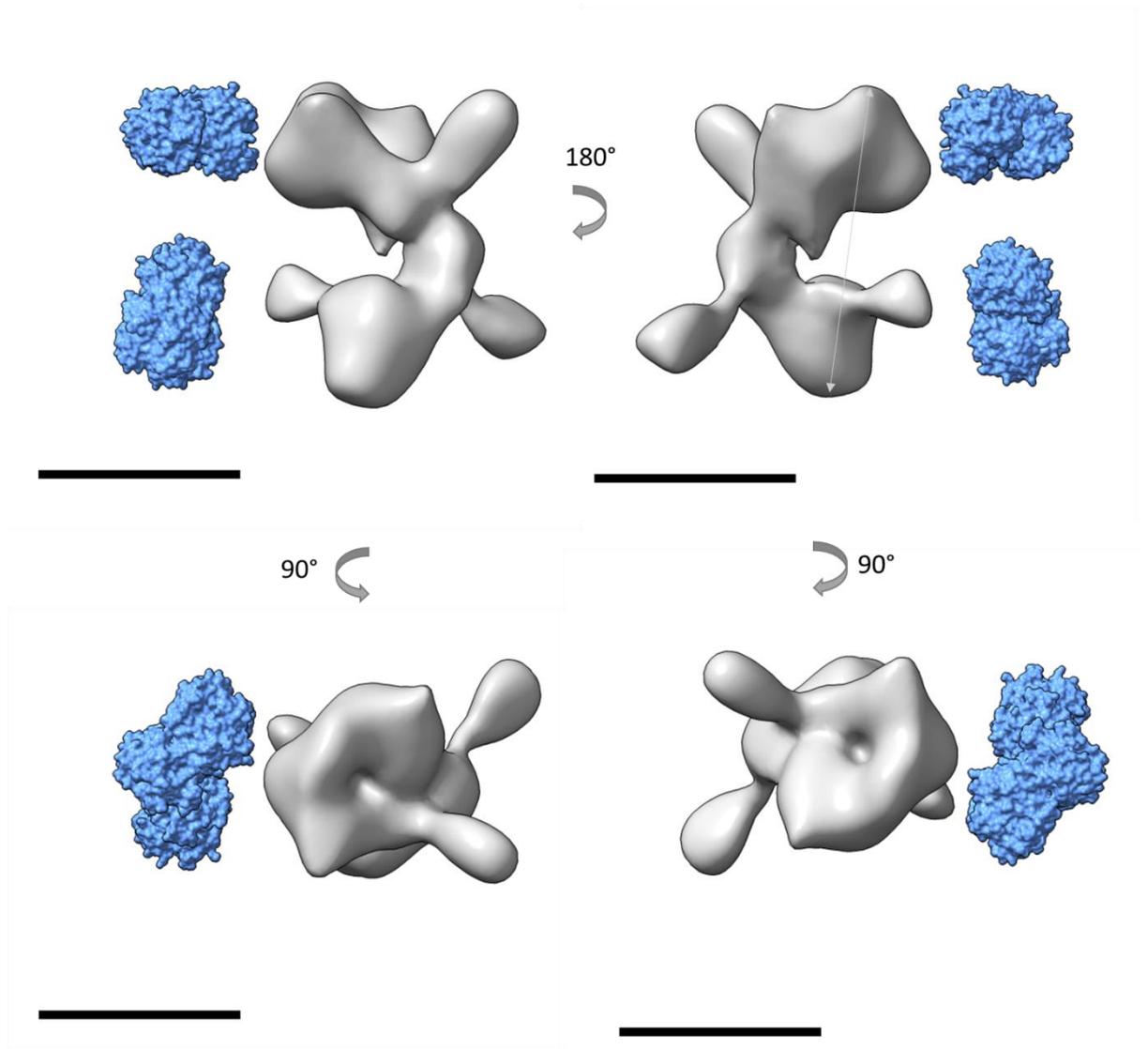


Figure 5.13. 3D density map of the *AtSS4* 2D classes shows a size that is consistent with a dimer of *AtSS4*. The 3D density map (grey) was created using the 2D class averages from Figure 5.12D and visualised using ChimeraX. Smaller blobs which were disconnected from the main volume (and are likely noise) were hidden in these images to simplify the display, by using the 'surface dust all s 16' command. The crystal structures of two monomers of the *AtSS4* catalytic domain (PDB: 6GNE, here visualised in blue) in different orientations are visualised next to the 3D density map (grey) for scale. The distance between the arrows in the second panel is approx. 160 Å. Scale bar = 100 Å, resolution of 3D density map = 22.5 Å.

5.3 Discussion

In this chapter, I have shown that SS4 exists in the high molecular weight range using SEC in both Arabidopsis and wheat leaves. In Arabidopsis leaves, the AtSS4 complex likely consists of multiple AtSS4 subunits, but there are possibly more transient interactions with other proteins. The size and stoichiometry of this AtSS4 complex is still unknown, but my experiments suggest that this might not be one large, stable complex. Rather, AtSS4 could form stable dimeric assemblies. It is possible that these dimers might assemble into higher order complexes with some flexibility, possibly facilitated by coiled coils formed by the N-termini of AtSS4 proteins.

5.3.1 SS4 could assemble into dimers and possibly higher order complexes

Assembly of multiple SS4 monomers could be a way to concentrate its substrates until a critical amount is reached for nucleation of an initiation point through crystallisation. Matilda Crumpton Taylor proposed in her thesis (Crumpton-Taylor, 2010) that ‘the most likely mechanism of granule initiation is via spontaneous crystallisation when the concentration of suitable glucans or malto-oligosaccharides within a volume of stroma reaches a critical level’. This could involve homogenous or heterogenous nucleation, and SS4 could be a protein that facilitates the heterogenous nucleation, enabling sufficient glucans in proximity for a crystallisation event.

Although SS4 proteins were detected at high molecular weight fractions both *in planta* and *in vitro*, the distribution pattern was much wider for the recombinant purified proteins. This discrepancy could be for a number of reasons, such as aberrant protein behaviour of SS4 *in vitro* without other protein partners, the presence of heterogenous protein fragments in the purification, or an atypical SEC elution pattern caused by a non-globular protein, or a combination of these reasons. The predicted flexible coils on the SS4 N-terminus could cause it to interact in an extended fashion with other SS4 monomers, causing atypical interactions and exposure of hydrophobic surfaces in the absence of other partner proteins.

A purer protein purification will help clarify some of the assembly behaviour of AtSS4 proteins *in vitro*. If the AtSS4 N-terminus is quite flexible, this could result in heterogenous aggregation and a range of assemblies, rather than one particular assembly. The SS4 degradation products could also influence heterogenous aggregation, with some of the species in the higher molecular weight fractions possibly consisting of a heterogenous assembly of different truncated versions. This is the issue raised by Raynaud *et al.* (2016) during their purification, and why they did not continue with the SEC using their full-length protein purification. For the experiments in this chapter, the

SEC functioned both as an analysis and purification technique, allowing the sample analysed on the NS-TEM to contain mostly full-length protein. With more sample, perhaps a second round of SEC or ion exchange chromatography could help improve the purity of the protein preparation.

The SS4 complex in the plant could also be influenced by the presence of glucan substrates attached to the SS4 proteins, which could add to the mass of the protein complex or have role in the aggregation of multiple SS4 proteins. In the *E coli* purifications, the SS4 proteins would not have these same glucan substrates. Future experiments should therefore include addition of glucan substrates (such as maltotriose used for the activity assay) to the recombinant SS4 proteins, and measurement of whether their oligomeric state changes using SEC or Blue Native PAGE.

It is also possible that both *in planta* and *in vitro*, there is an overestimation of the protein complex size, which is based on the column calibration of globular proteins. Based on the NS-TEM results, it is possible that in both contexts, SS4 is in fact mainly dimeric. Perhaps the coiled N-terminus of SS4 results in an extended structure that causes it to elute at high molecular weight fractions. However, it must be noted that Raynaud *et al.* (2016) found that using SEC, an N-terminally truncated *AtSS4* eluted as a dimer *in vitro* according to a standard calibration curve. While this conclusion is consistent with my NS-TEM results, it would suggest that *AtSS4* still runs as a globular protein on SEC, but the truncated *AtSS4* used by Raynaud *et al.* (2016) might also not behave the same as the full-length protein.

To investigate the oligomerisation of *AtSS4* in more detail and uncover the composition of its stable assembly unit, further structural studies using techniques that provide higher resolution than NS-TEM need to be carried out. The 2D class averages produced here show promise for further analysis and should be taken forward for cryo-electron microscopy. This would allow us to resolve whether the 2D class averages obtained here indeed show a dimeric state and would also enable us to resolve the nature of the protein interaction of the *AtSS4* N-terminus. Using BiFC with an *AtSS4* construct with either its coiled coils or its conserved region deleted, Raynaud *et al.* (2016) found that the coiled coils were not necessary for *AtSS4* dimerization, but the conserved region was. However, AlphaFold predicted an interaction between two SS4 N-termini to make up one coiled coil, with the conserved region coils not participating in this interaction (Figure 5.10). It is possible that the prediction of the orientation of the conserved regions is inaccurate, perhaps influenced by the fact that the catalytic domain was not included in the prediction, but running a prediction for the dimerization of full length SS4 proteins did not work. It is also possible that the results of (Raynaud *et al.*, 2016) are due to the *AtSS4* protein not folding or functioning properly without the conserved region, independent of a dimerization function. Perhaps it is necessary to have this linker region between the coils and the catalytic domain for a functional protein.

5.3.2 MRC may be a transient component of the SS4 granule initiation complex

Although Chapter 4 demonstrated that SS4 and MRC can interact with each other, the results in this chapter show no evidence of a stable or strong interaction between *AtMRC* and *AtSS4* in *Arabidopsis* leaves, suggesting that they may interact more transiently.

First, the SEC of leaf extracts shows no shift in the elution pattern of *AtSS4* in absence of *AtMRC* (Figure 5.1), although it is possible that there is a smaller shift which is undetectable with the resolution of the column. In future experiments, the fractions should be blotted with an anti-*AtMRC* antibody to resolve the contribution of *AtMRC* to this complex. Unfortunately, I was not able to purify a good anti-*AtMRC* antibody within this project.

Secondly, *AtMRC* was not detected in the AP-MS of *AtSS4* (Table 5.1). This AP-MS was done using *AtSS4* under a native promoter because I wanted to capture complex formation in a relatively native environment. However, *AtSS4* activity is relatively low compared to other starch synthases (Roldan *et al.*, 2007; Szydlowski *et al.*, 2009), and the *AtSS4* protein might be in quite low abundance. In the *AtPTST2*-YFP pull-down where *AtSS4* was captured (Seung *et al.*, 2018), they used *Arabidopsis* lines expressing *AtPTST2*-YFP under a p35S overexpression promoter, so perhaps the strong expression of the bait protein allowed for less stable interactions to be captured as well. To see if the interaction of *AtSS4* with *AtMRC* is present but just transient, crosslinking mass spectrometry experiments could be done, or a comparison of the current AP-MS with an AP-MS using overexpression of the *AtSS4* bait. The *AtSS4*-*AtSS4* interaction might be the most stable interaction, with additional transient interactions (such as with *AtMRC*) possibly not captured under levels of expression achieved with the native promoter. These interactions could happen in the context of a large *AtSS4* protein complex consisting of multiple *AtSS4* subunits (Figure 5.14A), or it could also be the case with just one extended *AtSS4* dimer (Figure 5.14B).

Further *in vitro* experiments with SS4 and MRC could also shed light on the role of MRC in the SS4 complex. With a recombinant purified MRC, I could test whether the presence of MRC influences the SEC elution pattern of recombinant SS4, or alters SS4 activity. Biophysical analyses could also be done to look at the strength of interaction between SS4 and MRC to test if the binding kinetics are consistent with a more transient interaction. Unfortunately, I was not able to obtain a stable soluble recombinant *AtMRC*, but the results in this chapter show potential for optimisation. This could involve finding an optimal concentration for the *in vitro* purification, as the *AtMRC* protein only started to precipitate after concentration. Storage conditions of the protein could also be optimised or a different vector with a solubility tag (e.g. GST-tag) could be used for expression.

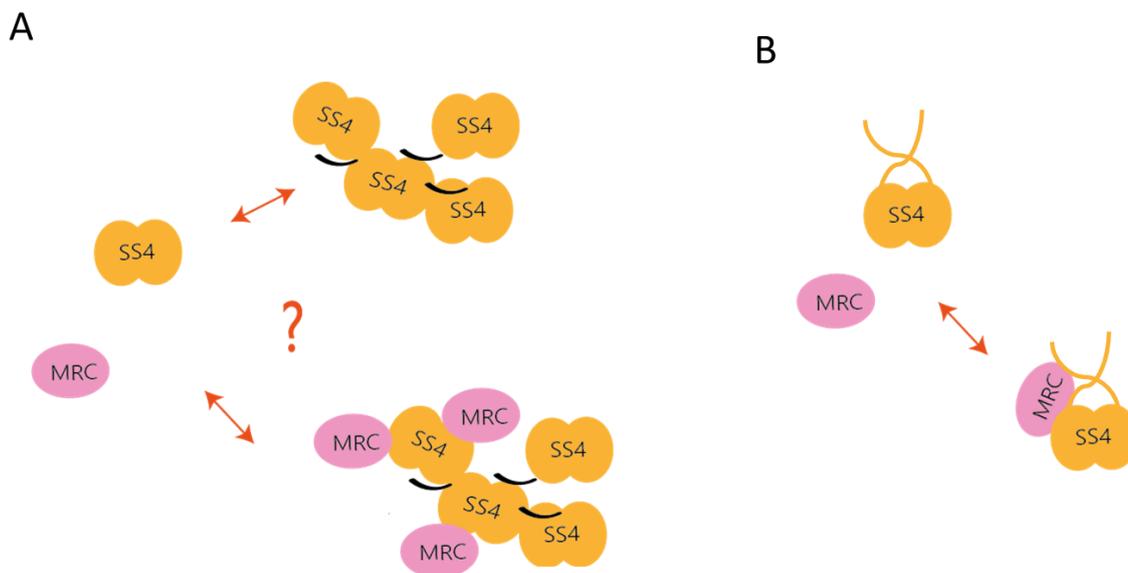


Figure 5.14. Hypothesised mechanisms of granule initiation by SS4. **A)** SS4 proteins could exist as stable dimers that could assemble into higher order structures to concentrate glucan substrates. This complex could also involve more transient interactions with proteins such as MRC. **B)** SS4 could act as a dimer with an extended structure that causes it to elute at HMW fractions in SEC, possibly interacting transiently with other proteins such as MRC.

5.3.3 Complex formation as a method of substrate concentration

As discussed in section 5.3.1, SS4 complex formation could promote the crystallisation of a starch granule initial through substrate concentration. A central role for *AtSS4* for crystallisation of an insoluble starch granule is supported by the results from a heterologous yeast system, where *AtSS4* consistently promoted the production of insoluble glucans in yeast lines expressing various combinations of *AtSS1* – *AtSS4* together with *ISA1/ISA2* (Pfister *et al.*, 2016). To really resolve the assembly of SS4 in plants and its potential substrate concentrating role in granule initiation, further work should focus on obtaining a native purification of the *AtSS4* complex and observing its oligomeric state using biochemical methods and electron microscopy.

Also, how oligomerisation might influence the activity of SS4 should be investigated, to see if oligomerisation enhances enzyme efficiency. The recombinant *AtSS4* and *TaSS4* purified in this chapter had activities of 3.7 turnovers/second and 0.5 turnovers/second respectively, using maltotriose as a substrate. This is only slightly lower than what was reported for barley *HvSS4* which had an *in vitro* activity of approx. 8.8 turnovers per second, using maltopentaose (DP5) as

an acceptor at 10 mM (Cuesta-Seijo *et al.*, 2016), but this is not a direct comparison as it was done with different substrates.

SS4 could be either a processive or distributive enzyme. If granule initiation priming really happens through oligomerisation of SS4 and nucleation through glucan substrate concentration, one might expect the activity of SS4 to be processive, holding on to its products for more than one round of glucan addition. For instance, cellulose synthases that concentrate their products in an oligomeric complex are processive enzymes (Haigler & Roberts, 2019). In peas, the granule bound GBSS is a processive enzyme, but SS2 is distributive (Denyer K, 1999). *In vitro* assays on barley starch synthases indicated that the soluble starch synthases including SS4 are likely distributive (Cuesta-Seijo *et al.*, 2016), but the authors specifically mention that for SS4, 'presence of products with a high degree of polymerization would be qualitatively compatible with a certain degree of processivity as explained for GBSSI, which we can thus not discard'. Even if SS4 is processive, it would not be expected to hold on to glucans for very long, as it is mostly active on short maltooligosaccharides rather than polysaccharides (Cuesta-Seijo *et al.*, 2016). One could also imagine that oligomerisation of SS4 could enable a pseudo-processive mechanism, whereby each individual SS4 enzyme is distributive, but the complex as a whole has a processive function.

