

# Analysis of aspects of the starch metabolic pathway in lower plants

by  
Ingrid Jacobs



UNIVERSITEIT  
STELLENBOSCH  
UNIVERSITY  
*Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the  
Faculty AgriSciences at Stellenbosch University*

*The financial assistance of the National Research Foundation (NRF) towards this research is hereby  
acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not  
necessarily to be attributed to the NRF.*

Supervisor: Prof James R Lloyd

December 2018

# Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2018

Copyright © 2018 Stellenbosch University

All rights reserved

# Abstract

Starch is an important plant storage polysaccharide that has been demonstrated to have a major influence on plant growth. Transitory starch is synthesized in the leaves of plants during the day as a product of photosynthesis and degraded at night to allow continued carbon allocation for growth and cellular processes. It is also produced and stored for longer periods of time in non-photosynthetic organs such as stems, tubers and seeds. The study of starch is important for several reasons – not only is it a vital part of the human diet, it is also utilised in many non-food applications such as the paper, textile, oil and pharmaceutical industries. The pathway of starch metabolism in higher plants has been studied for decades and of late, *Arabidopsis* has become the workhorse plant for many starch researchers due to the plethora of insertion mutants that are readily available. However, the predominant use of *Arabidopsis* as a model system has led to a narrow understanding of starch metabolism restricted to that of synthesis and degradation of leaf starch. There are ongoing attempts to translate the knowledge gleaned from *Arabidopsis* into studies on the storage organs of crop plants (e.g. rice and maize endosperm, potato tubers), as well as starch metabolism in lower plants (e.g. algae and mosses) to aid in elucidating the evolutionary development of starch metabolism in land plants.

This study investigated two aspects of starch metabolism in lower plants to determine whether the pathway of starch metabolism observed in higher plants is conserved. Firstly, a previously uncharacterised starch synthase from the red alga *Chondrus crispus* was examined due to the reported differences in substrate preference between red algal and green plant starch synthases, and the deviation in compartmentalisation of starch synthesis and storage in members of the red algae. The *C. crispus* starch synthase was analysed by means of multiple sequence alignment, site-directed mutagenesis and recombinant protein expression and purification. Features unique to red algal starch synthases were identified, including a C-terminal glycogen binding domain and sequence variations in important residues involved with substrate binding. During recombinant expression, the *C. crispus* protein was insoluble and accumulated in inclusion bodies. Attempts to recover active protein through optimisation of expression, the use of alternative expression systems and protein refolding were unsuccessful and biochemical characterisation of the starch synthase could not be performed.

Secondly, four putative orthologues of the *Arabidopsis* maltose excess (MEX) transporter were identified in the moss *Physcomitrella patens* and their functions examined through the generation of knockout mutant lines and complementation of *Escherichia coli* mutants defective for sugar transporters. Knockout mutants were successfully generated for the *P. patens MEX1a* gene, while complementation studies failed to produce active protein. Expression profiling in wild type *P. patens* suggest that the four *PpMEX* genes are differentially expressed depending on the developmental stage of the culture and may have specialised functions in various growth structures.

# Opsomming

Stysel is 'n belangrike polisakkaried wat gestoor word in plante waar dit dien as 'n noemenswaardige bron van koolstof vir plantegroei. Gedurende die dag word kortstondige stysel geproduseer in die blare van plante deur die proses van fotosintese en word dan deur die nag afgebreek om voortdurend koolstof te voorsien vir plantegroei en ander selprosesse. Dit word ook geproduseer en gestoor vir langer tydperke in plantorgane wat nie fotosintese ondergaan nie, soos stamme, knolle en sade. Stysel word bestudeer vir verskeie redes – nie net vorm dit 'n belangrike komponent van die menslike diëet nie, maar word ook gebruik in die papier-, tekstiele-, olie- en artsenykunde-industrië. Die roete van styselmetabolisme word al vir dekades ondersoek in hoër plante. Arabidopsis het onlangs die gekose sisteem geword van navorsers as gevolg van die oorfloed van mutante lyne wat vrylik beskikbaar is. Die oorheersende gebruik van Arabidopsis het egter 'n noue begrip van styselmetabolisme tot gevolg gehad wat beperk is tot dié van stysel in blare. Daar word voortdurend pogings aangewend om die kennis wat deur studies op Arabidopsis verkry is, toe te pas op die stoororgane van gewasse (bv. die kiemwit van rys en koring, en ook aartappelknolle), sowel as styselmetabolisme in laer plante (bv. alge en mos) om die evolusionêre ontwikkeling daarvan in landplante te ontrafel.

Hierdie studie het twee aspekte van styselmetabolisme in laer plante ondersoek om te bepaal of die pad van styselmetabolisme wat in hoër plante waargeneem is, bewaar word. Eerstens was 'n voorheen onbeskryfde stysel-sintase van die rooi alg *Chondrus crispus* bestudeer omdat daar vantevore berig is oor verskille in rooi alge en hoër plante met betrekking tot substraatvoorkeur en ligging van styselproduksie en -stoor in die sel. Die *C. crispus* stysel-sintase was ondersoek deur middel van 'multiple sequence alignment', 'site-directed mutagenesis' en rekombinante proteïenuitdrukking en –suiwering. Kenmerke wat uniek is tot die rooi alg sintase, soos 'n C-terminaal glikogeen-bindingsgebied en verskille in aminosure wat 'n rol speel in substraatbinding, was geïdentifiseer. Tydens rekombinant proteïenuitdrukking was die *C. crispus* proteïen onoplosbaar en het geaggregeer. Pogings om aktiewe proteïen te herkry deur optimalisering van uitdrukking, die gebruik van alternatiewe uitdrukkingstelsels en proteïen-hervouing was onsuksesvol en biochemiese karakterisering van die proteïen kon dus nie uitgevoer word nie.

Tweedens was vier putatiewe ortoloë van die Arabidopsis 'maltose excess (MEX)' vervoerder geïdentifiseer in die mos *Physcomitrella patens*. Die funksies van hierdie vier gene was ondersoek deur mutante mos lyne te genereer en *Escherichia coli* mutante, wat gebrekkig is vir sekere suikervervoerderproteïene, te komplementeer. Die generasie van 'n mutante lyn vir die *MEX1a* geen was suksesvol, terwyl die komplementering van *E. coli* misluk het as gevolg van 'n gebrek aan die uitdrukking van aktiewe rekombinante proteïen. Die uitdrukkingprofiel van die *PpMEX* gene in wildetipe *P. patens* dui aan dat die gene moontlik differensieel uitgedruk word en afhangend is van die ontwikkelingsfase van die mos. Dus het die vier gene moontlik gespesialiseerde funksies in verskillende mosstrukture.

# Acknowledgements

I would like to extend my gratitude to the following institutions and individuals, without whom this study would never have been a possibility, much less a reality.

To the National Research Foundation for funding my studies and providing me with financial access to achieve my degree.

To Prof Jens Kossman, for allowing me the opportunity to pursue my studies at the Institute for Plant Biotechnology.

To Prof Emile van Zyl and Dr Rose Cripwell from the Microbiology Department at Stellenbosch University for their kindness and assistance with yeast expression systems.

To Dr Paul Hills, Dr Christell van der Vyver, Dr Shaun Peters and Dr Bianke Loedolff from the Institute for Plant Biotechnology, for their valuable inputs and advice during the course of this study.

To my supervisor, Prof James Lloyd, for being patient and staying optimistic about my work even when I felt like giving up.

To my fellow colleagues at the Institute for Plant Biotechnology, Wildene le Roux, Janto Pieters, Felix van der Walt, Zanele Mgodana, Jonathan Jewell and all the other students and staff. Thank you for your kindness, friendship, support and guidance. Suffering together is much more bearable than suffering alone.

To my parents, who tried to understand all the ‘science lingo’ and offer advice where possible.

And lastly, a special thank you to my soon-to-be-husband, Kyle Hoffman, whose unfaltering support and love was the rock that carried me through.

# Table of Contents

Chapter 1: Starch – its metabolism and evolutionary development .....	1
1.1 General introduction.....	1
1.1.1 Starch.....	1
1.1.2 The importance of starch .....	1
1.1.3 Starch metabolism in higher plants.....	1
1.1.4 Overview of the evolution of land plants .....	7
1.1.5 Evolutionary development of starch metabolism .....	8
1.2 Outline of this study .....	10
Chapter 2: Heterologous expression and purification of a red algal starch synthase in <i>E. coli</i> .....	11
2.1 Introduction .....	11
2.1.1 Floridean starch structure and biosynthesis.....	11
2.1.2 Starch synthase structure .....	15
2.2 Aim and objectives .....	18
2.3 Materials and Methods .....	19
2.3.1 Microbial strains and vectors.....	19
2.3.2 Growth media .....	19
2.3.3 Sequence analysis.....	20
2.3.4 Gene synthesis .....	20
2.3.5 PCR amplification .....	20
2.3.6 Plasmid isolation.....	20
2.3.7 Restriction digests.....	20
2.3.8 DNA isolation from agarose gels .....	20
2.3.9 Site-directed mutagenesis .....	21
2.3.10 DNA sequencing.....	21
2.3.11 Bacterial transformation .....	21
2.3.12 Recombinant protein synthesis .....	21
2.3.13 Yeast transformation and protein expression .....	21
2.3.14 Protein extraction.....	22
2.3.15 Protein purification .....	22

2.3.16 Refolding denatured protein .....	22
2.3.17 SDS- and native PAGE.....	23
2.3.18 Plate-based screening for glycogen accumulation.....	23
2.4 Results .....	24
2.4.1 Sequence analysis.....	24
2.4.2 Construction of vectors.....	27
2.4.3 Site-directed mutagenesis .....	28
2.4.4 <i>E. coli</i> protein expression .....	29
2.4.5 Refolding of recombinant protein.....	32
2.4.6 Yeast protein expression.....	33
2.5 Discussion.....	36
Chapter 3: Analysis of <i>P. patens</i> maltose excess transporters and their role in starch metabolism .....	42
3.1 Introduction .....	42
3.1.1 Physcomitrella as a model system .....	42
3.2 Aim and objectives .....	44
3.3 Materials & Methods.....	45
3.3.1 Sequence analysis and phylogeny .....	45
3.3.2 <i>P. patens</i> strain, culture media and growth conditions.....	45
3.3.3 DNA and RNA extraction .....	45
3.3.4 Bacterial strains and vectors .....	46
3.3.5 Growth media .....	46
3.3.6 PCR amplification .....	46
3.3.7 DNA manipulation .....	47
3.3.8 Large-scale plasmid isolation .....	47
3.3.9 <i>P. patens</i> protoplast isolation and PEG-mediated transformation.....	47
3.3.10 Semi-quantitative RT-PCR.....	48
3.3.11 <i>E. coli</i> complementation analysis .....	48
3.4 Results .....	49
3.4.1 Sequence analysis.....	49
3.4.2 Construction of vectors.....	52

3.4.3 <i>E. coli</i> complementation analysis .....	54
3.4.4 Generation and screening of <i>P. patens</i> knockout mutants .....	56
3.4.5 <i>P. patens MEX</i> expression profile .....	56
3.5 Discussion.....	58
Chapter 4: General conclusion .....	63
Literature Cited.....	64



# List of Figures

Figure 1. Overview of diurnal starch metabolism in chloroplasts .....	4
Figure 2. Phylogenetic relationships between the major groups of extant plants.....	7
Figure 3. Conserved KXGGL (left) and KXGGL look-alike (right) motifs present in glycogen and starch synthases from bacteria ( <i>Escherichia coli</i> and <i>Agrobacterium tumefaciens</i> ) and plants ( <i>Arabidopsis thaliana</i> , <i>Oryza sativa</i> , <i>Hordeum vulgare</i> and <i>Solanum tuberosum</i> ).....	16
Figure 4. Multiple sequence alignment of the catalytic domain of glycogen and starch synthases from bacteria ( <i>Escherichia coli</i> and <i>Agrobacterium tumefaciens</i> ), red algae ( <i>Galdieria sulphuraria</i> , <i>Cyanidioschyzon merolae</i> and <i>Chondrus crispus</i> ), and higher plants ( <i>Arabidopsis thaliana</i> ).....	26
Figure 5. Construction of the YEpENO-BBH::CcSS vector and transformation into the <i>S. cerevisiae</i> $\Delta$ gsy2 mutant .....	27
Figure 6. Screening of <i>E. coli</i> colonies for the presence of the <i>C. crispus</i> starch synthase gene .....	29
Figure 7. Expression of recombinant CcSS protein in three <i>E. coli</i> strains.....	31
Figure 8. On-column refolding of recombinant CcSS extracted under denaturing conditions .....	32
Figure 9. Refolding of recombinant CcSS extracted under denaturing conditions through step-wise dialysis.....	33
Figure 10. Iodine staining of <i>S. cerevisiae</i> $\Delta$ gsy2 mutant carrying the BBH::CcSS plasmid to show glycogen accumulation.....	34
Figure 11. Expression and purification of recombinant CcSS protein in yeast.....	35
Figure 12. Starch/glycogen synthase activity of protein extracts from yeast .....	35
Figure 13. Representative maximum-likelihood (ML) tree showing relationships between members of the MEX protein family from different species of cyanobacteria and green plants.....	50
Figure 14. Schematic representation of the intron-exon structure of the <i>P. patens</i> MEX genes .....	52
Figure 15. PCR amplification of putative MEX orthologues identified in <i>P. patens</i> .....	53
Figure 16. Disruption of <i>PpMEX</i> genes by insertion of <i>nptII</i> -cassette during knockout vector construction .	53
Figure 17. Screen for correct orientation of <i>PpMEX</i> coding sequences in pBluescript during construction of complementation vectors.....	54
Figure 18. Complementation of <i>E. coli malF</i> and <i>ptsG</i> mutants with <i>P. patens</i> MEX coding sequences.....	55
Figure 19. Expression of <i>PpMEX</i> coding sequences in <i>E. coli malF</i> and <i>ptsG</i> mutant strains .....	55
Figure 20. Genotyping of <i>P. patens MEX1a</i> knockout mutant lines .....	56
Figure 21. Determination of the linear range of <i>actin5</i> ( <i>ACT5</i> ) amplification for subsequent sqRT-PCR analysis .....	57
Figure 22. Expression profile of <i>PpMEX</i> genes in protonemal and gametophytic tissue of wild type <i>P. patens</i> .....	57

# List of Tables

Table 1. The number of isoforms and types of starch metabolic enzymes.....	9
Table 2. Variations in the type of glucan observed in species of Rhodophyceae.....	12
Table 3. Strains and vectors used during this study.....	19
Table 4. Conserved residues or motifs in glycogen and starch synthases .....	28
Table 5. <i>Escherichia coli</i> strains and vectors used during this study. ....	46
Table 6. Putative MEX1 orthologues identified in <i>P. patens</i> . ....	49
Table 7. Similarity table of Arabidopsis and <i>P. patens</i> MEX proteins .....	51

## List of Abbreviations and non-SI units

<b>x g</b>	gravitational acceleration
<b>3-PGA</b>	3-phosphoglyceric acid
<b>aa</b>	amino acid
<b>ACT</b>	actin
<b>ADP</b>	adenosine diphosphate
<b>AGPase</b>	ADP-glucose pyrophosphorylase
<b>AMP</b>	adenosine monophosphate
<b>AMY</b>	alpha-amylase
<b>ara</b>	arabinose
<b>ATP</b>	adenosine triphosphate
<b>BAM</b>	beta-amylase
<b>BLAST</b>	basic local alignment search tool
<b>bp</b>	base pair
<b>CDP</b>	cytidine diphosphate
<b>Cys</b>	cysteine
<b>DBE</b>	debranching enzyme
<b>dH<sub>2</sub>O</b>	distilled water
<b>DNA</b>	deoxyribonucleic acid
<b>dNTP</b>	deoxyribonucleotide triphosphate
<b>DPE</b>	disproportionating enzyme
<b>DTT</b>	dithiothreitol
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>EF1<math>\alpha</math></b>	elongation factor 1-alpha
<b>GBSS</b>	granule-bound starch synthase
<b>GDP</b>	guanosine diphosphate
<b>Glu</b>	glutamate
<b>GST</b>	glutathione-S-transferase
<b>GSY</b>	glycogen synthase
<b>GT</b>	glycosyltransferase
<b>GWD</b>	glucan, water dikinase
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<b>his</b>	histidine
<b>IPTG</b>	isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>ISA</b>	isoamylase
<b>kDa</b>	kilo Dalton
<b>lac</b>	lactose

<b>LB</b>	lysogeny broth
<b>LDA</b>	limit-dextrinase
<b>LEW</b>	lysis-equilibration-wash
<b>LSF</b>	like sex four
<b>Lys</b>	lysine
<b>M</b>	molar
<b>Ma</b>	million years ago
<b>MEX</b>	maltose excess
<b>MOS</b>	malto-oligosaccharides
<b>nptII</b>	neomycin phosphotransferase
<b>OD<sub>600</sub></b>	optical density at 600 nm
<b>PAGE</b>	polyacrylamide gel electrophoresis
<b>PBS</b>	phosphate buffered saline
<b>PCR</b>	polymerase chain reaction
<b>PEG</b>	polyethylene glycol
<b>pGlcT</b>	plastidial glucose transporter
<b>Phe</b>	phenylalanine
<b>PMSF</b>	phenylmethylsulfonyl fluoride
<b>PTS</b>	phosphotransferase system
<b>PWD</b>	phosphoglucan, water dikinase
<b>RNA</b>	ribonucleic acid
<b>RT-PCR</b>	reverse transcription polymerase chain reaction
<b>SBE</b>	starch branching enzyme
<b>SC</b>	synthetic complete
<b>SD</b>	standard deviation
<b>SDM</b>	site-directed mutagenesis
<b>SDS</b>	sodium dodecyl sulfate
<b>SEX</b>	starch excess
<b>sqRT-PCR</b>	semi-quantitative reverse transcription polymerase chain reaction
<b>SS</b>	starch synthase
<b>TB</b>	terrific broth
<b>TE</b>	Tris-EDTA
<b>Thr</b>	threonine
<b>Tris-HCl</b>	trisaminomethane hydrochloride
<b>Tyr</b>	tyrosine
<b>UDP</b>	uridine diphosphate
<b>YPD</b>	yeast extract-peptone-dextrose
<b>YPDS</b>	yeast extract-peptone-dextrose-sorbitol

# Chapter 1: Starch – its metabolism and evolutionary development

## 1.1 General introduction

### 1.1.1 Starch

Starch is a storage polysaccharide that occurs in most plants where it is stored in the form of granules. It is a polyglucan consisting of two fractions, namely amylose and amylopectin – the amylose fraction being composed of linear alpha-1,4-glucan chains, while amylopectin contains shorter alpha-1,4-glucan chains linked together by alpha-1,6-branch points. Amylose and amylopectin are arranged together in a semi-crystalline structure to form insoluble starch granules. In most plants, the biosynthesis, storage and degradation of starch occurs within plastids, such as the chloroplasts of photosynthetic organs (e.g. leaves) or amyloplasts of non-photosynthetic organs (e.g. stems, tubers and seeds) (Zeeman et al., 2010; Streb and Zeeman, 2012; MacNeill et al., 2017).

### 1.1.2 The importance of starch

The study of starch is important for a number of reasons – not only is it a vital part of the human diet, it is also utilised in many non-food applications such as the paper, textile, oil and pharmaceutical industries. The amount of starch purified annually demonstrates its importance for industry – in the European Union alone during 2016, nearly 11 million tons of starch were isolated for industrial use ([www.starch.eu/the-european-starch-industry/#figures](http://www.starch.eu/the-european-starch-industry/#figures)). Early research on starch metabolism focused on understanding the mechanism of starch synthesis due to its significance as an industrial feedstock (Zeeman et al., 2010). More recently, research has shown the importance of starch in plant biomass accumulation (Sulpice et al., 2009) and how manipulation of its metabolism in leaves leads to an increase in plant yield (Gibson et al., 2011; Liu et al., 2016). Modifying starch metabolism has also been shown to influence abiotic stress tolerance in plants through various mechanisms, meaning that alterations in its metabolism could be used to produce plants that are more tolerant to abiotic stress (Thalmann and Santelia, 2017). The ingestion of starch is also implicated in human health, as it is quickly digested and can lead to a rapid spike in blood glucose which may lead to individuals developing insulin resistance. The incidence of this can be decreased by ingesting so-called ‘resistant’ starch as an alternative, which does not digest as quickly in the gut (Zhu et al., 2012) and has also been implicated in protecting against colorectal cancer (Nofrarias et al., 2007).

### 1.1.3 Starch metabolism in higher plants

The pathway of starch metabolism in higher plants has been studied for decades. Initially, research was aimed at mutations that lead to a visible change in seed morphology by affecting the starch content in storage organs. Recently, *Arabidopsis thaliana* (hereafter referred to as *Arabidopsis*) has become the workhorse plant for many starch researchers due to the plethora of insertion mutants that are readily available. It is important to note, however, that the predominant use of *Arabidopsis* as a model system has led to a narrow understanding of starch metabolism restricted to that of synthesis and degradation of leaf starch. There are ongoing attempts to

translate the knowledge gleaned from *Arabidopsis* into studies on the storage organs of crop plants (e.g. rice and maize endosperm, potato tubers) (Santelia and Zeeman, 2010; Lloyd and Kossmann, 2015; MacNeill et al., 2017). Finally, there has also been a significant effort to examine starch metabolism in one lower plant, the unicellular green algae *Chlamydomonas reinhardtii* (hereafter referred to as *Chlamydomonas*) (Ball and Morell, 2003), to aid in elucidating how starch metabolism evolved from unicellular organisms to multicellular vascular plants that form our main crops today. Interestingly, many of the steps appear to have been conserved during the evolution of vascular plants, although they are often catalysed by differing numbers of isoforms in multi-cellular plants compared with *Chlamydomonas* (Nougue et al., 2014). The following sections will elaborate on the pathways of starch synthesis and degradation in higher plants.

#### 1.1.3.1 Starch synthesis

The pathway of starch synthesis begins with the biosynthesis of ADP-glucose by the enzyme ADP-glucose pyrophosphorylase (AGPase) (Figure 1) which, in plants, is present as a heterotetramer composed of two large and two small subunits. The action of AGPase has been described as catalysing a rate-limiting step in starch synthesis and the enzyme is allosterically regulated, being activated by 3-phosphoglyceric acid (3-PGA) and inhibited by phosphate (Ghosh and Preiss, 1966). Its subcellular localisation is also dependent on species and plant tissue – in leaves and tubers, plastidial AGPases occur in chloroplasts (Okita et al., 1979) or amyloplasts (Kim et al., 1989) respectively, while in cereal endosperm AGPases are mostly cytosolic (Denyer et al., 1996; Beckles et al., 2001).

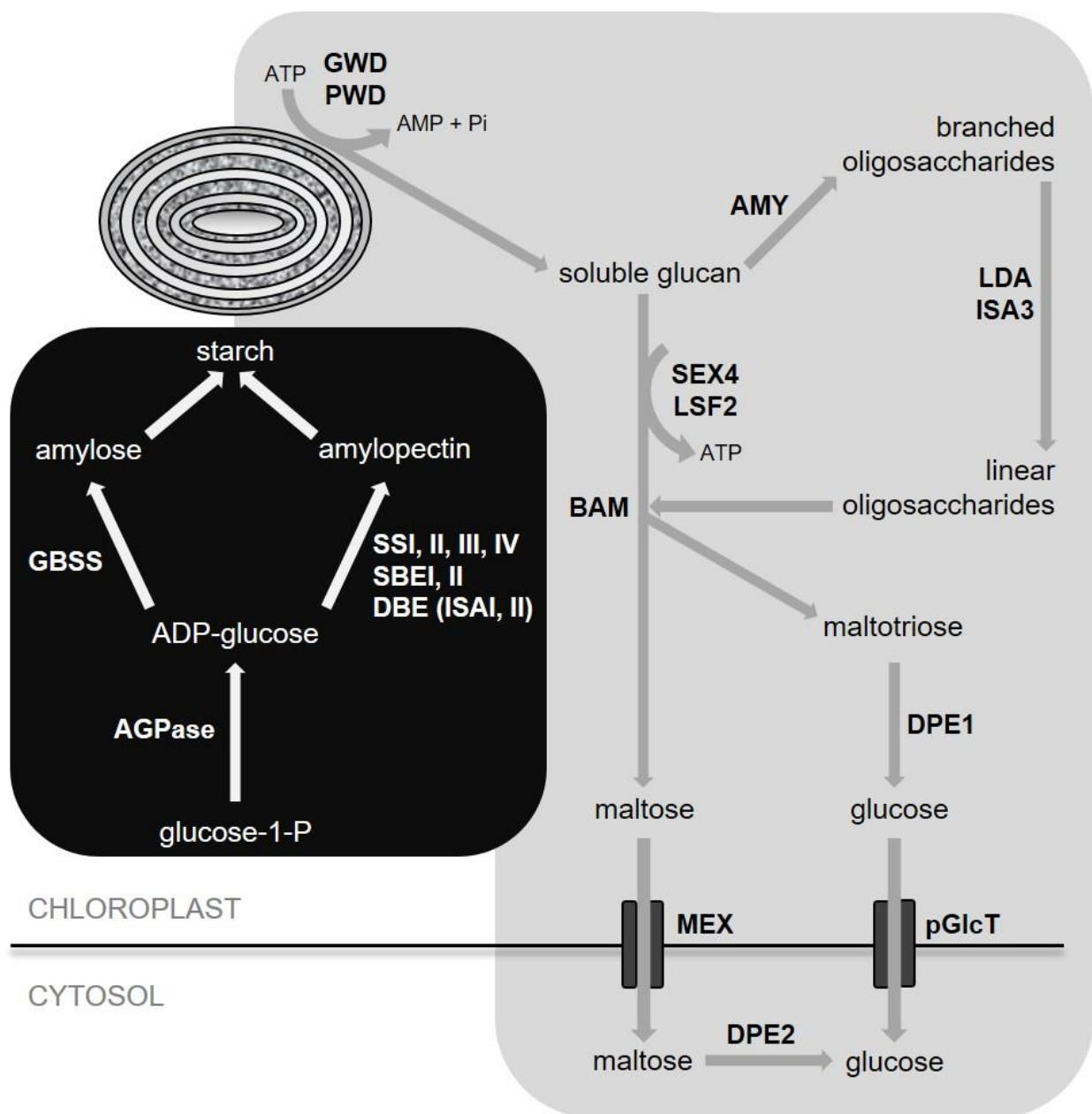
After the conversion of glucose-1-phosphate to ADP-glucose, the glucose moiety of this activated sugar is incorporated into linear  $\alpha$ -1,4 linked chains by the action of starch synthases (SS; ADP-glucose:(1 $\rightarrow$ 4)- $\alpha$ -D-glucan 4- $\alpha$ -D-glucosyltransferase; EC:2.4.1.21). In higher plants there are five classes of starch synthases designated GBSS (granule-bound starch synthase), SSI, SSII, SSIII and SSIV. The GBSS isoform is associated with the starch granule and is known to synthesize amylose. Mutants defective for GBSS have been found in a number of monocots (maize, rice, barley, wheat), dicots (potato, pea and amaranth) and green algae (*Chlamydomonas*), all of which show wild-type levels of starch with no production of amylose (Ono and Suzuki, 1957; Nelson and Rines, 1962; Murata et al., 1965; Konishi et al., 1985; Hovenkamp-Hermelink et al., 1987; Delrue et al., 1992; Denyer et al., 1995; Nakamura et al., 1995).

The other (soluble) starch synthase isoforms are implicated in the production of amylopectin, each having different properties and performing distinct roles in starch synthesis. Their distribution within the granular or stromal fractions of the plastid also varies depending on species, tissue and developmental stage (Zeeman et al., 2010). Mutants and transgenic plants impaired in particular isoforms have shown that they are responsible for synthesising different amylopectin chain lengths – SSI synthesises very short chains (Commuri and Keeling, 2001; Nakamura, 2002), while SSII synthesises chains of intermediate length (Craig et al., 1998; Edwards et al., 1999b; Umemoto et al., 2002) and SSIII synthesises longer chains (Wang et al., 1993; Gao et al., 1998; Fulton et al., 2002). In *Arabidopsis*, class III and IV starch synthases are also implicated in the initiation of starch granules. Double mutants (*ssIII-ssIV*) lacking both isoforms are rendered incapable of producing starch, whereas triple mutants (*ssI-ssII-ssIII* and *ssI-ssII-ssIV*) demonstrate that the presence of

either SSIII or SSIV is sufficient in promoting granule initiation and starch synthesis, even in the absence of other soluble starch synthases (Szydłowski et al., 2009). The exact role of the different starch synthase isoforms are, however, likely to vary between species as, for example, SSI or SSIIIa have been demonstrated to be indispensable for initiation of starch synthesis in rice endosperm (Fujita et al., 2011).

The  $\alpha$ -1,6-branchpoints within amylopectin are formed by starch branching enzymes (SBEs;  $\alpha$ -1,4-glucan:  $\alpha$ -1,4-glucan-6-glycosyltransferase; EC: 2.4.1.18) and occur concurrently with elongation of the glucan chains. Branching enzymes cut existing  $\alpha$ -1,4-glucan chains and transfer the severed chain to an intact chain via an  $\alpha$ -1,6-link. There are two classes of branching enzymes in higher plants designated SBEI (or SBE B) and SBEII (or SBE A) which preferentially transfer longer and shorter glucan chains respectively (Guan and Preiss, 1993; Takeda et al., 1993; Burton et al., 1995). There appears to be a synergistic interaction between the two classes as they exhibit overlapping specificities and can, at least in part, complement each other. For example, while the loss of SBEI has almost no effect, loss of SBEII results in an increase in the amylose content of tuber starch and the combined downregulation of SBEI and SBEII lead to an extremely high amylose content in potato (Safford et al., 1998; Jobling et al., 1999; Schwall et al., 2000). There is also increasing evidence for the formation of protein complexes between starch synthase and branching enzymes in starch-storage organs. As they act on a common substrate, multimeric complexes of these enzymes could serve to synchronise their activities in order to produce the complex structures present in starch granules (Zeeman et al., 2010). Enzyme complexes containing SSI, SSIII and SBEII have been reported in maize as well as complexes consisting of SBEI and SBEII in wheat (Tetlow et al., 2004; Tetlow et al., 2008).