5.3.4 Complex formation and substrate concentration in different species

In addition to detailed structural analysis of the *AtSS4* protein from Arabidopsis leaves, it is a big question as to how this compares to *TaSS4* in wheat leaves and endosperm. I have shown in this chapter that SS4 also exists at a high molecular weight in wheat leaves, but its biochemical regulation is not likely to be identical to Arabidopsis, as discussed in Chapter 4. This could be due to different SS4 protein interaction partners in the two species, or because of different granule initiation compensation mechanisms in the absence of SS4 (Hawkins *et al.*, 2021). Even more in contrast to SS4 in the leaves, *TaSS4* in the endosperm seems to have a role in limiting granule initiations rather than promoting them (Hawkins *et al.*, 2021). It will be especially interesting to see whether *TaSS4* acts in a similar potential large complex in the endosperm, and to compare the components of *TaSS4* complexes in the wheat endosperm with those in wheat leaves and Arabidopsis leaves to determine how it performs these contrasting roles.

Rice endosperm SS4 has also been detected at high molecular weight fractions (>700kDa) in SEC (Crofts *et al.*, 2015; Ida *et al.*, 2022), suggesting large protein complex formation of SS4 might be a common method of action for SS4 across species and tissues. SS/SBE large complexes have been studied through SEC and immunoblotting of endosperm extracts in various species (see

Introduction), both in wild type and mutants, where different complexes have been seen. However, still not much is known about the biochemical and structural mechanisms in which these large protein assemblies function.

Overall, in this chapter I have expanded our understanding of the biochemical behaviour of SS4 proteins, showing their tendency to assemble into high molecular weight structures both *in planta* and *in vitro*, but this assembly could be heterogenous *in vitro*. However, NS-TEM of recombinant proteins has shown that AtSS4 likely exists as a stable dimer, and it is possible that the dimers assemble into larger structures or that they have an extended shape that causes the elution at high molecular weight fractions both *in planta* and *in vitro*, raising many further questions to explore regarding the structural mechanism of SS4.

Chapter 6 - General Discussion

The aim of my work was to deepen our understanding of the biochemical mechanism of starch granule initiation, particularly regarding the central role of SS4 and its partner protein MRC. Granule initiation is proposed to act through a large protein complex with SS4 as the core component that provides enzymatic activity. Non-enzymatic granule initiation proteins may associate with this complex and provide the right environment for granule initiation to happen in a regulated way. Here I have characterised a previously unknown role for the durum wheat ortholog of MRC, expanding our understanding of granule initiation in a major starch crop. Similar to in *Arabidopsis*, in wheat MRC also has a role in promoting granule initiation in leaves. By contrast, MRC represses the initiation of B-type granules in the endosperm during early endosperm development, influencing the granule size distribution in mature grains (Chapter 3). I have therefore demonstrated that granule initiation proteins can perform specific and contrasting roles depending on the species and tissue, with big differences in MRC function between wheat leaf and endosperm. The direct interaction between SS4 and MRC seen for the *Arabidopsis* orthologs was also detected for the wheat orthologs, and I have shown that both the N- and C-terminal regions of wheat SS4 are involved in this interaction, suggesting the whole protein can mediate associations with other proteins (Chapter 4). Expression of wheat MRC could not complement the *Arabidopsis mrc* mutant, while expression of wheat SS4 could only partially complement the *ss4* mutant. Other protein interactions or species-specific protein interaction domains may result in different mechanisms of granule initiation in *Arabidopsis* and wheat.

We next need to understand how these pairwise protein interactions fit into the context of the granule initiation complex in the plant. *AtSS4* can form high molecular weight assemblies on its own, where the *AtSS4* dimer could be the stable unit, possibly as part of a large granule initiation complex (Chapter 5). The interactions with non-enzymatic granule initiation proteins such as *AtMRC* might be transient and understanding the dynamics of granule initiation proteins will be important to elucidate their exact mechanism.

In this chapter, I will discuss the implications of my findings and possible future research directions in relation to three main topics: the different roles of starch granule initiation proteins in various tissues and species, how the granule initiation complex could consist of both stable and transient protein interactions, and how transient interactions could be related to the characteristic punctate localisation of granule initiation proteins in the context of biomolecular condensates.

6.1 The role of starch granule initiation proteins can vary between species and tissues

The large diversity in storage starch granule morphologies among species at least partially reflects differences in the spatiotemporal patterns of granule initiation, suggesting differences in the mechanism of granule initiation between species. This could be facilitated by known proteins of granule initiation performing different roles depending on the species and tissue they are in, and by different compensation mechanisms in various tissues when certain protein functions are disrupted. *AtPTST2* and *AtMRC* are proposed to act in the same complex in Arabidopsis leaf granule initiation, but the opposite expression patterns between *TtBGC1* and *TtMRC* (and other granule initiation genes) in the endosperm through grain development suggest that these orthologs might not act together during granule initiation in the wheat endosperm. This is also consistent with the opposite phenotypes of *Ttbgc1* (fewer B-type granules, (Chia *et al.*, 2020)) and *Ttmrc* mutants (higher volume of B-type granules, Chapter 3) in durum wheat.

Obtaining more information on protein levels during grain development and how well they match transcript levels, and exploring whether wheat BGC1 and MRC interact and are part of the same complex in the developing wheat endosperm will be an important step in understanding B-type granule initiation. This can be achieved through biochemical approaches such as SEC, immunoblotting and proteomics throughout endosperm development. Applying these methods over developmental time might also reveal changes in protein interactions through time. Co-IP could determine whether different proteins are pulled down with the same bait over time, and AP-MS and SEC with immunoblotting of developing endosperm material could reveal whether there are changes in protein complex composition and size. As MRC and BGC1 are non-enzymatic and likely function in granule initiation through their influence on SS4, these interactions will also need to be further studied in the endosperm. The sources for glucan substrates for granule initiations in the endosperm should also be investigated, since the initiation of A-type granules would need to use *de novo* glucan substrates at the very start of starch synthesis, but the A-type granules can serve as a source of glucan substrates for B-type granule initiation.

The occurrence of the two distinct types of granules within the endosperm that differ in morphology presents an ideal material to study mechanisms that control granule morphology. Studies have also quantified differences between the structural and physicochemical properties of wheat A- and B-type granules, but the results were not always consistent (Ao & Jane, 2007; Li *et al.*, 2016; Ma *et al.*, 2015; Saccomanno *et al.*, 2022; Salman *et al.*, 2009; Shang *et al.*, 2020; Sun *et al.*, 2021). Many of these variations may be due to differences in the purity of A- and B-type granule preparations, and differences in genetic material.

My work, alongside previous work, identifies proteins and interactions that are important in B-type granule initiation specifically. Further characterisation of these proteins could advance our understanding of whether differences in A- and B-type granule morphology can be explained by similar differences in genetic control. It is possible that distinct proteins are responsible for controlling A- and B-type granule growth, and that such proteins may also contribute to the variation in granule morphology between species and tissues.

Alternatively, the difference in A- and B-type granule morphology may be inherent to the properties of the polymers within the granules. It has been proposed that the differences in amylopectin structure between A- and B-type granules, particularly the chain length distribution, could promote specific alignment of the amylopectin to assemble into the more disc-shaped A-type granules or more spherical B-type granules (Ao & Jane, 2007). However, the details of the amylopectin matrix arrangement within granules are still unknown. No changes in overall chain length distribution from WT were observed for the wheat endosperm starch in *mrc* mutants (Chapter 3), which had different A/B-type granule volume ratios than WT, or *ss4* mutants (Hawkins *et al.*, 2021), which had aberrant granule morphology compared to WT. Therefore, additional properties other than amylopectin structure may still underlie the differences in morphology between A- and B-type granules.

MRC has a similar role in wheat leaves compared to Arabidopsis leaves, as it also promotes granule initiation. However, the less severe *mrc* phenotype in wheat compared to Arabidopsis highlights differences in the roles that granule initiation proteins may have depending on the physiological context. Interestingly, wheat *ss4* mutants could have a more severe leaf starch phenotype than Arabidopsis *ss4*, as the total starch content in *ss4* wheat leaves was less than wild type (Hawkins *et al.*, 2021), whereas in Arabidopsis this was variable, and no difference was measured in one study (Seung *et al.*, 2018). Variability in mutant phenotypes across species was recently seen for *ptst2* mutants of wheat, barley and *Brachypodium*, with the most severe effects on starch granule initiation in leaves observed in *Brachypodium* (Watson-Lazowski *et al.*, 2022), and these effects were also distinct from the effects seen in Arabidopsis (Seung *et al.*, 2017) and rice (Zhang *et al.*, 2022).

Although all the studied granule initiation mutants have fewer (and sometimes larger) granules per chloroplast in wheat leaves (Chapter 3 for MRC, (Hawkins *et al.*, 2021) for SS4, (Watson-Lazowski *et al.*, 2022) for BGC1), they do not have the striking single large granule per chloroplast seen in the Arabidopsis mutants defective in any of these proteins. Watson-Lazowski *et al.* (2022) proposed two reasons for interspecies variation of granule initiation phenotypes: difference in sensitivity to defects in granule initiation in the leaves (*Brachypodium* and Arabidopsis are more reliant on leaf starch as a carbohydrate reserve than the other grass species), and differences in

the compensatory mechanisms of granule initiation in the absence of specific proteins. This could be due to a difference in the protein composition (the Pooidiae do not have PTST3 and SS5 orthologs), or differences in the function of the same set of proteins. There are also differences in the transcriptional regulation of starch metabolism genes between leaves of different species, for example some genes in wheat leaves are less regulated by the circadian clock than in Arabidopsis leaves (Rees *et al*, 2022).

The less tight regulation of starch metabolism in wheat leaves could explain why there is such a big variation in their starch content between experiments (Chapter 3), making starch content more sensitive to environmental changes. Environmental factors like temperature and light could have a big effect on starch granule morphology and starch content, but this has not been systematically studied yet. This large variation in phenotypes in wheat leaves makes it difficult to study quantitative starch traits. Leaf starches look similar among studied species, but there could be differences between species in the number of granules that are typically initiated per chloroplast. Wild-type wheat leaves had a broad range of granule numbers per chloroplast, reaching much higher numbers (up to 17 granules/chloroplast section) than has been seen for Arabidopsis. Since our measurements of granule number in wheat were done on 2D LM sections, the actual number of granules per chloroplast is likely even higher in wheat leaves. We also do not know whether the number of granules per chloroplast in wheat leaves increases over chloroplast development as seen for Arabidopsis leaves (Crumpton-Taylor *et al.*, 2012) and exploring this will reveal more about the dynamics of starch granule initiation in wheat leaves. Robust experimental design and statistical analyses need to be developed to study these highly variable starch phenotypes, as small sample sizes will be prone to errors.

6.2 Coordination of stable and transient protein interactions could facilitate the mechanisms of the granule initiation complex

Different granule initiation complexes could be acting in different tissues, or even within the same tissue since the protein interactions within the complex could be dynamic. The stable homo-dimeric SS4 could be at the heart of these complexes. Further biochemical analyses to investigate the kinetics of AtSS4 dimerization compared with its interaction with other proteins like AtMRC can reveal whether associations with other proteins are weaker and more amenable to dynamic interactions. We also need to understand whether multiple AtSS4 dimers associate to form larger oligomeric complexes, and if these associations are weaker than between two AtSS4 monomers in a stable dimeric unit. *In vitro* studies using purified proteins can help elucidate the inherent

binding kinetics of these proteins, and *in planta* proteomic studies can put these interactions into a larger context. Pulling down a native AtSS4 complex from Arabidopsis leaves and studying its structure using electron microscopy could reveal whether multiple dimeric units assemble together on a higher order. The elution pattern of this native AtSS4 complex on SEC should be compared to that observed in Arabidopsis protein extracts (Chapter 5). Overall, these experiments will provide more robust evidence on the possible dynamic configurations of the AtSS4 complex in plants.

To better understand complex dynamicity, future research should distinguish between stable complex members and transient interactors of the complex, as well as precisely which members of the granule initiation complex interact directly with each other, and which interactions are facilitated by other proteins. Cross-linking mass spectrometry (XL-MS) experiments could capture proteins that interact more transiently with the complex, like AtMRC and AtPTST2, and enable discovery of potentially more unidentified proteins involved in granule initiation, that could not be detected with previous conventional AP-MS methods. XL-MS could also enable us to explore the direct interaction surfaces between proteins, providing spatial restraint information (Graziadei & Rappsilber, 2022) which can be used for determining which proteins interact directly and for resolving the conformation of flexible domains for which structural information is difficult to obtain. In this way, XL-MS data could inform the model building procedure of the cryo-EM work proposed in Chapter 5.

The stoichiometry of different proteins within the complexes could also be studied using *in vitro* reconstitutions of a (partial) granule initiation complex consisting of AtSS4, AtMRC, AtPTST2, AtPTST3 and AtMFP1, or subsets of these. Comparing complex formation in different mutants that lack individual complex members could also aid our understanding of the interactions within the complex. This could provide information on the dynamicity of protein complexes, since it is possible that when one of the granule initiation proteins is missing, different compensatory mechanisms could be at play in different tissues and species, forming various alternative complexes. In rice amylopectin biosynthesis, the compensatory complexes formed differed depending on the specific SS and BEs which were missing, demonstrating flexibility in the complex formations (Ida *et al.*, 2022).

This project has focused on the role of AtMRC because of its direct association with AtSS4, but this cannot be seen outside of the context of AtPTST2, seeing as AtMRC was pulled down from Arabidopsis leaves using AtPTST2 as bait (Seung *et al.*, 2018). AtPTST2 is hypothesised to bring long glucan substrates (> DP10) with secondary structure to AtSS4 (Seung *et al.*, 2017), which are needed for crystallisation (Gidley & Bulpin, 1987). The ability of AtSS4 to promote crystalline glucan formation does not require AtPTST2 *per se*, since AtSS4 alone was sufficient to do this in a

heterologous yeast system (Pfister *et al.*, 2016). Seung *et al.* (2017) proposed that *AtSS4* dimerization might also be important for crystallisation. However, the *ptst2* mutants demonstrate that the precise regulation of granule initiation and crystallisation could require substrate coordination by *AtPTST2*. Initiation and crystallisation are perhaps two separate but connected processes coordinated by *AtSS4*, and how *AtMRC* and *AtPTST2* are involved remains to be elucidated.

A large protein complex may provide a mechanism to connect the different possible roles of *AtSS4* (see Chapter 1): elongating *de novo* glucan primers, elongating longer glucan chains, coordinating the timing and location of starch granule initiation and regulating granule morphology. A large complex could form a single starch granule initiation point within specific locations in the plastid. Within this complex, glucan primers (such as *de novo* synthesised maltose and longer maltooligosaccharides) could be elongated by the *AtSS4* glycosyltransferase activity, in the processes of primary and secondary granule initiation (Figure 1.3). This process could happen efficiently due to the proximity of many glucan substrates and *AtSS4* dimers in the plastid stroma, possibly enabling a pseudo-processive mechanism whereby elongated glucans can be passed on to another *AtSS4* protein. This activity could be enhanced by non-enzymatic proteins such as *AtMRC*, potentially providing a scaffold for oligomerisation. *AtPTST2* could promote the formation of crystalline products by *AtSS4* by delivering longer glucan substrates that have already developed secondary structure.

Once *AtSS4* has seeded initiation of a single starch granule, other starch synthases, BEs and DBEs can act together to synthesise a starch granule. *AtSS4* is then also involved in coordinating the anisotropic growth of Arabidopsis leaf starch granules. It is not known whether the *AtSS4* glycosyltransferase activity is connected to its control of granule morphology. To dissect this, we can study the starch granule phenotype of transgenic Arabidopsis lines expressing an inactive *AtSS4* in an *ss4* mutant background. The inactivation of *AtSS4* can be achieved by mutating a glutamate at position 929, which is a highly conserved residue in the catalytic domain of starch synthases. The equivalent conserved glutamate has been used to inactivate *AtGBSS* (E486A) (Seung *et al.*, 2015). This inactivation could dissect whether *AtSS4* glycosyltransferase activity is tightly connected to its control of granule initiation, granule morphology, or both.

6.3 How can we consolidate the punctate localisation of granule initiation proteins with the granule initiation complex?

Connecting the biochemistry of protein interactions and complex formation with the nature of the punctate localisation of these proteins within plastids will be key to understanding their mechanism. If transient and dynamic biochemical interactions are involved in regulating granule initiation complexes, the punctate localisation observed for granule initiation proteins could also be dynamic. The *AtSS4/AtMRC* puncta are less numerous and more distinct than the *AtPTST2/AtMFP1* puncta, and it is unknown whether some of the *AtPTST2/AtMFP1* puncta overlap with those of *AtSS4/AtMRC*. *AtSS4* was detected both in the stroma and attached to thylakoid membranes (Gamez-Arjona *et al.*, 2014; Seung *et al.*, 2018), but it is not known if their puncta represent one or both of these states. It is tempting to speculate that each of these puncta could represent a granule initiation event at a specific location within the plastid. This could be studied in more detail using a similar experimental set up to (Burgy *et al.*, 2021), where starch-depleted chloroplasts could be followed through the process of granule initiation, and the localisation of granule initiation proteins like *AtSS4* could be simultaneously studied. Observing whether the number of *AtSS4* puncta changes throughout plant development could also be informative, since chloroplasts of young leaves tend to have more starch granules than those of older leaves (Crumpton-Taylor *et al.*, 2012).

Punctate protein localisation has gained a lot of attention in recent years, due to its association with proteins that demonstrate liquid-liquid phase-separation (LLPS), also named biomolecular condensates (Banani *et al.*, 2017). These condensates are composed of dynamic protein assemblies that associate in high concentration and display liquid-like behaviour, enabling rapid assembly and disassembly of protein complexes. One example of plastidial LLPS is the pyrenoid, a membrane-less sub-compartment which forms a carbon concentrating mechanism for efficient carbon fixation in algal chloroplasts. In *Chlamydomonas reinhardtii* pyrenoids, multiple Rubisco complexes are connected by a network of tethering proteins called EPYC1, forming through LLPS (He *et al.*, 2020).

The dynamic behaviour and substrate/protein concentrating action of biomolecular condensates would be consistent with the proposed interactions in the granule initiation complex. Firstly, Banani *et al.* (2017) highlighted that a key difference between a molecular condensate and a canonical macromolecular protein complex is that condensates can vary in their composition, stoichiometry and size, typically concentrating 10 to several hundred different proteins and they are often much larger than typical macromolecular assemblies. Canonical macromolecular complexes are more static in their size and composition. If this is the case for the SS4 granule

initiation complex, dynamic associations could partially explain the broad SS4 elution patterns I observed using SEC (Figure 5.9). Secondly, LLPS behaviour tends to be facilitated by disordered regions on proteins. The broad SEC elution patterns could also be due to the SS4 proteins not eluting according to globular protein sizes, with a potentially extended and flexible N-terminus. The predicted disordered regions at the start of SS4 proteins (Chapter 5) could be involved in dynamic interactions, while the coiled coil domain (and perhaps also the glycosyltransferase domain) could be more involved in stable interactions, or even in LLPS as well, as coiled-coil domains have also been shown to facilitate LLPS (Fang *et al*, 2019). Finally, in LLPS, scaffolds are components that are essential for condensate formation (Banani *et al.*, 2017). Since AtMRC was not necessary for AtSS4 to form punctate localisation in protoplasts (Vandromme *et al.*, 2019), it is unlikely that AtMRC is required for LLPS, but its puncta-inducing behaviour in a transient *N. benthamiana* system (Chapter 4) suggests it may share some of these scaffolding protein properties.

If the granule initiation proteins are forming a biomolecular condensate through LLPS, it may complicate the purification of distinct complexes from plants. However, the transgenic lines *ss4/pAtSS4:AtSS4-eGFP-HA* and *mrc/pAtMRC:AtMRC-RFP-HA* generated in this thesis could enable their characterisation *in planta*. The dynamics of the SS4 and MRC puncta could be studied using techniques such as fluorescence recovery after photobleaching (FRAP). This is a commonly used technique to study the liquid-like mixing behaviour of proteins, where fluorescent proteins are bleached, and if they have liquid-like behaviour they will then reassemble back in the bleached area. Additionally, we should explore whether granule initiation proteins also localise to puncta in amyloplasts, and if so whether this could also be through LLPS. As amyloplasts do not have thylakoid membranes, it would be interesting to see if the localisation of MFP1 and other partner proteins is different from that in chloroplasts, and whether this could also shed some light on the differences between starch granule initiation in different tissues and species. Creating stable transgenic lines in wheat expressing fluorescently tagged granule initiation proteins would allow us to perform similar studies in wheat as in Arabidopsis. We can also study *in vitro* condensate formation by purifying fluorescently tagged SS4 and MRC proteins and observing their assembly behaviour under different concentrations.

References

- Abe N, Asai H, Yago H, Oitome NF, Itoh R, Crofts N, Nakamura Y, Fujita N (2014) Relationships between starch synthase I and branching enzyme isozymes determined using double mutant rice lines. *BMC Plant Biol* 14: 80
- Abt MR, Pfister B, Sharma M, Eicke S, Burgy L, Neale I, Seung D, Zeeman SC (2020) STARCH SYNTHASE5, a Noncanonical Starch Synthase-Like Protein, Promotes Starch Granule Initiation in Arabidopsis. *Plant Cell* 32: 2543-2565
- Abt MR, Zeeman SC (2020) Evolutionary innovations in starch metabolism. *Curr Opin Plant Biol* 55: 109-117
- Ahmed Z, Tetlow IJ, Ahmed R, Morell MK, Emes MJ (2015) Protein-protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heteromeric enzyme complexes in the synthesis of A and B granules. *Plant Sci* 233: 95-106
- Ao Z, Jane J (2007) Characterization and modeling of the A- and B-granule starches of wheat, triticale, and barley. *Carbohydrate Polymers* 67: 46-55
- Austin JR, 2nd, Frost E, Vidi PA, Kessler F, Staehelin LA (2006) Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. *Plant Cell* 18: 1693-1703
- Ball SG, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* 54: 207-233
- Banani SF, Lee HO, Hyman AA, Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18: 285-298
- Bathgate GN, Palmer GH (1973) THE IN VIVO AND IN VITRO DEGRADATION OF BARLEY AND MALT STARCH GRANULES. *Journal of the Institute of Brewing* 79: 402-406
- Bechtel DB, Zayas, I., Kaleikau, L., and Pomeranz, Y. (1990) Size-distribution of wheat starch granules during endosperm development. *Cereal Chemistry* 67: 59 - 63
- Bertoft E (2017) Understanding Starch Structure: Recent Progress. *Agronomy* 7
- Blennow A, Bay-Smidt, A.M., Wischmann, B., Olsen, C.E., and Møller, B.L. (1998) The degree of starch phosphorylation is related to the chain length distribution of the neutral and the phosphorylated chains of amylopectin. *Carbohydrate Research* 307: 45 - 54
- Borrill P, Ramirez-Gonzalez R, Uauy C (2016) expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiol* 170: 2172-2186
- Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371: 578-586
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34: 525-527
- Brummell DA, Watson LM, Zhou J, McKenzie MJ, Hallett IC, Simmons L, Carpenter M, Timmerman-Vaughan GM (2015) Overexpression of STARCH BRANCHING ENZYME II increases short-chain branching of amylopectin and alters the physicochemical properties of starch from potato tuber. *BMC Biotechnology* 15: 28
- Brust H, Orzechowski S, Fettke J, Steup M (2013) Starch Synthesizing Reactions and Paths: in vitro and in vivo Studies. *Journal of Applied Glycoscience* 60: 3-20
- Bull SE, Seung D, Chanez C, Mehta D, Kuon JE, Truernit E, Hochmuth A, Zurkirchen I, Zeeman SC, Gruissem W *et al* (2018) Accelerated ex situ breeding of GBSS- and PTST1-edited cassava for modified starch. *Science Advances* 4
- Burgy L, Eicke S, Kopp C, Jenny C, Lu KJ, Escrig S, Meibom A, Zeeman SC (2021) Coalescence and directed anisotropic growth of starch granule initials in subdomains of Arabidopsis thaliana chloroplasts. *Nat Commun* 12: 6944
- Carciofi M, Shaik SS, Jensen SL, Blennow A, Svensson JT, Vincze É, Hebelstrup KH (2011) Hyperphosphorylation of cereal starch. *Journal of Cereal Science* 54: 339-346

- Caspar T, Huber SC, Somerville CR (1985) Alterations in growth, photosynthesis, and respiration in a starchless mutant of *Arabidopsis thaliana* (L.) deficient in chloroplast phosphoglucomutase activity. *Plant Physiol* 79: 11 - 17
- Chen J, Chen Y, Watson-Lazowski A, Hawkins E, Barclay JE, Fahy B, Denley Bowers R, Corbin K, Warren FJ, Blennow A *et al* (2022a) The plastidial protein MRC promotes starch granule initiation in wheat leaves but delays B-type granule initiation in the endosperm. *bioRxiv*: 2022.2010.2007.511297
- Chen J, Hawkins E, Seung D (2021) Towards targeted starch modification in plants. *Curr Opin Plant Biol* 60: 102013
- Chen J, Watson-Lazowski A, Vickers M, Seung D (2022b) Gene expression profile of the developing endosperm in durum wheat provides insight into starch biosynthesis. *bioRxiv*: 2022.2010.2021.513215
- Chia T, Adamski NM, Saccomanno B, Greenland A, Nash A, Uauy C, Trafford K (2017) Transfer of a starch phenotype from wild wheat to bread wheat by deletion of a locus controlling B-type starch granule content. *J Exp Bot* 68: 5497-5509
- Chia T, Chirico M, King R, Ramirez-Gonzalez R, Saccomanno B, Seung D, Simmonds J, Trick M, Uauy C, Verhoeven T *et al* (2020) A carbohydrate-binding protein, B-GRANULE CONTENT 1, influences starch granule size distribution in a dose-dependent manner in polyploid wheat. *J Exp Bot* 71: 105-115
- Crofts N, Abe N, Oitome NF, Matsushima R, Hayashi M, Tetlow IJ, Emes MJ, Nakamura Y, Fujita N (2015) Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *J Exp Bot* 66: 4469-4482
- Crofts N, Nakamura Y, Fujita N (2017) Critical and speculative review of the roles of multi-protein complexes in starch biosynthesis in cereals. *Plant Sci* 262: 1-8
- Crumpton-Taylor M, 2010. Starch granule number and size in *Arabidopsis thaliana* leaves, John Innes Centre. University of East Anglia, Norwich, UK.
- Crumpton-Taylor M, Grandison S, Png KM, Bushby AJ, Smith AM (2012) Control of starch granule numbers in *Arabidopsis* chloroplasts. *Plant Physiol* 158: 905-916
- Crumpton-Taylor M, Pike M, Lu KJ, Hylton CM, Feil R, Eicke S, Lunn JE, Zeeman SC, Smith AM (2013) Starch synthase 4 is essential for coordination of starch granule formation with chloroplast division during *Arabidopsis* leaf expansion. *New Phytol* 200: 1064-1075
- Cuesta-Seijo JA, Nielsen MM, Ruzanski C, Krucewicz K, Beeren SR, Rydhal MG, Yoshimura Y, Striebeck A, Motawia MS, Willats WG *et al* (2016) In vitro Biochemical Characterization of All Barley Endosperm Starch Synthases. *Front Plant Sci* 6: 1265
- Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P *et al* (2014) Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biology* 15: 546
- Delatte T, Trevisan M, Parker ML, Zeeman SC (2005) *Arabidopsis* mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. *Plant J* 41: 815-830
- Denyer K WD, Motawia S, Møller BL, Smith AM. (1999) Granule-bound starch synthase I in isolated starch granules elongates malto-oligosaccharides processively. *Biochem J* 340 (Pt 1): 183-191
- Dhital S, Shrestha AK, Gidley MJ (2010) Relationship between granule size and in vitro digestibility of maize and potato starches. *Carbohydrate Polymers* 82: 480-488
- Erdos G, Pajkos M, Dosztanyi Z (2021) IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* 49: W297-W303
- Esch L, Ngai QY, Barclay JE, Seung D (2022) AtFZL is required for correct starch granule morphology in *Arabidopsis* chloroplasts. *bioRxiv*: 2022.2010.2020.512996
- Fang X, Wang L, Ishikawa R, Li Y, Fiedler M, Liu F, Calder G, Rowan B, Weigel D, Li P *et al* (2019) *Arabidopsis* FLL2 promotes liquid-liquid phase separation of polyadenylation complexes. *Nature* 569: 265-269

- Feike D, Seung D, Graf A, Bischof S, Ellick T, Coiro M, Soyk S, Eicke S, Mettler-Altmann T, Lu KJ *et al* (2016) The Starch Granule-Associated Protein EARLY STARVATION1 Is Required for the Control of Starch Degradation in *Arabidopsis thaliana* Leaves. *Plant Cell* 28: 1472-1489
- Fernandez O, Ishihara H, George GM, Mengin V, Flis A, Sumner D, Arrivault S, Feil R, Lunn JE, Zeeman SC *et al* (2017) Leaf Starch Turnover Occurs in Long Days and in Falling Light at the End of the Day. *Plant Physiol* 174: 2199-2212
- Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340: 245-246
- Funfgeld M, Wang W, Ishihara H, Arrivault S, Feil R, Smith AM, Stitt M, Lunn JE, Niittyla T (2022) Sucrose synthases are not involved in starch synthesis in *Arabidopsis* leaves. *Nat Plants* 8: 574-582
- Gamez-Arjona FM, Li J, Raynaud S, Baroja-Fernandez E, Munoz FJ, Ovecka M, Ragel P, Bahaji A, Pozueta-Romero J, Merida A (2011) Enhancing the expression of starch synthase class IV results in increased levels of both transitory and long-term storage starch. *Plant Biotechnol J* 9: 1049-1060
- Gamez-Arjona FM, Merida A (2021) Interplay Between the N-Terminal Domains of *Arabidopsis* Starch Synthase 3 Determines the Interaction of the Enzyme With the Starch Granule. *Front Plant Sci* 12: 704161
- Gamez-Arjona FM, Raynaud S, Ragel P, Merida A (2014) Starch synthase 4 is located in the thylakoid membrane and interacts with plastoglobule-associated proteins in *Arabidopsis*. *Plant J* 80: 305-316
- Gidley MJ, Bulpin PV (1987) Crystallisation of malto-oligosaccharides as models of the crystalline forms of starch: minimum chain-length requirement for the formation of double helices. *Carbohydrate Research* 161: 291 - 300
- Gietz RD, Woods RA (2001) Genetic transformation of yeast. *Biotechniques* 30: 816-820, 822-816, 828 passim
- Graf A, Schlereth A, Stitt M, Smith AM (2010) Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proc Natl Acad Sci U S A* 107: 9458-9463
- Graziadei A, Rappsilber J (2022) Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure* 30: 37-54
- Guo H, Liu Y, Li X, Yan Z, Xie Y, Xiong H, Zhao L, Gu J, Zhao S, Liu L (2017) Novel mutant alleles of the starch synthesis gene TaSSIVb-D result in the reduction of starch granule number per chloroplast in wheat. *BMC Genomics* 18
- Haigler CH, Roberts AW (2019) Structure/function relationships in the rosette cellulose synthesis complex illuminated by an evolutionary perspective. *Cellulose* 26: 227-247
- Hannah LC (2007) Starch Formation in the Cereal Endosperm. In: *Endosperm: Developmental and Molecular Biology*, Olsen O.-A. (ed.) pp. 179-193. Springer Berlin Heidelberg: Berlin, Heidelberg
- Hawkins E, Chen J, Watson-Lazowski A, Ahn-Jarvis J, Barclay JE, Fahy B, Hartley M, Warren FJ, Seung D (2021) STARCH SYNTHASE 4 is required for normal starch granule initiation in amyloplasts of wheat endosperm. *New Phytol* 230: 2371-2386
- He S, Chou HT, Matthies D, Wunder T, Meyer MT, Atkinson N, Martinez-Sanchez A, Jeffrey PD, Port SA, Patena W *et al* (2020) The structural basis of Rubisco phase separation in the pyrenoid. *Nat Plants* 6: 1480-1490
- Helle S, Bray F, Verbeke J, Devassine S, Courseaux A, Facon M, Tokarski C, Rolando C, Szydowski N (2018) Proteome Analysis of Potato Starch Reveals the Presence of New Starch Metabolic Proteins as Well as Multiple Protease Inhibitors. *Front Plant Sci* 9: 746
- Hennen-Bierwagen TA, Liu F, Marsh RS, Kim S, Gan Q, Tetlow IJ, Emes MJ, James MG, Myers AM (2008) Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. *Plant Physiol* 146: 1892-1908
- Howard T, Rejab NA, Griffiths S, Leigh F, Leverington-Waite M, Simmonds J, Uauy C, Trafford K (2011) Identification of a major QTL controlling the content of B-type starch granules in *Aegilops*. *J Exp Bot* 62: 2217-2228