Importantly, debranching enzymes (DBEs;  $\alpha$ -1,6-glucanohydrolase) are required for the ordered structure within amylopectin as they cleave excess  $\alpha$ -1,6 bonds. There are two types of DBEs in higher plants, namely isoamylase (ISA; EC: 3.2.1.68) and limit-dextrinase (LDA, or pullulanase; EC: 3.2.1.142). There are three classes of the ISA type – ISA1, ISA2 and ISA3. In dicotyledenous plants, ISA1 and ISA2 form a heteromultimeric protein complex (Hussain et al., 2003; Delatte et al., 2005; Wattedled et al., 2005), while in cereals ISA1 also exists as a homomultimer in addition to a heteromultimer with ISA2 (Utsumi and Nakamura, 2006; Kubo et al., 2010; Lin et al., 2013). Cereal endosperms, Arabidopsis leaves, green algae and potato tubers lacking ISA1 exhibit severe alterations to glucan structure with a reduction in granular starch and accumulation of phytoglycogen instead (James et al., 1995; Mouille et al., 1996; Nakamura et al., 1997; Dauvillée et al., 2001b; Burton et al., 2002; Bustos et al., 2004; Delatte et al., 2005; Wattedled et al., 2005). Similarly, this effect on glucan structure has been noted in Arabidopsis and potato when ISA2 is deficient (Zeeman et al., 1998b; Bustos et al., 2004; Delatte et al., 2005; Wattedled et al., 2005). Phytoglycogen is a water-soluble glycogen-like polymer with more short chains and branchpoints than amylopectin and is incapable of self-organising into the higher-order structures observed within starch granules. This accumulation of phytoglycogen is thought to occur due to excess branchpoints being introduced by SBEs which are not removed by the ISA1/ISA2 complex and thus prevent granule formation. ISA3 and LDA mainly seem to be involved in starch degradation (discussed below).



**Figure 1. Overview of diurnal starch metabolism in chloroplasts.** The pathway of starch synthesis is highlighted in black and that of starch degradation in grey. Starch synthases (GBSS, SSI, SSII, SSIII, SSIV) incorporate ADP-glucose into glucan chains, while SBEs introduce branch-points. Excess branches are cleaved by DBEs to allow self-organisation of amylose and amylopectin into starch granules. At night, the degradation of starch begins with solubilisation of the granule surface through reversible phosphorylation by GWD and PWD to allow subsequent attack by amylolytic enzymes (BAM, AMY, LDA/ISA3, DPE1) which produce maltose and glucose. Dephosphorylation by SEX4 and LSF2 is necessary to allow full degradation of glucan chains into soluble sugars. These sugars are then exported to the cytosol by transporters (MEX, pGlcT) where maltose is further converted to glucose by the action of DPE2. AGPase: ADP-glucose pyrophosphorylase, GBSS: granule-bound starch synthase, SS: starch synthase, SBE: starch-branching enzyme, DBE: debranching enzyme, ISA: isoamylase, GWD: glucan, water dikinase, PWD: phosphoglucan, water dikinase, BAM:  $\beta$ -amylase, AMY:  $\alpha$ -amylase, LDA: limit-dextrinase, SEX4: Starch Excess-4, LSF2: Like Sex Four-2 DPE: disproportionating enzyme, MEX: maltose excess transporter, pGlcT: plastidial glucose transporter.



### 1.1.3.2 Starch degradation

Depending on the time of day and the plant organ, starch degradation will occur. In leaves, starch is accumulated during the day and degraded at night to provide a continued supply of carbon while the plant cannot photosynthesise. In seeds and tubers on the other hand, starch degradation may be delayed for several months and occurs only once energy is required for germination and growth of sprouts. The pathway of starch degradation is best understood in leaves and cereal endosperms, while knowledge about the pathway in tubers, roots and non-cereal seeds is limited. A brief explanation of the general pathway of starch degradation in leaves is outlined below, as well as some differences that occur in starch breakdown in endosperms.

The first step in starch degradation is catalysed by glucan, water dikinase (GWD; ATP:  $\alpha$ -1,4-glucan, water phosphotransferase; EC: 2.7.9.4) which phosphorylates the C6 position of glucosyl residues in amylopectin, followed by phosphorylation of the C3 position by phosphoglucan, water dikinase (PWD; ATP: phosphor- $\alpha$ -1,4-glucan, water phosphotransferase; EC: 2.7.9.5) (Ritte et al., 2002; Ritte et al., 2006) (Figure 1). Loss of GWD in *Arabidopsis* (*sex1*), lotus, tomato and rice leads to a severe starch-excess phenotype in leaves (Yu et al., 2001; Nashilevitz et al., 2008; Vriet et al., 2010; Hirose et al., 2013), and additionally in potato leads to inhibition of cold-induced starch degradation in tubers (Lorberth et al., 1998). This phosphorylation serves to solubilise the surface of the starch granule and allow subsequent attack by amylolytic enzymes, primarily  $\beta$ -amylase,  $\alpha$ -amylase and isoamylase.

A crucial step in starch degradation is the removal of phosphate groups incorporated by GWD and PWD by the action of two phosphatases, namely SEX4 and LIKE SEX FOUR-2 (LSF2) (Kotting et al., 2009; Santelia et al., 2011). *Arabidopsis* mutants lacking SEX4 show a decreased rate of starch degradation and an accompanying accumulation of these phosphorylated malto-oligosaccharides (MOSs) during the night (Zeeman et al., 1998a; Kotting et al., 2005; Niittylä et al., 2006) while potato plants repressed in either gene also demonstrate reduced rates of starch degradation (Samodien et al., 2018). Dephosphorylation by SEX4 and LSF2 is required to allow the action of BAMs, which are inhibited by phosphate groups close to the non-reducing ends of glucan chains (Takeda and Hizukuri, 1981; Kotting et al., 2005).

The combined activity of the amylolytic enzymes is responsible for the breakdown of starch into simple glucose and maltose units. Maltose is liberated from the non-reducing ends of exposed linear glucan chains by exo-acting  $\beta$ -amylases (BAMs;  $\alpha$ -1,4-glucan malto-hydrolase; EC 3.2.1.2). In potato and *Arabidopsis*, research has shown that BAMs are essential for breakdown of transitory leaf starch as mutant plants with down-regulated expression exhibit a starch-excess phenotype in their leaves, indicating an inability to degrade starch normally (Scheidig et al., 2002; Fulton et al., 2008). Higher plants have a multitude of genes coding for BAMs – in *Arabidopsis* there are a total of nine genes, which are suspected to have overlapping as well as specialised roles. Some have been found localised to the phloem (Wang et al., 1995) while others even act as transcription factors (Reinhold et al., 2011). As BAMs are incapable of cleaving branchpoints or shorter MOSs, the action of other enzymes is required to fully degrade starch. The  $\alpha$ -1,6-branchpoints of amylopectin are hydrolysed by the action of DBEs (ISA3 and LDA), which release short linear MOS (Delatte et al., 2006; Zeeman et al., 2007). The shorter MOS are further acted upon by disproportionating enzyme 1 (DPE1;  $\alpha$ -1,4-glucan 4- $\alpha$ -

glucanotransferase; EC: 2.4.1.25) which cleaves a maltosyl moiety and transfers it to another glucan, thereby generating glucose and a longer glucan that can again be degraded by BAM. In cereal endosperm,  $\alpha$ -amylases (AMYS; 4- $\alpha$ -D-glucan glucanohydrolase; EC 3.2.1.1.) in the aleurone layer are essential for the breakdown of starch during seed germination. Anti-sense inhibition of AMYI-1 (an AMY1-type protein) in the endosperm of rice leads to delayed seedling germination and growth (Asatsuma et al., 2005). However, the role of AMYS in leaf starch degradation remains unclear and seems to differ between species. Of the three AMY types in *Arabidopsis* (AMY1, AMY2 and AMY3) only AMY3 is localised to the chloroplast and thought to be involved in transitory leaf starch degradation, whereas AMY1 and AMY2 are likely cytosolic (Stanley et al., 2002). Starch degradation in *Arabidopsis* mutants lacking only AMY3, as well as AMY1 and AMY2 respectively, remains unaffected (Yu et al., 2005). Simultaneous removal of AMY3 and ISA3 increases severity of the starch excess phenotype observed in *isa3* mutants, indicating the contribution of AMY3 to normal starch degradation (Streb et al., 2012). In rice leaves, AMYI-1 was demonstrated to be chloroplastic and play a pivotal role in remobilisation of transitory starch (Asatsuma et al., 2005).

After the breakdown of starch into maltose and glucose, these sugars are exported across the plastid membrane to the cytosol by glucose- or maltose-specific transporters (Weber et al., 2000; Niittylä et al., 2004). *Arabidopsis mex1* mutants defective for the plastidial maltose transporter (MEX1) exhibit a severe retardation of growth in addition to accumulation of high levels of starch and maltose (Niittylä et al., 2004), indicating that maltose is the major product of starch breakdown exported from the chloroplast. Recent research has also demonstrated the importance of the plastidic glucose translocator (pGlcT) in the export of starch degradation products. While single mutants (*pglct-1*) showed no visible growth defects, double mutants (*pglct-1/mex1*) lacking both transporters exhibited a more severe growth defect than either single mutant, alongside reductions in starch turnover and photosynthetic activities and extreme chloroplast abnormalities (Cho et al., 2011). In the cytosol, maltose is acted upon by disproportionating enzyme 2 (DPE2) to synthesise glucose (Chia et al., 2004).

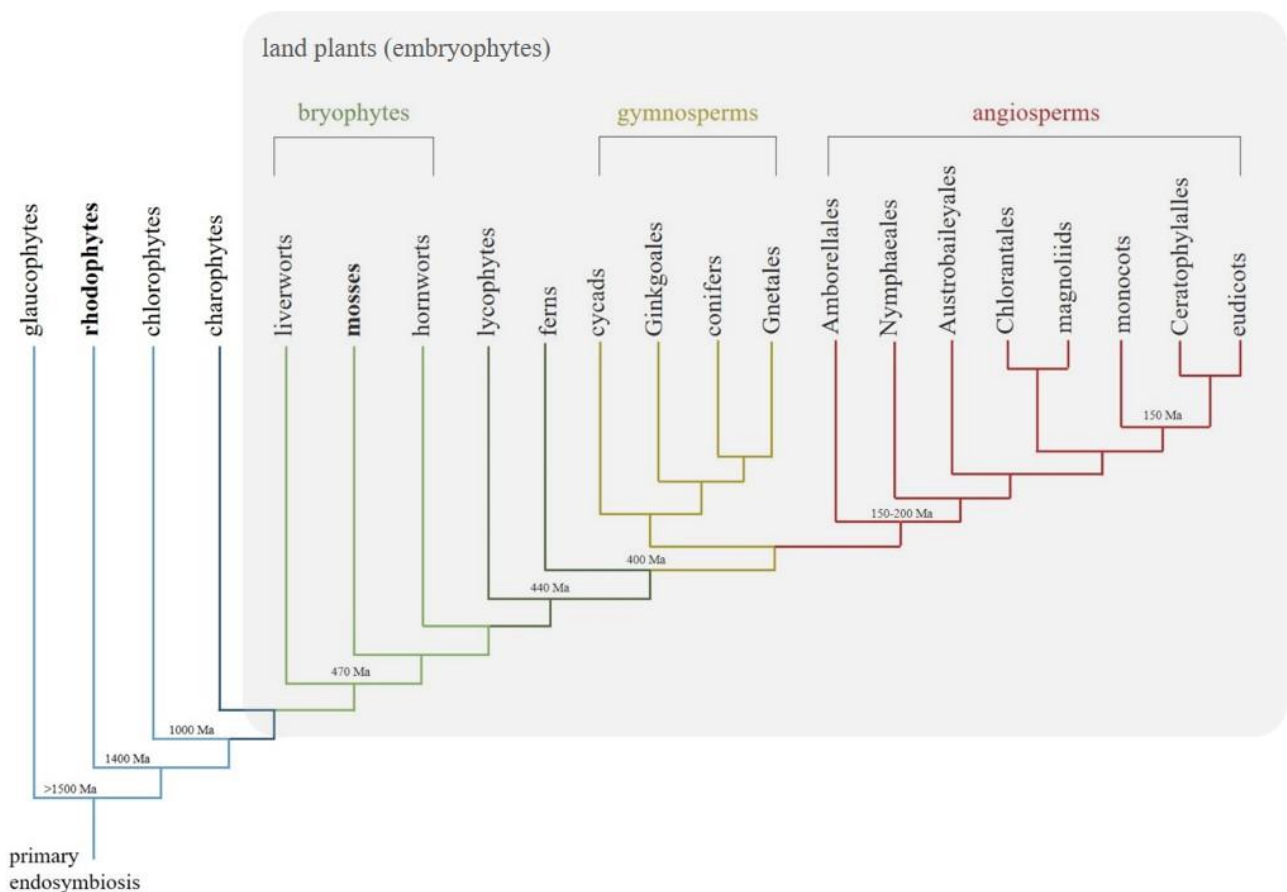
Glucose molecules in the cytosol become phosphorylated and can be used either for sucrose synthesis or enter glycolysis. Cytosolic glucose that originates from starch degradation products is produced by two separate routes. One of these involves the production of glucose in the plastid and its transport into the cytosol, while the other involves the production of maltose and its transport into the cytosol, followed by metabolism by DPE2. The maltose route appears to be the predominant one in *Arabidopsis* leaves as mutations in either DPE2 (Chia et al., 2004) or the maltose transporter (Niittylä et al., 2004) have a much greater effect on starch degradation than mutations in DPE1 (Critchley et al., 2001) or pGlcT (Cho et al., 2011).

The pathway of endosperm starch degradation occurs in a vastly different environment and thus shows some marked differences to that of leaf starch. The starch-degrading enzymes present in endosperm include  $\alpha$ -amylase, BAM, DBE (specifically LDA) and  $\alpha$ -glucosidase (maltase; EC: 3.2.1.3). Here,  $\alpha$ -amylase is generally accepted to play a central role in starch degradation while  $\beta$ -amylase seems to be partly redundant. This is demonstrated by the ability of some barley cultivars which almost completely lack BAM to still exhibit normal seedling growth (Kaneko et al., 2000). After the liberation of maltose and MOS by  $\alpha$ -amylase,  $\beta$ -

amylase and LDA, maltase is responsible for their conversion to glucose (Stanley et al., 2011). There is also evidence suggesting a synergistic interaction between maltase and  $\alpha$ -amylase to directly liberate glucose from the surface of the starch granule (Sun and Henson, 1991). This is in stark contrast to leaf starch degradation, where maltose is exported to the cytosol and then hydrolysed by DPE2.

#### 1.1.4 Overview of the evolution of land plants

Despite the advances made in understanding the pathway of starch biosynthesis in model plants such as *Arabidopsis*, there is much that remains unclear about starch metabolism in other plants. Understanding starch metabolism in lower plants such as mosses and algae is key in unravelling the evolutionary development of multicellular vascular plants from their once unicellular ancestor. Land plants (embryophytes) evolved from multicellular freshwater algae (presumably related to the charophytes) as a monophyletic sister group to green algae (chlorophytes) (Figure 2).



**Figure 2. Phylogenetic relationships between the major groups of extant plants.** Estimated time of divergence during plant evolution is indicated in millions of years ago (Ma). Adapted from Pires and Dolan (2012).

The earliest microfossil evidence for plants on land extends back to 470 million years ago (Ma), with spore and tissue fragments resembling that of extant liverworts (Wellman et al., 2003; Gensel, 2008; Rubinstein et al., 2010). The colonisation of land was facilitated by the progressive development of several key features subsisting in extant land plants. First, the last common ancestor of the bryophytes developed two distinct multicellular generations in its life cycle, namely gametophytes (haploid) and sporophytes (diploid). There is

evidence that development of the sporophyte generation was, in part, achieved by recruitment of genes and genetic mechanisms specific to gametophytes in ancestral plants into new regulatory roles in the sporophyte generation (see Pires and Dolan, 2012 and references within).

Secondly, vascularisation and the development of leaf-structures occurred approximately 440 Ma in the last common ancestor of lycophytes and ferns, which allowed movement away from damp areas and colonisation of areas with more restricted access to water (Crane and Kenrick, 1997). The development of roots was also first observed in the lycophyte clade, although it is suggested that root development occurred at least twice, as evidence for root systems in other vascular plants is only found later (Gensel et al., 2001; Raven and Edwards, 2001). Lastly, the evolution of seeds and flowers was likely the most critical development in land plants. The first seed plants appeared approximately 400 Ma, with the gymnosperms becoming the most dominant flora between 260 and 70 Ma (Willis and McElwain, 2002). This was followed by development of flowers in the angiosperms (approximately 150 – 200 Ma), that have dominated the world flora from 100 – 65 Ma until the present day (Willis and McElwain, 2002).

### **1.1.5 Evolutionary development of starch metabolism**

It has been theorised that the development of plastids and the ability to synthesise starch arose from an endosymbiotic event between an ancestral cyanobacterium and a heterotrophic eukaryotic host (Moreira et al., 2000; Rodríguez-Ezpeleta et al., 2005). This primary endosymbiotic event led to the generation of three lineages capable of oxygenic photosynthesis, namely Glaucophyta (freshwater unicellular algae), Rhodophyceae (red algae) and Chloroplastida (green algae and higher plants). These are collectively referred to as the Archaeplastida, and are marked by plastids harbouring only two membranes (Adl *et al.*, 2005). A later, secondary endosymbiotic event gave rise to derived plastidial lines (e.g. brown algae, cryptophytes, haptophytes, alveolates) containing plastids with more than two membranes (Keeling, 2009).

It is evident that some gene transfer occurred between the host and endosymbiont after the primary endosymbiotic event, as the enzymes involved in starch metabolism show a clear phylogenetic relationship to enzymes of both eukaryotic and cyanobacterial origin (Patron and Keeling, 2005; Deschamps et al., 2008a). The bacterially-derived genes were first expressed in the plastid, and after incorporation into the host's nuclear genome were expressed in the cytosol of the eukaryotic host as they lacked transit peptides. Here, they were thought to have resulted in the production of cytosolic starch (as is evident in members of the Rhodophyceae and Glaucophyta). This shift from cytosolic glycogen to starch production presumably took place as semi-crystalline storage polyglucans provide a more permanent carbon sink than glycogen (Ball et al., 2011). The synthesis of a more starch-like polyglucan was also made possible by the addition of enzymes responsible for amylose synthesis (GBSS) and others responsible for the crystalline structure of starch from the cyanobacterial endosymbiont (Patron and Keeling, 2005). Indeed, upon loss of isoamylase activity (responsible for the crystalline nature of starch) a marine cyanobacterium capable of producing both glycogen and starch has been demonstrated to shift to a glycogen-dominant mechanism of carbohydrate storage (Cenci et al., 2013).

**Table 1. The number of isoforms and types of starch metabolic enzymes present in rhodophytes (red algae; *Chondrus crispus*, *Cyanidioschyzon merolae*), chlorophytes (green algae; *Ostreococcus tauri*, *Chlamydomonas reinhardtii*), bryophytes (moss; *Physcomitrella patens*) and higher plants (*Arabidopsis thaliana*). The number of isoforms were obtained from Ball et al. (2015) and Stander (2015). GBSS - granule-bound starch synthase; DPE – disproportionating enzyme; GWD – glucan water dikinase; PWD – phosphoglucan water dikinase.**

Enzyme	Rhodophytes		Chlorophytes		Bryophytes	Higher plants
	<i>C. merolae</i>	<i>C. crispus</i>	<i>O. tauri</i>	<i>C. reinhardtii</i>	<i>P. patens</i>	<i>A. thaliana</i>
ADP-glucose pyrophosphorylase	-	-	3	5	9	6
ADP-glucose starch synthase	-	-	5	7	11	5
GBSS	-	-	1	2	2	1
UDP-glucose starch synthase	1	1	-	-	-	-
Starch branching enzyme	1	1	2	3	3	3
Limit dextrinase (pullulanase)	1	1	1	1	1	1
Isoamylase	2	2	3	3	3	3
$\alpha$ -amylase	1	-	3	3	7	3
Starch phosphorylase	1	1	3	2	6	2
DPE1	-	-	1	1	3	1
$\beta$ -amylase	1	1	2	3	7	9
DPE2	2	1	1	1	1	1
GWD-PWD	1	1	3	2	5	2
Glucan phosphatase	1	1	1	1	5	1
Maltose transporter (MEX)	-	-	1	1	4	1

The divergence of the Chloroplastida is marked by a relocation of the starch metabolic pathway from the cytosol to the plastid as starch enzymes acquired transit peptides. During this period, a series of gene duplications accompanied by specialisation occurred (Deschamps et al., 2008c), explaining the high number of isoforms of starch metabolic enzymes present in the Chloroplastida as compared to the Rhodophyceae (see Table 1). To date, no plant has been identified that synthesizes starch in both compartments indicating that the formation of plastidial starch came alongside the loss of cytosolic starch in green algae and higher plants (Ball et al., 2011). There is a clear increase in complexity of the starch biosynthetic pathway from red algae to green algae, mosses and angiosperms. For example, red algae such as *Chondrus crispus* have only one starch synthase, whereas *Chlamydomonas* (green alga), *Physcomitrella patens* (moss) and *Arabidopsis* (angiosperm) have nine, thirteen, and six different starch synthase isoforms respectively (Table 1). Much like glycogen metabolism in eukaryotes and bacteria, the starch metabolic pathway in red algae is maintained by fewer than 12 genes as only one isoform exists for each enzymatic reaction. In contrast, the biochemical pathway in Chloroplastida consists of a complex network of 30 – 40 polypeptides encoded by separate genes.

This is an interesting observation, as the complexity of starch granule architecture was thought to be attributed to the complexity of the starch biosynthetic pathway. However, members of the Rhodophyceae are capable of accumulating complex starch granules with a metabolic pathway no more complex than that of glycogen metabolism. Also interesting to note is that all steps of the starch metabolic pathway are the same in Chloroplastida and Rhodophyceae, with the only notable difference being the absence of AGPase and an ADP-glucose requiring starch synthase, and the presence of a UDP-glucose requiring starch synthase.

## 1.2 Outline of this study

This project consists of two parts/chapters, each of which considers a different aspect of starch metabolism in plants. The first experimental chapter (Chapter 2) focuses on a class of enzymes, namely starch synthases, which are involved in the production of starch. The second experimental chapter (Chapter 3) of this project forms part of a larger study which examines whether the pathway of starch metabolism observed in higher plants is conserved in the moss *P. patens*.

# Chapter 2: Heterologous expression and purification of a red algal starch synthase in *E. coli*

## 2.1 Introduction

### 2.1.1 Floridean starch structure and biosynthesis

The red algae (Rhodophyceae) form a monophyletic clade which consists of several major classes. They differ from green plants as they produce what is called ‘floridean starch’, named after the class Florideophyceae from which it was originally isolated. Floridean starch has been an interesting focus point of research as: (1) it is synthesized and stored in the cytosol like animal, fungal and bacterial glycogen, rather than in plastids like green algae and higher plants; and (2) red algal starch synthase enzymes utilize UDP-glucose, the same substrate as for other non-plant eukaryotic glycogen synthesis, unlike bacterial and green plant starch synthases which utilize ADP-glucose. While some floridean starches may share structural and physico-chemical characteristics similar to those of starch from higher plants (Viola et al., 2001; Yu et al., 2002), they generally exhibit a shorter chain length, higher branching frequency and lower gelatinization temperatures (Yu et al., 2002; Stadnichuk et al., 2007).

Floridean starch was first thought to completely lack amylose and consist of a polymer that is more similar to amylopectin than glycogen (Peat et al., 1959; Meeuse et al., 1960; Manners and Wright, 1962; Manners and Sturgeon, 1982; Viola et al., 2001). Although this holds true for members of the Florideophyceae, later studies revealed the presence of amylose in species of the Rhodellophyceae, Bangiophyceae and Porphyridiophyceae classes which produce a semi-amylopectin type of glucan (Table 2) similar to that found in some cyanobacteria (McCracken and Cain, 1981; Nakamura et al., 2005; Shimonaga et al., 2006; Shimonaga et al., 2008). A gene encoding a GBSS-like enzyme, identified and isolated from *Porphyridium purpureum* (Porphyridiophyceae) (Shimonaga et al., 2006), shares high homology with higher plant GBSSI confirming the ability of some red algae to manufacture amylose. Like the Florideophyceae, members of the Cyanidiophyceae such as *Cyanidium caldarium* and *Cyanidioschyzon merolae* were also found to lack amylose and produce a glycogen-type polyglucan instead of amylopectin (Frederick, 1967; Shimonaga et al., 2008).

The type of storage polyglucan produced is also in some cases dependent on environmental conditions. The unicellular red alga *Galdieria maxima* is an acidophile found in hot sulphuric springs where it is capable of switching between heterotrophic and autotrophic growth depending on environmental conditions. During autotrophic growth *G. maxima* produces only trace amounts of carbohydrate-like compounds which fail to resemble storage polysaccharides, with high quantities of starch-like granules only being produced under heterotrophic growth conditions (Stadnichuk et al., 2007). These polysaccharides chemically resemble starch although they appear to be closer in nature to cyanobacterial phytyglycogen than other typical floridean starches with regards to their degree of branching (Stadnichuk et al., 2007).

Interestingly, the ability to synthesise amylose appears to be limited to later diverging lineages and absent in more primitive red algal species such as the Cyanidiophyceae and Florideophyceae classes (Table 2). This

could indicate that accumulation of a glycogen-type storage polyglucan was the ancestral state in the first red algae with the subsequent development of amylose synthesis by later acquisition of GBSS through lateral gene transfer in the Porphorydiophyceae. However, the presence of amylose in some cyanobacteria and GBSS in all Archaeplastida suggest rather that the first diverging red algae produced a semi-amylopectin starch containing amylose (Nakamura et al., 2005; Plancke et al., 2008; Shimonaga et al., 2008). This ancestral ability to synthesise cytosolic starch was later lost in Cyanidiophyceae and replaced by glycogen synthesis as a result of the hot acidic environments which members of this class inhabit. Evidence for this assumption is based on the average melting temperature of floridean starches which falls far below the optimal growth temperature of Cyanidiophyceae and would thus be selected against due to its instability at high temperatures (Shimonaga et al., 2008). While the melting temperature of floridean starches has been demonstrated to range between 36 - 43°C (Shimonaga et al., 2008), the hot acidic environments these red algae reside in can reach up to 56°C (Brock, 1978; Gross and Oesterhelt, 1999). This view is further strengthened by the presence of GWD in the genome of *C. merolae* as a remnant of the starch degradation pathway and its absence in most bacteria and eukaryotes that are unrelated to plants (Matsuzaki et al., 2004; Shimonaga et al., 2008), although there are some exceptions as in the case of *Toxoplasma gondii* that acquired GWD through horizontal gene transfer (Coppin et al., 2005). Therefore, the common ancestor of all plants is thought to have produced starch-like structures, while the storage polyglucan produced by the first diverging red algal ancestor was more reminiscent of the semi-amylopectin and amylose-containing type starch present in extant Porphorydiophyceae like *P. purpureum* and *P. sordidum*.

**Table 2. Variations in the type of glucan observed in species of Rhodophyceae (red algae).** Classes are arranged by estimated time of divergence from the last common ancestor in ascending order with Cyanidiophyceae representing the most primitive class based on Qiu et al. (2016) and Yang et al. (2016).

Species	Class	Type of glucan	Source
<i>Cyanidium caldarium</i>	Cyanidiophyceae	Glycogen	(Shimonaga et al., 2006)
<i>Galdieria sulphuraria</i>	Cyanidiophyceae	Glycogen	(Shimonaga et al., 2008)
<i>Galdieria maxima</i>	Cyanidiophyceae	Glycogen	(Stadnichuk et al., 2007)
<i>Cyanidioschyzon merolae</i>	Cyanidiophyceae	Semi-amylopectin	(Hirabaru et al., 2010)
<i>Gracilaria lemaneiformis</i>	Florideophyceae	Semi-amylopectin	(Yu et al., 2002)
<i>Gracilariopsis sp.</i>	Florideophyceae	Semi-amylopectin	(Yu et al., 2002)
<i>Gracilaria chilensis</i>	Florideophyceae	Semi-amylopectin	(Yu et al., 2002)
<i>Rhodorus marinus</i>	Stylonematophyceae	Semi-amylopectin, amylose	(Shimonaga et al., 2008)
<i>Rhodella violaceae</i>	Rhodellaphyceae	Semi-amylopectin, amylose	(Shimonaga et al., 2008)
<i>Porphyridium purpureum</i>	Porphyridiophyceae	Semi-amylopectin, amylose	(Shimonaga et al., 2006)
<i>Porphyridium sordidum</i>	Porphyridiophyceae	Semi-amylopectin, amylose	(Shimonaga et al., 2008)

The unusual subcellular localisation of floridean starch synthesis and storage to the cytosol raises the question as to whether its starch metabolism pathway resembles that found in plants or glycogen metabolism in bacteria,



fungi and animals. The majority of research examining starch metabolism in red algae has focused on the substrate used for chain elongation during starch synthesis. These have demonstrated the ability of starch synthases present in red algal extracts to incorporate both ADP-glucose and UDP-glucose under *in vitro* conditions (Frederick, 1967; Nagashima et al., 1971). Similarly in higher plants, the ability of GBSS to utilise both ADP-glucose and UDP-glucose *in vitro* has been demonstrated, although the use of UDP-glucose *in vivo* seems unlikely due to the lack of UDP-glucose pools in the plastid (Preiss, 1988; Kleczkowski, 1994). The starchless phenotype observed in AGPase mutants further supports the use of ADP-glucose as the sole substrate for plant starch synthesis *in vivo* (Lin et al., 1988; Zabawinski et al., 2001; Rösti et al., 2007). This was initially assumed to also be true for red algae and floridean starch synthesis. Partially purified starch synthase from the red alga *Serraticardia maxima* exhibited the highest activity for ADP-glucose (compared with UDP-glucose and GDP-glucose) *in vitro* (Nagashima et al., 1971). Likewise, crude extracts of *P. purpureum* showed incorporation of ADP[<sup>14</sup>C-glucose] into starch, although specificity for UDP[<sup>14</sup>C-glucose] was not tested (Sheath et al., 1979). Lower incorporation of UDP[<sup>14</sup>C-glucose] than ADP[<sup>14</sup>C-glucose] into glycogen was also reported for crude extracts from *Gracilaria gracilis* (Sesma and Iglesias, 1998). However, later research revealed that the starch synthase from *Gracilaria tenuistipitata* was highly sensitive and showed a 8.6 – 16.2 fold higher incorporation of UDP[<sup>14</sup>C-glucose] as compared to ADP[<sup>14</sup>C-glucose] when protein was extracted under optimised conditions in the presence of antioxidants and protease inhibitors (Nyvall et al., 1999). Similar to higher plant starch synthases, the activity of the *G. tenuistipitata* UDP-glucose utilising starch synthase was also increased by the addition of citrate (Nyvall et al., 1999). Furthermore, starch synthases of *G. tenuistipitata*, *Gracilaria sordida* and *Gracilariopsis lemaneiformis* showed significantly higher activity for UDP-glucose as compared to ADP-glucose when assayed in the presence of citrate and excess glycogen (Nyvall, 2000). Based on these results, it was suggested that the low substrate specificity for UDP-glucose observed in previous studies (Nagashima et al., 1971; Sesma and Iglesias, 1998) was due to the selective loss of UDP-glucose utilising starch synthases under the protein extraction conditions employed. The *G. tenuistipitata* starch synthase was the first starch synthase known to be specific for UDP-glucose. It is assumed to be a soluble starch synthase as only low levels of activity could be detected in association with starch granules (Nyvall, 2000). This corroborates the observation of a semi-amylopectin type starch lacking amylose in species closely related to *G. tenuistipitata* (Yu et al., 2002) as GBSS is embedded within the starch granule in green algae and higher plants.