- Hwang SK, Koper K, Satoh H, Okita TW (2016) Rice Endosperm Starch Phosphorylase (Pho1) Assembles with Disproportionating Enzyme (Dpe1) to Form a Protein Complex That Enhances Synthesis of Malto-oligosaccharides. *J Biol Chem* 291: 19994-20007
- Ida T, Crofts N, Miura S, Matsushima R, Fujita N (2022) Starch Biosynthetic Protein Complex Formation in Rice ss2a be2b (+) Double Mutant Differs from Their Parental Single Mutants. *J Appl Glycosci* (1999) 69: 23-33
- International Wheat Genome Sequencing C (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361
- Ishihara H, Alseekh S, Feil R, Perera P, George GM, Niedzwiecki P, Arrivault S, Zeeman SC, Fernie AR, Lunn JE *et al* (2022) Rising rates of starch degradation during daytime and trehalose 6-phosphate optimize carbon availability. *Plant Physiol* 189: 1976-2000
- Jane J, Chen YY, Lee LF, McPherson AE, Wong KS, Radosavljevic M, Kasemsuwan T (1999) Effects of Amylopectin Branch Chain Length and Amylose Content on the Gelatinization and Pasting Properties of Starch. *Cereal Chemistry Journal* 76: 629-637
- Jane JL, Kasemsuwan T, Leas S, Zobel H, Robyt J (1994) Anthology of starch granule morphology by scanning electron microscopy. *Starch - Starke* 46: 121 - 129
- Jeong SY, Rose A, Meier I (2003) MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure. *Nucleic Acids Res* 31: 5175-5185
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583-589
- Jung Y-J, Nogoy FM, Lee S-K, Cho Y-G, Kang K-K (2018) Application of ZFN for Site Directed Mutagenesis of Rice SSIVa Gene. *Biotechnology and Bioprocess Engineering* 23: 108-115
- Kaur B, Ariffin F, Bhat R, Karim AA (2012) Progress in starch modification in the last decade. *Food Hydrocolloids* 26: 398-404
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907-915
- Kim S, Lee DS, Choi IS, Ahn SJ, Kim YH, Bae HJ (2010) Arabidopsis thaliana Rubisco small subunit transit peptide increases the accumulation of Thermotoga maritima endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. *Transgenic Res* 19: 489-497
- Kitahara K, Hamasuna K, Nozuma K, Otani M, Hamada T, Shimada T, Fujita K, Suganuma T (2007) Physicochemical properties of amylose-free and high-amylose starches from transgenic sweetpotatoes modified by RNA interference. *Carbohydrate Polymers* 69: 233-240
- Kozlov SS, Blennow A, Krivandin AV, Yuryev VP (2007) Structural and thermodynamic properties of starches extracted from GBSS and GWD suppressed potato lines. *International Journal of Biological Macromolecules* 40: 449-460
- Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P *et al* (2017) Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci U S A* 114: E913-E921
- Langeveld SMJ, Van wijk R, Stuurman N, Kijne JW, de Pater S (2000) B-type granule containing protrusions and interconnections between amyloplasts in developing wheat endosperm revealed by transmission electron microscopy and GFP expression. *Journal of Experimental Botany* 51: 1357 - 1361
- Leterrier M, Holappa LD, Broglie KE, Beckles DM (2008) Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications. *BMC Plant Biol* 8: 98
- Li H, Gidley MJ, Dhital S (2019) High-Amylose Starches to Bridge the "Fiber Gap": Development, Structure, and Nutritional Functionality. *Compr Rev Food Sci Food Saf* 18: 362-379
- Li L-F, Zhang Z-B, Wang Z-H, Li N, Sha Y, Wang X-F, Ding N, Li Y, Zhao J, Wu Y *et al* (2022) Genome sequences of five Sitopsis species of Aegilops and the origin of polyploid wheat B subgenome. *Molecular Plant* 15: 488-503
- Li L, Jiang H, Campbell M, Blanco M, Jane J-I (2008) Characterization of maize amylose-extender (ae) mutant starches. Part I: Relationship between resistant starch contents and molecular structures. *Carbohydrate Polymers* 74: 396-404

- Li M, Daygon VD, Solah V, Dhital S (2021) Starch granule size: Does it matter? *Crit Rev Food Sci Nutr*: 1-21
- Li W, Gao J, Wu G, Zheng J, Ouyang S, Luo Q, Zhang G (2016) Physicochemical and structural properties of A- and B-starch isolated from normal and waxy wheat: Effects of lipids removal. *Food Hydrocolloids* 60: 364-373
- Li Y, Andrade J (2017) DEApp: an interactive web interface for differential expression analysis of next generation sequence data. *Source Code Biol Med* 12: 2
- Li Z, Guo K, Lin L, He W, Zhang L, Wei C (2018) Comparison of Physicochemical Properties of Starches from Flesh and Peel of Green Banana Fruit. *Molecules* 23
- Lindeboom N, Chang PR, Tyler RT (2004) Analytical, Biochemical and Physicochemical Aspects of Starch Granule Size, with Emphasis on Small Granule Starches: A Review. *Starch - Stärke* 56: 89-99
- Linden JC, Schilling N, Brackenhofer H, Kandler O (1975) Asymmetric labelling of maltose during photosynthesis in $^{14}\text{CO}_2$. *Zeitschrift für Pflanzenphysiologie*: 176 - 181
- Liu F, Romanova N, Lee EA, Ahmed R, Evans M, Gilbert EP, Morell MK, Emes MJ, Tetlow IJ (2012) Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. *Biochem J* 448: 373-387
- Lomako J, Lomako WM, Whelan WJ (2004) Glycogenin: the primer for mammalian and yeast glycogen synthesis. *Biochim Biophys Acta* 1673: 45-55
- Lopez-Rubio A, Flanagan BM, Gilbert EP, Gidley MJ (2008) A novel approach for calculating starch crystallinity and its correlation with double helix content: a combined XRD and NMR study. *Biopolymers* 89: 761-768
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550
- Lu KJ, Pfister B, Jenny C, Eicke S, Zeeman SC (2018) Distinct Functions of STARCH SYNTHASE 4 Domains in Starch Granule Formation. *Plant Physiol* 176: 566-581
- Lundquist PK, Mantegazza O, Stefanski A, Stuhler K, Weber APM (2017) Surveying the Oligomeric State of Arabidopsis thaliana Chloroplasts. *Mol Plant* 10: 197-211
- Ma S, Zheng XL, Wang XX, Shang JY, Bao QD, Li L (2015) Effect of A- and B-type granules on the physical properties of starch from six wheat varieties. *Quality Assurance and Safety of Crops & Foods* 7: 531-536
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG *et al* (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* 51: 885-895
- Malinova I, Alseekh S, Feil R, Fernie AR, Baumann O, Schottler MA, Lunn JE, Fettke J (2017) Starch Synthase 4 and Plastidal Phosphorylase Differentially Affect Starch Granule Number and Morphology. *Plant Physiol* 174: 73-85
- Malinova I, Qasim HM, Brust H, Fettke J (2018) Parameters of Starch Granule Genesis in Chloroplasts of Arabidopsis thaliana. *Front Plant Sci* 9: 761
- Mason JM, Arndt KM (2004) Coiled coil domains: stability, specificity, and biological implications. *ChemBiochem* 5: 170-176
- Matsushima R (2015) Morphological Variations of Starch Grains. In: *Starch*, pp. 425-441.
- Matsushima R, Maekawa M, Fujita N, Sakamoto W (2010) A rapid, direct observation method to isolate mutants with defects in starch grain morphology in rice. *Plant Cell Physiol* 51: 728-741
- Matsushima R, Maekawa M, Kusano M, Kondo H, Fujita N, Kawagoe Y, Sakamoto W (2014) Amyloplast-localized SUBSTANDARD STARCH GRAIN4 protein influences the size of starch grains in rice endosperm. *Plant Physiol* 164: 623-636
- Matsushima R, Maekawa M, Kusano M, Tomita K, Kondo H, Nishimura H, Crofts N, Fujita N, Sakamoto W (2016) Amyloplast Membrane Protein SUBSTANDARD STARCH GRAIN6 Controls Starch Grain Size in Rice Endosperm. *Plant Physiol* 170: 1445-1459
- Matsushima R, Maekawa M, Sakamoto W (2015) Geometrical Formation of Compound Starch Grains in Rice Implements Voronoi Diagram. *Plant Cell Physiol* 56: 2150-2157

- Matsushima R, Yamashita J, Kariyama S, Enomoto T, Sakamoto W (2013) A Phylogenetic Re-evaluation of Morphological Variations of Starch Grains among Poaceae Species. *Journal of Applied Glycoscience* 60: 37-44
- Mengin V, Pyl ET, Alexandre Moraes T, Sulpice R, Krohn N, Encke B, Stitt M (2017) Photosynthate partitioning to starch in *Arabidopsis thaliana* is insensitive to light intensity but sensitive to photoperiod due to a restriction on growth in the light in short photoperiods. *Plant Cell Environ* 40: 2608-2627
- Mergner J, Frejno M, List M, Papacek M, Chen X, Chaudhary A, Samaras P, Richter S, Shikata H, Messerer M *et al* (2020) Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 579: 409-414
- Merida A, Fettke J (2021) Starch granule initiation in *Arabidopsis thaliana* chloroplasts. *Plant J* 107: 688-697
- Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat Methods* 19: 679-682
- Miura S, Crofts N, Saito Y, Hosaka Y, Oitome NF, Watanabe T, Kumamaru T, Fujita N (2018) Starch Synthase IIa-Deficient Mutant Rice Line Produces Endosperm Starch With Lower Gelatinization Temperature Than Japonica Rice Cultivars. *Frontiers in Plant Science* 9
- Müller LM, Gol L, Jeon J-S, Weber APM, Davis SJ, von Korff M (2018) Temperature but not the circadian clock determines nocturnal carbohydrate availability for growth in cereals. *bioRxiv*
- Myers AM, James MG, Lin Q, Yi G, Stinard PS, Hennen-Bierwagen TA, Becraft PW (2011) Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. *Plant Cell* 23: 2331-2347
- Nakamura Y (2015) Initiation Process of Starch Biosynthesis. In: *Starch*, pp. 315-332.
- Nakamura Y, Kainuma K (2022) On the cluster structure of amylopectin. *Plant Mol Biol* 108: 291-306
- Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7: 61-80
- Nie G, Hendrix, D.L., Webber, A.N., Kimball, B.A., and Long, S.P. (1995) Increased accumulation of carbohydrates and decreased photosynthetic gene transcript levels in wheat grown at an elevated CO₂ concentration in the field. *Plant Physiol* 108: 975-983
- Nielsen MM, Ruzanski C, Krucewicz K, Striebeck A, Cenci U, Ball SG, Palcic MM, Cuesta-Seijo JA (2018) Crystal Structures of the Catalytic Domain of *Arabidopsis thaliana* Starch Synthase IV, of Granule Bound Starch Synthase From CLg1 and of Granule Bound Starch Synthase I of *Cyanophora paradoxa* Illustrate Substrate Recognition in Starch Synthases. *Front Plant Sci* 9: 1138
- Niittyla T, 2003. The role of malto-oligosaccharides in starch metabolism of *Arabidopsis thaliana* (L.) leaves. University of East Anglia.
- Noda T, Kimura T, Otani M, Ideta O, Shimada T, Saito A, Suda I (2002) Physicochemical properties of amylose-free starch from transgenic sweet potato. *Carbohydrate Polymers* 49: 253-260
- O'Connell JD, Zhao A, Ellington AD, Marcotte EM (2012) Dynamic reorganization of metabolic enzymes into intracellular bodies. *Annu Rev Cell Dev Biol* 28: 89-111
- Olinares PD, Ponnala L, van Wijk KJ (2010) Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol Cell Proteomics* 9: 1594-1615
- Parker ML (1985) The relationship between A-type and B-type starch granules in the developing endosperm of wheat. *Journal of Cereal Science* 3: 271 - 278
- Peng C, Wang Y, Liu F, Ren Y, Zhou K, Lv J, Zheng M, Zhao S, Zhang L, Wang C *et al* (2014) FLOURY ENDOSPERM6 encodes a CBM48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. *Plant J* 77: 917-930
- Pfister B, Sanchez-Ferrer A, Diaz A, Lu K, Otto C, Holler M, Shaik FR, Meier F, Mezzenga R, Zeeman SC (2016) Recreating the synthesis of starch granules in yeast. *Elife* 5
- Pfister B, Shields JM, Kockmann T, Grossmann J, Abt MR, Stadler M, Zeeman SC (2022) Tuning heterologous glucan biosynthesis in yeast to understand and exploit plant starch diversity. *BMC Biol* 20: 207

- Pfister B, Zeeman SC (2016) Formation of starch in plant cells. *Cell Mol Life Sci* 73: 2781-2807
- Purushotham P, Ho R, Zimmer J (2020) Architecture of a catalytically active homotrimeric plant cellulose synthase complex. *Science* 369: 1089–1094
- Qu J, Ma C, Feng J, Xu S, Wang L, Li F, Li Y, Zhang R, Zhang X, Xue J *et al* (2016) Transcriptome Dynamics during Maize Endosperm Development. *PLoS One* 11: e0163814
- Ragel P, Streb S, Feil R, Sahrawy M, Annunziata MG, Lunn JE, Zeeman S, Merida A (2013) Loss of starch granule initiation has a deleterious effect on the growth of arabidopsis plants due to an accumulation of ADP-glucose. *Plant Physiol* 163: 75-85
- Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A *et al* (2018) The transcriptional landscape of polyploid wheat. *Science* 361
- Raynaud S, Ragel P, Rojas T, Merida A (2016) The N-terminal Part of Arabidopsis thaliana Starch Synthase 4 Determines the Localization and Activity of the Enzyme. *J Biol Chem* 291: 10759-10771
- Rees H, Rusholme-Pilcher R, Bailey P, Colmer J, White B, Reynolds C, Ward SJ, Coombes B, Graham CA, de Barros Dantas LL *et al* (2022) Circadian regulation of the transcriptome in a complex polyploid crop. *PLoS Biol* 20: e3001802
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29: 24-26
- Roldan I, Wattedled F, Mercedes Lucas M, Delvalle D, Planchot V, Jimenez S, Perez R, Ball S, D'Hulst C, Merida A (2007) The phenotype of soluble starch synthase IV defective mutants of Arabidopsis thaliana suggests a novel function of elongation enzymes in the control of starch granule formation. *Plant J* 49: 492-504
- Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* 42: 38-48
- Saccomanno B, Berbezy P, Findlay K, Shoesmith J, Uauy C, Viallis B, Trafford K (2022) Characterization of wheat lacking B-type starch granules. *J Cereal Sci* 104: 103398
- Saccomanno B, Chambers AH, Hayes A, Mackay I, McWilliam SC, Trafford K (2017) Starch granule morphology in oat endosperm. *Journal of Cereal Science* 73: 46-54
- Sadali NM, Sowden RG, Ling Q, Jarvis RP (2019) Differentiation of chromoplasts and other plastids in plants. *Plant Cell Rep* 38: 803-818
- Saito M, Tanaka T, Sato K, Vrinten P, Nakamura T (2017) A single nucleotide polymorphism in the "Fra" gene results in fractured starch granules in barley. *Theor Appl Genet* 131: 353-364
- Salman H, Blazek J, Lopez-Rubio A, Gilbert EP, Hanley T, Copeland L (2009) Structure–function relationships in A and B granules from wheat starches of similar amylose content. *Carbohydrate Polymers* 75: 420-427
- Samodien E, Jewell JF, Loedolff B, Oberlander K, George GM, Zeeman SC, Damberger FF, van der Vyver C, Kossmann J, Lloyd JR (2018) Repression of Sex4 and Like Sex Four2 Orthologs in Potato Increases Tuber Starch Bound Phosphate With Concomitant Alterations in Starch Physical Properties. *Frontiers in Plant Science* 9
- Sato Y, Takehisa H, Kamatsuki K, Minami H, Namiki N, Ikawa H, Ohyanagi H, Sugimoto K, Antonio BA, Nagamura Y (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res* 41: D1206-1213
- Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180: 519-530
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9: 676-682
- Seung D (2020) Amylose in starch: towards an understanding of biosynthesis, structure and function. *New Phytol* 228: 1490-1504
- Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, Lu KJ, Zanella M, Zeeman SC (2017) Homologs of PROTEIN TARGETING TO STARCH Control Starch Granule Initiation in Arabidopsis Leaves. *Plant Cell* 29: 1657-1677

Seung D, Echevarria-Poza A, Steuernagel B, Smith AM (2020) Natural Polymorphisms in Arabidopsis Result in Wide Variation or Loss of the Amylose Component of Starch. *Plant Physiol* 182: 870-881

Seung D, Lu KJ, Stettler M, Streb S, Zeeman SC (2016) Degradation of Glucan Primers in the Absence of Starch Synthase 4 Disrupts Starch Granule Initiation in Arabidopsis. *J Biol Chem* 291: 20718-20728

Seung D, Schreier TB, Burgy L, Eicke S, Zeeman SC (2018) Two Plastidial Coiled-Coil Proteins Are Essential for Normal Starch Granule Initiation in Arabidopsis. *Plant Cell* 30: 1523-1542

Seung D, Smith AM (2019) Starch granule initiation and morphogenesis-progress in Arabidopsis and cereals. *J Exp Bot* 70: 771-784

Seung D, Soyk S, Coiro M, Maier BA, Eicke S, Zeeman SC (2015) PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in Arabidopsis. *PLoS Biol* 13: e1002080

Seung D, Thalmann M, Sparla F, Abou Hachem M, Lee SK, Issakidis-Bourguet E, Svensson B, Zeeman SC, Santelia D (2013) Arabidopsis thaliana AMY3 is a unique redox-regulated chloroplastic alpha-amylase. *J Biol Chem* 288: 33620-33633

Shang J, Li L, Zhao B, Liu M, Zheng X (2020) Comparative studies on physicochemical properties of total, A- and B-type starch from soft and hard wheat varieties. *Int J Biol Macromol* 154: 714-723

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1: 2856-2860

Shewry PR, Hey SJ (2015) The contribution of wheat to human diet and health. *Food Energy Secur* 4: 178-202

Singh A, Compart J, Al-Rawi SA, Mahto H, Ahmad AM, Fettke J (2022) LIKE EARLY STARVATION 1 alters the glucan structures at the starch granule surface and thereby influences the action of both starch-synthesizing and starch-degrading enzymes. *Plant J* 111: 819-835

Smirnov MN, Smirnov VN, Budowsky EI, Inge-Vechtomov SG, Serebrjakov NG (1967) Red pigment of adenine-deficient yeast *Saccharomyces cerevisiae*. *Biochemical and Biophysical Research Communications* 27: 299-304

Smith AM, Zeeman SC (2006) Quantification of starch in plant tissues. *Nat Protoc* 1: 1342-1345

Smith AM, Zeeman SC (2020) Starch: A Flexible, Adaptable Carbon Store Coupled to Plant Growth. *Annu Rev Plant Biol* 71: 217-245

Soh HN, Sissons MJ, Turner MA (2006) Effect of Starch Granule Size Distribution and Elevated Amylose Content on Durum Dough Rheology and Spaghetti Cooking Quality. *Cereal Chemistry Journal* 83: 513-519

Stitt M, Zeeman SC (2012) Starch turnover: pathways, regulation and role in growth. *Curr Opin Plant Biol* 15: 282-292

Streb S, Zeeman SC (2012) Starch metabolism in Arabidopsis. *Arabidopsis Book* 10: e0160

Suh DS, Verhoeven T, Denyer K, Jane J-I (2004) Characterization of Nubet and Franubet barley starches. *Carbohydrate Polymers* 56: 85-93

Sun X, Sun Z, Saleh ASM, Zhao K, Ge X, Shen H, Zhang Q, Yuan L, Yu X, Li W (2021) Understanding the granule, growth ring, blocklets, crystalline and molecular structure of normal and waxy wheat A- and B- starch granules. *Food Hydrocolloids* 121

Sun Y, Jiao G, Liu Z, Zhang X, Li J, Guo X, Du W, Du J, Francis F, Zhao Y *et al* (2017) Generation of High-Amylose Rice through CRISPR/Cas9-Mediated Targeted Mutagenesis of Starch Branching Enzymes. *Frontiers in Plant Science* 8

Szewcowska M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M *et al* (2013) Metabolic fluxes in an illuminated Arabidopsis rosette. *Plant Cell* 25: 694-714

Szydlowski N, Ragel P, Hennen-Bierwagen TA, Planchot V, Myers AM, Merida A, d'Hulst C, Wattedled F (2011) Integrated functions among multiple starch synthases determine both amylopectin chain length and branch linkage location in Arabidopsis leaf starch. *J Exp Bot* 62: 4547-4559

- Szydłowski N, Ragel P, Raynaud S, Lucas MM, Roldan I, Montero M, Munoz FJ, Ovecka M, Bahaji A, Planchot V *et al* (2009) Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases. *Plant Cell* 21: 2443-2457
- Takahata Y, Tanaka M, Otani M, Katayama K, Kitahara K, Nakayachi O, Nakayama H, Yoshinaga M (2010) Inhibition of the expression of the starch synthase II gene leads to lower pasting temperature in sweetpotato starch. *Plant Cell Rep* 29: 535-543
- Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* 32: 204-206
- Tetlow IJ, Beisel KG, Cameron S, Makhmoudova A, Liu F, Bresolin NS, Wait R, Morell MK, Emes MJ (2008) Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. *Plant Physiol* 146: 1878-1891
- Tetlow IJ, Bertoft E (2020) A Review of Starch Biosynthesis in Relation to the Building Block-Backbone Model. *Int J Mol Sci* 21
- Tetlow IJ, Emes MJ (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life* 66: 546-558
- Tetlow IJ, Emes MJ (2017) Starch Biosynthesis in the Developing Endosperms of Grasses and Cereals. *Agronomy* 7: 81
- Tetlow IJ, Wait R, Lu Z, Akkasaeng R, Bowsher CG, Esposito S, Kosar-Hashemi B, Morell MK, Emes MJ (2004) Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. *Plant Cell* 16: 694-708
- Toyosawa Y, Kawagoe Y, Matsushima R, Crofts N, Ogawa M, Fukuda M, Kumamaru T, Okazaki Y, Kusano M, Saito K *et al* (2016) Deficiency of Starch Synthase IIIa and IVb Alters Starch Granule Morphology from Polyhedral to Spherical in Rice Endosperm. *Plant Physiol* 170: 1255-1270
- Tuncel A, Kawaguchi J, Ihara Y, Matsusaka H, Nishi A, Nakamura T, Kuhara S, Hirakawa H, Nakamura Y, Cakir B *et al* (2014) The rice endosperm ADP-glucose pyrophosphorylase large subunit is essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. *Plant Cell Physiol* 55: 1169-1183
- Uauy C, Wulff BBH, Dubcovsky J (2017) Combining Traditional Mutagenesis with New High-Throughput Sequencing and Genome Editing to Reveal Hidden Variation in Polyploid Wheat. *Annu Rev Genet* 51: 435-454
- Vandromme C, Spriet C, Dauvillee D, Courseaux A, Putaux JL, Wychowski A, Krzewinski F, Facon M, D'Hulst C, Wattebled F (2019) PII1: a protein involved in starch initiation that determines granule number and size in Arabidopsis chloroplast. *New Phytol* 221: 356-370
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A *et al* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50: D439-D444
- Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D *et al* (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588: 277-283
- Wang W, Hostettler CE, Damberger FF, Kossmann J, Lloyd JR, Zeeman SC (2018) Modification of Cassava Root Starch Phosphorylation Enhances Starch Functional Properties. *Frontiers in Plant Science* 9
- Warren FJ, Gidley MJ, Flanagan BM (2016) Infrared spectroscopy as a tool to characterise starch ordered structure--a joint FTIR-ATR, NMR, XRD and DSC study. *Carbohydr Polym* 139: 35-42
- Watson-Lazowski A, Raven E, Feike D, Hill L, Barclay JE, Smith AM, Seung D (2022) Loss of PROTEIN TARGETING TO STARCH 2 has variable effects on starch synthesis across organs and species. *J Exp Bot* 73: 6367-6379
- Win J, Kamoun S (2004) pCB301-p19: A Binary Plasmid Vector to Enhance Transient Expression of Transgenes by Agroinfiltration. <http://www.KamounLab.net>
- Xiao J, Li C, Xu S, Xing L, Xu Y, Chong K (2015) JACALIN-LECTIN LIKE1 Regulates the Nuclear Accumulation of GLYCINE-RICH RNA-BINDING PROTEIN7, Influencing the RNA Processing of

- FLOWERING LOCUS C Antisense Transcripts and Flowering Time in Arabidopsis. *Plant Physiol* 169: 2102-2117
- Yu TS, Zeeman SC, Thorneycroft D, Fulton DC, Dunstan H, Lue WL, Hegemann B, Tung SY, Umemoto T, Chapple A *et al* (2005) alpha-Amylase is not required for breakdown of transitory starch in Arabidopsis leaves. *J Biol Chem* 280: 9773-9779
- Yun MS, Kawagoe Y (2010) Septum formation in amyloplasts produces compound granules in the rice endosperm and is regulated by plastid division proteins. *Plant Cell Physiol* 51: 1469-1479
- Zeeman SC, Kossmann J, Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* 61: 209-234
- Zeeman SC, Thorneycroft D, Schupp N, Chapple A, Weck M, Dunstan H, Haldimann P, Bechtold N, Smith AM, Smith SM (2004) Plastidial alpha-glucan phosphorylase is not required for starch degradation in Arabidopsis leaves but has a role in the tolerance of abiotic stress. *Plant Physiol* 135: 849-858
- Zeeman SC, Tiessen A, Pilling E, Kato KL, Donald AM, Smith AM (2002) Starch synthesis in Arabidopsis. Granule synthesis, composition, and structure. *Plant Physiol* 129: 516-529
- Zeng D, Liu T, Ma X, Wang B, Zheng Z, Zhang Y, Xie X, Yang B, Zhao Z, Zhu Q *et al* (2020) Quantitative regulation of Waxy expression by CRISPR/Cas9-based promoter and 5'UTR-intron editing improves grain quality in rice. *Plant Biotechnology Journal* 18: 2385-2387
- Zhang L, Li N, Zhang J, Zhao L, Qiu J, Wei C (2022) The CBM48 domain-containing protein FLO6 regulates starch synthesis by interacting with SSIVb and GBSS in rice. *Plant Mol Biol* 108: 343-361
- Zhang L, Ren Y, Lu B, Yang C, Feng Z, Liu Z, Chen J, Ma W, Wang Y, Yu X *et al* (2016) FLOURY ENDOSPERM7 encodes a regulator of starch synthesis and amyloplast development essential for peripheral endosperm development in rice. *J Exp Bot* 67: 633-647
- Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nature Protocols* 1: 641-646
- Zhang X, Myers AM, James MG (2005) Mutations affecting starch synthase III in Arabidopsis alter leaf starch structure and increase the rate of starch synthesis. *Plant Physiol* 138: 663-674
- Zhang Y, He P, Ma X, Yang Z, Pang C, Yu J, Wang G, Friml J, Xiao G (2019) Auxin-mediated statolith production for root gravitropism. *New Phytol* 224: 761-774
- Zhao X, Andersson M, Andersson R (2018) Resistant starch and other dietary fiber components in tubers from a high-amylose potato. *Food Chemistry* 251: 58-63
- Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang YG *et al* (2020) Triticum population sequencing provides insights into wheat adaptation. *Nat Genet* 52: 1412-1422
- Zhu F (2015) Isolation, Composition, Structure, Properties, Modifications, and Uses of Yam Starch. *Comprehensive Reviews in Food Science and Food Safety* 14: 357-386
- Zhu X, Li S, Pan S, Xin X, Gu Y (2018) CSI1, PATROL1, and exocyst complex cooperate in delivery of cellulose synthase complexes to the plasma membrane. *Proc Natl Acad Sci U S A* 115: E3578-E3587
- Ziegler GR, J.A. C, J. R (2005) Spherulitic crystallization in starch as a model for starch granule initiation. *Biomacromolecules* 6: 1547 - 1554
- Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7

Appendices

Appendix 1: List of Primers

A) Primers used for cloning, sequencing and Arabidopsis genotyping

Primer	Amplification	Sequence	Purpose
JC31	linkerHA f	GGCGGAGGTGGTTCTGGCGG	LinkerHA PCR fragment
JC32	linkerHA r	tcaagcgtaatctggaacatcatatgggtaACTTC CACCACCGCCTGAAC	LinkerHA PCR fragment
JC33	pK7- Xbal_pAtSS4_F	GGTACCCGGGGATCCTCTAGAACTTATG GGAATTCTAGATGTAATG	In-Fusion cloning
JC34	pK7- Xbal_pAtSS4_R	TCGTCGTCATGTTTAATTAGATACCTAAA CGAAG	In-Fusion cloning
JC35	pAtSS4_AtSS4_F	CTAATTAACATGACGACGAAGCTATCG AGCTTC	In-Fusion cloning
JC36	pAtSS4_AtSS4_R	CACCTCCGCCCGTGCATTAGGAACAGC TCTTG	In-Fusion cloning
JC37	AtSS4_Linkers_F	TAATCGCACGGGCGGAGGTGGTTCTGG C	In-Fusion cloning
JC38	AtSS4_Linkers_R	TTGAACGATCTCAAGCGTAATCTGGAAC ATC	In-Fusion cloning
JC39	HA_tNOS-pK7- HindIII_F	TTACGCTTGAGATCGTTCAAACATTTGGC AA	In-Fusion cloning
JC40	HA_tNOS-pK7- HindIII_R	ATCCAAGCTCAAGCTAAGCTTGATCTAG TAACATAGATGACACCGC	In-Fusion cloning
JC41	pK7- Xbal_pAtMRC_F	GGTACCCGGGGATCCTCTAGACAAATAC TACTGGAAGAACC	In-Fusion cloning
JC42	pK7- Xbal_pAtMRC_R	AGAAACCCATCCTGAGAGCATAATTATG CAATCGC	In-Fusion cloning
JC43	pAtMRC_AtMRC_ F	TGCTCTCAGGATGGGTTTCTCTCAGGCC A	In-Fusion cloning
JC44	pAtMRC_AtMRC_ R	CACCTCCGCCTAAAATTACACCTTCCATG CATGT	In-Fusion cloning
JC45	AtMRC_Linkers_F	TGTAATTTTAGGCGGAGGTGGTTCTGGC	In-Fusion cloning
JC46	AtMRC_Linkers_R	TTGAACGATCTCAAGCGTAATCTGGAAC ATC	In-Fusion cloning
JC47	pK7kanR	GCGGTTCTGTCTCAGTTCCAAA	sequencing
JC48	pAtSS4-AtSS4	AAAGACTGTGAAGCTCCTTTTG	sequencing
JC49	AtSS4-2	GGGTAGTAGTATTCATGGCAGC	sequencing
JC50	AtSS4-3	GGCTCGAAAAGTCTGATGCA	sequencing
JC51	AtSS4-4	TCTCAAACATTGTAACAACCGTG	sequencing
JC52	NOS-tag	ACCCGCCAATATATCCTGTCA	sequencing
JC53	pAtMRC-AtMRC	AGGATTACTGTGTTGGTTTTCGA	sequencing
JC54	AtMRC-2	TGGGTTTCTCTCAGGCCATT	sequencing
JC55	AtMRC-3	GAGAAGCCAATGAAACCATGGA	sequencing
JC56	AtMRC-4	TCAACGTGGTCATGGGAAGA	sequencing
JC57	AtMRC-5-R	CCTACCTTCTAAATTCGGCG	sequencing
JC58	AtSS4-Linker- repeat F	aggttcaGGCGGAGGTGGTTCTGGC	Q5 insertion correction of a deletion
JC59	AtSS4-Linker- repeat R	ccaccgccCGTGCATTAGGAACAGCTCTT G	Q5 insertion correction of a deletion
JC62	AtSS4- eGFP_IF_insert_f	ataCCCATATGATGTTCCAGATTAC	Q5 insertion
JC63	AtSS4- eGFP_IF_insert_r	TACTTGACAGCTCGTCC	Q5 insertion