This information led to a revised view of floridean starch synthesis for two reasons. Firstly, it was thought to possibly proceed via two pathways, one taking place in the plastid using ADP-glucose and the other occurring in the cytosol utilising UDP-glucose (Viola et al., 2001). Although red algal starch synthases showed a much higher specificity for UDP-glucose and this accounted for majority of the floridean starch produced, there were still reported incidences of ADP-glucose incorporation (see above) coupled with the reported purification and characterisation of AGPase (the enzyme which catalyses production of ADP-glucose) in *G. gracilis* (Sesma and Iglesias, 1998). It was suggested that two separate enzymes are responsible for the different pathways (Nyvall et al., 1999; Viola et al., 2001). However, analysis of genome sequences from four red algae, namely *C. merolae*, *G. sulphuraria*, *P. purpureum* and *Chondrus crispus* (Matsuzaki et al., 2004; Bhattacharya et al.,

2013; Collén et al., 2013; Schonknecht et al., 2013) has revealed both the lack of AGPase and the presence of only a single starch synthase gene per red algal genome. Moreover, based on phylogenetic analyses these red algal starch synthases belong to the same glycosyltransferase family as synthases from derived plastidial lines such as the glaucophyte *Cyanophora paradoxa*, which has been demonstrated to synthesise cytosolic starch via a UDP-glucose based pathway (Plancke et al., 2008; Ball et al., 2015). Thus, it seems plausible that red algal starch synthases preferentially utilise UDP-glucose as substrate similarly to their derived sister lineages, and that the ADP-glucose utilising activity observed in early studies can be attributed to the non-specificity of the same enzyme as the red algal genomes harbour only a single starch synthase encoding gene. However, a detailed biochemical analysis of red algal starch synthases is lacking.

The pathway of starch metabolism in red algae shares many similarities with that of green algae and higher plants, despite a number of key differences. Photosynthetic carbon assimilation takes place in the plastid (rhodoplast in red algae) and the photosynthetic products are exported to the cytosol via the triose-phosphate translocator (Linka et al., 2008). In the cytosol UDP-glucose is incorporated into glucan chains to form storage polysaccharides, which vary from glycogen to semi-amylopectin containing starch (with or without amylose) depending on the species of red algae. The pathway of floridean starch degradation appears to follow that of green algae and higher plants. From the sequenced genomes, it is clear that red algae (with the exception of the Cyanidiophyceae which produce glycogen) possess GWD or PWD and starch phosphatase (SEX4)-like genes. Based on sequence data, it appears that in some species like *P. purpureum* and *C. merolae* GWD activity is sufficient at solubilising the surface of starch granules as a gene encoding PWD is absent (Ball et al., 2015), however, *in vivo* studies have not yet been performed to corroborate this. It is interesting to note the presence of an  $\alpha$ -1,4-glucan lyase (EC 4.2.2.13) in the rhodoplast of *G. lemaneiformis* as observed by immunogold/transmission electron microscopy (Yu and Pedersén, 1993). This enzyme cleaves 1,5-anhydrofructose from  $\alpha$ -1,4-linked glucan chains or MOSs, although its role in red algae remains unclear, as its compartmental separation from starch granules prompts questions surrounding the source of MOSs (Yu et al., 1993; Yu et al., 1995; Nyvall, 2000).

To date, biochemical characterisation of the UDP-glucose starch synthase enzymes present in red algae is lacking – particularly in commercially important species such as *C. crispus* which is a rich source of carrageenan (commonly used as a thickener and stabiliser in food, printing and cosmetics). In addition to its importance to a number of industries, a large amount of knowledge and resources are available for *C. crispus* including its genome sequence (Collén et al., 2013), mitochondrial genome sequence (Leblanc et al., 1995), transcriptomic data (Collén et al., 2006; Collén et al., 2007), population ecology (Wang et al., 2008; Krueger-Hadfield et al., 2011), stress metabolism (Collén and Davison, 1999), pathogen interactions (Bouarb et al., 1999) and the effect of UV-radiation (Kräbs et al., 2004). Part of this project will involve attempting to biochemically characterise the starch synthase of this model species in order to shed light on the evolutionary development of starch metabolism.

### 2.1.2 Starch synthase structure

All glycogen and starch synthases belong to the glycosyltransferase superfamily of enzymes (EC 2.4.x.y). More specifically, they are classified as retaining-type glycosyltransferases with a GT-B fold. Proteins with a GT-B fold consist of two similar Rossmann-like  $\alpha$ - $\beta$ - $\alpha$  domains which fold together to form a deep cleft containing an ADP-binding pocket (Coutinho et al., 2003). In this way, binding of the donor and acceptor molecules takes place simultaneously and the binding sites are brought in close proximity to form the active site where transfer of the glucosyl moiety from ADP- or UDP-glucose to existing glucan chains occurs (Bowles et al., 2005; Sheng et al., 2009). These crystal structures of the inactive ‘open’ conformation as well as the catalytically active ‘closed’ conformation upon binding of ADP-glucose have been observed in *E. coli* glycogen synthase (Sheng et al., 2009).

Bacterial glycogen synthases share 30 – 36% protein sequence identity with plant starch synthases, and display similarities in overall structure and topology (Yep et al., 2004). They exhibit a highly conserved catalytic domain spanning a 60 kDa region which, in bacterial glycogen synthases comprise majority of the enzyme, while only forming part of plant and algal starch synthases at the C-terminus (Letierrier et al., 2008; Liu et al., 2015). Predictions of structural relationships between the catalytic domains of plants (rice, barley, potato) have shown significant similarities to that of *Agrobacterium tumefaciens* and *E. coli* glycogen synthases (Buschiazzo et al., 2004; Sheng et al., 2009; Momma and Fujimoto, 2012; Cuesta-Seijo et al., 2013; Nazarian-Firouzabadi and Visser, 2017). This catalytic domain consists of a glycosyltransferase 5 (GT5) and 1 (GT1) domain, the latter of which is common among enzymes belonging to the GT-B family and is, therefore, suspected to be involved in nucleotide binding (Nazarian-Firouzabadi and Visser, 2017).

The N-terminal region of the GT5 domain contains a KXGGL motif which is highly conserved among different starch and glycogen synthases. It has been reported that this motif plays a direct role in substrate binding in the *E. coli* glycogen synthase as the lysine residue has been demonstrated to be a site of ADP/ADP-glucose binding, possibly through ionic interactions with the phosphate moiety (Furukawa et al., 1990; Furukawa et al., 1993). The two glycine residues have also been demonstrated to be critical for enzyme activity (Yep et al., 2004; Yep et al., 2006). In bacterial glycogen synthases and most higher plant starch synthases the motif has a threonine amino acid at the second position (KTGGL) (Figure 3). There are some exceptions such as Arabidopsis SSIII and SSIV where there is an amino acid substitution to valine (KVGGL) and SSI from barley where the second amino acid is substituted to serine (KSGGL) (Figure 3). A KVGGL motif is conserved in SSIII of both monocots and dicots (rice, barley, sorghum, *Brachypodium distachyon*, wheat, Arabidopsis, soybean), where it has been suggested that the valine residue plays a role in the preference of SSIII to produce longer glucan chains than SSI or SSII (Mishra et al., 2017). The variable amino acid at the second residue may thus determine preference for glucan primers of specific lengths. Of these, threonine and serine share similar chemical properties as both are polar neutral amino acids, with valine being chemically distinct as it is a non-polar aliphatic amino acid. These variations in amino acids (particularly those with different chemical properties) could contribute to differences in catalytic activity with regards to efficiency of substrate binding and preferential chain elongation between synthases.

A second KXGGL ‘look-alike’ motif is present towards the C-terminal of the GT5 domain, but this is less well conserved and shows differences in the first and second amino acid positions between different starch synthase isoforms in higher plants. For example, in SSI there is a variation to GTGGL, in SSII to AVGGL, in SSIII and SSIV to KTGGL, and in GBSSI to STGGL (Figure 3). The bacterial glycogen synthases also show a variation to RTGGL. The similarities between the KXGGL and the look-alike motif suggests that they may perform similar functions and both be implicated in the binding of ADP- or UDP-glucose (Edwards et al., 1999a). Other conserved residues and motifs are found in glycogen and starch synthases, some of which are thought to facilitate binding by stabilising the interaction between substrate and enzyme (Imparl-Radosevich et al., 1999; Buschiazzo et al., 2004; Yep et al., 2006), and others that serve to maintain stability of the domains upon formation of the three-dimensional protein structure (Wang et al., 2014).

Athaliana_SSIII	HIAVEMAPIAKVGGGLGDVWTS	Athaliana_SSIII	VPVFKTGGLEDTVFDVDHDK
Athaliana_SSIV	HIAAEMAPVAKVGGGLGDVVAG	Athaliana_SSIV	IPIAFKTGGLENDVFDIDDDT
Ecoli_GS	HVCSEMFPLIKTGGGLADVIGA	Ecoli_GS	LPLVFKTGGLEADTVSDCSLEN
Atumefaciens_GS	SVSSEIYPLIKTGGGLADVGA	Atumefaciens_GS	IPVWARTGGLEADTVIDANHAA
Osativa_GBSSI	FVGAEMAPWSKTGGGLGDVLGG	Osativa_GBSSI	PCACASTGGLEADTVIEGKTGF
Athaliana_GBSSI	FIGAEVGPWSKTGGGLGDVLGG	Athaliana_GBSSI	VPIVASTGGLEADTVKDGYTGF
Stuberosum_GBSSI	FVGTEVGPWSKTGGGLGDVLGG	Stuberosum_GBSSI	VPICASTGGLEADTVKEGYTGF
Athaliana_SSII	LVAEACAPFSKTGGGLGDVAGA	Athaliana_SSII	IPVHAVGGLRDTVQQFDPYS
Athaliana_SSI	FVTSEAAPYSKTGGGLGDVCGS	Athaliana_SSI	IPVHGTGGLEADTVENFNPPYA
Hvulgare_SSI	FVTGEAAPYAKSGGGLGDVCGS	Hvulgare_SSI	VPVHGTGGLEADTVETFNPPFG

**Figure 3. Conserved KXGGL (left) and KXGGL look-alike (right) motifs present in glycogen and starch synthases from bacteria (*Escherichia coli* and *Agrobacterium tumefaciens*) and plants (*Arabidopsis thaliana*, *Oryza sativa*, *Hordeum vulgare* and *Solanum tuberosum*).** Amino acid sequences were obtained from the UniProtKB database and multiple sequence alignment was performed using Clustal Omega (Sievers et al., 2011). GBSS: granule-bound starch synthase, SS: starch synthase.

All soluble starch synthases of higher plants display a variable extension or ‘arm’ towards their N-terminal which accounts for some of the discrepancies in molecular weight observed between the different isoforms (Nazarian-Firouzabadi and Visser, 2017). It appears that these N-terminal extensions may play a role in protein-protein interactions, localisation of the synthase to particular regions of the plastid, or aid in the binding of starch. The occurrence of multimeric protein complexes of starch metabolic enzymes has been reported in both maize and wheat. During specific stages of wheat endosperm development, the formation of complexes between starch synthases and starch branching enzymes have been observed through co-immunoprecipitation studies (Tetlow et al., 2008). Protein complexes consisting of SBEIIb, SBEI and starch phosphorylase have also been reported to occur in wheat (Tetlow et al., 2004). In addition to complexes between starch synthases and starch branching enzymes, complexes between different starch synthases isoforms (SSI and SSIII) were also detected in maize (Hennen-Bierwagen et al., 2008). These could be important in altering the *in vivo* functions of starch enzymes, particularly as their interaction seems dependent on different developmental stages. The localisation of *Arabidopsis* SSIV to specific parts of the thylakoid membrane has recently been demonstrated to be facilitated by its N-terminal extension which interacts with the fibrillin proteins of plastoglobules associated with thylakoids (Gámez-Arjona et al., 2014; Raynaud et al., 2016). Attaching to the

thylakoid membrane could be important for granule initiation and regulation of the number of starch granules, functions that are attributed to SSIV. The N-terminal extension of Arabidopsis and kidney bean SSIII also displays sequence similarity to that of carbohydrate binding molecules. By analysing truncated forms of SSIII, it was demonstrated that the N-terminal extension conferred a higher affinity for the glucan acceptor substrates amylose and amylopectin (Senoura et al., 2007; Valdez et al., 2008). The varying functions of these N-terminal extensions, therefore, appear to contribute to the specific activities of the different soluble starch synthase isoforms.

GBSSI proteins all possess a highly conserved ~20 amino acid C-terminal extension which is likely involved in amylose synthesis. Truncated and chimeric forms of starch synthases constructed from potato GBSSI and SSII have provided valuable insight into the functions of such extensions. Many of the chimeric proteins were either inactive or only had low levels of activity, possibly due to misfolding, indicating that the shuffled parts failed to correctly interact with the remainder of the sequence (Edwards et al., 1999a). In the case of active chimeric and truncated proteins, there was no influence on catalytic activity upon removal of the N-terminal extension. It seems that only the catalytic domain (containing the KXGGL and KXGGL look-alike motifs) is responsible for activity, while the extension functions as a non-catalytic accessory. Even the specific activity of GBSSI can be attributed to the catalytic domain and the unique C-terminal extension present in this isoform.

In addition to their various functions and localizations the starch synthase isoforms exhibit different biochemical properties. For instance in potato, GBSSI has a higher  $K_m$  to ADP-glucose than soluble SS isoforms, meaning that amylose synthesis becomes reduced before amylopectin synthesis when ADP-glucose pools are depleted in the plastid (Edwards et al., 1999b). GBSSI displays a much lower affinity for the acceptor molecule amylopectin (Edwards et al., 1999a), but exhibits a higher affinity than SSII for the acceptor maltotriose (Denyer et al., 1999). This preference toward acceptor molecules is a reflection of the distinct roles that the different starch synthase isoforms play in elongation of glucan chains to a certain length (as mentioned previously).

Starch synthases can be classified as either processive or distributive in their mode of action. Processive enzymes bind to the polyglucan acceptor molecule and remain bound after the addition of each new glucosyl residue, rapidly generating long chains. In contrast, distributive enzymes release the acceptor molecule upon addition of a glucosyl residue, resulting in the elongation of many chains over a longer period of time. Few studies have examined the mode of action of starch synthases. One such study in pea revealed that SSII acts as a distributive synthase, while GBSSI displays a processive mode of action (Denyer et al., 1999). Interestingly, later experiments revealed that potato GBSSI was capable of shifting from a distributive mode (in the absence of amylopectin) to a processive mode of action with a much higher rate of reaction when amylopectin was present (Denyer et al., 1999; Edwards et al., 1999b).

It is clear that, despite their overall similarity in sequence and structure, there are many differences between starch synthases with regards to specific activity, substrate affinity, mode of action and interactions with other molecules via N-terminal extensions. Access to the sequences of red algal starch synthases provides the

opportunity to analyse this enzyme by a number of methodologies. Experiments similar to that of Edwards et al., (1999a) which construct truncated and chimeric proteins from red algal and higher plant starch synthases could aid in identifying regions of the synthases that are essential for substrate specificity of either ADP- or UDP-glucose. In addition, it could enable engineering of a starch synthase with altered biochemical properties that could be useful for synthesising starch with improved industrial properties.

## 2.2 Aim and objectives

The aim of this chapter is to examine the function of a starch synthase from the red alga *C. crispus*.

- Objective 1: The purification of recombinant protein with the same sequence as *C. crispus* starch synthase.
- Objective 2. Examination of the kinetics of this red algal starch synthase through biochemical characterisation using enzyme-based activity assays.
- Objective 3: The substrate specificity and domain functions of the red algal starch synthase will be examined by altering the peptide sequence using site-directed mutagenesis to target important residues.

## 2.3 Materials and Methods

### 2.3.1 Microbial strains and vectors

**Table 3. Strains and vectors used during this study.**

Strain	Genotype	Source
<i>E. coli</i>		
DH5 $\alpha$	F <sup>-</sup> <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> $\phi$ 80 <i>dlacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> )U169, <i>hsdR17</i> ( <i>rK</i> <sup>-</sup> <i>mK</i> <sup>+</sup> ), $\lambda$ <sup>-</sup>	Thermo Fisher
TOP10	F <sup>-</sup> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 araD139</i> $\Delta$ ( <i>araleu</i> )-7697 <i>galU galK rpsL</i> (Str <sup>R</sup> ) <i>endA1 nupG</i>	Invitrogen
BL21CodonPlus(DE3)-RIPL	B F <sup>-</sup> <i>ompT hsdS</i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) <i>dcm</i> <sup>+</sup> Tet <sup>R</sup> <i>gal</i> $\lambda$ (DE3) <i>endA Hte</i> ( <i>argU proL Cam</i> <sup>r</sup> ) ( <i>argU ileY leuW Str/Spec</i> <sup>R</sup> )	Agilent Technologies
BL21AI	B F <sup>-</sup> <i>ompT gal dcm lon hsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) ( <i>malB</i> <sup>+</sup> ) <sub>K-12</sub> ( $\lambda$ <sup>S</sup> ) <i>araB::T7RNAP-tetA</i>	Invitrogen
BL21(DE3) $\Delta$ <i>glgCAP</i> <sup>1</sup>	B F <sup>-</sup> <i>dcm ompT hsdS</i> ( <i>r<sub>b</sub><sup>-</sup>m<sub>b</sub><sup>-</sup></i> ) <i>gal</i> $\lambda$ (DE3) $\Delta$ <i>glgCAP::kan</i>	(Moran-Zorzano et al., 2007)
<b>Bacterial vectors:</b> pRSET-A (Invitrogen), pG-KJE8 (Takara Bio Inc), pENTR <sup>TM</sup> /D-TOPO <sup>®</sup> (Invitrogen), pMDC32 and pMDC85 (Curtis and Grossniklaus, 2003).		
<i>S. cerevisiae</i>		
$\Delta$ <i>gsy2</i> <sup>2</sup>	MATa MAL2-8C SUC2 <i>gsy2</i> $\Delta$ <i>ura3-52</i>	(Farkas et al., 1991)
<b>Yeast vector:</b> YEpENO-BBH <sup>3</sup>		

### 2.3.2 Growth media

Lysogeny Broth (LB) was prepared with 1.0% (w/v) peptone, 0.5% (w/v) yeast extract and 1.0% (w/v) NaCl. Antibiotics were prepared as 1000-fold stock solutions, filter-sterilised (0.2  $\mu$ m) and added to cooled media just before use. Synthetic complete media lacking uracil (SC-ura) was prepared using 0.67% (w/v) yeast nitrogen base without amino acids (Sigma-Aldrich), 0.192% (w/v) yeast synthetic drop-out medium supplements without uracil (Sigma-Aldrich) and 2% (w/v) D-glucose. Yeast extract-peptone-dextrose (YPD) liquid media was prepared using 1% (w/v) yeast extract, 2% (w/v) peptone and 2% (w/v) D-glucose. YPD with sorbitol (YPDS) liquid media was prepared as YPD with the addition of 1 M D-sorbitol. Glucose was prepared as a 20% (w/v) stock solution, filter-sterilised (0.2  $\mu$ m) and added to cooled solutions after

<sup>1</sup> Kindly provided by Dr Christophe D'Hulst (Lille University, France).

<sup>2</sup> Obtained from the Institute for Wine Biotechnology (Department of Viticulture and Oenology, Stellenbosch University).

<sup>3</sup> Kindly provided by Prof Emile van Zyl (Department of Microbiology, Stellenbosch University).

autoclaving. The pH of SC-ura and YPD was adjusted to 6 using KOH. For solid media, 1.5% (w/v) bacteriological agar was added to LB and 2% (w/v) bacteriological agar to YPD or SC-ura media before autoclaving.

### 2.3.3 Sequence analysis

Multiple sequence alignments were performed using MEGA-X (Kumar et al., 2018). Protein size was predicted using ProtParam (Gasteiger et al., 2005).

### 2.3.4 Gene synthesis

The *C. crispus* starch synthase gene was synthesised by Epoch Life Science Inc (Texas, USA) based on the available NCBI reference protein sequence (XP\_005718355) and optimised for expression in *E. coli* (see Appendix A: for the optimised nucleotide sequence).

### 2.3.5 PCR amplification

For screening purposes, reactions contained 1X Green GoTaq® Reaction Buffer (Promega), 1.25 U GoTaq® DNA Polymerase (Promega), 0.2 mM dNTP mix, 0.1 µM of each primer and 25–50 ng DNA in a 50 µL total reaction volume. For construct preparation Q5 High Fidelity DNA polymerase (New England Biolabs) was used. Each 50 µL reaction contained 1X Q5® Reaction Buffer (New England BioLabs), 1 U Q5® High-Fidelity DNA Polymerase (New England BioLabs), 0.5 µM of each primer and 25–50 ng DNA. All oligonucleotides were synthesized by Inqaba Biotechnical Industries (Pty) Ltd. (Pretoria, South Africa).

### 2.3.6 Plasmid isolation

Unless otherwise stated, plasmid isolations were performed on 5 mL overnight *E. coli* cultures (LB containing the appropriate antibiotics, 37°C) using the Wizard® Plus SV Minipreps DNA Purification System (Promega). Plasmid isolation from 10 mL *S. cerevisiae* overnight cultures (YPD, 30°C) was preceded by a cell-disruption step which consisted of vigorous vortexing for 5 min after the addition of acid-washed glass beads (425–600 µm diameter, Sigma-Aldrich).

### 2.3.7 Restriction digests

All restriction enzymes and buffers used were obtained from New England BioLabs. Digest reaction components, incubation time and heat inactivation steps were determined using the online NEBcloner® tool (v1.3.3, <http://www.nebcloner.neb.com/#!/redigest>). A 50 µL digest reaction contained 0.5–1 µg DNA, 1 X restriction buffer and 10 U of restriction enzyme. Reactions were incubated for a minimum of 2 hrs at the recommended temperature. Heat inactivation steps were performed at 65°C or 80°C for 20 min.

### 2.3.8 DNA isolation from agarose gels

Restriction digests, DNA and PCR products were electrophoresed on 1% (w/v) agarose gels containing 0.5 µg/mL ethidium bromide at 100 V for 60–90 min. DNA was purified from agarose using the Wizard® SV Gel and PCR Clean-Up System (Promega).



### 2.3.9 Site-directed mutagenesis

Primers were designed using the NEBaseChanger™ v1.2.7 (New England BioLabs®, <https://nebasechanger.neb.com/>) online design software. Site-directed mutagenesis (SDM) was performed using the Q5® Site-Directed Mutagenesis Kit (NEB) according to the manufacturer's instructions.

### 2.3.10 DNA sequencing

Sequencing was performed by the Central Analytical Facility (CAF) at Stellenbosch University.

### 2.3.11 Bacterial transformation

Chemically-competent *E. coli* cells were prepared as described by Inoue et al. (1990) and transformed via a standard heat-shock method (Sambrook and Russel, 2001). Selection for transformants occurred overnight at 37°C on LB containing the appropriate antibiotics.

### 2.3.12 Recombinant protein synthesis

Colonies harbouring plasmids designed for protein expression were grown in an overnight culture (5 mL LB containing the appropriate antibiotics, 37°C) before being used to inoculate 100 mL of LB (containing antibiotic reduced to half strength) and incubated (37°C) until the culture reached the desired OD<sub>600</sub>. Strains containing a vector without insert were used as a control. Protein expression was induced by the addition of IPTG (final concentration 0.5mM) or arabinose (final concentration 0.2% (w/v)) depending on bacterial strain. The culture was then placed at either 4, 25 or 37°C with shaking. Two millilitre aliquots of each culture were removed at 0, 4, 6 and 20 hours post induction and pelleted by centrifugation at 16 000 x g, 4°C for 5 min before being stored at -20°C until use. After 20 hours growth the remainder of the 100 mL culture was pelleted and stored.

### 2.3.13 Yeast transformation and protein expression

Electro-competent cells were prepared and transformed following a modified protocol by Becker and Lundblad (1994) and Cho et al (1999). A 5 mL overnight culture (YPD, 30°C) was pelleted by centrifuging at 1500 x g for 2 min. The pellet was washed twice with dH<sub>2</sub>O and then gently suspended in 640 µL of dH<sub>2</sub>O. After the addition of 80 µL of 10X TE buffer (100 mM Tris-HCl, 10 mM EDTA, pH 7.5) and 80 µL of 1 M LiAc (pH 7.5), the cells were incubated at 30°C for 45 min with gentle agitation. Next, 20 µL of 1M DTT was added and the cells incubated at 30°C for a further 15 min. The cells were washed once with cold dH<sub>2</sub>O and once with cold Electroporation Buffer (1 M sorbitol, 20 mM HEPES), before being suspended in 250 µL Electroporation Buffer. Approximately 10 µg of linearised vector DNA and insert DNA was added to 50 µL of electro-competent cells in a chilled 0.2 cm electroporation cuvette. The sample was pulsed (1.4 kV, 200 Ω, 25 µF) and 1 mL YPDS was immediately added. The cells were placed back on ice for 5 min, before being incubated (30°C) for 1 hr. Transformants were selected on SC-ura agar plates (30°C, 2 – 3 days) and screened for the correct vector via PCR. Overnight cultures (5 mL SC-ura, 30°C) of confirmed transformants were used to inoculate 200 mL SC-ura and incubated (30°C with shaking, 16–18 hrs) to allow expression of the recombinant

protein. Cells were harvested by centrifugation (1 500 x g, 4°C, 5 min) and washed once with dH<sub>2</sub>O before being stored at -20°C until protein extraction.

### 2.3.14 Protein extraction

Protein was extracted from bacterial cultures under native conditions as described by the *Purification of His-tag Proteins: User Manual* (Macherey-Nagel, 2014). Bacterial cell pellets were thawed on ice and suspended in 2 mL Lysis-Equilibration-Wash (LEW) Buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0). The cells were lysed with 1 mg/mL lysozyme on ice for 30 min, and then sonicated on ice in 10 x 15 s bursts with a 15 s cooling period between each burst. The resulting crude lysate was centrifuged at 10 000 x g for 30 min to separate the cellular debris and insoluble protein fraction from the soluble protein fraction. The cleared supernatant (soluble protein fraction) was aliquoted to a clean tube and the remaining pellet (insoluble protein fraction) was suspended in 2 mL LEW Buffer. To isolate proteins from inclusion bodies under denaturing conditions, cell pellets were treated as before. After separation of the protein fractions the pellet was washed twice with LEW Buffer before being suspended in 2 mL Denaturing Solubilisation Buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 8 M urea, pH 8.0) and stirred on ice for 1 hr to dissolve inclusion bodies. The sample was centrifuged at (10 000 x g, 30 min, 20°C) to remove cellular debris and the supernatant carefully transferred to a clean tube.

Protein was extracted from *S. cerevisiae* cultures (SC-ura, 30°C, overnight) using a modified protocol from Dunn and Wobbe, (1993). Cell pellets were slowly thawed on ice and suspended in 3 mL glass bead disruption buffer (20 mM Tris-HCl pH 7.9, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 5% glycerol, 1 mM DTT, 0.3 M ammonium sulfate, 1 X Protease Inhibitor Cocktail (Sigma-Aldrich), 1 mM PMSF) before the addition of 80 µL of isoamylalcohol. Four volumes of chilled, acid-washed glass beads (425–600 µm diameter, Sigma-Aldrich) were added and the suspension was vortexed at maximum speed in 5 x 30 s bursts with 15 s cooling periods on ice between each burst. The supernatant was decanted and centrifuged (12 000 x g, 60 min, 4°C) to remove cellular debris. All protein samples were stored at 4°C until SDS-PAGE analysis that same day.

### 2.3.15 Protein purification

Recombinant his-tagged protein was purified from crude protein extracts using Protino<sup>®</sup> Ni-TED Packed Columns (Macherey-Nagel) according to the manufacturer's instructions. The flow-through from the various protein fractions was collected in microfuge tubes and stored at 4°C until use for SDS-PAGE and native PAGE analysis.

### 2.3.16 Refolding denatured protein

A modified protocol of Lemerrier et al. (2003) was followed for on-column refolding of recombinant protein. After immobilisation and washing of protein on a Protino<sup>®</sup> Ni-TED Packed Column (Macherey-Nagel) according to the manufacturer's instructions, the column was sequentially washed with Denaturing Solubilisation Buffer (section 2.3.14) containing a decreasing concentration gradient of urea (6 M, 5 M, 4 M, 3 M, 2 M, 1 M) with 5 mL of buffer per wash. This was followed by a final wash step using 5 mL LEW Buffer (section 2.3.14), and then elution of the protein using 2 mL Elution Buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl,

250 mM imidazole, pH 8.0). The different fractions of flow-through from the column were collected and stored at 4°C until further analysis.