JC64	AtMRC-backboneC_r	TCTAGAGGATCCCCGGGTACC	IF backbone amplification
JC67	AtSS4-cTP_cdEC_corr_f	CGAAAACCTGTATTTTCAGGGATGTAAG ATGCGTCAAC	Gibson assembly
JC20	AtSS4_rev	caggctgaaatcttctcTTAGGTCCGGTTTG GTAC	Gibson assembly
JC17	phase2vector_fwd	GAGAAGATTTTCAGCCTG	Gibson assembly
JC18	phase2vector_rev	CTGAAAATACAGGTTTTTCG	Gibson assembly
JC72	AtMRC-cdEC-seq-mid	AAGAAAAGCTGCGCGAGAC	sequencing
JC74	AtSS4-cdEC-seq-mid1	AGCGAGACTGATGAACGGAT	sequencing
JC75	AtSS4-cdEC-seq-mid2	TATCGCAGCAGAAATGGCAC	sequencing
JC84	AtMRC-CTP_cdEC_pOPIN E_f	AGGAGATATACCATGGTGAACCACAAAC AAAAAAGC	In-Fusion cloning
JC85	AtMRC-CTP_cdEC_pOPIN E_r	GTGATGGTGATGTTTCAGAATGACTCCC TCCATACAC	In-Fusion cloning
DSB174	TaMRC1 cd midseq	AGAAGGTGACCGAGGTTGAG	Sequencing/Genotyping
Z1901_T35S	T35S reverse	gggttcttatatgctcaacacatgag	Sequencing/Genotyping
DSA_pProEX_L	pProEXf	TTTTTGCGCCGACATCATAACG	for sequencing insert on pProEX
DSA_pProEX_R	pProExr	GCAGGCTCTAGATTCGAAAGCG	for sequencing insert on pProEX
DSB112	TaSS4cd Midseq 2	TTACTTCATCGAACCGCAGC	Sequencing/Genotyping
Z1385_eYFP_seq_rev	YFP reverse	ACA CGC TGA ACT TGT GGC CGT	Sequencing/Genotyping
DSB352	mrc-3 genotyping f	AGTCTCCACCACCATTGGAC	Genotyping
DSB353	mrc-3 genotyping r	TTCAGCATCTGAAGTCCGAAC	Genotyping
JC80	mrc-3_LP_3	TGCGTTCTCCAGGAGTTTCT	Genotyping
JC81	mrc-3_RP_3	GCTTGAGCGGAAATATCGG	Genotyping
SAIL LB1	SAIL LB1	GCCTTTTCAGAAATGGATAAATAGCCTT GCTTCC	Genotyping

B) KASP markers for genotyping the wheat mutants. All primer sequences are given 5' to 3'. The wild-type allele primers had the VIC/HEX tail (GAAGGTCGGAGTCAACGGATT) on the 5' ends, while the mutant allele primers had the FAM tail (GAAGGTGACCAAGTTCATGCT) on the 5' ends. The nucleotide(s) that discriminate the wild-type from the mutated base (for wild-type and mutant primers) or the homeologous SNPs (for common primers) are indicated in capital letters. These primers were designed by and this table was compiled by David Seung.

Gene	Line and mutation	Wild-type allele	Mutant allele	Common
TtMRC-A1	Kronos3272 (Q258*)	agcaacagttaggagctgC	agcaacagttaggagctgT	cctcgattcattgatctggcg
	Kronos598 (L289F)	ccatcctaaactctgcttctcaaG	ccatcctaaactctgcttctcaaA	gaagcaacagttaggGAGcT
	Kronos4681 (Q550*)	catcaggctcagatgctcgC	catcaggctcagatgctcgT	gcaagatcgccagtgagC
6B pseudogene	Kronos4305	ttgagaagcagagtttaggatgG	ttgagaagcagagtttaggatgA	acgtttgaagtcagtgataataccaA
	Kronos3078	aggattcagagctttctgatacaC	aggattcagagctttctgatacaT	tgaagctcaGcaatttcactgc

Appendix 2: gBlocks synthesised

Linker
HA-tag

Linker-eGFP-HA

GGCGGAGGTGGTTCTGGCGGTGGAGGTTCAAGGCGGTGGTGAAGTATGGTGAGCAAGGGCGAGGAGCT
GTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAG
CCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAAC
ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGC
TGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCAC
ATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGtaccatgatg
atgtccagattacgctga

Linker-RFP-HA

GGCGGAGGTGGTTCTGGCGGTGGAGGTTCAAGGCGGTGGTGAAGTATGGCCTCCTCCGAGGACGTCATC
AAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGTCCGTGAACGGCCACGAGTTCGAGATCGAGGGC
GAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCTG
CCCTTCGCTGGGACATCCTGTCCCCTCAGTTCAGTACGGCTCCAAGGCTACGTGAAGCACCCCGCCGAC
ATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACCTCGAGGACG
GCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG
CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAG
CGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAGCTGAAGGACGGCGG
CCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAG
ACCGACATCAAGCTGGACATCACCTCCACAACGAGGACTACACCATCGTGAACAGTACGAGCGCGCCG
AGGGCCGCACTCCACCGGCGCCTAACTaccatgatg
atgtccagattacgctga

AtSS4-cTP_cdEC

ATGTGTAAGATGCGTCAACAGCGGGGTTCTGACTCTTCCAAACGTCAGGAAGTTAAGAAGGGCTCCCCA
AACCGATATTATCAATAAATTCGGGCTTGCAATCCAATAACGACGAAGAAAGTGATTTAGAGAACGGTAGC
GCTGATTCAGTGCCTTCACTGAAGAGTGATGCCGAAAAAGGCTCGTCGATCCACGGTTCAATAGACATGAA
TCACGCTGACGAAAATTTAGAGAAAAAGACGACATTCAGACCACTGAAGTCAACCGTCGTAAATCAAAGA
CTGCGAAAAAGAAAGGAGAGTCTATTCATGCAACGATCGACATTGGCCATGATGATGGCAAGAACCTTGA
CAACATTACCGTACCAGAGGTCGCTAAAGCATTGAGTCTGAACAAGAGTGAAGGAGAACAATTTCTGAT
GGTCAATTCGGGGAGCTGATGACAATGATACGGTCTGCTGAAAAGAACAATACTTCGTCTTGATGAGGCC
GTGCGACAGCGTTGGACGACCTTAATAAAATTTCTGAGCGACAAGGAGGCGCTGCAAGGCGAGATCAATGT
TCTGGAAATGAAGCTGAGCGAGACTGATGAACGGATTAAAAAGTGCAGCCCAAGAAAAAGCTCACGTGGA
GTTGTTGGAGGAACAGCTTAAAAAGCTTCGTCATGAAATGATTTCTCAATTGAGTCAGACGGTTACGTCT
TAGCACTGTCAAAAAGAATTAGAAACACTTAAGCTGGAGAATCTTTCTGAGAAATGATATTGAAATGCTT
AAATCCGAGCTGGATTCTGTAAGGACACAGGCGAACGCGTTGTTGTCCTGGAGAAAGAGTGCTCGGGAT
TAGAGTCTTCAGTAAAGGACCTTGAATCGAAACTGTCCGTCAGCCAGGAAGACGTTTCTCAGTTGTCTACG
TTGAAGATCGAGTGTACGGATTTGTGGGCAAAAGTCGAAACGTTGCAACTGTTGCTGGACAGAGCTACAA
AACAGGCGGAACAAGCGGTCATCGTGTACAGCAAAATCAGGACCTGCGCAACAAAGTAGACAAAATCGA
GGAATCGCTGAAAGAAGCTAACGTATACAAGGAGTCGTCGAGAAAATCCAACAATACAATGAGTTAATG
CAACATAAGGTCACGCTTCTGGAGGAGCGGTTGGAGAAGTCAGACGCTGAGATCTTTTCTACGTCCAATT
ATACCAGGAATCTATAAAAAGAATTCCAAGAAACTCTTGAGTCTTTGAAAGAGGAGAGTAAAAAAAAGTCA
AGAGATGAGCCGGTCGATGACATGCCCTGGGATTACTGGTCAAGACTTCTTCTTACGGTTGACGGGTGGCT
GCTTGAGAAGAAGATCGCATCCAATGACGCGGATCTGCTGCGTGATATGGTTTTGGAAGAAAGACCGCCGT
ATTCACGATACGTACATTGACGTCAAAGATAAGAACGAGCGGGACGCGATATCCGCTTTTTTGAACCTTGT
GAGTTCTCTACCAGCAGCGGATTATATGTAGTACATATCGCAGCAGAAATGGCACCCGTTGCGAAAGTCG
GGGGCTTGGGGGACGTCGTTGCAGGACTGGGAAAAGCCCTTCAGCGGAAAGGACATCTTGTAGAGATCA
TTTTGCCAAAGTATGACTGTATGCAATATGACCGGGTCCGTGATCTGCGGGCATTGGATACGGTCTGTCGAG
TCATACTTCGACGGCAAATTTGACAAAAACAAGATATGGATAGGTACCGTGGAGGGATTGCCCGTTCACCT
TATAGAGCCACAGCATCCTAGTAAGTTTTTTTTGGCGGGCCAATTTTACGGAGAACAGGACGATTTTTCGTC
GGTTCAGTTACTTTAGCCGGGCTGCCTTGAATTGTTGCTTACGTCAGGAAAAAAACCCGACATAATCCACT
GCCATGATTGGCAAACGGCTTTCTGGCCCCCTTTACTGGGATCTTTATGCACCGAAGGGTCTTGATTTCG
CCCGATCTGTTTTACTTGTACATAATTTTGAATATCAGGGCACAGCCAGCGCATCAGAGTTGGGGTCTGT
GGCCTTGATGTCAACCAGTTGAATCGTCCGGATAGAATGCAGGACCACTCTCCGGTGACCGGGTAAACCC
GGTAAAGGGAGCTATTATCTTTTCAAATATCGTTACGACAGTCTCACCGACGTATGCACAAGAAGTTCGTA
CTGCGGAGGGGGTAAGGGACTTCATTCGACCTTGAATTTCCAATCGAAGAAGTTCATAGGAATTTTGAAC
GGTATTGACACCGACTCCTGGAACCCAGCCACGGATCCATTCTTGAAGGCCAGTTCAACGCAAAGGATCT
TCAGGGTAAAGAAGAGAATAAGCATGCTCTGCGGAAGCAACTTGGGCTTTCATCAGCCGAGAGTCGCCGC
CCCCTGGTAGGATGCATACCCGGTTAGTACCTCAAAAAGGCGTCCATCTGATACGGCACGCGATATATCG
TACTCTGGAATTAGGAGGACAATTTGTCCTTCTGGGAGTTGCCCCGTCCACATATCCAACGCGAGTTCGA
AGGGATAGAACAACAATTCAAGTCGCACGACCACGTCCGGCTTTTGTGAAATATGATGAGGCGCTTTCCC
ATACCATATACGACCCAGCGACCTGTTTATAATCCCATCTATTTTTGAGCCCTGCGGGTTGACGCAGATGA
TTGCGATGCGGTATGGTTCCATCCCTATTGCTCGAAAACCGGAGGCTTGAACGACAGCGTATTTGATATC
GATGATGACACGATTCCAACGCAGTTCCAAAACGGATTACATTCCAAACTGCAGATGAACAGGGCTTTAA
CTATGCACTGGAACGTGCATTTAACCAATTATAAGAAAAGACGAGGAGAAGTGGATGAGATTAGTTGAAAAG
GTGATGAGCATCGACTTCTCGTGGGGATCGTCGGCTACGCAGTACGAAGAGTTATACACACGCTCCGTCTC
TAGAGCCCGGGCTGTACCAAACCGGACCTAA

AtMRC-ctp_cdEC_att

att sites (not used in final cloning)

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGTGAACCACAAACAAAAAGCTTAGTTGCTTTTCCAAGCAT
AACCCGGCGCAAAAGACATCTTTTACTTAGTGTACAATCAGTGTGCATAATACACGCCAAATATAAACGACAATGG
GTCCGCTGAATCGGGCAACGTCCTTTTGTATAAACTTTTCGCACGCACCCATCGCTTAGAAAAGACAACTAACCAACAC
AGCGTGTATCCTGACGATGATGACCTTCCCTATTCCAACCTGGGTGTATTGGAGTCGGACCTTGAAGCCGCCCTTGTGC
CGCTGCTGAAACCGGAAGAAGACCTTCATGATGCAGAGAGAAAAGCTTCTTAGTGACAAAAATAAATTGAATAGAGCA
AAAGAAGAATTGGAGAAGCGGGAAAAGACGATAAGTGAAGCTAGTCTGAAGCACGAAAGTTTGCAAGAAGAGTTGA
AGCGCGCAAATGTGGAATTGGCATCACAGGCTCGGGAGATTGAAGAACTTAAACACAAGTTGCGTGAACGGGATGA
GGAACGGGCCGCACTGCAGTCTTCGCTGACTTTAAAGGAAGAGGAAGTTGAGAAGATGCGGCAAGAGATTGCAAATA
GAAGTAAAGAGGTATCAATGGCAATATCGGAATTCGAAAGCAAGTCTCAACTGTTGTGAAGGCGAACGAGGTGGTT
AAACGTCAGGAGGGAGAGATTTACGCACTTCAGCGTGCCCTGGAGGAGAAGGAAGAAGAGCTTGAGATATCAAAAG
CCACAAAAAAGTTAGAGCAAGAAAAGCTGCGCGAGACGGAAGCCAACCTTAAGAAACAGACTGAGGAGTGGTTAAT
AGCGCAGGACGAGGTCAATAAACTTAAAGAAGAGACCGTAAAGCGGCTTGGAGAAGCGAACGAGACTATGGAAGAT
TTCATGAAAGTGAAGAAGTTATTAACAGATGTTAGATTTGAATTAATATCATCTCGTGAGGCCCTGGTCTTCAGCCGTG
AGCAGATGGAAGAGAAGGAAGTTCTGCTTGAGAAACAAGTTGAAGAGTTAGAAGAACAACGGAAGTCTGTATTATCA
TATATGCAAAGTCTTAGAGATGCCACACGGAGGTGGAATCTGAACGCGTGAAGTTGCGCGTCTGGAAGCGAAGAA
TTTTGCACTTGAACGTGAGATCTCAGTGCAGAAGGAGTTGTTGGAGGATTTGCGCGAAGAGCTGCAGAAAAGAAAAAC
CACTGCTGGAGCTGGCTATGCACGATATATCTGTAATACAAGATGAGTTGTACAAAAAGGCGAACGCTTTCCAGGTCT
CGCAGAACCTGCTTCAAGAAAAGGAAAGTTCCTTGTGGAAGCTAAGCTTGAATCCAGCACCTTAAGAGCGAACAA
GCTAGTCTTGAAGTCTTTTACAGGAGAAGGACGAAGAATTAGCCGAAGCCCGAATAAATTAGGCGAGGTAAATCA
AGAAGTAACTGAACTTAAGGCTCTGATGATATCGCGCGAGGACCAGTTGATGGAAGCTACAGAAATGTTGAAGGAAA
AAGACGTTTCTTATAGAAATTGAAGGTGAGTTGGGAAGCTCTAAATTGAAAGTACAGAGGCAGAGATGGTGGTA
GAACGTATAGCCGAATTGACCAACCGTCTGCTTATGTCAACTACCAATGGTCAAACAGAACGCCATGAGAATCAAC
AATGAAATCTCTATTGATTCGATGCAGCAACCTTTGGAGAAGCCACACGACGATTACGGGATGGAAAATAAGAGATTG
GTGATGGAAGTCTTTTACCCGGGAGAACTTAAAGATGAAAGAGATGGAAGTGTGGCCGTCCAGCGCTCTTACT
TTCAAAGACGAAGAAATAAATGTTGTCATGGGCCGTTAGAAGCTAAAGAGCAAGAATTGAAGAACTGAAAGAGG
AGACTATCAATGATTCCGAAGACCTTAAAGTCCTTACGCGCTTGCCCAAGAGCGTGTGGGGGAAAAGACTATGGGTG
ACCTTGCATAGAAATGTTGCAGCTTGAAGCTGCTAATCTGGAAGTCGAGGCCGCAACTTCTGCCTTGCAGAAATTGG
CTAAGATGAGTACCGAGTTGTTAACCAAGCAGACATGTCTATAGAAGCAGATACTACGCATACGGTGTGCGGAAAC
GCGGATACAGCGAGGGTTCCAACGAGTGTCTTGGGGAAGTTAAACGGAGGTGGTGTGAGACTTTGGTCCCTGACAGA
AAAGCTGTTAGAGAATGCGGGCATAGTCGAGGCACCTCCACGTGTATGGAGGGAGTCATTCTGTAAAGACCCAGCTTT
CTTGTAACAAGTGGTCCCC

Appendix 3: Standard media components

Media	Ingredients
Laemmli buffer (10 ×)	1 M DTT, 30% glycerol, 20% SDS, 0.5 M Tris HCl (pH 6.8), 0.05% bromophenol blue
SOC	20 g/L tryptone, 5 g/L yeast extract, 580 mg/L NaCl, 186 mg/L KCl, 2.03 g/L MgCl ₂ (6H ₂ O), 2.56 g/L MgSo ₄ (7H ₂), 3.6 g/L glucose
LB liquid	10 g/L peptone from casein, 5 g/L yeast extract, 10 g/L NaCl
LB agar	10 g/L peptone from casein, 5 g/L yeast extract, 10 g/L NaCl, 11 g/L agar
TE buffer	1 M Tris (pH 8), 0.5 M EDTA (pH 8)
MS agar (3% sucrose, 0.8% agar)	4.41 g/L Murashige & Skoog Medium including vitamins, 30 g/L sucrose, 8 g/L agar
YPAD liquid	10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 0.02 g/L adenine
SD+ (synthetic dextrose minimal media)	6.9 g/L Yeast Nitrogen Base without Amino Acids (Formedium), 20 g/L glucose
SD+ agar	6.9 g/L Yeast Nitrogen Base without Amino Acids (Formedium), 20 g/L glucose, 20 g/L agar
TBS	20 mM Tris HCl pH 7.5, 150 mM NaCl
TBST	20 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) Tween 20

Appendix 4: *Aegilops speltoides* MRC gene model.