For protein refolding by step-wise dialysis, a protocol by Yang et al. (2011) was adapted. Large cell pellets (100 mL culture) were washed twice with Wash Buffer I (WBI: 20 mM Tris-HCl, pH 8.0), suspended in 5 mL Lysis Buffer (50 mM Tris-HCl, 0.25 mg/mL lysozyme, pH 8.0) and sonicated on ice as before. The sample was centrifuged (30 000 x g, 4°C, 45 min) to pellet inclusion bodies and washed twice with Wash Buffer II (WBII: 50 mM Tris-HCl, 50 mM NaCl, 2% Triton X-100, 1.5 mM  $\beta$ -mercaptoethanol, 1.6 M urea, pH 8.0) and twice with WBI. The purified inclusion bodies were pelleted by centrifugation (30 000 x g, 4°C, 30 min) and dissolved in 5 mL Inclusion Body Solubilization Reagent (ThermoFisher) according to the manufacturer's instructions. This protein extract was diluted 40-fold into 50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 50 mM NaCl, 10 mM  $\beta$ -mercaptoethanol and centrifuged at 30 000 x g, 4°C for 45 min. The precipitated protein was suspended in 5 mL Extraction Buffer (50 mM Tris-HCl, 50 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 8 M urea, pH 8.0) and transferred to dialysis tubing (Sigma). The sample was dialysed at 4°C against a 40-fold volume of Dialysis Buffer (40 mM Tris-HCl, 150 mM NaCl, 1 mM ZnCl<sub>2</sub>, 10 mM  $\beta$ -mercaptoethanol, 100 mM L-arginine, pH 8.0), which was changed every 3 hrs and contained decreasing concentrations of urea (6 M, 4 M, 2 M, 0 M).

### 2.3.17 SDS- and native PAGE

Protein samples were prepared for SDS-PAGE by the addition of 1 volume sample buffer (2X) (4% (w/v) SDS, 20% (w/v) glycerol, 120 mM Tris-HCl (pH 6.8), 0.02% (w/v) bromophenol blue) and heating at 95°C for 5 min, followed by centrifugation at 16 000 x g, 4°C for 10 min. All samples were electrophoresed on an 8% (w/v) polyacrylamide gel at 120 V for 90 min. Gels were stained with Coomassie staining solution [(0.25% (w/v) Coomassie Brilliant Blue R-250, 45% (v/v) methanol, 10% (v/v) acetic acid) at 37°C for 2 hrs and destained overnight at room temperature in 45% (v/v) methanol, 10% (v/v) acetic acid].

For native PAGE, protein samples were prepared by the addition of 6X loading buffer (50% (v/v) glycerol, 0.02% (w/v) bromophenol blue). Glycogen/starch synthase activity was assessed using discontinuous native PAGE on 8% (w/v) polyacrylamide gels containing either 0.1% (w/v) amylopectin or glycogen. After electrophoresis for 2 hrs at 4°C, the gel was incubated overnight at room temperature with gentle agitation in activity buffer (50 mM Tricine-NaOH, 0.5 M Na-citrate, 2 mM EDTA, 1 mM either ADP-glucose or UDP-glucose, 2 mM DTT, pH 8.0). Activities were visualised through incubation in Lugol's solution (10% (w/v) potassium iodide, 5% (w/v) iodine).

### 2.3.18 Plate-based screening for glycogen accumulation

A 5  $\mu$ L aliquot of *S. cerevisiae* overnight culture was spotted onto solid SC-ura media. The resulting colonies were stained with iodine crystal vapour by inverting plates over iodine crystals for 1 min.

## 2.4 Results

### 2.4.1 Sequence analysis

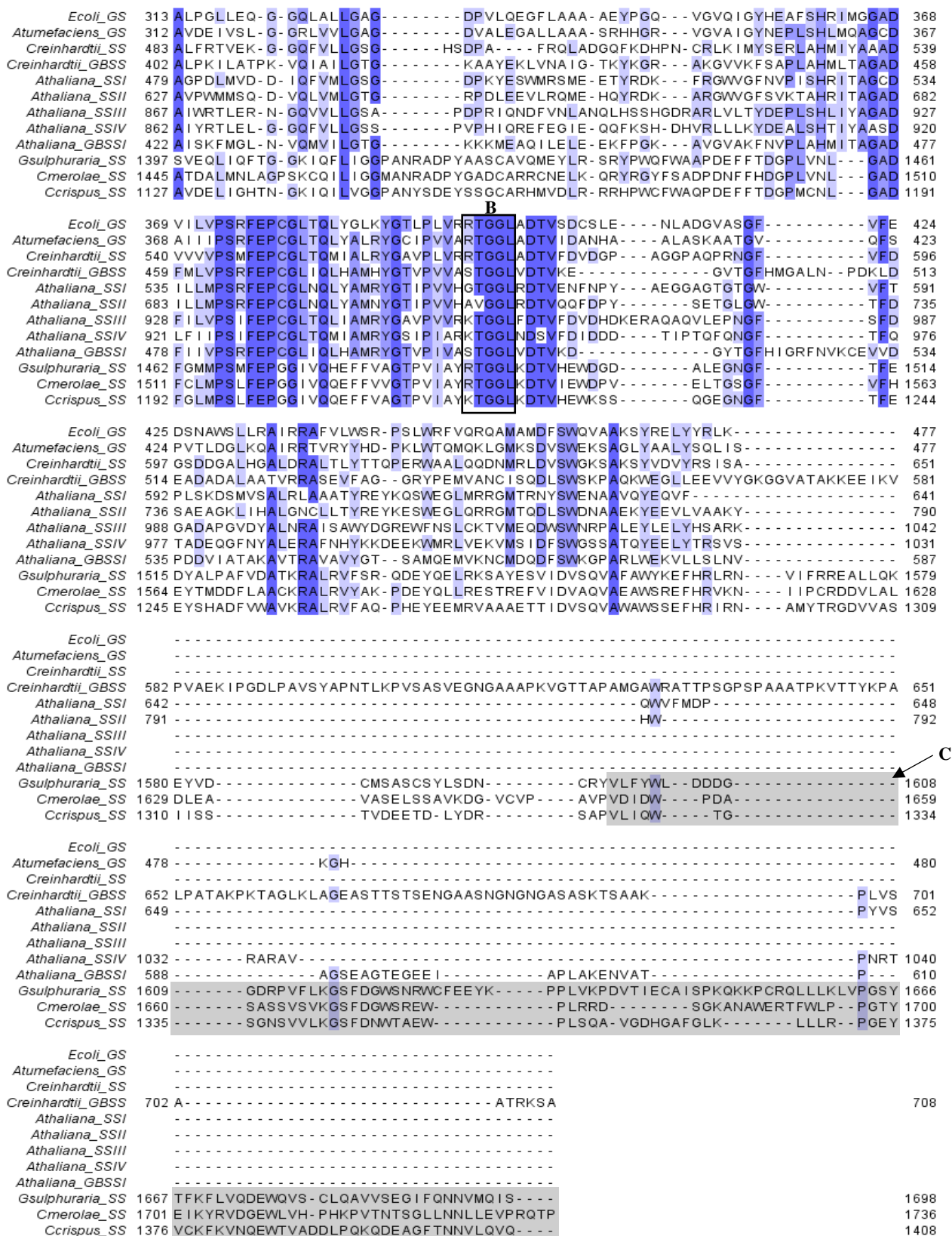
The *C. crispus* starch synthase is predicted to have a molecular weight of 157.2 kDa (as determined by ProtParam) and lacks a chloroplast transit peptide (as determined by ChloroP). The protein sequence (XP\_005718355) was aligned to those of other algal, bacterial and higher plant starch/glycogen synthases to examine similarities and differences between them (see Appendix C: for full alignment). The red algal starch synthases share many of the features present in green algae and higher plants, including the core catalytic domain containing the KXGGL and KXGGL look-alike motifs, as well as the N-terminal arm extension.

The majority of the sequence similarity between the starch synthases lies in the catalytic domain although there is a noticeable variation in the KXGGL motif towards the N-terminus of the catalytic domain which, in the red algal species, shows an amino acid substitution to SIGGL (*C. merolae*) or TAGGL (*C. crispus* and *G. sulphuraria*) (Figure 4). Serine and threonine are hydroxylic amino acids while the second position is occupied by the aliphatic amino acids isoleucine or alanine. In the bacterial glycogen synthases and green plant starch synthases used in this study the motif is KTGGL, with the exception of Arabidopsis SSIII and SSIV and *C. reinhardtii* SS (KVGGL) (Figure 4). In this case, the basic lysine residue is conserved in the first position and the second position varies between a hydroxylic threonine and aliphatic valine amino acid.

The KXGGL look-alike motif towards the C-terminal of the catalytic domain appears to harbour less conservation between species and more similarities within phylogenetic groups. In *G. sulphuraria* and *C. merolae*, the look-alike motif is similar to that of the bacterial glycogen synthases and *C. reinhardtii* soluble SS (RTGGL), while in *C. crispus* it resembles that of Arabidopsis SSIII and SSIV (KTGGL) (Figure 4). Both groups thus have a basic residue (arginine and lysine) followed by a hydroxylic residue (threonine). The Arabidopsis and *C. reinhardtii* GBSS both have a hydroxylic residue at the first position (STGGL) and Arabidopsis SSI has an aliphatic residue (GTGGL) (Figure 4). Arabidopsis SSII is the only protein that shows variation at the second position and has two aliphatic amino acid residues (AVGGL) (Figure 4).

Some of the starch synthases, such as the *C. reinhardtii* GBSS exhibit C-terminal extensions (Figure 4). In the case of red algal starch synthases this extension demonstrates similarity to known glycogen binding domain sequences from the AMP-activated protein kinase  $\beta$ -subunit.



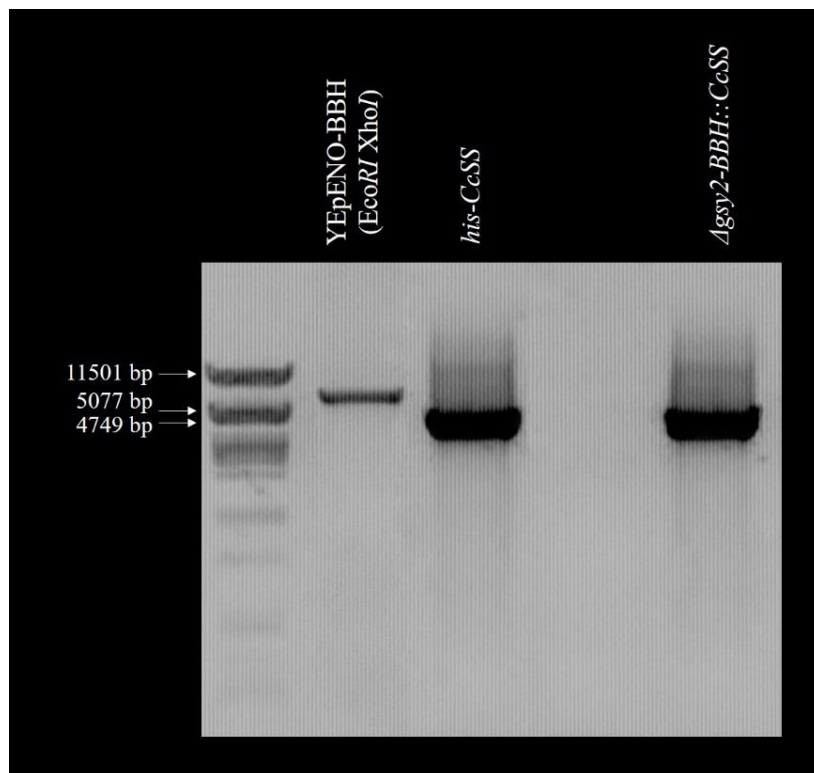


**Figure 4. Multiple sequence alignment of the catalytic domain of glycogen and starch synthases from bacteria (*Escherichia coli* and *Agrobacterium tumefaciens*), red algae (*Galdieria sulphuraria*, *Cyanidioschyzon merolae* and *Chondrus crispus*), and higher plants (*Arabidopsis thaliana*). Colouring indicates percentage identity between sequences. The conserved KXGGL (A) and KXGGL look-alike (B) motifs and the glycogen binding domain (C) are outlined. GS: glycogen synthase, SS: starch synthase, GBSS: granule-bound starch synthase.**

## 2.4.2 Construction of vectors

A gene encoding an identical protein to the starch synthase from *C. crispus* was synthesised to be optimal for expression in *E. coli* and ligated into the pRSET-A bacterial expression vector to produce the pRSET-A::CcSS construct. It was ligated in-frame downstream of a poly-histidine tag to allow for purification using affinity chromatography.

The CcSS gene (including the N-terminal 5'-histidine-tag) was amplified from the pRSET-A::CcSS construct with primers that add a 27-30 bp region flanking either side of the gene and which is homologous to the promoter or terminator sequences present in the YEpENO-BBH yeast expression vector (Figure 5). This was carried out to allow for homologous recombination by the endogenous *S. cerevisiae* system during transformation. The YEpENO-BBH vector was linearised using *EcoRI* and *XhoI* (Figure 5), and both the linearised vector and PCR product was purified before being co-transformed into the *S. cerevisiae*  $\Delta$ gsy2 mutant. Plasmid isolated from  $\Delta$ gsy2 colonies were screened for the presence of the CcSS gene using PCR with primers that bind within the gene (Figure 5).



**Figure 5. Construction of the YEpENO-BBH::CcSS vector and transformation into the *S. cerevisiae*  $\Delta$ gsy2 mutant.** The linearised YEpENO-BBH vector and an amplicon of the CcSS gene and upstream histidine tag (*his-CcSS*) containing flanking regions of homology to the vector were co-transformed into *S. cerevisiae* and ligated through homologous recombination. Plasmid from the resulting colonies were screened for the presence of the insert ( $\Delta$ gsy2 YEpENO-BBH::CcSS).

### 2.4.3 Site-directed mutagenesis

A literature search was performed to find conserved residues or motifs within all known starch and glycogen synthases that are important for either activity or substrate specificity of the enzyme. The position of the corresponding conserved residues in the *C. crispus* SS were identified using the protein sequence alignment in Figure 4. Residues that were targeted for site-directed mutagenesis in the red algal starch synthase were identified and selected based on conserved residues found in other studies (Table 4).

**Table 4. Conserved residues or motifs in glycogen and starch synthases.** The functions of residues in *A. tumefaciens* (*Agt*) and *E. coli* were determined in other studies (highlighted in bold), and the corresponding residues were identified in *Arabidopsis* (*At*) and three red algal species (*G. sulphuraria*, *C. merolae* and *C. crispus*) through multiple sequence alignment.

<i>AgtGS</i>	<i>EcGS</i>	<i>AtSSIII</i>	<i>AtGBSS</i>	<i>GsSS</i>	<i>CmSS</i>	<i>CcSS</i>	Function
Glu376	<b>Glu377</b>	Glu936	Glu486	Glu1470	Glu1519	Glu1200	Essential for enzyme activity (Yep et al., 2004)
<b>Tyr354</b>	<b>Tyr355</b>	Tyr914	Phe464	Thr1451	His1500	Thr1181	Interacts with adenine ring of ADP-glucose (Buschiazzo et al., 2004; Yep et al., 2006)
<b>Lys15</b> <b>Thr16</b>	<b>Lys15</b> <b>Thr16</b>	Lys608 Val609	Lys98 Thr99	Thr1096 Ala1097	Ser1146 Ile1147	Thr829 Ala830	KXGGL motif which interacts with ADP-glucose and stabilises the substrate (Furukawa et al., 1990; Furukawa et al., 1993; Buschiazzo et al., 2004)

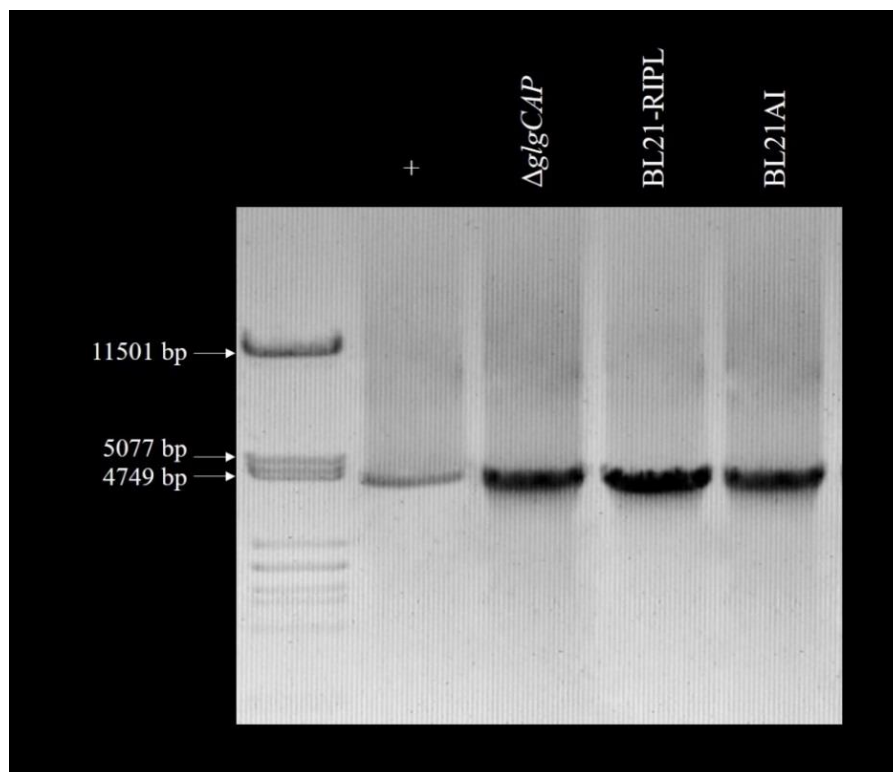
In *E. coli*, a change of the Glu377 residue to a Cys377 residue is known to inactivate the enzyme (Yep et al., 2004). The corresponding Glu1200 residue in *CcSS* was targeted and substituted to Cys1200 to generate an inactive mutant enzyme (E1200C) that could act as negative control during biochemical analyses. To examine possible sites of substrate specificity in the enzyme the conserved KXGGL motif was targeted. This motif is suspected to play a role in the catalytic mechanism by interacting with and stabilising binding of the sugar nucleotide substrate (ADP- or UDP-glucose) (Furukawa et al., 1990; Furukawa et al., 1993; Buschiazzo et al., 2004). The first two residues of the KXGGL motif (or TAGGL in *C. crispus*) were mutagenized to Lys829 and Thr830 respectively, to generate a mutant enzyme (TA829-30KT) that resembles the bacterial and green plant glycogen/starch synthases (KTGGL) which preferentially utilise ADP-glucose. In *A. tumefaciens* glycogen synthase the Tyr354 residue has been demonstrated to interact with the adenine ring of ADP-glucose (Buschiazzo et al., 2004). Furthermore, in all glycogen and starch synthases which preferentially utilise ADP-glucose, this position is occupied by aromatic amino acids (Tyr or Phe) while promiscuous enzymes or UDP-glucose specific enzymes show no particular trend (Buschiazzo et al., 2004). As none of the red algal starch synthases exhibit an aromatic amino acid at this position, the corresponding Thr1191 residue in *CcSS* was mutagenized to a Tyr residue (T1181Y) to further examine possible sites of substrate specificity.

Back-to-back primers were designed to implement the necessary base pair changes and the pRSETA::*CcSS* construct was used as template DNA during PCR amplification. The template DNA was degraded, and the PCR product ligated to form the mutagenized vectors which were transformed into NEB 5-alpha competent cells. Plasmid was isolated from transformants and sequenced to confirm the appropriate base changes.



#### 2.4.4 *E. coli* protein expression

Several *E. coli* strains were used during this study. The BL21(DE3)  $\Delta$ *glgCAP* strain (Moran-Zorzano et al., 2007), which lacks the whole glycogen biosynthetic machinery, was used to avoid background glycogen synthase activity during starch synthase activity assays. The BL21CodonPlus(DE3)-RIPL strain was used to optimise protein expression as it contains extra copies of tRNA genes which aid in expression of recombinant proteins from AT- or GC-rich genomes. Lastly, the BL21AI strain was used to allow tighter regulation of protein expression through the *ara* operon, as opposed to the ‘leaky’ promoter of the *lac* operon (present in DE3 strains). Strains were transformed with the pRSET-A::*CcSS* construct and resulting colonies were screened for the correct vector using colony PCR (Figure 6).

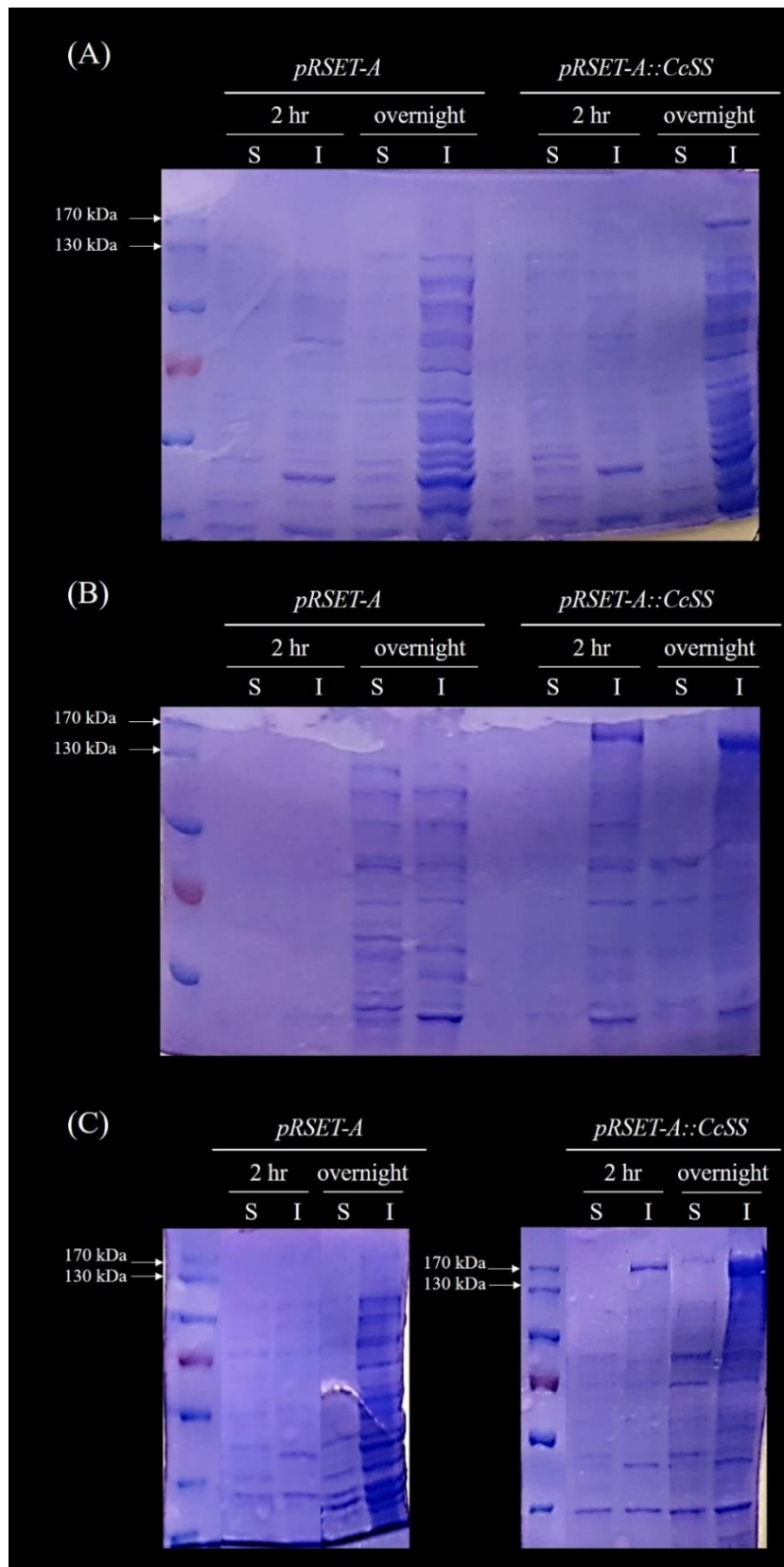


**Figure 6. Screening of *E. coli* colonies for the presence of the *C. crispus* starch synthase gene.** The pRSET-A::*CcSS* construct was transformed into three different *E. coli* strains (BL21(DE3)- $\Delta$ *glgCAP*, BL21CodonPlus(DE3)-RIPL and BL21AI). The resulting colonies were screened through PCR using primers that bind within the *CcSS* gene. The initial construct was used as positive control (+).

Protein expression was first attempted with the BL21(DE3)  $\Delta$ *glgCAP* strain at 37°C. A vector without insert (pRSET-A) was used as a negative control. After 2 hours of expression, no recombinant protein could be detected during SDS-PAGE analysis of the combined soluble and insoluble protein fractions (Figure 7). However, after allowing expression to continue overnight a protein similar in size to the predicted molecular weight of the *CcSS* (157.2 kDa) was observed, which was absent in the control. It was, however, only observed in the insoluble protein fraction. As insoluble protein indicates misfolding and subsequent aggregation in

bacterial inclusion bodies, the use of two other *E. coli* strains optimised for protein expression was implemented (BL21CodonPlus(DE3)-RIPL and BL21AI). Expression in these two strains showed a similar trend to the BL21(DE3)  $\Delta glgCAP$  strain. In both cases, the recombinant protein was visible in the insoluble fraction after 2 hrs of expression, as well as after overnight expression, but was not observed to be present in the soluble fraction (Figure 7). Following this, the use of chaperone proteins was employed to aid in correct folding of the recombinant protein. Each of the three *E. coli* strains were co-transformed with the pRSET-A::CcSS construct and the pG-KJE8 chaperone plasmid which expresses several chaperone proteins. The expression experiments were repeated but again, the recombinant protein accumulated in the insoluble fraction and no soluble protein was visible (results not shown).

The expression conditions were then altered to attempt expression of the recombinant protein in the soluble fraction. Both the  $OD_{600}$  of the culture before induction and the temperature at which expression took place were lowered to promote slower bacterial growth and limit excessive protein production (San-Miguel et al., 2013). The experiment was repeated using all three strains with induction occurring once cultures reached an  $OD_{600}$  of 0.6 (as before) while the induction temperature was adjusted to 4, 25 and 37°C respectively. Cultures of each strain that had reached an  $OD_{600}$  of 0.1 were also induced and incubated at the three different temperatures. However, none of the alterations yielded recombinant protein in the soluble fraction (results not shown).

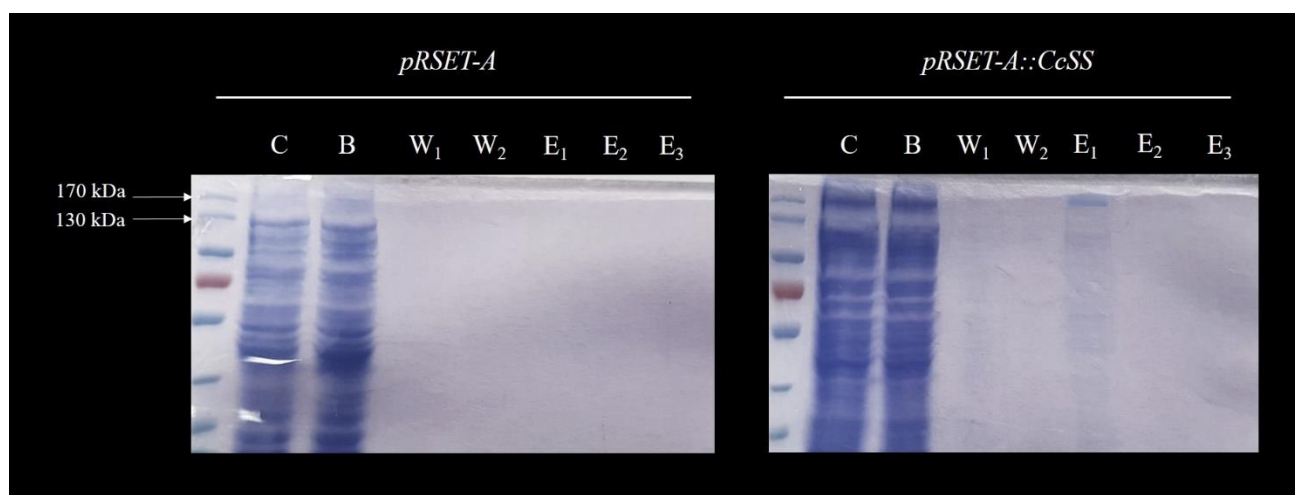


**Figure 7. Expression of recombinant CcSS protein in three *E. coli* strains:** (A) BL21(DE3)  $\Delta$ *glgCAP*, (B) BL21CodonPlus(DE3)-RIPL and (C) BL21AI. Strains carrying the *pRSET-A::CcSS* plasmid were grown in LB medium at 37°C and expression was induced at mid-log phase ( $OD_{600}$  0.6). The soluble (S) and insoluble (I) fractions of samples harvested 2 hrs post-induction as well as after overnight induction were resolved using SDS-PAGE. The recombinant protein was observed between 130 – 170 kDa only in the insoluble fraction of all strains.

### 2.4.5 Refolding of recombinant protein

As expression of the *C. crispus* SS in *E. coli* resulted in insoluble protein that accumulated in inclusion bodies, the protein was extracted and refolding was attempted using two different methods to try and recover starch synthase activity. Protein was expressed in BL21(DE3)*AggCAP* cells harbouring the pRSETA::*CcSS* construct (37°C, 20 hrs) and purified from inclusion bodies under denaturing conditions.

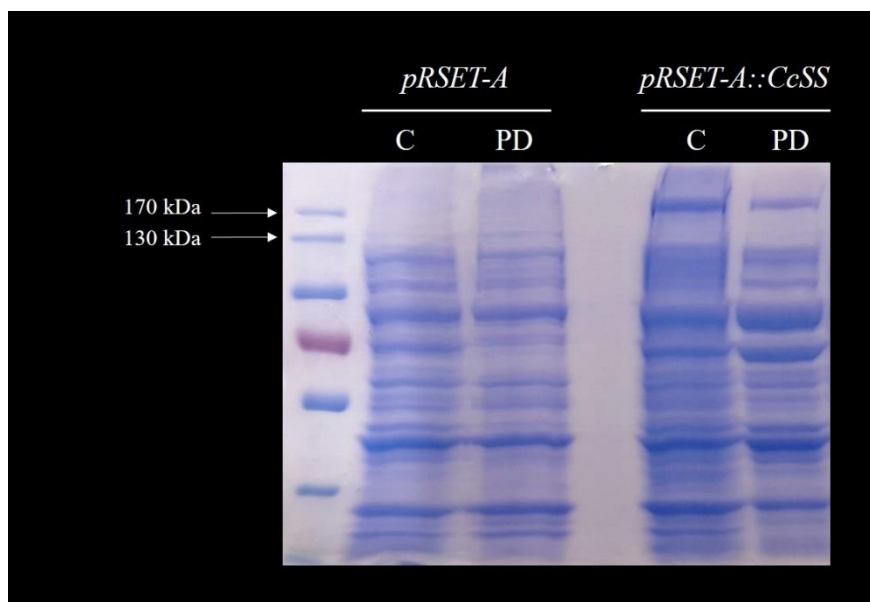
For on-column refolding, the protein was bound to a his-tag specific column through affinity chromatography and washed to remove other proteins. The immobilised *CcSS* protein was then washed with decreasing concentrations of urea to allow on-column renaturation and subsequently eluted from the column. The flow-through of the various fractions were collected and analysed using SDS-PAGE. The denatured *CcSS* is clearly visible in the crude protein extract prior to loading onto the column and is represented by a band at approximately 150 kDa in size (Figure 8). This band is absent in the empty vector control. The majority of the recombinant protein did not bind to the column and was eluted during the binding step, as a band appearing equal in intensity to that observed in the crude extract is visible in the flow-through. However, a faint band corresponding to the size of the *CcSS* protein is visible in the first elution step, indicating that some recombinant protein did bind to the column and was subjected to the urea wash. This protein elution was further analysed for starch synthase activity using native PAGE, but no activity could be detected (results not shown).



**Figure 8. On-column refolding of recombinant *CcSS* extracted under denaturing conditions.** The *E. coli* BL21CodonPlus(DE3)-RIPL strain carrying the pRSET-A::*CcSS* plasmid was grown in LB medium at 37°C and expression was induced at mid-log phase (OD<sub>600</sub> 0.6). Protein was extracted after 20 hrs post-induction. The recombinant protein was immobilised on a his-tag specific column and washed sequentially with decreasing concentrations of urea and the different protein fractions were resolved through SDS-PAGE. Low levels of recombinant *CcSS* were detected in the first elution between 130 and 170 kDa in size. C: Crude extract, B: Binding, W1+2: Wash 1 and 2, E1+2+3: Elution 1, 2 and 3.

Renaturation of the recombinant protein was also attempted by dialysis of the denatured, purified protein against 100 000-fold of buffer containing decreasing concentrations of urea to allow for the slow removal of

denaturing agents over a period of 3 days. The presence of the CcSS protein before and after dialysis was confirmed through SDS-PAGE and enzyme activity was assessed using native PAGE. A band of approximately 150 kDa in size representing the recombinant protein was observed in the crude protein extract (before dialysis), as well as in the protein sample post-dialysis (Figure 9). No activity could be detected in either the pre- or post-dialysis protein samples (results not shown). As the protein sample seemed to precipitate and turn cloudy when the urea concentration decreased below 4 M, the experiment was repeated with the addition of arginine to the dialysis buffer to facilitate renaturation by increasing protein solubility (Tsumoto et al., 2004). The protein sample remained clear until reaching a urea concentration of around 2 M, after which protein precipitation occurred and the sample turned cloudy. Although the presence of arginine appeared to slightly improve protein solubility, no improvement in activity could be detected with native PAGE analysis of the protein sample after dialysis (results not shown).

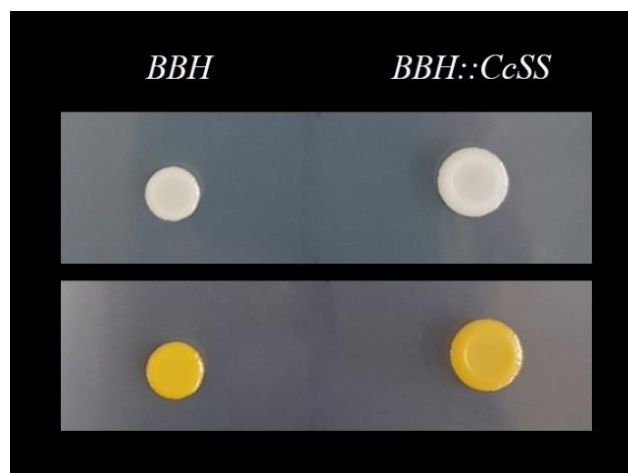


**Figure 9. Refolding of recombinant CcSS extracted under denaturing conditions through step-wise dialysis.** The *E. coli* BL21CodonPlus(DE3)-RIPL strain carrying the pRSET-A::CcSS plasmid was grown in LB medium at 37°C and expression was induced at mid-log phase (OD<sub>600</sub> 0.6). Protein was extracted after 20 hrs post-induction and dialysed against large volumes of buffer containing decreasing concentrations of urea. The extracted protein was resolved using SDS-PAGE before and after step-wise dialysis. C: crude extract, PD: post-dialysis.

#### 2.4.6 Yeast protein expression

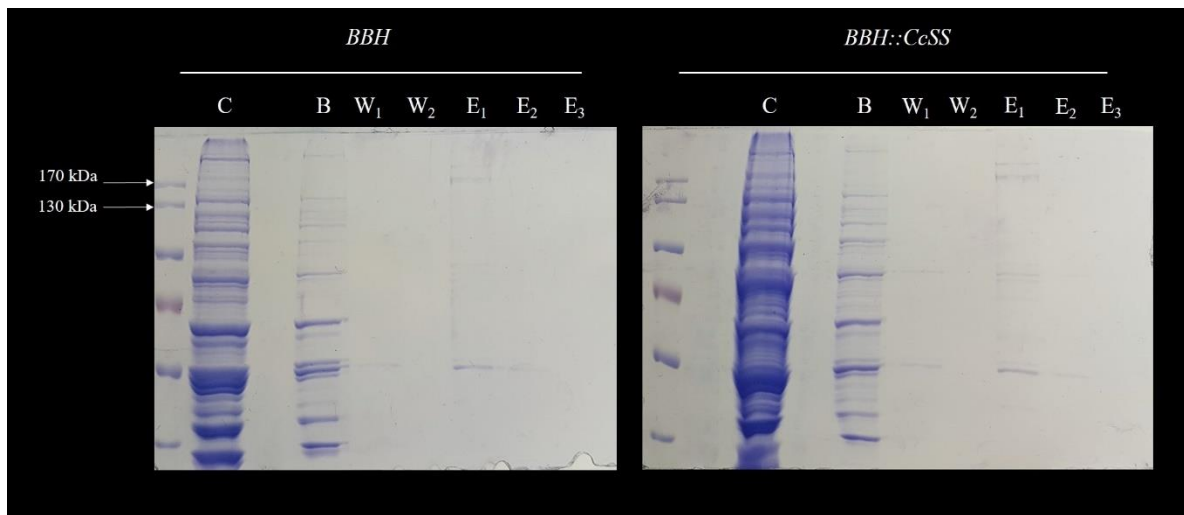
As recombinant protein expression in *E. coli* proved difficult, an alternative expression system was considered – *Saccharomyces cerevisiae* was used as it is a eukaryote harbouring the necessary post-translational modification machinery to ensure proper folding of the eukaryotic recombinant protein. The endogenous yeast glycogen synthase also utilises UDP-glucose as substrate, similar to what has been reported for red algal starch synthases. The *S. cerevisiae* *Δgcy2* strain which lacks the major glycogen synthase isoform (GSY2) was used

for yeast recombinant protein expression to avoid background activity during enzymatic activity assays. The *Δgsy2* strain carrying the episomal BBH:*CcSS* vector was spotted onto SC-ura agar and the resulting colonies were exposed to iodine vapour to examine glycogen accumulation due to recombinant starch synthase activity. There was no visible difference in polyglucan accumulation between cultures containing BBH:*CcSS* and the control, as both stained a similar yellow colour after exposure to iodine vapours (Figure 10). The colonies containing the expression vector appeared to be larger than the control, but this can be attributed to the discrepancy in cell density of the culture aliquots that were pipetted onto the growth media.

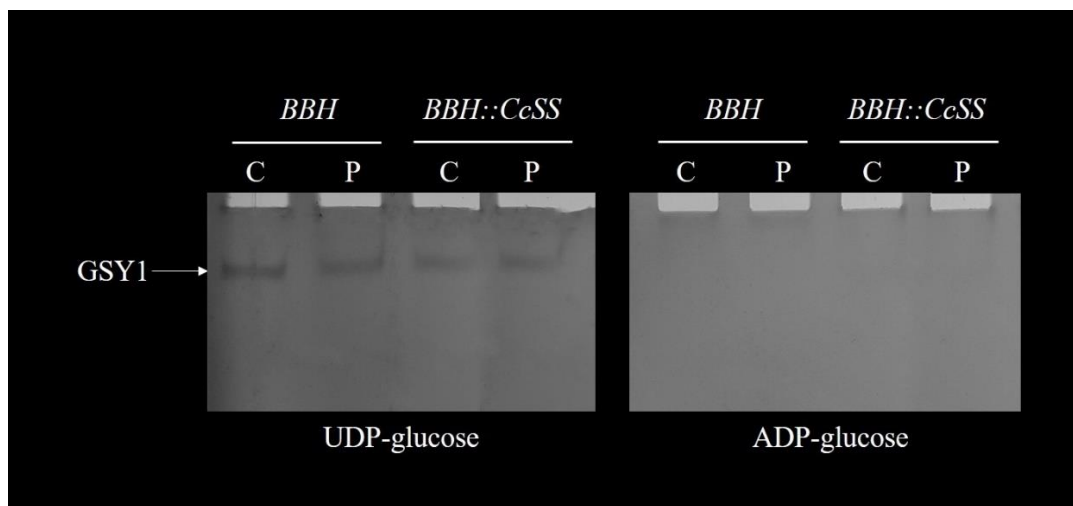


**Figure 10. Iodine staining of *S. cerevisiae* *Δgsy2* mutant carrying the BBH:*CcSS* plasmid to show glycogen accumulation.** Cultures containing a vector without insert were used as control. There was no increase in glycogen accumulation in colonies carrying the BBH:*CcSS* plasmid.

The recombinant protein was then expressed in the *S. cerevisiae* *Δgsy2* strain carrying the BBH:*CcSS* plasmid. Following protein extraction and purification using a his-tag specific column, the various fractions were resolved using SDS-PAGE and native PAGE. No recombinant protein was observed in either of the protein fractions (Figure 11). Although there appears to be a protein of the predicted size between 130 – 170 kDa in the first elution, a protein of corresponding size is also present in the vector without insert control. To ascertain whether there might be recombinant protein at a concentration too low to visualise, the crude extract as well as the eluted protein sample was analysed for starch synthase activity using native PAGE. Again, no starch synthase activity could be detected for the recombinant *CcSS* protein (Figure 12). The presence of staining in both the control and extracted protein samples incubated with UDP-glucose indicates that the activity is not due to the recombinant protein but can rather be attributed to the endogenous *S. cerevisiae* glycogen synthase isoform 1 (GSY1) (Figure 12).



**Figure 11. Expression and purification of recombinant CcSS protein in yeast.** The *S. cerevisiae*  $\Delta gsy2$  strain carrying the BBH::CcSS plasmid was grown in SC-ura medium at 30°C for 20 hrs. The protein extract was purified using a his-tag specific column and the various protein fractions were resolved using SDS-PAGE. No recombinant protein was detected. C: Crude extract, B: Binding, W1+2: Wash 1 and 2, E1+2+3: Elution 1, 2 and 3.



**Figure 12. Starch/glycogen synthase activity of protein extracts from yeast.** Protein was extracted and purified from a culture of the *S. cerevisiae*  $\Delta gsy2$  strain carrying the BBH::CcSS plasmid grown in SC-ura medium at 30°C for 20 hrs. The crude protein extract (C) and purified protein (P) were resolved using native-PAGE containing 0.1% (w/v) glycogen and incubated with 1 mM of either ADP- or UDP-glucose. Activity was only detected for the endogenous *S. cerevisiae* glycogen synthase (GSY1) incubated with UDP-glucose.

## 2.5 Discussion

In contrast to starch synthesis and storage in green plants, floridean starch is synthesised and stored in the cytosol of red algae and red algal starch synthases have been reported to preferentially utilise UDP-glucose over ADP-glucose as substrate for starch synthesis (Nyvall et al., 1999; Nyvall, 2000; Viola et al., 2001; Yu et al., 2002; Shimonaga et al., 2006; Stadnichuk et al., 2007; Shimonaga et al., 2008; Hirabaru et al., 2010). Owing to the differences in compartmentalisation and substrate preference, the amino acid sequences of these enzymes were compared to attempt to identify essential regions that could explain these variations. Multiple sequence alignment of the *C. crispus* starch synthase with starch/glycogen synthases from bacteria, green plants and other red algae revealed, as expected, that the majority of the sequence similarity lies within the catalytic domain of these enzymes. The red algal starch synthases uniquely contained a glycogen binding domain present at the C-terminus (Figure 4), which shows sequence similarity to the glycogen binding domain of the AMP-activated protein kinase  $\beta$ -subunit. When a metabolic stress in the form of low ATP levels is sensed, this kinase regulates the energy balance in the cell by activating processes which generate ATP and repressing those that consume it. The glycogen binding domain allows the kinase to sense energy reserves in the form of glycogen which, if available, inhibits the catalytic activity of the kinase (McBride et al., 2009). The reason for the presence of a glycogen binding domain in the red algal starch synthase is, however, unclear. It is possible that regulatory mechanisms for starch synthesis exist in red algae that are not present in higher plants, such that starch synthesis is suppressed when the glycogen binding domain senses the presence of high levels of accumulated starch. Alternatively, the glycogen binding domain could aid in binding of the starch synthase to starch that is being synthesized in the cytosol, thereby facilitating elongation of the glucan chains.

To facilitate isolation and biochemical characterisation of a red algal starch synthase, a synthesised gene encoding a protein identical to the *C. crispus* starch synthase was expressed in *E. coli*. However, expression and purification of the recombinant *C. crispus* starch synthase proved difficult as the synthase accumulated as insoluble protein in bacterial inclusion bodies. Several strategies were implemented to attempt to remedy this and express active recombinant protein in the soluble fraction. Firstly, two alternative *E. coli* protein expression strains were employed without any shift from insoluble to soluble protein production (Figure 7). Secondly, strains were co-transformed with a plasmid expressing several chaperone proteins to aid in correct folding of the recombinant protein. The chaperone plasmid contained copies of the *dnaK*, *dnaJ*, *grpE*, *groES* and *groEL* genes as their protein products are known to work synergistically and simultaneous expression of the DnaK-DnaJ-GrpE and Gro-EL-GroES chaperone teams has been demonstrated to be effective in recovery of correctly folded and soluble recombinant Cryj2 protein (a Japanese cedar pollen allergen) in *E. coli* (Nishihara et al., 1998). Although expression of the chaperone proteins could be observed during visualisation of the protein fractions, no soluble recombinant protein could be detected.

Thirdly, the expression conditions were optimised with regards to temperature, duration of expression and the initial culture density before induction of expression. In previous studies, high yields of recombinant P5 $\beta$ R2 (progesterone 5 $\beta$ -reductase) have been obtained by inducing expression in *E. coli* during early-log phase and maintaining expression at a low temperature (4°C) for 48 – 72 hrs (San-Miguel et al., 2013). However,



replicating these experimental conditions with *E. coli* expressing the *C. crispus* starch synthase failed to produce any soluble protein (Figure 7).

Fourthly, recovery of active recombinant *C. crispus* starch synthase was attempted by purification of the protein under denaturing conditions and subsequent refolding using two methods that have proved effective for other recombinant proteins, namely on-column refolding (Lemercier et al., 2003) and step-wise dialysis (Yang et al., 2011). The principle behind these methods is that the addition of a high concentration of a strong denaturant (e.g. urea or guanidinium chloride) and reducing agents (e.g. dithiothreitol or 2-mercaptoethanol) during protein extraction will denature the protein by removing undesirable interactions and disulphide bonds (Yamaguchi and Miyazaki, 2014). If followed by the gradual removal of the denaturing agent, it will allow for refolding of the recombinant protein into its native three-dimensional structure. The more rapid on-column refolding method involves immobilisation of the recombinant protein on a column and sequential washing of the column before elution of the protein. The *C. crispus* starch synthase successfully bound to the column and remained bound during the urea wash steps (Figure 8), but the eluted protein did not display any starch/glycogen synthase activity during native PAGE analysis (Figure 9). Similarly, renaturation of the *C. crispus* starch synthase through slow step-wise dialysis (Figure 9) over a period of three days was not successful in producing active protein. Even the addition of arginine to aid in stabilisation during refolding and prevent aggregation and precipitation (Tsumoto et al., 2004) of the recombinant protein did not result in the recovery of active protein.

Finally, as the *C. crispus* starch synthase is of eukaryotic origin, a eukaryotic yeast expression system was implemented as opposed to the prokaryotic *E. coli* bacterial system. *S. cerevisiae* was selected as it harbours the necessary eukaryotic post-translational modification machinery required for folding of some eukaryotic proteins into their native structures, and the endogenous *S. cerevisiae* glycogen synthases (GSY1 and GSY2) utilise UDP-glucose as substrate (Farkas et al., 1991). As a *S. cerevisiae* double mutant lacking both glycogen synthase genes could not be procured, the  $\Delta$ gsy2 mutant lacking activity of the major isoform was used for expression studies. Even though no recombinant *C. crispus* protein could be observed following expression in *S. cerevisiae*  $\Delta$ gsy2 (Figure 11), the protein extract was still analysed for starch/glycogen synthase activity. However, no activity could be detected for the recombinant starch synthase (Figure 12) and yeast colonies carrying the vector did not show an increase in glycogen accumulation as compared to wild type (Figure 10).

As *E. coli* expression during this study resulted only in the production of insoluble recombinant protein, further optimisation of bacterial expression can be conducted to attempt production of soluble protein and thus circumvent the need for the time-consuming process of protein refolding or the use of an alternative expression system like *S. cerevisiae* which has a much lower protein yield. One course of action is to express the *C. crispus* starch synthase under the control of a weaker promoter, such as the T5 promoter. The T7 promoter used during this study is designed for overexpression which can result in protein accumulation to a level that is toxic to the bacterial cells and leads to protein aggregation in inclusion bodies (Cabrita and Bottomley, 2004). Alternately, a different protein tag for purification could be implemented as some tags might interfere with protein folding and solubility. The glutathione-S-transferase (GST)-fusion tag has a dual function as an

affinity tag to simplify protein purification as well as aiding in solubilisation of the protein (Esposito and Chatterjee, 2006), and as such could rescue solubility of the recombinant protein. Another possibility is to construct a truncated form of the *C. crispus* starch synthase which includes only the catalytic domain and removes the N-terminal extension as well as the C-terminal glycogen binding domain (Figure 4). These extensions have previously been demonstrated to be nonessential for catalytic activity in potato GBSSI and SSII and the core catalytic domains alone displayed near wild type levels of activity (Edwards et al., 1999a). By expressing the catalytic domain (466 aa) of the *C. crispus* protein without its N-terminal (817 aa) or C-terminal (127 aa) extensions the solubility of the protein could be improved and enable rudimentary biochemical characterisation. In addition to isolation of the catalytic domain, future experiments which involve construction of truncated and chimeric forms of the *C. crispus* starch synthase and other green plant starch synthases could be performed. Such experiments have been performed previously using potato GBSSI and SSII, which analysed the function of the different domains in determining catalytic properties, thermo-sensitivity, specific activity or product specificity (Edwards et al., 1999a). Replicating this experiment with a red algal and higher plant starch synthase could aid in identifying regions important for the reported difference in substrate preference between the two groups of enzymes.

The absence of starch synthase activity in the case of refolded *E. coli*-expressed protein can be attributed to incorrect folding of the recombinant protein. The presence of disulfide bonds in the native protein structure could present a hurdle during refolding as the reducing agents (dithiothreitol or 2-mercaptoethanol) present in the refolding buffer interfere with interactions between cysteine residues. As the three-dimensional structure of the *C. crispus* starch synthase has not been solved, it is difficult to determine whether such disulfide bonds are present and required for formation of active protein. Prediction software that calculate the probability of disulfide bond formation between cysteine residues in the *C. crispus* amino acid sequence display conflicting results. While DiANNA (v1.1, Ferre and Clote, 2005a; Ferre and Clote, 2005b; Ferre and Clote, 2006) predicts a total of 9 disulfide bonds in the *C. crispus* protein, DISULFIND (Ceroni et al., 2006) predicts that none of the cysteine residues form disulfide bonds with confidence intervals ranging between 7-9 (where 0 = low and 9 = high). Therefore, it is unclear whether the inability to form disulfide bridges is a factor that contributes to inactivity of the refolded protein. However, a number of granule-bound and soluble starch synthases from higher plants have been successfully expressed in *E. coli*, including those from potato, rice, maize and kidney bean (Edwards et al., 1995; Knight et al., 1998; Cao et al., 1999; Edwards et al., 1999a; Senoura et al., 2004; Fujita et al., 2008), thereby indicating that recombinant expression of these starch synthases is possible.

During refolding, the protein often takes on an intermediate structure from where it can assume the active three-dimensional structure or misfold and aggregate into inclusion bodies (Cabrita and Bottomley, 2004). Some of the variables that influence which pathway of folding is followed include the concentration of the target protein and components of the refolding buffer, such as pH, redox conditions, ionic strength and the presence of additives. Thus, the process of protein refolding is often time-consuming and difficult to optimise due to the vast number of variables that can dictate success and refolding conditions that are favourable to a particular protein are not necessarily transferable to another. Protein refolding kits, such as the QuickFold™

(AthenaES<sup>®</sup>) or Pierce<sup>™</sup> (Thermo Scientific<sup>™</sup>) kits, are commercially available and aid in streamlining the optimisation process by including a variety of buffer conditions which enable rapid determination of the optimal buffer for the particular protein. Although costly, one of these commercial kits may prove useful as an initial diagnostic tool to determine the requirements of the *C. crispus* starch synthase for successful refolding.

The lack of activity observed in recombinant protein produced by both the *E. coli* and *S. cerevisiae* expression systems in this study can also be explained by a lack of knowledge surrounding the substrate and primer requirements of red algal starch synthases. For example, the endogenous *S. cerevisiae* glycogen synthases, similar to other eukaryotic glycogen synthase enzymes, are activated in the presence of glucose-6-phosphate (Farkas et al., 1991; Roach, 2002). The activity of some plant starch synthases are also stimulated in the presence of citrate (Boyer and Preiss, 1979; Cao et al., 1999; Nakamura et al., 2014), while others require the presence of polyglucan primers in the form of amylopectin, glycogen or cluster dextrin to initiate starch synthesis (Fujita et al., 2006; Nakamura et al., 2014; Hayashi et al., 2018). Analogous to higher plant starch synthases, previous biochemical studies involving crude protein extracts from red algae demonstrated that the incorporation of UDP-glucose, but not ADP-glucose, into glucan chains was stimulated by citrate and excess glycogen (Nagashima et al., 1971; Sesma and Iglesias, 1998; Nyvall et al., 1999; Nyvall, 2000). There is also uncertainty surrounding the substrate preference of red algal starch synthases between ADP- and UDP-glucose (Viola et al., 2001). As the *C. crispus* starch synthase in this study was assayed for activity in the presence and absence of citrate, glycogen/amylopectin, glucose-6-phosphate, and both UDP- or ADP-glucose, it is possible that red algal starch synthases have additional requirements that are currently poorly understood.

In this study, site-directed mutagenesis was performed on the *C. crispus* starch synthase to aid in elucidating important residues that play a role in the specificity for ADP- or UDP-glucose in glycogen/starch synthases. An enzyme mutated to be inactive was generated to serve as a negative control during biochemical analysis through substitution of Glu1200 with a Cys residue (E1200C). This Glu residue is fully conserved in all glycogen and starch synthases studied to date and is reported to be essential for enzyme activity (Yep et al., 2004). In *E. coli*, substitution of the corresponding residue with a Cys residue has been demonstrated to abolish glycogen synthase activity (Yep et al., 2004). Further, the conserved KXGGL motif, which has been implicated in interacting with and stabilising binding of the substrate ADP-glucose in both *E. coli* and *A. tumefaciens* glycogen synthases (Furukawa et al., 1990; Furukawa et al., 1993; Buschiazzo et al., 2004), was targeted for site-directed mutagenesis. In the glycogen and starch synthases analysed during this study, the first two residues of this motif vary while the 'GGL' pattern is conserved. In bacterial and green plant synthases, the first two residues are occupied by lysine and threonine/valine (KTGGL or KVGGL) (Figure 4). However, in the red algal starch synthases the first and second positions are occupied by hydroxylic and aliphatic amino acids respectively, and the motif reads TAGGL (*C. crispus* and *G. sulphuraria*) or SIGGL (*C. merolae*). Due to the differences in properties of the motif residues between red algal and bacterial/green plant synthases, the Thr829 and Ala830 residues of the *C. crispus* starch synthase were substituted to Lys829 and Thr830 respectively (TA829-30KT). These residues were mutagenized to resemble the ADP-glucose utilising enzymes in order to determine whether this would result in a change in substrate preference of the enzyme. In addition,

the Tyr354 (*Agrobacterium tumefaciens*) and Tyr355 residues (*Escherichia coli*) of bacterial glycogen synthases have been reported to have interactions with the adenine ring of ADP-glucose (Buschiazzo et al., 2004; Yep et al., 2006), and were subsequently targeted for site-directed mutagenesis. In the case of *E. coli* glycogen synthase, the substitution of tyrosine to alanine, leucine or methionine results in a 31 – 33-fold decrease of the apparent affinity for ADP-glucose as substrate (Yep et al., 2006). In addition, substitution with phenylalanine, tryptophan and histidine only slightly lowers the apparent affinity for ADP-glucose (1.6 – 6-fold respectively). Despite the decrease in affinity, both the wild type and all mutant enzymes significantly preferred ADP-glucose to GDP-, UDP- and CDP-glucose, with the latter two purine sugars being utilised the least efficiently (Yep et al., 2006). Although substitution of this tyrosine residue evidently alters the affinity for ADP-glucose, it does not completely reverse its specificity for this sugar nucleotide. In Arabidopsis starch synthases, all of which also preferentially utilise ADP-glucose, the corresponding position is occupied by either phenylalanine (GBSS, SSI and SSII) or tyrosine (SSIII and SSIV) (Figure 4, Table 4). These residues correlate with the tyrosine residue present in wild type bacterial glycogen synthases, as well as the almost negligible decrease in ADP-glucose apparent affinity observed upon substitution to phenylalanine. In the red algal species, which are predicted to utilise UDP-glucose based on sequence data and early biochemical studies, the position is occupied by threonine (*G. sulphuraria* and *C. crispus*) or histidine (*C. merolae*) (Figure 4, Table 4). While the effect of substitution to His was studied in *E. coli*, the consequence of a change to threonine such as in the case of the red algal starch synthases remains to be examined. Thus, the Thr1181 residue of the *C. crispus* starch synthase was mutagenized to a tyrosine residue to generate a mutant enzyme (T1181Y) that resembles the green plant starch synthase sequences. Although this residue is suspected to play a role in the recognition of the adenine ring of ADP-glucose, it is possible that the amino acid itself is irrelevant and the specificity rather lies in the structure of the substrate binding pocket formed in the closed conformation of the enzyme. Due to time constraints and an inability to produce active, soluble wild type *C. crispus* starch synthase, the effect of the mutagenized residues could not be determined. However, vectors that express a range of *C. crispus* starch synthase mutant enzymes with possible altered properties were developed for future use.

The reported preference of red algal starch synthases for UDP-glucose as substrate presents the opportunity to use this protein as a biotechnological tool in increasing starch synthesis and thereby plant growth in green plants, particularly crop plants. Transitory starch is synthesized in the leaves of plants during the day as a product of photosynthesis and degraded at night to allow continued carbon allocation for growth and cellular processes. It is also produced and stored for longer periods of time in non-photosynthetic organs such as stems, tubers and seeds. Starch has been demonstrated to have a major influence on plant growth, as mutant plants where starch is lacking grow less well than control plants because they are unable to sustain growth during the dark period and exhibit symptoms of carbon starvation (Stitt and Zeeman, 2012). Green plant starch synthases utilise ADP-glucose as substrate, and a lesser ability to use UDP-glucose *in vitro* has been reported for GBSS although the use of UDP-glucose *in vivo* is unlikely due in part to the lack of high UDP-glucose levels within plastids. The cytosol is, however, rich in UDP-glucose. By expressing the cytosolic UDP-glucose *C. crispus* starch synthase in a higher plant, a second carbon pool could be synthesized and possibly increase starch accumulation during the light period and mobilisation thereof during the dark. This alternative of starch pool

could, therefore, positively influence plant growth by buffering a plant against lack of carbon during cloudy weather. If transferred to the edible parts of crop plants such as potato, it has the potential to improve food security by improving plant yield through increasing starch amounts which are often one of the major sinks within storage organs. Furthermore, if the synthesized starch has a high amylose content, it would possess health benefits for human consumption as recalcitrant starch would be digested over a longer period of time, thereby preventing rapid blood glucose spikes which may lead to the development of insulin resistance and other associated diseases (Nofrarias et al., 2007; Zhu et al., 2012).

# Chapter 3: Analysis of *P. patens* maltose excess transporters and their role in starch metabolism

## 3.1 Introduction

### 3.1.1 Physcomitrella as a model system

During the last several years, the moss *Physcomitrella patens* (hereafter referred to as *Physcomitrella*) has emerged as an attractive model system to examine aspects surrounding the cell, developmental and evolutionary biology of plants. Its value as a model organism lies within its simple growth requirements and propagation, short life cycle, potential for targeted gene manipulation, and conservation of molecular pathways present in higher plants. *Physcomitrella* is capable of completing its entire life cycle within the span of only two months while growing *in vitro* on a minimal medium consisting of sterile sand, soil or agar containing simple salts (Wettstein, 1924; Engel, 1968; Nakosteen and Hughes, 1978). Its lifecycle also features two generations, namely a diploid sporophyte phase and a haploid gametophyte stage, the latter of which is the dominant life stage (Wettstein, 1924; Rensing et al., 2008). This provides the potential to examine the manifestation of recessive traits in a visible phenotype during gametophyte growth. Arguably the most valuable feature of *Physcomitrella* is its extremely high frequency of homologous recombination, which allows for insertion or disruption of genes at targeted loci (Schaefer and Zryd, 1997).