The *Aegilops speltoides* gene model for MRC-A1 was identified by Andy Chen. The genome of *Aegilops speltoides* isolate TS01 was searched using BLASTn, using the *TaMRC-A1* coding sequence as the query. A match was identified on Chromosome 6 in the region from 180007045 to 180009267, with 97.5% sequence identity. Sequences highlighted in yellow represent the two putative exons of MRC, red text represents the start and stop codons, blue text represents splice junctions.

```
>CM038146.1:180003045-180012267 Aegilops speltoides isolate TS01
chromosome 6, whole genome shotgun sequence
TAATGGCTCTCTCCACGAATGAAGAGGGGGTGGAGGGGTATAAAATAGCCTCCACACAAAATCCAACCGTTACA
CACATTTGAGCAATCTCGGTGGGACCAAAGTGGACAACACGGTGGAGACCGATAGTTCAAATAATGTGAACGTT
GGGATTTTTGGTGGGACCGAAAATACAATCTCGGTTAGTCCGATTCTGCAATGACAAGAGCATGCCAATTTG
GTGGGACCGACTTCATCATCTCGGTGACACCGATTTGATGATGTAGGTTCTGAAGCTTGACAACATCATGTTTG
GTGAGACCGAGTTCGTAATCGAAAATCCGAGTTTTCTAGGGTTTGGCTTATGGTGGGTTTTCATCTCGGTG
TGATTGAGTTTGGAAATATCGGTGGGACCGAGACTACTTATGAGTGTATATGGCGGTGAAAGTTCTTGATAGTT
TAGGAGATGTATCACTAAGCACTTGAGCAACAGATCATCATCATAACCTCATCCCTTTTAATAGTATTGGCTT
TTCTTATGGACTCAATGTGATCTTGGATCACTAAACCAAAAATGGAGTCTTGAGCTTTTGCCAATGCTTGCCT
TAGCATTTTGGGGGTCCACTTCTTGATCCTTGCCATACCCATATCAATGAACTCTCTGAAATGATTCTCAAGT
AACTATTAGTTC AATGATATATATGTTGTTATTAATTACCAAAACCACTGGGATTAGTTGCACTTTTCGGTTTG
TAAGTTGGTGCCTTCTAAACAAAGCGTCGTGGCGGGATCTACATGTGAAGCGAGTACATAGCTGCTTCGAAGC
AGCAAATGAAGGAGTGGGATGAAGGAGTTCATTTGATCTAGGTGTCATACCTAGTGCATCGGGACCAATGA
AGATCTTCGTGACAATCTTTGGTGAATTGCCTTGCCAAAAGAATCCAGATTTTACAAGAGGACCAAGCACAT
CAAGAGACGCTTCAATTCATTCCGGGACCAAGTCTAGTGGGAGACATAGAGATTTGCAAGATACATACGGAT
CTGAATGTTGCAGACCCGTTGACTAAGCCTCTCTCACGAGCAAAACATGATTAGCACCAAGACTCCATGGGTG
TTAGAATCATTACTATGTAATCTAGATTATTGACTCTAGTGCAAGTGGGAGACTGAAGGAAATATGCCCTAGA
GGCAATAATAAAGTTATTATTTATTTCTCATATCATGATAAATGTTTATTATTCATGCTAGAATTGTATTAA
CCGAAAACATGATAGATGTGTGAATACGTAGACAAACATATAGTCACTAGTATGCCTCTACTTGACTAGCTCA
TTAATCAAAGATGGTTATGTTTCCCTAACCATAGACATGTGTTGTCATTTGATTAATGGGATCACATCATTATG
AGAATGATGTGATTGACATGACCCATTCCGTTAGCCTAGCACTTGATCGTTTAGTATATTGCTATTGCTTTCT
TCATGACTTATACAAAGTTCCGCAACTATGAGATTGTGCAACTCCGTTTACGGAAGAACACTTTGTGTGCTAC
CGGACTCACAAACGAATCGGGTGATTATAAAGGTGCTCTCTGGTGTTTAGAAGGTACATGTTGGGTTGGCATAA
TTTCGAGATTAGGTTTTGTCACTCCGATTATCGGAGAGGTATCTCTGGGCCCTCTTGGTAATACTCATCACCTA
AGCCTTGCAAGCATTGTAACATAATGAGTTAGTTATAAGATGATGTATTACGGAACGAGTAAAGAGACTTTCCG
GGAACGAGATTGAACTAGGTATTGGATACCGACGATCGAATCTCGGGCAAGTAACATACCGATGACAAAGGGA
ACAACGTATGTTGTTATGCGGTTTGACCGATAAAGATCTTCGTAGAAATATGTACGAACCAATATGGGCATCCA
GGTCCCGCTATTGGTTATTGACCGAGAATGGTTCCTAGGTCATGTCTACATAGTTCTCGAACCCGTAGGGTCCG
CACACTTAACGTTACGATGACAGTTTTATTATGAGTTTACAAGTTTTGATGTACCTAAGTTTGTTCGGAGTCC
CAGATGTGATCACGGACATGACGAGGAGTCTCGAAATGGTGCAGACATAAAGATTGATATATTGGACGACTAT
ATTCCGGACACCTAAGTGTTAGGGTGATTTTGGAGAAAATCGGAGTTGGGAGGGTTACCGAACCCCGGAGAGTA
TTGGGCCTTATGGGCCTTAGGGAAAGGAGAGAGGGCGGCCAAGATGGGCCGCGCGCTCCCTCCCGGTCCGA
ATTGACTAGAGGGGGTGGCGCTTCTTTCCCTTCCCCCTTCCCTCCTAGTAGGAGTAGGAAAGGAGGAGTCCTA
CTCCTACTAGGAGGAGGATTCTCCTCCTTGGCACGCCAAGGGCCGGTGGCTCCCTTTGCTCCTTTATATC
TCGTGGGCGGGGGGCATAGACACACAAGTTGATCTACGGATCGTTCCTTAGCCGTGTGCGGTGCCCTCCACCA
TATTCCACCTCGGTTCATATTGTGCGGAGTTTAGGCGAAGCCTGCGTCGGTGAACATCATCATCGTCACCAC
GCCGTCGTCTTGTGGAACATCATCTCCGGAGCTTTCTTTGGATCGGAGCCGAGATCGCCATCGAAATAACGTG
TCTTTGAACTCGGAGGCCCGACGTTCCGGTGTGTTGGATCGGTGGATCGTGAAGGACGACACTACATCAACCGC
GTTGTGTTAACGCTTTCGTTACGGTTACAGAGTGCCTGGACGAACACTCTCCCTCTCGTTGCTATGCCATCA
CCATGATCTTGGTGTGCGTAGGAAATTTTTTGAATTAACACTTCCCAACACCGACATAGAGGAGTCCGGC
CACCTCCATCCCAAGGATGCAATTTTTTTTTTTTGGGGTGAAGGATGCAGTTCTTCGATAAGAAGTGTTTTCC
AATAGCATATGGATGTTGACCTTTTATCCTAGTATTCCATTTTCGAGTCAAAATAGTCCCTAACATTATAACAA
TTTTCATGTGTTACAAACTCCTAACAAATTTTCATGTGTTATATAAACTCATTTCTTTAAGGCAAACAAAGTG
TACCTTAAGTTTCATGTGTTATACTCATAAAATTTGGAACAGAAAAGAAAAAGGTAGTAAGACGAGTGAACGGG
GAAGAAAAGCAGTAGAAGGCAAACGACGCAGCTCTCTCTCACGCTTCTCCCGTGGTGCAGCTTGCAGTCCACA
CGCGGGCACGCGGCTGGGCGCGCCGGTTCCACCACCTCATCTCCGCACTCCCTCTGCCTCGTATCTCGTCCG
CTTCTCCACACCCCGCAGGAGCATTGCCAGCCGTCCGATCGCGCCCGGGCGGCGGTGGTTCCCTCTCCCAT
GTTCCGCGGCCATGCGCCTCTCCATAGGCTCCCCATCCCCGTGCGCGCCGGCGGCGGTGGCCGCGCTCTCCG
```

CAGCACATCCCCGTCGTGCCGTACCGCCAGTCATGTGAGCGCCCGCTGATCTTTTTCTTCCTTTTCTCATATCG
CTGTTTTCGTGGTACCACGCTGCTCACTGTTACATGGACTGCTCGCGTTCGTGTTTTCCCGATTCCCGTGCCCGT
CCACACGTGTTTGAAGTAGAAGGATACTAGATTTGGTGTCTAATTCATGTTCTGCTAGTACTAGTACTTTTT
TAAAAAACTTTTCTGGAATTGGTTCGATTGTGATAAATTCAGTAACTGCACCTGGCTGAACAAATCTTGA
TTGGAGAACGGCCTATGAACTCAAAAAATTATACTGAACAGATGAAATGTTTATGCAGAGGTATGCTTGAGA
TCAAATTTTCATCGGTTATGATACTTCACCTTATATGACAGTGAATTTCTGAAGTTCAGTGTACTGTCTTTCAG
TTCGTCGATTTACAACAATTTTTTACGTGCTTAGTTTTGAGGAAAGGATATTCCTCAGATTGCTTCACTAGGTT
GTGACCATTTTCCTTATCCTAATATCCTACTTATGCATTGTTTCTGCAACTCTCTCAGTTATGTTTCAGGCAG
AAGCTGAGTTTTATGGTGGCATTTCAGACTCAGCATCTGAAATATGCTCCTCGCTTGATCAAAATCAGCCGTAA
AAGGTATTAGATCAAAATACAACCTGATGGTGATAATGGAACGACTGAGCCAGCTAGAGAGTTGCTGGAGCGGCT
ATTTGCGAAGACACAAAAGTTTAGACACTGGTGTCTTCATGATAGTGAACCTGAGCGTGAGCATTGAGGTCCTG
AAGTCTGAATTCGAGGGTGCCTTGTCTATCCTCAGAAAACAAAGAGAGGGATCTTCGCAGCGCAGAGAAGAGGG
TTTCCGATGATCGGATAAGATTGAGCAAGACGAAGCAGGATCTTGATCAGAGAGAGGAAGCGATCCGCAAAGC
TTATGTAAGGCAACAAGGAATAGAGAAAAGCACTGAAAAAGGCAAGTAGAGATCTGGCGTTGCGAGTGAAGCAG
ATCAGTAATCTGAAGCTTTTGGTTGAGGGGCAAGACAGGACTATTGCCAGTTTACAAGCTTTGCTTTTCTCAGA
AGGTAACCTGAAGTGGAAAATCTCAAACGAGATATGTTCAAGAAGAACGAGGAAGCAGACCTGATGCGTTTCTCAGA
GATCAGGTCCAAAGAACAGTTGGTTCTTACAGCTAATCAAGCTATTGCGCAGCAAGAAGCAACAGTTAGGGAG
CTGCAAAGTGAAATTAAGAAAGACGATCGATATCGCCAGATCAAAATGAATCGAGGAAAACCTAATGAAGAGA
AACTGAAAGTTGCTGAACAGGAACCTTGAGAAGCAGAGTTTAGGATGGTTAGCAGCACAAACAAGAGTTAAAGGA
ACTTGCACAACTGGCATTCAAAGATACAGATGATATCAATGGTATTATCACTGACTTCAAACGTGTGAGGTCT
CTGCTAGATGCTGTACGCTCTGAATTAATCTCTTCAAAGATGCTTTCGCTTCTCTCGCAGACAAATAGAAG
ATCAAGCGGTTTCAGTTGCAGGAACAAGTACAGGAACCTCGAGGACCAAAGGTTATTACTGATGTCTTACACCCA
TGATTTGGAGGCTGCTAAACTGGAGATTCAAGGGAAGACACAGGAGCTCAGTTACGCACAGTCTCGTTGTCAT
GAACCTGAATCACAGTTACTTCAGGAAAGGGAGAAGGTTCGAGTCTCTAGAAGCCGAATTAGCCAAAGAAAAAC
AGAGCTTAGAACATAGAACTGAAGAAGTAGGCTTTCTCAGAAGGAGCTTGTTCAGAAAAGAAAATGAGTGCAC
CAAATCACAAGAACCTTGTTAAAGTAAAAGAGTTTGGAGCTGTTAGAAGCCAGACAGGAAGTCCAAGATATGAAG
TTAAAGGTAGAGTCTATTCAATTGGCTGTTCAAGAAAAGGATTCAGAGCTTTCTGATACACAGAGCAGACTAA
CTGAAGTCAGCAGTGAATTTGCTGAGCTTCAGCAGTTGCTAAATAGCAAGAAGGATCAACTGCTTTCAGGCTAG
AACTGAATTAGATGATAAAGAGCAACATATAGAAAACACTGGAGAGTGAAGTTGGATAGCATAACGGCTCAGATGC
TCGCAAGCTGAATCCATGGTTCAAAGGATGGCTGATCTCACTGGCGATCTTGTAGTTCCGTAAAAGCCGGAG
AAATGGACATCTATACATTACTGGATGATGAAATTTCAAGCACTGGTACAGCCCTCGAGTCCAATTTGCATAA
GCATAATCAACTGGAGGCTGACATAGAGATGTTAAGAGAATGCTTGCGGCATAAGGACATGGAGTTGAGAGCT
GCTCATGAAGCACTTGATGCCAAAGATCACGAGCTGAAGGCAGTACTTAGAAAAGTGGGATGTGAAGGAGCGGG
AAGTACGTGAGTTAGAAGAGTTACCGGATCCAGTGCCACAAATGAACTTGCTGGTTTTTCCAGTGAGACAAC
AGAGGACGGCATTGTAGGAGAGATGGAGCTCCAGAGCTTCAAATGAAAGCTGTGGAGGTCGAAGCACTTGCT
GCTACGACTGCATTGAGGAAGCTAGCGGATATGACTAAGGATTTCTTCAAACACGGCAAAGCTGATTCTGGTA
TTGACTTGGTTGCATCAGAGAGTCAGAAAATCAGTAAATGTGATCCTAAAAATGGAAGTACACAAGAAGACGGA
TGTGATTCTTGAAGCTGAAAAAGAAAATAGTTAGGCTCTTCTCATTGACAAAAACAGATTGTCACTGATGACATA
ATAAACGATATTGAGGAAATGATAGCTTCAAACCTGAAGCATGTAGTCTTCCAATTTCTATCAAGATAGCTTCTCAGA
GTAGAGATATACCAGATTGATCTTTTGAACATTTATGGACAGTGTATGTCGCCCCAGAAGGATGAGATCTTCTCT
GGTTGATATCACAACCTGCCATTTTGAAGAAAGGGTAACATGTTGAGCAGAAGCTGGTCATCTGATCCTTTGTTT
TCCTTTTGTAAATGTACCTCAAACCTATTCCTCAGATCTTTGTTCAATGTGTTCTCCTAAATATACATGGGAAG
TTATTGATATGGCAATTGGTGCTGGTTTTTCGTTCATGCCAAAACTACTACTGAGATTGTGGTGAATGTTTCCAT
GTGCAGCAAAGTTGGTGCAGCGATGTTGCAGCGTGTGTTTATTTTTACTGTTTTCATTGCAACTTCAAATAGTT
TCAAATCCCAAAGCTACTAAAGTAGATGGCTATCAAGTACTCTCTCTTTTTTCTTCAAAGAAAGGTTAAATACT
CTTGTCTTTTCCACTTGCTTTTGAAGAGCTCCGTAGTTCAAATTTACACAGGATGCTTGGAGCAATTTACACAGGA
TGCTTGGAGCATCAAACCTGTTAGAAAACAAATGGGCCTAGTCCATGTACAATTTCTGAAATCTCAAATAGAGACCC
ATAATTAAGGGATAATTAGATTTATGCCCTAGTTGTGTCTCACTCAGCTGTTTTTACCCTAATTTCCAAGA
GCCACCGCTCTCTCCAAGTCACTTCGCTCCTTATACTTTTGCATTTGACCGTTGATCTTCAAGTTTGA
AACCTCATAACTAATTCATACTAAATCAAAAAAATGCAAATAAGATTTCAAATGTTTCGGAAGAAACATCACCT
ATGTGTCAGTGTCAATTTGCATTCATGAAAAAGTGTGGAAAGTGCACATCTGAGTTTTAGCTCTTATGATAC
CACCATGAATAGTAAAATGTAGAAAAAATAATCAAAAAATCAAACAAATTTGGTGGCAAAGAATGACAA
ATGTTGTAAGTGTCTTCCAAGTTTTATCAGGGAATGGCATCCGTGGATGTCGTCGCAAAAAAATCAGCACT
CCAAAATAGATTTTTTTTTTGGCCACGACTTTCACGAATGTCGTTCCCTCATGAACTTGGCAAGCACTTAAAC
ATTTGTCATTTCTTACCACCAATTTTTTGAATTTTTTATTTTTTTTAAATATACTATTTCATGGTGGTAT
CATAAGAGCTAAAACCTCGGATGAGCACTTTCCAACACTTTTTTCATGAGTGCAAATGACAATGGCATATAGGTG
ATGTTTTTCCGAACATTTTGTATCTTATTTGCATTTTTTATTTAGTATGAATTAGTTATGAATTTTTTCAA
CTGATGGTCAAAGGTCATAGGGCAAAAGCATAAGAGGAGCAAAGTGAATGGATAGAACCAGGTTGACTTTTTGGA
AATTATGGGTAAACAACCGAGTGGGTACAACCAGGGGCATAAATCTAATTTATCCATAATTAATGGCAAGT
GGTGGTGTAAAGTTTAGAGGAAGTGTTCACCGCTTTGTATAGTGGGTTCTCTTTTTTTTTTCGAGAACTTCC