*Physcomitrella* is also important in evolutionary biology as it is a member of the bryophytes (liverworts, hornworts and mosses), who are remnants of the first diverging lineages of embryophytes to move to land (Figure 2). It was the first bryophyte of which the complete genome was sequenced (Rensing et al., 2008), followed by the recent sequencing of the genome of two other mosses, *Sphagnum magellanicum* and *Sphagnum fallax* (The Sphagnum Project; Shaw et al., 2016), and a liverwort, *Marchantia polymorpha* (Bowman et al., 2017). Access to the genomes of these bryophytes is an essential step in reconstructing the evolutionary events that occurred as they occupy the ideal phylogenetic position between extant land plants and the sister group of aquatic green algae they share a common ancestor with. Characteristics that were vital to the conquest of land in the first embryophytes can be inferred by their presence in bryophytes and angiosperms, and their concurrent absence in aquatic algae.

The transition to a terrestrial habitat involved overcoming challenges in the form of fluctuations in temperature, variations in the availability of water, and greater exposure to radiation. Thus, dramatic changes in plant morphology and the addition of mechanisms to protect against abiotic stresses were required. Fossil evidence suggests that the first embryophytes are thought to have structurally resembled extant bryophytes and harbour a dominant haploid gametophyte stage (Kenrick and Crane, 1997). Sexual reproduction was also restricted as male gametes probably required water for motility (Kenrick and Crane, 1997). Analysis of the *Physcomitrella* genome suggests that a large-scale gene (or possibly whole genome) duplication occurred at least once (Rensing et al., 2007; Rensing et al., 2008), although there is also evidence that supports the occurrence of two ancestral whole genome duplications (Rensing et al., 2013). These gene or genome duplications are viewed as the source material for diversification and the emergence of evolutionary novelties (Lynch and Conery, 2000).

Despite differences in the number of isoforms, the general pathway of starch metabolism seems to be conserved between algae, mosses and higher plants. A comparative bioinformatic study of green algal genomes revealed the presence of at least one isoform for each of the steps involved in starch synthesis and degradation (Deschamps et al., 2008b). For this reason, starch metabolism has been extensively studied in organisms with simpler genomic, molecular and physiological characteristics than higher plants, such as the unicellular green alga *Chlamydomonas*. Many of the enzymes and the regulatory processes that govern their activities in starch metabolism are found in both vascular plants and *Chlamydomonas*. For example, the activity of AGPase is allosterically regulated and activated by 3-PGA while being inhibited by phosphate, in both this alga and in higher plants (Iglesias et al., 1994; Kleczkowski, 1999). Likewise, mutations in genes encoding GBSS lead to the absence of amylose in starch in both *Arabidopsis* and *Chlamydomonas* (Delrue et al., 1992), while mutation in either *Isa1* or *Isa2* lead to accumulation of phytoglycogen (Dauvillée et al., 2001a; Delatte et al., 2005; Wattedled et al., 2005). However, some differences in the starch metabolic pathway have been observed between *Chlamydomonas* and higher plants. Like *Arabidopsis*, mutations in the gene encoding DPE results in the accumulation of MOS in *Chlamydomonas* (Wattedled et al., 2003); however, unlike *Arabidopsis dpe1* mutants, the *Chlamydomonas* mutant also shows a severe starch deficiency (Critchley et al., 2001; Zabawinski et al., 2001; Wattedled et al., 2003). This suggests that DPE serves a dual purpose in *Chlamydomonas* and is involved in starch biosynthesis in addition to its role in starch degradation.

A second difference between *Chlamydomonas* and higher plants in starch metabolism appears to be in the specificity of the plastidial maltose export (MEX) protein. Recently, a *Chlamydomonas* mutant exhibiting altered carbon metabolism was isolated. It was determined that a disruption of the *CrMEX1* gene, an orthologue of the *Arabidopsis MEX1* gene (Niittylä et al., 2004), was responsible for the over-accumulation of starch granules and increased lipid levels (Jang et al., 2015). This is similar to the phenotype observed in *Arabidopsis mex1* mutants which demonstrates a starch-excess phenotype (Niittylä et al., 2004). In addition, the *Arabidopsis* mutant accumulates high levels of maltose in its leaves which is consistent with the idea that maltose is the major end-product of starch degradation and is transferred across the chloroplast envelope to the cytosol by MEX1 (Niittylä et al., 2004; Stettler et al., 2009). However, neither the *Chlamydomonas* mutant nor wildtype strain demonstrate any observable change in maltose levels during starch degradation (Findinier et al., 2017). Upon complementation of the *Arabidopsis mex1* mutant with the algal orthologue the protein demonstrates a low or non-existent capacity to transport maltose over the plastid membranes. Intriguingly, the complemented mutants displayed a wildtype growth phenotype, accompanied by accumulation of both starch and maltose in leaves of the transgenic plants (Findinier et al., 2017). Complementation of an *E. coli malF* mutant, that lacks an active maltose transporter, with the algal protein also did not restore bacterial growth on maltose as the sole carbon source, while use of the higher plant gene did (Findinier et al., 2017). This indicates that the lower plant protein does not transport maltose, while the higher plant one does.

Interestingly, both the *Arabidopsis* and *Chlamydomonas* proteins were capable of reverting growth of an *E. coli ptsG* mutant lacking the endogenous glucose transporter activity (Findinier et al., 2017). This indicates that the MEX proteins from both green algae and *Arabidopsis* can transport glucose, a functionality that has

not yet been explored but is still compatible with the current view of starch degradation that glucose is a minor end-product of this pathway. It is important to note that a plastidial glucose transporter has been identified in *Arabidopsis*, but its role in starch degradation appears minimal as mutants display no visible growth defects (Cho et al. 2011).

These data support the theory that *Arabidopsis* MEX proteins transport both maltose and glucose, while the algal MEX protein appears to solely transport glucose. It is possible that the protein first evolved as a glucose transporter and later gained functionality as a maltose transporter after divergence of land plants from the lineage. The yield of maltose as the major product of starch breakdown and the ability of the MEX protein to transport maltose from the plastid could have been another important adaptation that was required during the conquest of land. If this is indeed the case, it would be interesting to examine whether the same effect is observed in a bryophyte such as *Physcomitrella*.

Few studies have been performed on starch metabolism in *Physcomitrella*, and knowledge around differences in the biochemical pathway with respect to higher plants is limited. As *Physcomitrella* has undergone genome duplications and thus boasts a large repertoire of isoforms for starch metabolic enzymes, it is difficult to examine the functions and characteristics of each isoform. Particularly in the case of starch synthases, where there are as many as eleven putative orthologues to *Arabidopsis* (Table 1), a vast amount of work would be required. In contrast, there are only four putative MEX orthologues to characterise in *Physcomitrella* which seems a more surmountable task. Analysing the maltose or glucose transport capacity of the *Physcomitrella* MEX proteins would aid in understanding how the functionality of the MEX transporter has changed during plant evolution and perhaps provide more insight to the development of starch metabolism in the first embryophytes.

### 3.2 Aim and objectives

The aim of this chapter is to analyse whether the MEX transporter plays the same role in starch degradation in *Physcomitrella* as it does in *Arabidopsis* leaves.

- Objective 1: The function of the proteins in *Physcomitrella* starch metabolism will be examined by producing knockout mutants. Constructs designed to remove each gene from the genome through homologous recombination will be produced and transformed into *Physcomitrella* mutants. Knockout mutants will be isolated and examined in terms of their phenotype.
- Objective 2: The sugar specificity of the *Physcomitrella* putative MEX transporters will be studied by complementation in *E. coli* mutants that lack specific transporters and are unable to grow on media containing maltose or glucose as the sole carbon source. The ability of the *Physcomitrella* MEX transporters to restore transport of maltose or glucose and thus growth in these *E. coli* mutants will be examined.



## 3.3 Materials & Methods

### 3.3.1 Sequence analysis and phylogeny

Protein sequences were obtained from the NCBI RefSeq database and the *P. patens* genome was obtained from Phytozome (v12.1, Goodstein et al., 2012) CD-HIT (Li et al., 2001; Li et al., 2002; Li and Godzik, 2006; Huang et al., 2010), Mafft (v7.402, Katoh and Toh, 2008), BMGE (v1.12, Criscuolo and Gribaldo, 2010) and MEGA-X (Kumar et al., 2018) were used for phylogenetic analyses. GraphPad Prism (v6) was used for statistical analyses and generating graphs.

### 3.3.2 *P. patens* strain, culture media and growth conditions

The *P. patens* Gransden strain (Engel, 1968) was used as the wild type strain during this study. All methods and materials used to culture *P. patens* were previously described by Collier and Hughes (1982) and Cove *et al.* (2009). *P. patens* was grown on a variation of solid BCD medium (Ashton et al., 1979), hereafter referred to as BCDT medium, which consisted of 5 mM C<sub>4</sub>H<sub>6</sub>O<sub>6</sub>, 1 mM MgSO<sub>4</sub>, 1.84 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM KNO<sub>3</sub>, 9.93 μM H<sub>3</sub>BO<sub>3</sub>, 0.103 μM Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 0.266 μM CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.191 μM ZnSO<sub>4</sub>·7H<sub>2</sub>O, 1.97 μM MnCl<sub>2</sub>·4H<sub>2</sub>O, 0.169 μM KI, 0.22 μM CuSO<sub>4</sub>, 1 mM CaCl<sub>2</sub> and 45 μM FeSO<sub>4</sub>. The pH of the medium was adjusted to 5.8 with KOH and 0.55% (w/v) agar added before autoclaving. CaCl<sub>2</sub> and FeSO<sub>4</sub> were prepared separately as stock solutions and filter-sterilised prior to addition to the cooled media. Routine monthly sub-culturing of *P. patens* was conducted by dividing established colonies and transferring the harvested tissue to fresh media. Plates were sealed with Micropore™ surgical tape (3M™, USA-MN) and maintained in a growth room at a temperature of 25 ± 2°C and a 16 hr light/8 hr dark photoperiod using cool white fluorescent lights (50 μM photons m<sup>-2</sup> s<sup>-1</sup>; OSRAM L 58W/640, Germany).

### 3.3.3 DNA and RNA extraction

DNA was extracted from *P. patens* tissue following a protocol adapted from (Edwards et al., 1991). Briefly, plant tissue was flash-frozen and ground in liquid nitrogen. Approximately 100 mg of powdered tissue were added to 400 μL extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% (w/v) SDS) and vortexed for 5 s. After pelleting the debris (16 000 x g, 5 min, 4°C) 300 μL of supernatant was transferred to 300 μL of isopropanol and mixed by inversion. Following a 20 min incubation at room temperature, the sample was centrifuged (16 000 x g, 15 min) and the supernatant discarded. The pellet was washed with 70% (v/v) ethanol and dissolved in 100 μL nuclease-free water. The sample was centrifuged (16 000 x g, 5 min) to pellet impurities and the supernatant transferred to a clean tube.

Total RNA was extracted from *P. patens* powdered tissue using the Maxwell® 16 LEV Plant RNA Kit (Promega) and the Maxwell® 16 Instrument (Promega). RNA was eluted in a final volume of 30 μL of nuclease-free water. Nucleic acid concentration and integrity was assessed spectrophotometrically and visualised by agarose gel electrophoresis. First-strand cDNA synthesis was performed on 1 μg of RNA per reaction using the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Thermo Scientific).

### 3.3.4 Bacterial strains and vectors

Table 5. *E. coli* strains and vectors used during this study.

Strain	Genotype	Source
DH5a	F <sup>-</sup> endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 $\phi$ 80dlacZ $\Delta$ M15 $\Delta$ (lacZYA-argF)U169, hsdR17(rK <sup>-</sup> mK <sup>+</sup> ), $\lambda$	(Thermo Fisher)
10-beta	F <sup>-</sup> , araD139 $\Delta$ (ara-leu)7697 <i>fhuA lacX74 galK</i> ( $\phi$ 80 $\Delta$ (lacZ)M15) <i>mcrA galU recA1 endA1 nupG rpsL</i> (Str <sup>R</sup> ) $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> )	(New England BioLabs)
TST6 ( <i>malF</i> ) <sup>4</sup>	F <sup>-</sup> ( <i>araD139</i> ) <sub>B/r</sub> $\Delta$ ( <i>argF-lac</i> )169 $\lambda$ - <i>flhD5301</i> $\Delta$ ( <i>fruK-yeiR</i> )725( <i>fruA25</i> ) <i>relA1 rpsL150</i> (Str <sup>R</sup> ) <i>rbsR22 malF55::Tn10</i> $\Delta$ ( <i>fimB-fimE</i> )632( <i>:IS1</i> ), <i>deoC1</i>	(Silhavy)
JW1087-2 ( <i>ptsG</i> ) <sup>4</sup>	F <sup>-</sup> , $\Delta$ ( <i>araD-araB</i> )567 $\Delta$ <i>lacZ</i> 4787( <i>:rrnB-3</i> ) $\lambda$ - $\Delta$ <i>ptsG763::kan rph-1</i> $\Delta$ ( <i>rhaD-rhaB</i> )568 <i>hsdR514</i>	(Baba et al., 2006)
<b>Vectors:</b> pMiniT <sup>TM</sup> 2.0 vector (New England BioLabs), pBluescript SK (+) vector (Agilent Technologies), pMBL11a vector (Kamisugi and Cuming, 2009), pQE-30:: <i>AtMEX</i> <sup>4</sup> and pQE-30:: <i>CrMEX</i> <sup>4</sup>		

### 3.3.5 Growth media

General growth of *E. coli* was performed on LB (section 2.3.2) and Terrific Broth (TB). TB was prepared using 1.2% (w/v) peptone, 2.4% (w/v) yeast extract and 0.4% (v/v) glycerol before sterilisation by autoclaving. Complementation studies were performed on M9 minimal medium (20% (v/v) M9 salts, 0.2% (v/v) 1M MgSO<sub>4</sub>, 0.01% (v/v) 1M CaCl<sub>2</sub>) containing either 0.5% (w/v) D-maltose or D-glucose as carbon source. M9 salts were prepared using 6.4% (w/v) Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 1.5% (w/v) KH<sub>2</sub>PO<sub>4</sub>, 0.25% (w/v) NaCl and 0.5% (w/v) NH<sub>4</sub>Cl and autoclaved. Sugars, MgSO<sub>4</sub> and CaCl<sub>2</sub> were prepared as stock solutions which were filter-sterilised (0.2  $\mu$ m) and added to cooled media before use. For solid media, 1.5% (w/v) bacteriological agar was added before autoclaving.

### 3.3.6 PCR amplification

For screening purposes, reactions contained 1X Green GoTaq® Reaction Buffer (Promega), 1.25 U GoTaq® DNA Polymerase (Promega), 0.2 mM dNTP mix, 0.1  $\mu$ M of each primer and 25–50 ng DNA in a 50  $\mu$ L total reaction volume. Where the PCR product was used for cloning, reactions contained 1X Phire<sup>TM</sup> Plant PCR Buffer (Thermo Scientific), high-fidelity Phire<sup>TM</sup> Hot Start II DNA Polymerase (Thermo Scientific), 0.5  $\mu$ M of each primer and 25–50 ng DNA in a 50  $\mu$ L total reaction volume. All oligonucleotides were synthesized by Inqaba Biotechnical Industries (Pty) Ltd. (Pretoria, South Africa).

<sup>4</sup> Kindly provided by Dr David Dauvillée (Lille University, France).

### 3.3.7 DNA manipulation

Electrophoresis, DNA purification, plasmid isolation, restriction digests, preparation of competent cells and heat-shock transformations are described under Chapter 2 section 2.3 (Materials and Methods). Linearised plasmids were dephosphorylated when necessary with FastAP Thermosensitive Alkaline Phosphatase (ThermoScientific) as per the manufacturer's instructions. Ligation reactions were performed using T4 DNA Ligase (Promega) and Rapid Ligation Buffer (Promega) according to the manufacturer's instructions.

### 3.3.8 Large-scale plasmid isolation

Plasmid isolation methodology adapted from Birnboim and Doly (1979) and Sambrook et al. (1989) was used. Briefly, plasmid containing cells were incubated in 200 mL TB with appropriate antibiotics overnight (37°C). Cells were collected by centrifugation (4000 x g, 20 min) before the pellet was suspended in 8 mL Solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA) and incubated on ice (5 min). Then, 16 mL Solution II (0.2 M NaOH, 1% (w/v) SDS) and 12 mL Solution III (3 M KoAc pH 4.8) was added with a 5 min incubation step on ice after the addition of each solution. The sample was filtered through Miracloth (Merck) twice before precipitating DNA by the addition of 0.7 volumes of isopropanol. After incubating for 30 min at room temperature, the sample was centrifuged (1500 x g, 30 min) and the pellet dissolved in 1 mL dH<sub>2</sub>O. RNA was precipitated for 10 min at room temperature after the addition of 1 mL 5 M LiCl. The sample was centrifuged (16 000 x g, 10 min) and the supernatant transferred to a clean tube. After the addition of 1 volume of chloroform:isoamylalcohol:phenol (24:1:25), the sample was vigorously vortexed for 30 s and centrifuged (18 000 x g, 3 min). An aliquot of 400 µL of the top phase was transferred to a clean tube and the previous step repeated with chloroform:isamylalcohol (24:1). Then, 400 µL of the top phase was mixed with 0.1 volumes of 8 M LiCl and 2.5 volumes of ethanol and incubated at room temperature (10 min). Plasmid DNA was pelleted by centrifuging (16 000 x g, 10 min) and washed twice with 70% (v/v) ethanol before being resuspended in 100 µL TE/RNase (10 mM Tris-HCl, 1 mM EDTA pH 8.0, 0.002% (v/v) RNase).

### 3.3.9 *P. patens* protoplast isolation and PEG-mediated transformation

*P. patens* knockout mutants were produced following a protocol modified from Cove et al. (2009), Hoffmann and Charlot (2009) and Liu and Vidali (2011). *P. patens* protonemal tissue was harvested and homogenised in water through vortexing with glass beads (0.5 mm). Solid BCDT medium plates were overlaid with autoclaved cellophane disks (A.A. Packaging Limited, UK) and 2 mL of protonemal homogenate was pipetted onto each plate. The tissue was grown for a minimum of 7 days before being collected from the cellophane disks and transferred to 9 mL of an 8% (w/v) D-mannitol solution. To this, 3 mL of a solution containing 2% (w/v) Driselase (Sigma-Aldrich) and 8.5% (w/v) D-mannitol was added and the mixture incubated at room temperature with gentle agitation. After 1 hr, the mixture was filtered through a 100 µm cell strainer (BD Falcon™, 08-7771-19) to isolate the protoplasts. The solution was centrifuged (250 x g, 5 min) and the protoplasts gently washed twice with PRM-L (liquid BCDT medium containing 8.5% (w/v) D-mannitol). Total numbers of protoplasts were quantified from a 40 µL aliquot using a hemocytometer, and the pellet carefully suspended in MMM solution (8.5% (w/v) mannitol, 15mM MgCl<sub>2</sub>, 0.1% (w/v) MES pH 6) to a final concentration of 1.6 million protoplasts/mL.

A 600  $\mu\text{L}$  aliquot of suspended protoplasts was added to approximately 60  $\mu\text{g}$  of linearised vector construct and mixed by gentle swirling. Thereafter, 700  $\mu\text{L}$  of a PEG solution (7% (w/v) mannitol, 100 mM  $\text{CaNO}_3$ , 10 mM Tris-HCl pH 7.2, 40% (w/v) PEG 4000) was added and mixed by gentle swirling. The solution was incubated at room temperature (30 min) with gentle agitation before being heat-shocked ( $45^\circ\text{C}$ , 5 min) in a water bath and left to recover at room temperature (30 min). The protoplasts were overlaid with 6.5 mL PRM-L and incubated overnight at room temperature in the dark. The following day, most of the supernatant was removed and protoplasts were gently suspended in 2 mL freshly-autoclaved and cooled PRM-T (solid BCDT medium containing 8% (w/v) D-mannitol), and immediately pipetted onto PRM-B (solid BCDT medium containing 6% (w/v) D-mannitol) plates overlaid with cellophane disks. The plates were gently swirled to spread the PRM-T before setting. Plates were then sealed and incubated at room temperature with constant light for 1 week.

After 1 week of regeneration, the cellophane disks were transferred to solid BCDT plates containing 50  $\mu\text{g}/\text{mL}$  G418 (Sigma-Aldrich) for another week to select for transformants. Following this, the disks were transferred to solid BCDT plates without any antibiotic selection to allow for the loss of plasmids from unstable transformants. After another week, disks were again transferred to BCDT plates containing 50  $\mu\text{g}/\text{mL}$  G418 to select for stable transformants. *P. patens* colonies that survived this second round of selection were sub-cultured and screened.

### 3.3.10 Semi-quantitative RT-PCR

First-strand cDNA synthesis was performed as before on 1  $\mu\text{g}$  of RNA and 2  $\mu\text{L}$  (approximately 100 ng) of cDNA was used in a 50  $\mu\text{L}$  amplification reaction. The linear range of amplification was determined by the addition of 5  $\mu\text{L}$  of 5 mM EDTA during the last 2 s of the elongation phase after 5, 10, 15, 25 and 30 cycles, with a total run of 35 cycles.

### 3.3.11 *E. coli* complementation analysis

Confirmed *E. coli* transformants were grown overnight ( $37^\circ\text{C}$ ) in LB containing the appropriate antibiotics. The cells were pelleted by centrifugation (1500 x g, 15 min,  $4^\circ\text{C}$ ) and washed twice with Phosphate Buffered Saline (PBS: 0.8% (w/v) NaCl, 0.02% (w/v) KCl, 0.144% (w/v)  $\text{Na}_2\text{HPO}_4$ , 0.024% (w/v)  $\text{KH}_2\text{PO}_4$ , pH 7.4). Cells were suspended in PBS and subsequently used to inoculate M9 minimal liquid medium to a final cell density ( $\text{OD}_{600}$ ) of 0.05. M9 minimal media contained either 0.5% (w/v) D-maltose or D-glucose as sole carbon-source and 0.5 mM IPTG. Cultures were grown ( $37^\circ\text{C}$ ) for a period of 24 hrs, after which the  $\text{OD}_{600}$  was recorded. Samples were repeated in triplicate.

## 3.4 Results

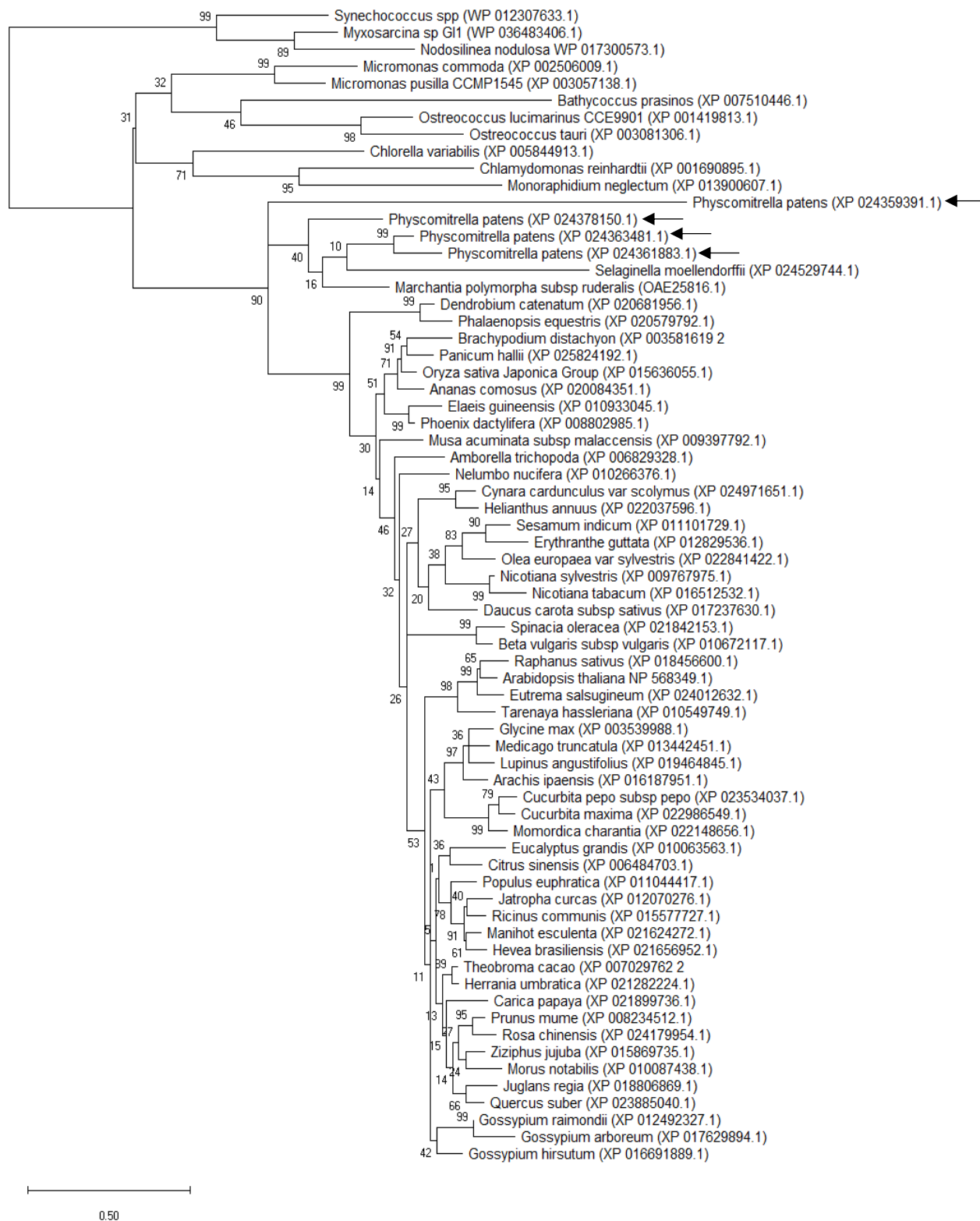
### 3.4.1 Sequence analysis

A BLAST search was performed using the Arabidopsis MEX (*AtMEX*) protein sequence (Q9LF50) against the *P. patens* genome on Phytozome (v12.1, Goodstein et al., 2012) to identify four putative orthologues which were arbitrarily designated *PpMEX1a*, *PpMEX1b*, *PpMEX1c* and *PpMEX1d*. Of these, the *PpMEX1a*, *PpMEX1c* and *PpMEX1d* gene sequences showed 53% identity to the Arabidopsis *MEX1*, while *PpMEX1b* shares only 29% identity. *PpMEX1a* and *PpMEX1b* were both predicted to have chloroplastic signal peptides and multiple transmembrane domains, indicating that they are likely membrane proteins (Table 6). *PpMEX1c* and *PpMEX1d* showed conflicting results as they both lack signal peptides but do, however, contain putative transmembrane domains suggesting that they are possible membrane proteins (Table 6).

**Table 6. Putative *AtMEX1* orthologues identified in *P. patens*.** The presence of chloroplast signal peptides and transmembrane domains was predicted using ChloroP (v1.1, Emanuelsson et al., 1999) and TMHMM (v2.0, Sonnhammer et al., 1998; Krogh et al., 2001) respectively.

Designation	Sequence ID (accession number)	Identity to <i>AtMEX1</i>	E-value	Sequence length (aa)	Predicted chloroplast signal peptide (aa)	Predicted transmembrane domains
<i>AtMEX1</i>	<b>Q9LF50</b>	-	-	<b>415</b>	<b>47</b>	<b>9</b>
<i>PpMEX1a</i>	Pp3c6_18710 (XP_024378150.1)	53%	1e-112	427	45	10
<i>PpMEX1b</i>	Pp3c21_10850 (XP_024359391.1)	29%	7e-16	350	65	2
<i>PpMEX1c</i>	Pp3c23_14070 (XP_024361883.1)	52%	4e-77	250	-	6
<i>PpMEX1d</i>	Pp3c24_890 (XP_024363481.1)	53%	5e-105	380	-	9

To confirm that the *P. patens* and Arabidopsis MEX proteins are true orthologues, proteins showing similarity to the *P. patens* proteins were identified using BLASTP against the RefSeq database with a threshold of e-value lower than 1e-10. Duplicate sequences were removed and the remaining sequences were clustered using CD-HIT (Li et al., 2001; Li et al., 2002; Li and Godzik, 2006; Huang et al., 2010) with an identity threshold of 90%. The representative sequences of each cluster were aligned using Mafft (v7.402, Katoh and Toh, 2008) and highly variable sites were removed using BMGE (v1.12, Criscuolo and Gribaldo, 2010). A maximum likelihood (ML) tree was constructed using MEGA-X (Kumar et al., 2018) with default parameters and General Reversible Chloroplast (cpREV) substitution model as MEX proteins have been demonstrated to be targeted to the chloroplast membrane (Niittylä et al., 2004; Findinier et al., 2017). The tree was manually inspected, and paralogs shared only by closely related species were removed. The remaining set of proteins was aligned, trimmed and used to construct the final ML tree as before with 100 bootstrap replicates.



**Figure 13. Representative maximum-likelihood (ML) tree showing relationships between members of the MEX protein family from different species of cyanobacteria and green plants. The tree is drawn to scale with branch lengths measured in the number of substitutions per site. Boot-strap support values from 100 replicates are indicated at nodes. The *P. patens* MEX proteins are indicated with arrows.**

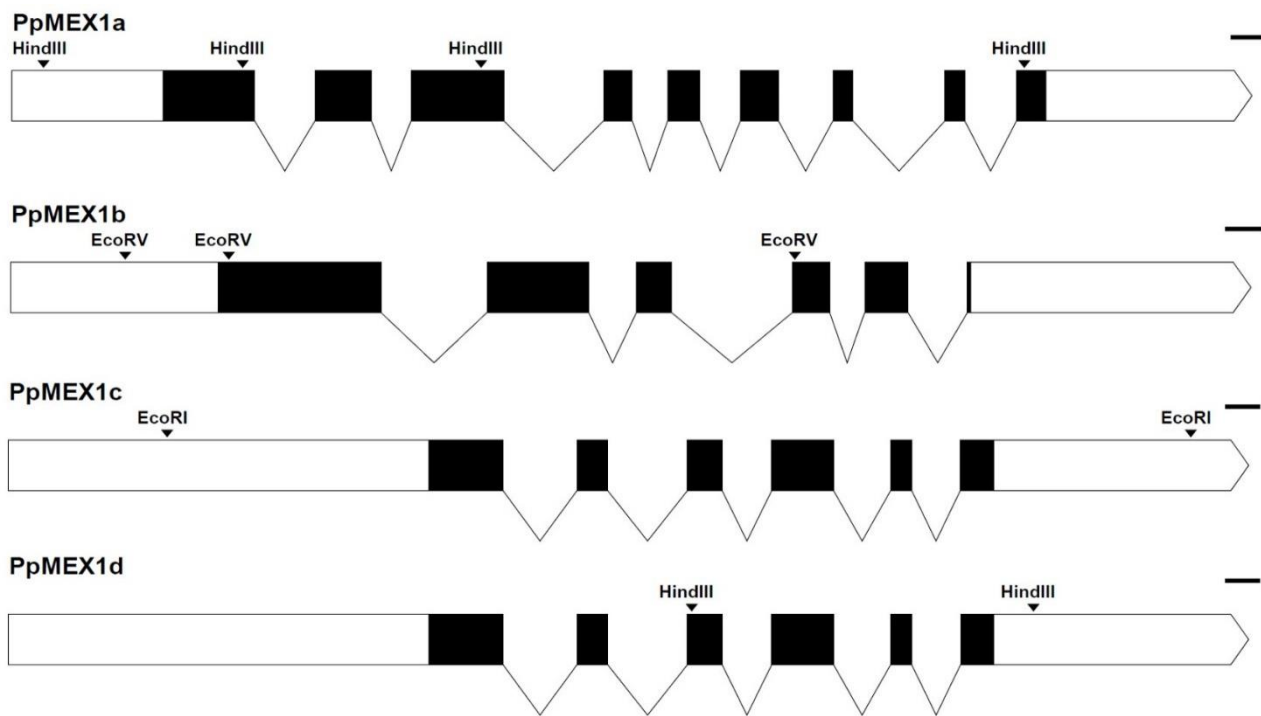
The resulting tree shows relatively high bootstrap values at nodes that separate the organisms into four main clusters, namely 1) cyanobacteria, 2) green algae, 3) bryophytes and lycophytes, and 4) higher plants (Figure 13). It appears that MEX proteins from all green plants (those in clusters 2, 3 and 4) are orthologous and originate from a single gene, possibly inherited from cyanobacteria by a common ancestor. The phylogenetic relationships between the MEX proteins mostly coincide with the species tree and proteins from closely-related species tend to group together. The *P. patens* MEX1c and MEX1d proteins appear to be most closely related to each other and share similarity with proteins from the liverwort *Marchantia polymorpha* and the lycophyte *Selaginella moellendorffii*. The *PpMEX1b* protein shows the least amount of similarity to the other *P. patens* proteins.

To quantify the sequence similarity between *Arabidopsis* and *P. patens* MEX proteins, pairwise local alignments using the Smith-Waterman were conducted and the percentage similarities for each alignment were compared. As with the phylogenetic analysis, *PpMEX1b* shows the lowest similarity values to the *Arabidopsis* (46.3%) and other *P. patens* MEX proteins (53.2%, 53.3% and 53.3%), while *PpMEX1c* and *PpMEX1d* show the highest similarity to each other (86.5%) (Table 7).

**Table 7. Similarity table of *Arabidopsis* and *P. patens* MEX proteins.** Similarity values represent the percentage of aligned amino acids and were calculated by pairwise Smith-Waterman local alignment.

	Arabidopsis thaliana MEX1	Physcomitrella patens MEX1a	Physcomitrella patens MEX1b	Physcomitrella patens MEX1c	Physcomitrella patens MEX1d
Arabidopsis thaliana MEX1 (NP_568349.1)	100	64.8	46.3	70.3	67.8
Physcomitrella patens MEX1a (XP_024378150.1)	64.8	100	53.2	66.6	66.9
Physcomitrella patens MEX1b (XP_024359391.1)	46.3	53.2	100	53.3	53.3
Physcomitrella patens MEX1c (XP_024361883.1)	70.3	66.6	53.3	100	86.5
Physcomitrella patens MEX1d (XP_024363481.1)	67.8	66.9	53.3	86.5	100

The intron-exon structure of the four *PpMEX* genes was analysed by aligning the available gene and transcript sequences using Splign (Kapustin et al., 2004; Kapustin et al., 2008) and schematically representing the intron-exon boundaries. *PpMEX1c* and *PpMEX1d* show an identical pattern with regards to their intron-exon structures (Figure 14).



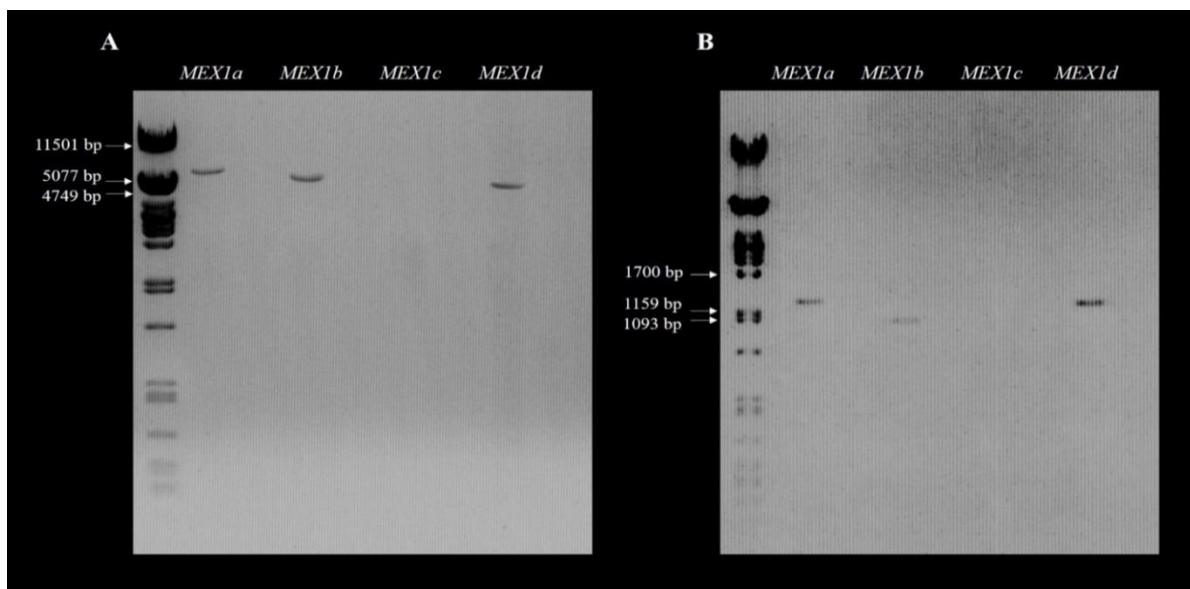
**Figure 14. Schematic representation of the intron-exon structure of the *P. patens* *MEX* genes.** Both 5' and 3' untranslated regions are indicated by unfilled rectangles, while black rectangles represent coding regions within the genes. The position of restriction sites utilised during the construction of knockout vectors is indicated for each gene. Scale bars represent 100 bp.

### 3.4.2 Construction of vectors

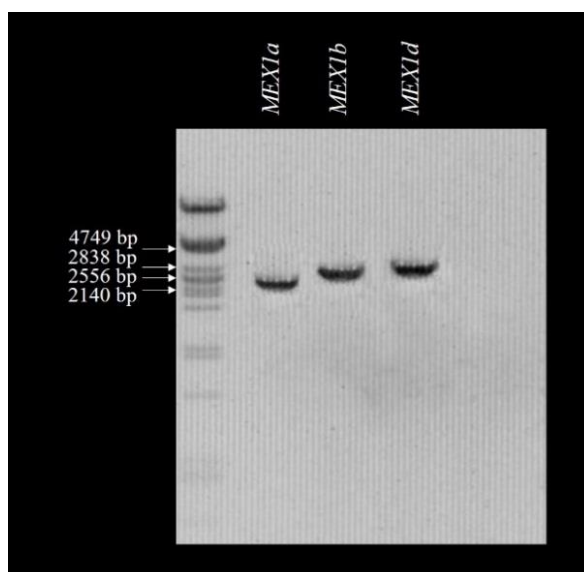
#### 3.4.2.1 *P. patens* knockout vectors

*P. patens* *MEX* gene sequences were amplified from wild type genomic DNA using primers that amplify both the gene and 500 bp flanking each side of the predicted start and stop codons. All *PpMEX* sequences, except for *PpMEX1c*, were successfully amplified and matched the expected size (Figure 15A). These amplicons were ligated into the pMiniT™ 2.0 vector and the *PpMEX* genes disrupted by insertion of a *neomycin phosphotransferase* (*nptII*)-expression cassette after restriction with *HindIII* (*PpMEX1a*, *PpMEX1d*) and *EcoRV* (*PpMEX1b*) restriction enzymes (Figure 14). Disruption of *PpMEX* genes by the *nptII*-cassette was confirmed via PCR using primers that bind within the *PpMEX* and *nptII* genes (Figure 16).





**Figure 15. PCR amplification of putative *MEX* orthologues identified in *P. patens*.** *PpMEX* gene sequences were amplified from (A) wild type genomic DNA and the corresponding coding sequences were amplified from (B) wild type cDNA. Three *PpMEX* sequences successfully amplified, while the *PpMEX1c* sequence did not amplify from either gDNA or cDNA.

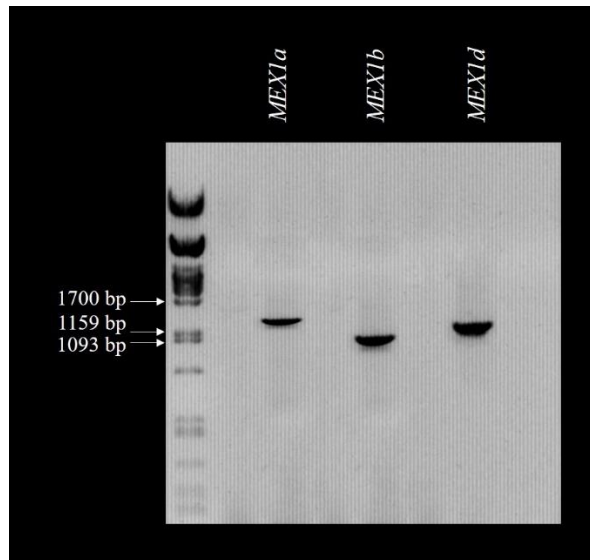


**Figure 16. Disruption of *PpMEX* genes by insertion of *nptII*-cassette during knockout vector construction.** Plasmids were screened through PCR amplification with primers that bind within the *nptII* and *PpMEX* genes. Insertions were observed for all three *PpMEX* gene constructs.

#### 3.4.2.2 *E. coli* complementation vectors

The *PpMEX* coding sequences were amplified from *P. patens* wild type cDNA using primers that amplify from the predicted start codon to the predicted stop codon and simultaneously add 5'-*XbaI* and 3'-*XhoI* restriction sites. Like the gene sequences described above, all *PpMEX* coding sequences except for *PpMEX1c* were successfully amplified (Figure 15B). The PCR product was purified and ligated into the pMiniT™ 2.0 vector before being transformed into 10-beta Competent *E. coli*. Plasmids were isolated and the *PpMEX* coding

sequences were excised using *XbaI* and *XhoI*. The fragments were purified from agarose gel and ligated into linearised (*XbaI* and *XhoI*) and dephosphorylated pBluescript SK (+). The ligation mixture was transformed into DH5 $\alpha$  competent cells and transformants were selected via blue/white screening. Correct orientation of the insert was confirmed using colony PCR with primers that bind within the *PpMEX* coding sequences and the promoter site of the vector (Figure 17).



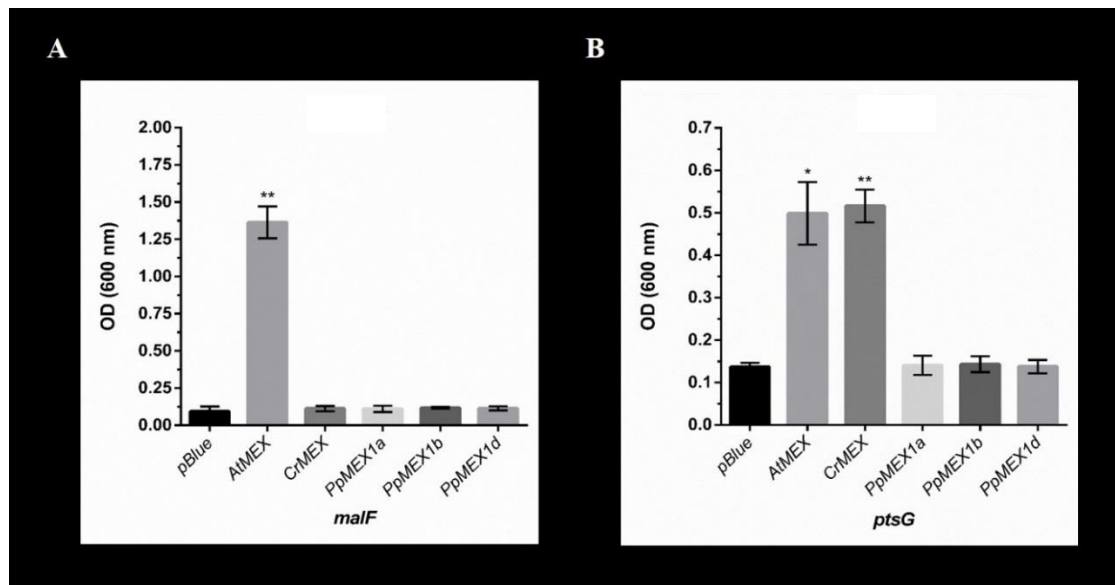
**Figure 17. Screen for correct orientation of *PpMEX* coding sequences in pBluescript during construction of complementation vectors.** Plasmids were screened with primers that bind within the promoter and at the 3'-end of the *PpMEX* coding sequences.

### 3.4.3 *E. coli* complementation analysis

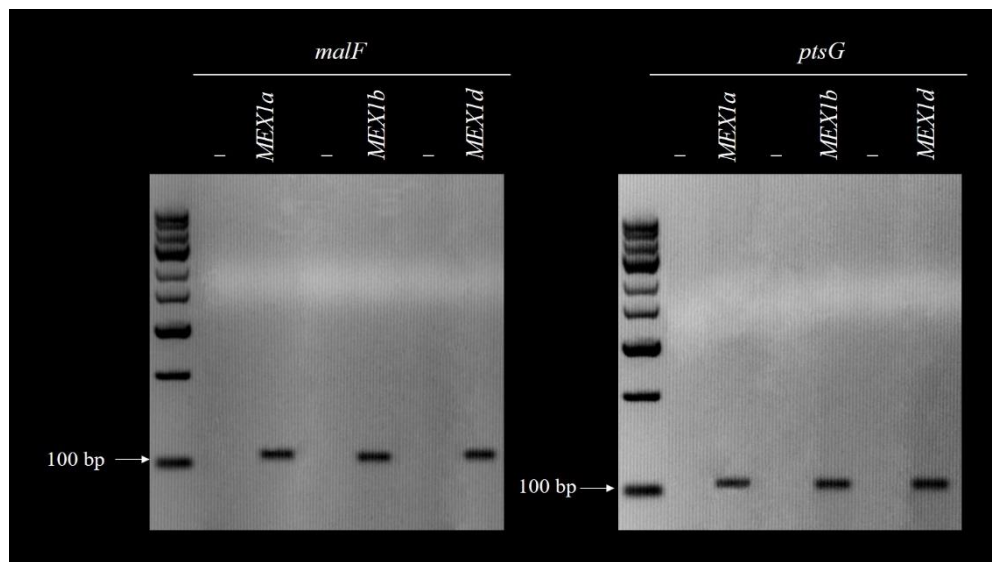
The TST6 (*malF*) and JW1087-2 (*ptsG*) *E. coli* mutant strains were used to analyse the ability of the putative *PpMEX* transporters to translocate different sugars. The TST6 strain lacks the maltose/maltodextrin transport system permease protein (*malF*), an important component of the transporter complex involved in maltose/maltodextrin import, and is unable to grow on media where maltose is the sole carbon source, while the JW1087-2 strain lacks the EIICB component (*ptsG*) of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) which is involved in glucose transport and is incapable of growth where glucose is the sole carbon source. The strains were transformed with pBluescript vectors containing the *PpMEX* coding sequences (*PpMEX1a*, *PpMEX1b*, and *PpMEX1d*) and cultivated in M9 minimal medium containing either maltose or glucose. The bacterial density (OD<sub>600</sub>) of each culture was measured after 24 hrs of growth to assess whether the *PpMEX* transporters are capable of reverting bacterial growth on these carbon sources. Vectors containing the coding sequences from *Arabidopsis* and *Chlamydomonas MEX* genes (pQE-30::*AtMEX* and pQE-30::*CrMEX*; Findinier et al., 2017), as well as a vector without insert were used as controls.

Complementation with the *Arabidopsis* and *C. reinhardtii MEX* genes enabled growth on glucose, while only the *Arabidopsis MEX* protein was capable of reverting growth on maltose (Figure 18). None of the *PpMEX* genes were able to support growth of the *malF* or *ptsG* mutant *E. coli* strains in maltose or glucose, respectively

(Figure 18). To confirm that the lack of bacterial growth was not due to a lack of expression from the vector constructs, RT-PCR analysis was performed on RNA extracted from each culture during mid-log phase (in LB medium) to detect the *PpMEX* transcripts. All of the strains showed expression for the relevant *PpMEX* coding sequences (Figure 19).



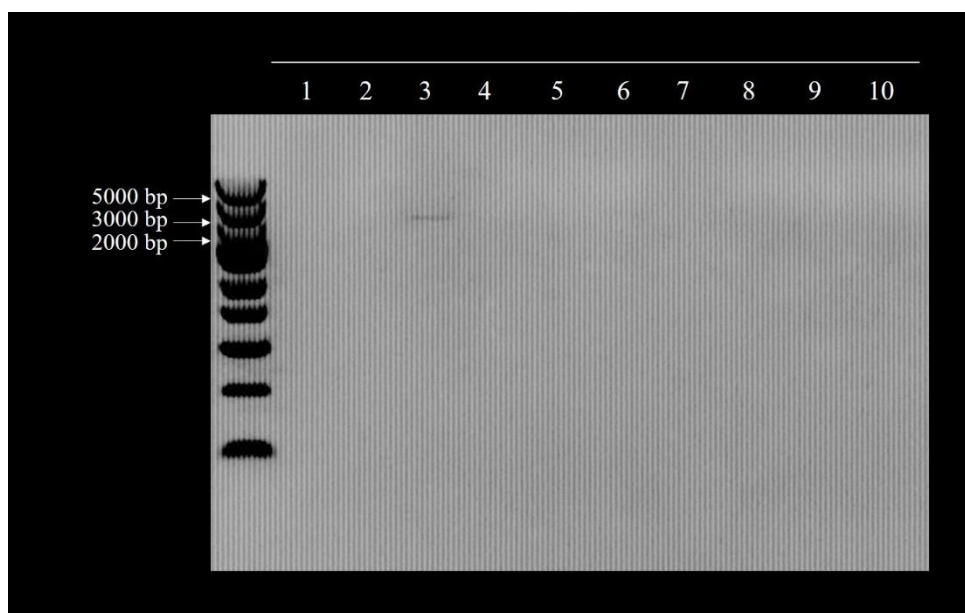
**Figure 18. Complementation of *E. coli malF* and *ptsG* mutants with *P. patens MEX* coding sequences.** Cultures were grown in M9 minimal medium containing either 0.5% (w/v) maltose (A) or glucose (B). Vectors containing the *MEX* coding sequences from Arabidopsis (*AtMEX*) and *C. reinhardtii* (*CrMEX*) were used as positive controls. Bacterial density was measured as OD<sub>600</sub> after 24 hrs of growth. Results are expressed as the mean  $\pm$  SD of three independent cultures (Student's t-test). \* P-value < 0.05; \*\* P-value < 0.01.



**Figure 19. Expression of *PpMEX* coding sequences in *E. coli malF* and *ptsG* mutant strains.** RT-PCR analysis was performed on RNA extracted at mid-log phase from bacteria harbouring plasmids containing the *PpMEX* coding sequences. Expression was observed for all *PpMEX* genes in both the *malF* and *ptsG* strains. A control reaction where reverse transcriptase was omitted was included for each strain (-).

### 3.4.4 Generation and screening of *P. patens* knockout mutants

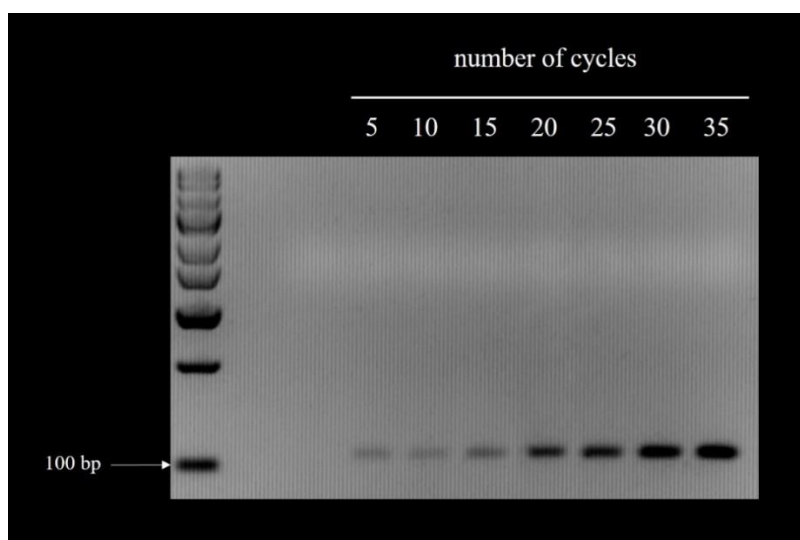
Knockout mutant lines for the *PpMEX* genes were generated by transformation of protoplasts with the knockout vectors and subsequent homologous recombination of the endogenous gene with the version disrupted by a *nptII*-cassette. Transformed protoplasts were screened for resistance to G418 (transferred by *nptII*) by alternating growth on medium containing antibiotic selection and medium without selection for two rounds. Transformants that survived the second round of selection were considered stable transformants (due to the permanent integration of the disrupted gene into the genome). Only colonies of the *PpMEX1a* knockout mutant line survived and a total of 10 individual colonies were selected for genotyping. *P. patens* tissue was used directly in PCR amplification to screen for disruption of the *PpMEX1a* gene using primers that bind within the *nptII* cassette and the *PpMEX1a* gene. An amplicon of the expected size was observed only for colony 3, indicating successful disruption of the *PpMEX1a* gene in this line (Figure 20).



**Figure 20. Genotyping of *P. patens* *MEX1a* knockout mutant lines.** Ten individual colonies were screened for disruption of *PpMEX1a* using primers that bind within the *nptII* and *PpMEX1a* genes. Colony 3 showed successful disruption of the *PpMEX1a* gene by the *nptII* cassette.

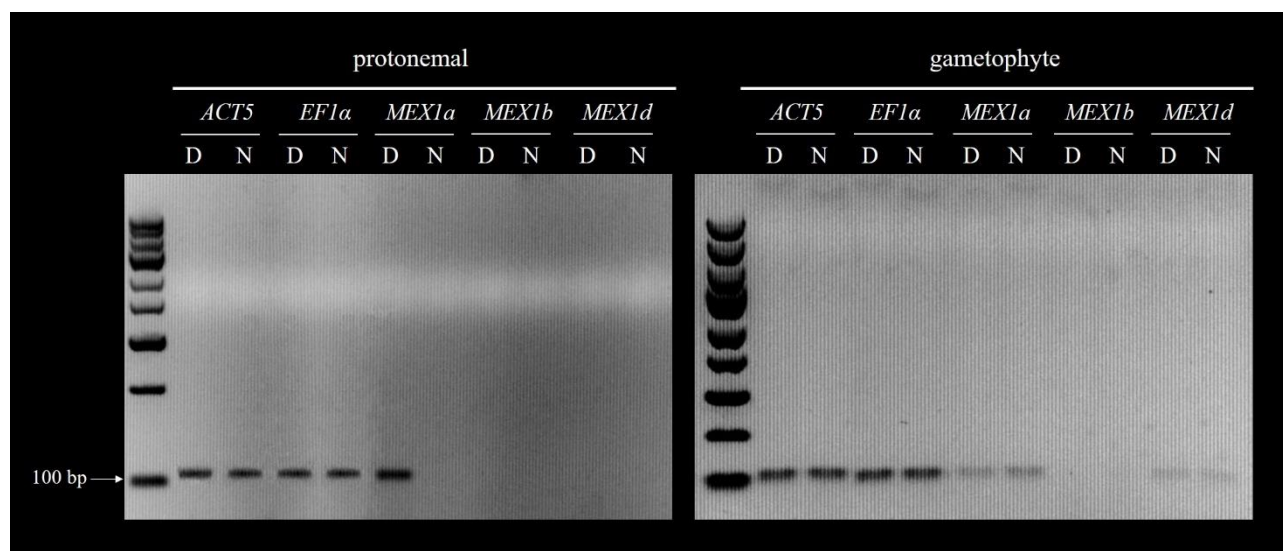
### 3.4.5 *P. patens* *MEX* expression profile

Semi-quantitative RT-PCR was used to determine the expression of *PpMEX* genes in wild type *P. patens*. RNA was extracted from both protonemal and gametophyte tissue of approximately 4-week old cultures at the start of the light period (day) and at the end of the light period (night). *Actin 5* (*ACT5*) and *elongation factor 1-alpha* (*EF1 $\alpha$* ) were used as reference genes (Le Bail et al., 2013). Primers that amplify approximately 100 bp fragments spanning exon-exon junctions were designed for each of the four *PpMEX* genes and the two reference genes. The linear range for *ACT5* was determined to be 25 cycles of amplification (Figure 21) and thus subsequent sqRT-PCR analysis of *PpMEX* gene expression was conducted at 25 cycles.



**Figure 21. Determination of the linear range of *actin5* (*ACT5*) amplification for subsequent sqRT-PCR analysis.** The linear range was determined to occur during 25 cycles of amplification.

In protonemal tissue, *PpMEX1a* showed high expression during the end of the light period (Figure 22). No expression of other *PpMEX* genes could be detected in protonemal tissue at either of the timepoints. In gametophytic tissue, *PpMEX1a* and *PpMEX1d* showed expression during the day and night. In the case of *PpMEX1a*, expression seems to be slightly higher during the night, while expression of *PpMEX1d* seems to be slightly higher during the day. No expression could be detected for *PpMEX1b*. As neither the gene sequence nor the coding sequence for *PpMEX1c* could be amplified during initial vector construction (section 3.4.2), the efficacy of the sqRT-PCR primer pair for this gene could not be evaluated effectively and thus cannot be ruled out as a contributing factor to the lack of expression observed for this gene (results not shown).



**Figure 22. Expression profile of *PpMEX* genes in protonemal and gametophytic tissue of wild type *P. patens*.** RNA was extracted from 4-week old cultures at the end of the day (D) and at the end of the night (N). Semi-quantitative PCR was performed to assess expression of *PpMEX* genes. *Actin 5* (*ACT5*) and *elongation factor 1- $\alpha$*  (*EF1 $\alpha$* ) were used as reference genes. Expression of *PpMEX1a* and *PpMEX1d* was observed at both time points in gametophytic tissue. Expression of *PpMEX1a* was also observed at the end of the day period in protonemal tissue.

### 3.5 Discussion

In this study, the presence of orthologues of the Arabidopsis MEX1 transporter (Niittylä et al., 2004) was investigated in the moss *P. patens*. Four putative orthologues were identified during the initial homology search against the *P. patens* genome and these genes were arbitrarily designated *PpMEX1a*, *PpMEX1b*, *PpMEX1c* and *PpMEX1d*. Phylogenetic analysis of the *PpMEX* protein sequences suggests that the *P. patens* MEX transporters are indeed orthologues of Arabidopsis MEX1 and that all green algal and plant MEX proteins identified share a common ancestor which likely inherited and derived this gene from cyanobacterial origin (Figure 13). *PpMEX1b* appears to be the least phylogenetically related to the other *PpMEX* proteins (Figure 13) and exhibits a markedly lower nucleotide identity to that of the Arabidopsis MEX1 and the lowest amino acid similarity to both the Arabidopsis MEX1 and other *PpMEX* sequences (Table 6; Table 7). The substantial difference between *P. patens* and higher plants with regards to the number of MEX isoforms can be attributed to the occurrence of at least one whole genome duplication event in *P. patens* between 30 and 60 million years ago (Rensing et al., 2007). Following this genome duplication event, there was a significant increase in the number of genes involved in metabolism, including those involved in the pathway of starch synthesis and degradation (Rensing et al., 2008; Rensing et al., 2013). This is evident when comparing the number of isoforms of starch metabolic enzymes between *P. patens* and other plants (Table 1). Divergence of *PpMEX1b* and clustering of the other three *PpMEX* proteins during phylogenetic analysis (Figure 13) could indicate that *P. patens* obtained two MEX genes following one of its reported whole genome duplications, one of which evolved into *PpMEX1b* and the other of which formed the common ancestor gene to *PpMEX1a*, *PpMEX1c* and *PpMEX1d* and further diverged due to more duplication events. This assumption is further supported by the high amino acid similarity (86.5%) between *PpMEX1c* and *PpMEX1d* (Table 7) and their apparent identical exon-intron structure (Figure 14) which could be an artefact of a duplication event.