AATCTATTCATTTTCAATCATGGCAGTACAACGAACACCAGAAAATAAAAAATAATTACATCCAGATTCGTAGAC
CACCTAATGACTACTACAAGCACTGAAGCGAGCCGAAGGCGCGCCACCATCATCGCCCCCTCCCTTGTCGGAGC
CGGACACAACCTTGTTGTAGTAGACAGTCGGGAAGTCGTGATAAGATCGAAAGGATCAACACACCAGAACA
GCAACCGCCCGGATGAAGACAACTTAGATCGAAAGGATCCAACCTCCACACTAAGTGATGTGTTGAGAAGA
GAAATAGAAGACCACACATGCGCTTGCCTTGCCTCGCCTGGCCGTGCACAACCTTCCTTTTGTAGTTTAATTTT
TTGATGTCTTGGCAGACAAGTTAATTATTTCTTGTGTGGTAAGTATATGACTTAGAAAACCAAGTTGGTTTGAG
ATCGTGATCACGACATGATAACCGCTTCTGGTCCGTTACCAGCCGCGGCAAAAAGACACAACGAAAACATCTAGG
GTTTTGTACATCTTGCAACTTACACTGCCACAGTAGTCTATTCATCCCGAACGCCAACGTGCATCGGCGTG
CGCGAGAGGAGGTCTCTGGAAGCGTTTCGTCTTGCATTTTGCACCGGGAGAGGACGAATTAGGTTTTTGGGA
AGTGCTCTGCGTGACTGTCCACGTTCTTCATCACGAGTTGTCTTCTGTCCAAGTCGGGCAGCACTACTCATTG
TCATCTACAACAACGTTAGCAAAAAGATTGTGTCGAACATCATCATCAACAACATCGCTCCTGCAGTAGCTAAC
GAACAGTACATCCATTGTAATCTGTTTCATGTCTCTATTTGTAGTTATTGTTACATATGTGATGATGCCGTGCA
TGTTATCTGGTTTTGTCTAGTATGCTAGATTATTGCATGCTATTTTTTTGTAGTACCATTTATGAATTATTTACT
GGAATTAATATTGGATTTGCCTAATATTTCCAACACAAAACGTGTACAAGCTAAGAACTTGAGGCTAGAGTTTA
ACTTTGAATCATCGAACCAACTATTTAAATCCATGGCAGACTAAGGCATAGCCTAGTGGTGGGAAGGGGCTGA
TGCCTTCCCCCACCAGATTCAAGGCATGGTACTTGCAAAATTGGGTTTTGTTGCACCAATTATACTGTAGGGG
GTTCTCTTACAGTCTTTTTCAGTCAAAAAAAAAAATATATTTAAATCCATGCCGAAGAACTCATGGTCTTGATCA
TTCTAGCACGCATATAAGTGGTTAACAAAAAAGCACGTCGTCATCTCTTCGGAGACTATTGATATCGATAAAT
GTGACTCAAAAATGGAAGTACACAAGAACATGGTGATCAGACATTTGTAGCCCCGTGTTGTACCTCGACAAG
CGACAGTGACAACAGGGTTGATCAA

Appendix 5: Calibration of HMW standards on Superose 6 Increase 10/300

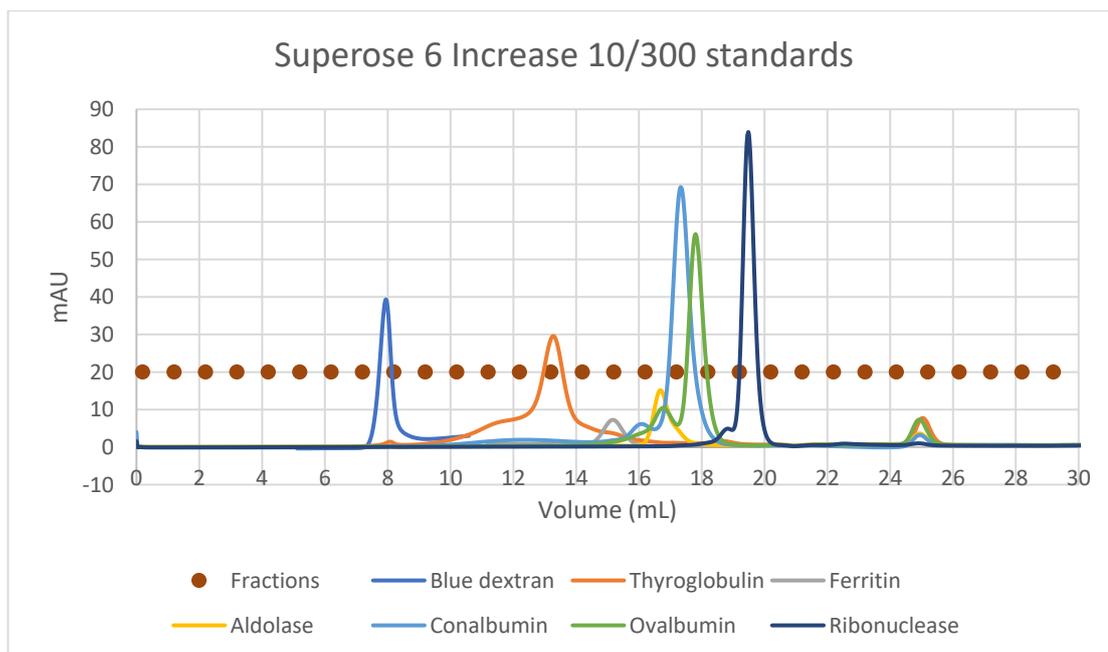


Figure 1. Elution curves of high molecular weight (HMW) standards (GE Healthcare Life Sciences). These were run on the Superose 6 Increase 10/300 column, in running buffer (100 mM Tris-HCl; 5% v/v glycerol; 150 mM KCl; 5 mM DTT). The spaces between the dots indicate individual 1 mL fractions.

Table 1. HMW standards properties on Superose 6 10/300

Standard	Elution peak (mL)	K_{av} (partition coefficient)	M_r (kDa)	Log M_r
Blue dextran	7.94			
Thyroglobulin	13.26	0.331257783	669	2.825426
Ferritin	15.16	0.449564134	440	2.643453
Aldolase	16.67	0.54358655	158	2.198657
Conalbumin	17.33	0.584682441	75	1.875061
Ovalbumin	17.8	0.613947696	44	1.643453
Ribonuclease A	19.49	0.719178082	13.7	1.136721

Figure 2. Superose 6 10/300 calibration curve. Thyroglobulin falls outside the linear range of this curve.

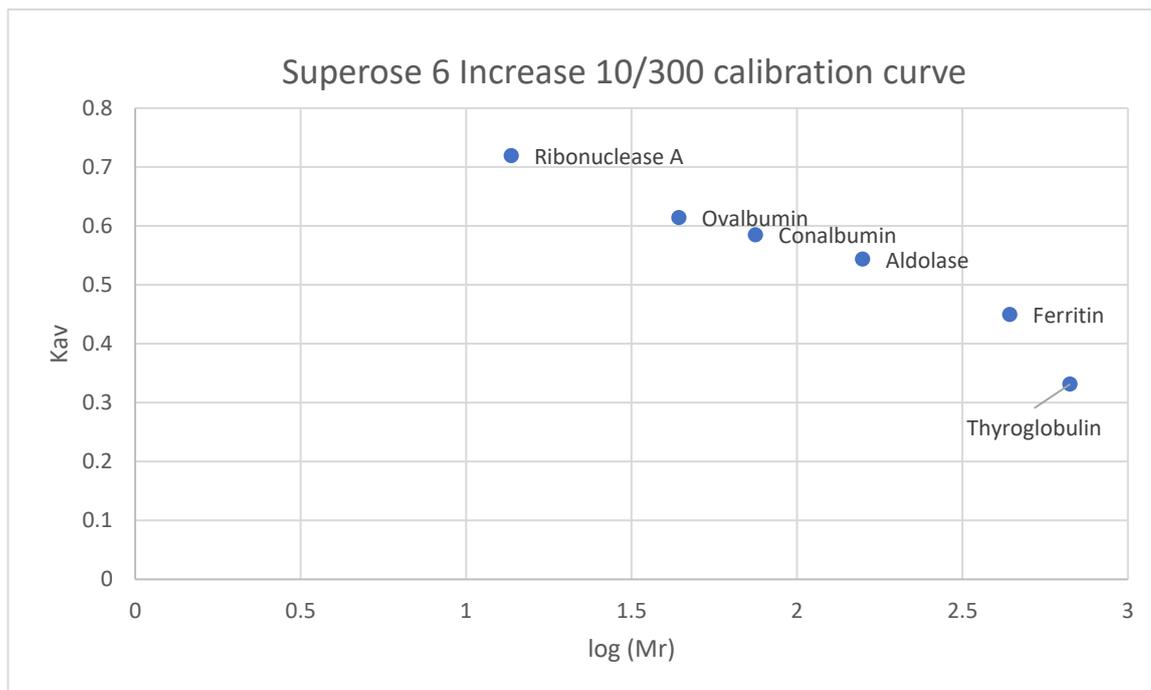


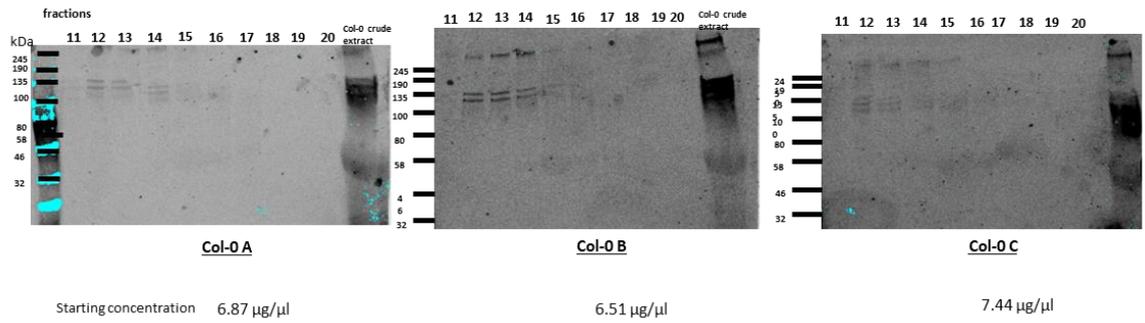
Table 2. Peak elution volumes of HMW standards and their corresponding fractions on Superose 6 10/300. The corresponding molecular weight is calculated using the K_{av} from table 1, but this is likely inaccurate at higher molecular weights than ferritin.

Standard	Fraction	Volume (ml)	Calculated molecular weight (kDa)
	1	0.2 - 1.2	
	2	1.2 - 2.2	
	3	2.2 - 3.2	
	4	3.2 - 4.2	
	5	4.2 - 5.2	
	6	5.2 - 6.2	
	7	6.2 - 7.2	
blue dextran	8	7.2 - 8.2	> 150 MDa
	9	8.2 - 9.2	65 MDa - 150MDa
	10	9.2 - 10.2	29 MDa - 65 MDa
	11	10.2 - 11.2	12 MDa - 29 MDa
	12	11.2 - 12.2	5MDa - 12 MDa
thyroglobulin	13	12.2 - 13.2	2 MDa- 5 MDa
thyroglobulin	14	13.2 - 14.2	1 MDa - 2 MDa
ferritin	15	14.2 - 15.2	450 kDa - 1 MDa
ferritin	16	15.2 - 16.2	196 kDa - 450 kDa
aldolase	17	16.2 - 17.2	85 - 196 kDa
conalbumin, ovalbumin	18	17.2 - 18.2	37 - 85 kDa
	19	18.2 - 19.2	16 - 37 kDa
ribonuclease A	20	19.2 - 20.2	7 - 16 kDa
	21	20.2 - 21.2	
	22	21.2 - 22.2	
	23	22.2 - 23.2	
	24	23.2 - 24.2	

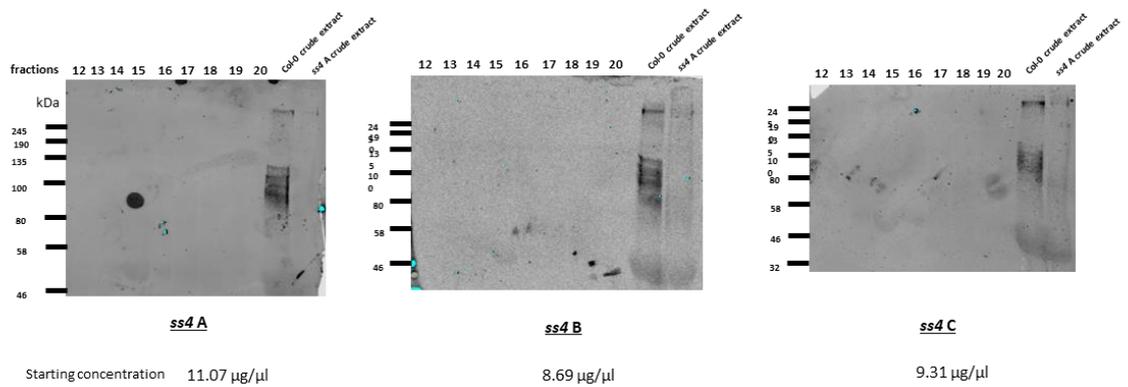
Appendix 6: Arabidopsis SEC immunoblot replicates

These immunoblots are biological replicates of the ones shown in Figure 5.1A, all blotted with an anti-AtSS4 antibody.

Col-0 fractions 11 - 20



ss4 fractions 12 - 20



mrc fractions 12 - 20