The Arabidopsis MEX1 protein is a chloroplastic membrane protein that has been demonstrated to translocate maltose from the plastid to the cytosol (Niittylä et al., 2004). To determine whether the *P. patens* MEX proteins share this subcellular localisation, the *PpMEX* sequences were analysed for the presence of N-terminal signal peptides and transmembrane domains to indicate that the proteins are targeted to the chloroplast and embedded within the membrane. While all of the *PpMEX* sequences possess transmembrane domains and are thus likely enveloped within a membrane, only *PpMEX1a* and *PpMEX1b* harbour an N-terminal chloroplast signal peptide like the Arabidopsis MEX1 (Table 6). The use of fluorescence protein tags and visualisation by confocal microscopy in other studies has demonstrated that the Arabidopsis and *C. reinhardtii* MEX proteins localise to the chloroplast envelope membrane (Niittylä et al., 2004; Findinier et al., 2017). Thus, further *in vivo* analysis of the subcellular localisation of these *PpMEX* proteins through fusion with a green fluorescence protein (GFP)-tag would be useful in determining whether these proteins are indeed embedded within the chloroplastial membrane like their counterparts in other green plants.

Of the four *P. patens* MEX proteins, both the genes and coding sequences could only be amplified for *PpMEX1a*, *PpMEX1b* and *PpMEX1d* (Figure 15). The lack of *PpMEX1c* amplification can be explained by four scenarios: either 1) the PCR cycling conditions were not properly optimised for amplification of

*PpMEX1c*; 2) the primers used during amplification did not bind the sequences effectively or the primer pairs were incompatible with regards to e.g. melting temperatures; 3) *PpMEX1c* is a pseudogene and could be the result of an error occurring during reconstruction of the *P. patens* genome; or 4) the *PpMEX* genes are expressed in a tissue-specific manner and the tissue type where *PpMEX1c* is expressed was not included in the sample from which nucleic acids were extracted. To counter the first scenario, a gradient PCR with a range of 12°C was performed to find the optimal annealing temperature of the primers. Furthermore, two different high-fidelity DNA polymerases were utilised to attempt recovery of the sequences, but no amplification could be observed. As some discrepancies could exist between the sequences deposited on Phytozome (Goodstein et al., 2012) and the actual gene and coding sequences resulting in ineffective binding of the primers, a variety of primers were designed that bind within different regions of the sequences to reduce the likelihood of this occurring. These primers were also used in several combinations, but neither of the combinations yielded amplification. Although the lack of amplification of the *PpMEX1c* gene sequence suggests that it is likely a pseudogene. In disagreement with this, there is expression data reported for *PpMEX1c* mRNA transcripts (Phytozome). As these transcripts were only detected in sporophyte tissue, it explains the inability to amplify a *PpMEX1c* coding sequence as the cDNA used as template in this study was synthesised from RNA that was extracted from a mixture of gametophytic and protonemal tissue. However, it still does not resolve the failure in isolating the *PpMEX1c* gene sequence.

To investigate the function of the MEX proteins in *P. patens*, knockout vectors containing a version of the *PpMEX* gene sequences disrupted by a *nptII*-cassette were constructed and used to transform wild type *P. patens*. Only *PpMEX1a* knockout mutants survived both rounds of screening by antibiotic selection and one mutant line showed successful disruption of the *PpMEX1a* gene (Figure 20). At the time of writing, there was not enough tissue available to confirm that expression of *PpMEX1a* had been repressed. Although incorporation of the disrupted *PpMEX1a* gene into the genome was confirmed, it is possible that the construct integrated into a random region and did not replace the endogenous copy of the *PpMEX1a* gene. Therefore, methods of transcript expression analysis such as sqRT-PCR or qPCR are essential in ascertaining that expression has been repressed. The possible phenotypic differences between wild type *P. patens* and the knockout mutant line could also not be examined due to the lack of tissue. To fully analyse any phenotypic effect that disruption of the *PpMEX1a* gene may manifest in *P. patens*, the mutant line will have to be grown on a variety of growth media containing, for example, different sugars as these compounds elicit different phenotypes in wild type *P. patens* cultures. The growth of various tissue types and structures, such as protonema, gametophytes and sporophytes, will need to be compared to the wild type structures.

As a precursor to the analysis of *PpMEX* gene expression in knockout mutant lines, a preliminary expression profile of the *PpMEX* genes in wild type *P. patens* was composed for comparison. In this study, protonemal and gametophytic tissue growing on BCD solid medium was harvested at the end of the light period (day) at the onset of diurnal starch degradation, and at the end of the dark period (night) when starch degradation comes to a halt. At these time points, starch metabolic genes should theoretically be expressed at their highest and lowest, respectively. High expression of *PpMEX1a* was observed at the end of the day in protonemal tissue,

while low expression of both *PpMEX1a* and *PpMEX1d* was observed at both time points in gametophytic tissue (Figure 22). These observations correlate with the available microarray data on Phytozome which indicates that *PpMEX1a* shows low expression in gametophores, leaflets and sporophytes, while high expression has been observed in protonema (specifically protonemal tissue growing on BCD solid medium). Growth of protonemal tissue under stress conditions such as high temperatures or altered light conditions (extended darkness, red light, or continuous light) is correlated with low to no expression of *PpMEX1a*. In the case of *PpMEX1d*, the microarray data demonstrates that low levels of expression have been observed with protonemal growth on Knop medium or BCD medium supplemented with ammonia, but that there is no expression during protonemal growth on BCD medium which corroborates the lack of expression observed in protonemal tissue during this study (Figure 22). *PpMEX1d* expression in gametophytic tissue (Figure 22) is also supported by the microarray data which shows high *PpMEX1d* transcript levels in gametophores and sporophytes. It would therefore be valuable to expand the expression profile of the *PpMEX* genes to include a wider variety of tissue types, growth media and growth conditions as these components all appear to play a role in their expression.

The lack of expression of *PpMEX1b* or *PpMEX1c* in this study is also corroborated by the corresponding microarray data as *PpMEX1b* shows very low to no expression in protonemal tissue and only low levels of expression in gametophores growing on BCD medium, while *PpMEX1c* expression is limited to sporophytes. It is possible that expression levels of *PpMEX1b* were too low to detect using sqRT-PCR analysis as the coding sequence was successfully amplified, albeit in low amounts, during end-point PCR using cDNA synthesised from a mixture of gametophytic and protonemal RNA (Figure 15). A more sensitive method (such as qPCR) might be preferable for detecting expression of the *PpMEX* genes, particularly when they are expressed at low levels such as in the case of *PpMEX1b*. As only protonemal and gametophytic tissue were analysed in this study, the absence of *PpMEX1c* expression can be attributed to the lack of sporophyte tissue although it is possible that the primers designed to detect *PpMEX1c* expression did not effectively bind or amplify the desired region. In this case, however, the former seems a more likely explanation. Regardless, as neither the gene nor the coding sequence of *PpMEX1c* could be amplified during the initial isolation of the *PpMEX* sequences, the ability of the primer pair to amplify the desired region of the coding sequence could not be confirmed using the initial sequences as template (as was the case for the other *PpMEX* sequences).

Observation of the expression of *PpMEX1a* in protonemal tissue during this study, as well as the microarray data showing that *PpMEX1c* expression is limited to sporophyte tissue could suggest that the different *PpMEX* proteins have been specialised to perform their functions during specific life stages or are targeted to particular growth structures. It is unknown whether the *PpMEX* transporters have redundant or overlapping functions. Therefore, it would be interesting to observe whether expression of the other *PpMEX* genes change upon repression of one *PpMEX* gene and whether these transporters are interchangeable and capable of compensating for the loss of function of their counterparts. Furthermore, generation of double and triple mutants would be beneficial in determining the specific functions of the *PpMEX* transporters or if and how they work synergistically during starch degradation in *P. patens*. Analysis of the expression of the different



*PpMEX* genes at different time points or within different tissues could be performed by isolating the promoter sequences that drive their expression and fusing these to a GUS reporter gene. *P. patens* mutants transformed with these constructs would aid in visualisation of where gene expression is localised.

Once *P. patens* knockout mutant lines have been generated for all the *PpMEX* genes, it would be essential to analyse differences in their starch and metabolite content to determine the effect of loss of function of the *PpMEX* proteins. In *Arabidopsis* and *C. reinhardtii*, mutant plants deficient in MEX activity demonstrate a starch excess phenotype, characterised by the accumulation of high levels of starch due to an inability to degrade this polyglucan and severely impaired growth (Niittylä et al., 2004; Jang et al., 2015; Findinier et al., 2017). Intriguingly, although the starch excess phenotype in *Arabidopsis* is accompanied by the accumulation of high levels of plastidial maltose (Niittylä et al., 2004; Stettler et al., 2009), neither wild type *C. reinhardtii* nor *MEX1* mutants contain detectable levels of maltose (Findinier et al., 2017). It would be interesting to examine whether the *PpMEX* proteins share the function of their *Arabidopsis* or *C. reinhardtii* counterparts and whether loss of these genes lead to starch accumulation with or without maltose accumulation.

Lastly, this study examined the ability of the *PpMEX* transporters to translocate sugars other than maltose, as has been observed for the *Arabidopsis* and *C. reinhardtii* MEX proteins. Upon complementation of *E. coli* mutant strains lacking the endogenous maltose (*malF*) or glucose (*ptsG*) transporters with the *Arabidopsis MEX1* coding sequence, the ability to grow on minimal media containing either maltose or glucose as the sole carbon source was restored (Findinier et al., 2017). This implies a dual function for the *Arabidopsis* protein in transporting both maltose and glucose, which correlates with the current view of the pathway of starch degradation where maltose is the major end-product exported to the cytosol, while smaller amounts of glucose are also exported from the plastid (Weber et al., 2000; Niittylä et al., 2004; Stettler et al., 2009). In contrast, the *C. reinhardtii* protein was only capable of reverting growth on glucose, not maltose (Findinier et al., 2017), suggesting that the ability to transport maltose could be a function that developed in the common ancestor of land plants after divergence of the green algal lineage. The *P. patens* coding sequences that were successfully amplified (*PpMEX1a*, *PpMEX1b* and *PpMEX1d*) were used to complement the *E. coli malF* and *ptsG* mutants. As expected, strains expressing the *Arabidopsis* and *C. reinhardtii MEX* coding sequences exhibited growth in glucose (*Arabidopsis* and *C. reinhardtii*) and maltose (*Arabidopsis*), but no significant change in bacterial growth could be detected in either glucose or maltose for any of the *PpMEX* genes (Figure 18).

To determine whether the lack of bacterial growth was due to a lack of expression, RT-PCR analysis was performed on the cultures during mid-log phase to detect *PpMEX* transcripts. Expression was observed for all *PpMEX* genes in both the *malF* and *ptsG* strains (Figure 19), indicating that the constructs were successfully expressing the sequences. However, it is possible that the mRNA transcripts were not translated into protein, the protein was misfolded and subsequently inactive or that protein was not targeted to the outer *E. coli* membrane. I utilised pBluescript as the vector driving expression of the *PpMEX* genes under the control of the lactose promoter, while the *Arabidopsis* and *C. reinhardtii MEX* sequences were ligated into pQE-30 under the control of the T5 promoter. It is possible that expression under the control of the T5 promoter would be more favourable for producing active protein. Due to the lack of a protein affinity tag in the pBluescript vector

or antibodies capable of binding the *PpMEX* transporters, analysis of *PpMEX* protein expression could not be performed and it is unclear whether the mRNA transcripts were indeed translated into soluble protein. Future experiments should focus on ligating these *PpMEX* coding sequences into the pQE-30 vector like the *Arabidopsis* and *C. reinhardtii* sequences to determine whether MEX activity is improved. The pQE-30 vector also has an N-terminal histidine tag to allow for protein analysis. It would also be valuable to add a native *E. coli* membrane peptide to the *PpMEX* coding sequences to target expression of the *PpMEX* proteins to the outer *E. coli* membrane to ensure that they are embedded in the membrane to allow uptake of sugars present in the growth media.

In a previous study, complementation of the *Arabidopsis mex1* mutant with the *C. reinhardtii MEX* sequence resulted in a wild type growth phenotype, but both starch and maltose still accumulated in the leaves of the transgenic plants indicating a low or absent capacity to revert maltose transport (Findinier et al., 2017). Replicating this experiment with the *P. patens MEX* sequences, in addition to complementation of the *E. coli* mutants, would provide valuable insight into the evolutionary development of MEX transporters by determining whether some or all of the *PpMEX* proteins have the capacity to transport maltose and/or glucose.

## Chapter 4: General conclusion

This study aimed to investigate two aspects of the starch metabolic pathway in plants. During both parts, a focus was placed on the comparison of starch metabolism between lower plants (algae and mosses) and higher plants. The aim of the first experimental chapter (Chapter 2) was to examine the function of a starch synthase from the red alga *C. crispus* through recombinant expression and biochemical characterisation. No active soluble protein could be recovered as the recombinant protein was insoluble and accumulated as inclusion bodies, despite various attempts to optimise expression and purification. However, key features were identified in the *C. crispus* and other red algal starch synthases which could contribute to the reported preference of these enzymes for UDP-glucose as substrate, unlike bacterial and green plant starch synthases which utilise ADP-glucose. A series of vectors containing mutated versions of the *C. crispus* starch synthase were constructed through site-directed mutagenesis of key residues identified during a literature search and multiple sequence alignment of starch synthases from different organisms. The second experimental chapter (Chapter 3) aimed to analyse whether the maltose transporter plays the same role during starch degradation in the moss *P. patens* as it does in Arabidopsis leaves by generating knockout mutants and performing complementation studies in *E. coli*. Phylogenetic analysis suggests that the *MEX* gene in the common ancestor of all plants has a cyanobacterial origin and displays a possible pattern of gene duplication in *P. patens* that resulted in four *MEX* genes as opposed to the one or two isoforms present in other green plants. Expression analysis of the *P. patens* *MEX* genes correlated with the available microarray data from previous studies and suggest that the proteins perform specialised functions in different tissue types. However, complementation of *E. coli* mutant strains was unsuccessful and *P. patens* knockout mutant lines could not be analysed due to time constraints.

Although this study answered some of the questions that were raised, many more remain unanswered and future studies are imperative in accomplishing this task. In order to draw conclusions about regions of substrate specificity in ADP- or UDP-glucose utilising starch synthases, the expression of the *C. crispus* starch synthase needs to be optimised to obtain active protein for biochemical characterisation. Although *in silico* comparison of sequence data can be useful for extrapolating features based on similarities and differences between proteins, this type of analysis is limiting, and experimental data is required to corroborate any deductions. Likewise, the effects on starch metabolism and growth in *P. patens* knockout mutant lines remains to be analysed and complementation studies are still incomplete.

## Literature Cited

- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, et al** (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**: 399–451
- Asatsuma S, Sawada C, Itoh K, Okito M, Kitajima A, Mitsui T** (2005) Involvement of  $\alpha$ -amylase I-1 in starch degradation in rice chloroplasts. *Plant Cell Physiol* **46**: 858–869
- Ashton N, Cove D, Featherstone D** (1979) The isolation and physiological analysis of mutants of the moss, *Physcomitrella patens*, which over-produce gametophores. *Planta* **144**: 437–442
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H** (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* **2**: 1–11
- Le Bail A, Scholz S, Kost B** (2013) Evaluation of reference genes for RT qPCR analyses of structure-specific and hormone regulated gene expression in *Physcomitrella patens* gametophytes. *PLoS One* **8**: e70998
- Ball S, Colleoni C, Arias M** (2015) Transition from glycogen to starch metabolism in cyanobacteria and eukaryotes. In Y Nakamura, ed, *Starch Metab. Struct.* Springer, Japan, pp 93–158
- Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C** (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot* **62**: 1775–1801
- 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
- Becker D, Lundblad V** (1994) Introduction of DNA into yeast cells. *Curr. Protoc. Mol. Biol.* **27**: 1934–3639
- Beckles D, Smith A, Ap Rees T** (2001) A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. *Plant Physiol* **125**: 818–827
- Bhattacharya D, Price D, Chan C, Qiu H, Rose N, Ball S, Weber A, Arias M, Henrissat B, Coutinho P, et al** (2013) Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* **4**: 1941
- Birnboim H, Doly J** (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513–1523
- Bouarb K, Potin P, Correa J, Kloareg B** (1999) Sulfated oligosaccharides mediate the interaction between a marine red alga and its green algal pathogenic endophyte. *Plant Cell* **11**: 1653–1650
- Bowles D, Isayenkova J, Lim E, Poppenberger B** (2005) Glycosyltransferases: managers of small molecules. *Curr Opin Plant Biol* **8**: 254–263

- Bowman J, Kohchi T, Yamato K, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, et al** (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**: 287–304
- Boyer C, Preiss J** (1979) Properties of citrate-stimulated starch synthesis catalysed by starch synthase I of developing maize kernels. *Plant Physiol* **64**: 1039–1042
- Brock T** (1978) The genus *Cyanidium*. In T Brock, ed, *Thermophilic Microorg. Life High Temp.* Springer-Verlag, New York, pp 255–302
- Burton R, Bewley J, Smith A, Bhattacharyya M, Tatge H, Ring S, Bull V, Hamilton W, Martin C** (1995) Starch branching enzymes belonging to distinct families are differentially expressed during pea embryo development. *Plant J* **7**: 3–15
- Burton R, Jenner H, Carrangis L, Fahy B, Fincher G, Hylton C, Laurie D, Parker M, Waite D, Van Wegen S, et al** (2002) Starch granule initiation and growth are altered in barley mutants that lack isoamylase activity. *Plant J* **31**: 97–112
- Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM** (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *EMBO J* **23**: 3196–3205
- Bustos R, Fahy B, Hylton C, Seale R, Nebane N, Edwards A, Martin C, Smith A** (2004) Starch granule initiation is controlled by a heteromultimeric isoamylase in potato tubers. *Proc Natl Acad Sci U S A* **101**: 2215–2220
- Cabrera L, Bottomley S** (2004) Protein expression and refolding - a practical guide to getting the most out of inclusion bodies. *Biotechnol Annu Rev* **10**: 31–50
- Cao H, Imparl-Radosevich J, Guan H, Keeling P, James M, Myers A** (1999) Identification of the soluble starch synthase activities of maize endosperm. *Plant Physiol* **120**: 205–216
- Cenci U, Chabi M, Ducatez M, Tirtiaux C, Nirmal-Raj J, Utsumi Y, Kobayashi D, Sasaki S, Suzuki E, Nakamura Y, et al** (2013) Convergent evolution of polysaccharide debranching defines a common mechanism for starch accumulation in cyanobacteria and plants. *Plant Cell* **25**: 3961–3975
- Ceroni A, Passerini A, Vullo A, Frasconi P** (2006) DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res* **34**: W177-181
- Chia T, Thorneycroft D, Chapple A, Messerli G, Chen J, Zeeman S, Smith S, Smith A** (2004) A cytosolic glucosyltransferase is required for conversion of starch to sucrose in *Arabidopsis* leaves at night. *Plant J* **37**: 853–863
- Cho K, Yoo Y, Kang H** (1999) delta-Integration of endo/exo-glucanase and beta-glucosidase genes into the yeast chromosome for direct conversion of cellulose to ethanol. *Enzyme Microb Technol* **25**: 25–30

- Cho M, Lim H, Shin D, Jeon J, Bhoo S, Park Y, Hahn T** (2011) Role of the plastidic glucose translocator in the export of starch degradation products from the chloroplasts in *Arabidopsis thaliana*. *New Phytol* **190**: 101–112
- Collén J, Davison I** (1999) Stress tolerance and reactive oxygen metabolism in the intertidal red seaweeds *Mastocarpus stellatus* and *Chondrus crispus*. *Plant Cell Environ* **22**: 1143–1151
- Collén J, Guisle-Marsollier I, Léger J, Boyen C** (2007) Response of the transcriptome of the intertidal red seaweed *Chondrus crispus* to controlled and natural stresses. *New Phytol* **176**: 45–55
- Collén J, Hervé C, Guisle-Marsollier I, Léger J, Boyen C** (2006) Expression profiling of *Chondrus crispus* (Rhodophyta) after exposure to methyl jasmonate. *J Exp Bot* **57**: 3869–3881
- Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, Barbeyron T, Michel G, Noel B, Valentin K, et al** (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A* **110**: 5247–5252
- Collier PA, Hughes KW** (1982) Life cycle of the moss, *Physcomitrella patens*, in culture. *J Tissue Cult Methods* **7**: 19–22
- Commuri P, Keeling P** (2001) Chain-length specificities of maize starch synthase I enzyme: studies of glucan affinity and catalytic properties. *Plant J* **25**: 475–486
- Coppin A, Varré JS, Lienard L, Dauvillée D, Guérardel Y, Soyer-Gobillard MO, Buléon A, Ball S, Tomavo S** (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *J Mol Evol* **60**: 257–267
- Coutinho P, Deleury E, Davies G, Henrissat B** (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**: 307–317
- Cove D, Perroud P, Charron A, McDaniel S, Khandelwal A, Quatrano R** (2009) The moss *Physcomitrella patens*: a novel model system for plant development and genomic studies. *Cold Spring Harb Protoc* **2**: 69–104
- Craig J, Lloyd J, Tomlinson K, Barber L, Edwards A, Wang T, Martin C, Hedley C, AM S** (1998) Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. *Plant Cell* **10**: 413–426
- Crane P, Kenrick P** (1997) Diverted development of reproductive organs: a source of morphological innovation in land plants. *Plant Syst Evol* **206**: 161–174
- Criscuolo A, Gribaldo S** (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**: 210
- Critchley J, Zeeman S, Takaha T, Smith A, Smith S** (2001) A critical role for disproportionating enzyme

in starch breakdown is revealed by a knock-out mutation in Arabidopsis. *Plant J* **26**: 89–100

- Cuesta-Seijo J, Nielsen M, Marri L, Tanaka H, Beeren S, Palcic M** (2013) Structure of starch synthase I from barley: insight into regulatory mechanisms of starch synthase activity. *Acta Crystallogr.* **69**: 1013–1025
- Curtis M, Grossniklaus U** (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**: 462–469
- Dauvillée D, Colleoni C, Mouille G, Buléon A, Gallant D, Bouchet B, Morell M, D’Hulst C, Myers A, Ball S** (2001a) Two loci control phytyloglycogen production in the monocellular green alga *Chlamydomonas reinhardtii*. *Plant Physiol* **125**: 1710–1722
- Dauvillée D, Colleoni C, Mouille G, Morell M, D’Hulst C, Wattedled F, Liénard L, Delvallé D, Ral J, Myers A, et al** (2001b) Biochemical characterization of wild-type and mutant isoamylases of *Chlamydomonas reinhardtii* supports a function of the multimeric enzyme organization in amylopectin maturation. *Plant Physiol* **125**: 1723–1731
- Delatte T, Trevisan M, Parker M, Zeeman S** (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
- Delatte T, Umhang M, Trevisan M, Eicke S, Thorneycroft D, Smith S, Zeeman S** (2006) Evidence for distinct mechanisms of starch granule breakdown in plants. *J Biol Chem* **281**: 12050–12059
- Delrue B, Fontaine T, Routier F, Decq A, Wieruszkeski J, Van Den Koornhuysse N, Maddelein M, Fournet B, Ball S** (1992) Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin. *J Bacteriol* **174**: 3612–3620
- Denver K, Barber L, Burton R, Hedley C, Hylton C, Johnson S, Jones D, Marshall J, Smith A, Tatge H, et al** (1995) The isolation and characterisation of novel low-amylose mutants of *Pisum sativum* L. *Plant, Cell Environ.* **18**: 1019–1026
- Denyer K, Dunlap F, Thorbjornsen T, Keeling P, Smith A** (1996) The major form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. *Plant Physiol* **112**: 779–785
- Denyer K, Waite D, Motawia S, Moller B, Smith A** (1999) Granule-bound starch synthase I in isolated starch granules elongates malto-oligosaccharides processively. *Biochem J* **340**: 183–191
- Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Putaux JL, Buléon A, Haebel S, Ritte G, Steup M, Falcón LI, et al** (2008a) Metabolic symbiosis and the birth of the plant kingdom. *Mol Biol Evol* **25**: 536–548
- Deschamps P, Haferkamp I, D’Hulst C, Neuhaus H, Ball S** (2008b) The relocation of starch metabolism to

chloroplasts: when, why and how. *Trends Plant Sci* **13**: 574–582

- Deschamps P, Moreau H, Worden AZ, Dauvillée D, Ball SG** (2008c) Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. *Genetics* **178**: 2373–2387
- Dunn B, Wobbe R** (1993) Preparation of protein extracts from yeast. *In* F Ausubel, R Brent, R Kingston, D Moore, J Seidman, J Smith, K Struhl, eds, *Curr. Protoc. Mol. Biol.* John Wiley & Sons, New York, p 13.3.1-13.3.9
- Edwards A, Borthakur A, Bornemann S, Venail J, Denyer K, Waite D, Fulton D, Smith A, Martin C** (1999a) Specificity of starch synthase isoforms from potato. *Eur J Biochem* **266**: 724–736
- Edwards A, Fulton D, Hylton C, Jobling S, Gidley M, Rossner U, Martin C, Smith A** (1999b) A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers. *Plant J* **17**: 251–261
- Edwards K, Johnstone C, Thompson C** (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* **19**: 1349
- Edwards S, Marshall J, Sidebottom C, Visser R, Smith A, Martin C** (1995) Biochemical and molecular characterization of a novel starch synthase from potato tubers. *Plant J* **8**: 283–294
- Emanuelsson O, Nielsen H, Von Heijne G** (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**: 978–984
- Engel PP** (1968) The induction of biochemical and morphological mutants in the moss *Physcomitrella patens*. *Am J Bot* **55**: 438–446
- Esposito D, Chatterjee D** (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* **17**: 353–358
- Farkas I, Hardy T, Goebbt MG, Roach PJ** (1991) Two glycogen synthase isoforms in *Saccharomyces cerevisiae* are coded by distinct genes that are differentially controlled. *J Biol Chem* **266**: 15602–15607
- Ferre F, Clote P** (2006) DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res* **34**: W182-185
- Ferre F, Clote P** (2005a) DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res* **33**: W230-232
- Ferre F, Clote P** (2005b) Disulfide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics* **21**: 2336–2346
- Findinier J, Tunçay H, Schulz-Raffelt M, Deschamps P, Spriet C, Lacroix J, Duchêne T, Szydlowski N, Li-Beisson Y, Peltier G, et al** (2017) The *Chlamydomonas mex1* mutant shows impaired starch



mobilization without maltose accumulation. *J Exp Bot* **68**: 5177–5189

**Frederick J** (1967) Glucosyltransferase isozymes in algae. *Phytochemistry* **6**: 1041–1045

**Fujita N, Goto S, Yoshida M, Suzuki E, Nakamura Y** (2008) The function of rice starch synthase I expressed in *Escherichia coli*. *J Appl Glycosci* **55**: 167–172

**Fujita N, Satoh R, Hayashi A, Kodama M, Itoh R, Aihara S, Nakamura Y** (2011) Starch biosynthesis in rice endosperm requires the presence of either starch synthase I or IIIa. *J Exp Bot* **62**: 4819–4831

**Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y** (2006) Function and characterization of starch synthase I using mutants in rice. *Plant Physiol* **140**: 1070–1084

**Fulton D, Edwards A, Pilling E, Robinson H, Fahy B, Seale R, Kato L, Donald A, Geigenberger P, Martin C, et al** (2002) Role of granule-bound starch synthase in determination of amylopectin structure and starch granule morphology in potato. *J Biol Chem* **277**: 10834–10841

**Fulton D, Stettler M, Mettler T, Vaughan C, Li J, Francisco P, Gil M, Reinhold H, Eicke S, Messerli G, et al** (2008) Beta-AMYLASE4, a noncatalytic protein required for starch breakdown, acts upstream of three active beta-amylases in *Arabidopsis* chloroplasts. *Plant Cell* **20**: 1040–1058

**Furukawa K, Tagaya M, Inouye M, Preiss J, Fukui T** (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. *J Biol Chem* **265**: 2086–2090

**Furukawa K, Tagaya M, Tanizawa K, Fukui T** (1993) Role of the conserved Lys-x-Gly-Gly sequence of the ADP-glucose-binding site in *Escherichia coli* glycogen synthase. *J Biol Chem* **268**: 23837–23842

**Gómez-Arjona F, Raynaud S, Ragel P, Mérida A** (2014) Starch synthase 4 is located in the thylakoid membrane and interacts with plastoglobule-associated proteins in *Arabidopsis*. *Plant J* **80**: 305–316

**Gao M, Wanat J, Stinard P, James M, Myers A** (1998) Characterization of dullI, a maize gene coding for a novel starch synthase. *Plant Cell* **10**: 399–412

**Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, Bairoch A** (2005) Protein identification and analysis tools on the ExpASY server. In JM Walker, ed, *Proteomics Protoc. Handb.* Humana Press, pp 571–607

**Gensel P** (2008) The earliest land plants. *Annu Rev Ecol Evol Syst* **39**: 459–477

**Gensel P, Kotyk M, Basinger J** (2001) Morphology of above- and below-ground structures in Early Devonian (Pragian-Emsian). In P Gensel, D Edwards, eds, *Plants Invade L. Evol. Perspect. Environ. Perspect.* Columbia University Press, New York, pp 83–102

**Ghosh H, Preiss J** (1966) Adenosine diphosphate glucose pyrophosphorylase. A regulatory enzyme in the biosynthesis of starch in spinach leaf chloroplasts. *J Biol Chem* **241**: 4491–4504

**Gibson K, Park J, Nagai Y, Hwang S, Cho Y, Roh K, Lee S, Kim D, Choi S, Ito H, et al** (2011) Exploiting

leaf starch synthesis as a transient sink to elevate photosynthesis, plant productivity and yields. *Plant Sci* **181**: 275–281

**Goodstein D, Shu S, Howson R, Neupane R, Hayes R, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al** (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**: D1178–D1186

**Gross W, Oesterhelt C** (1999) Ecophysiological studies on the red alga *Galdieria sulphuraria* isolated from southwest Iceland. *Plant Biol* **1**: 694–700

**Guan H, Preiss J** (1993) Differentiation of the properties of the branching isozymes from maize (zea-mays). *Plant Physiol* **102**: 1269–1273

**Hayashi M, Crofts N, Oitome N, Fujita N** (2018) Analyses of starch biosynthetic protein complexes and starch properties from developing mutant rice seeds with minimal starch synthase activities. *BMC Plant Biol* **18**: 59

**Hennen-Bierwagen T, Liu F, Marsh R, Kim S, Gan Q, Tetlow I, Emes M, James M, Myers A** (2008) Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. *Plant Physiol* **146**: 1892–1908

**Hirabaru C, Izumo A, Fujiwara S, Tadokoro Y, Shimonaga T, Konishi M, Yoshida M, Fujita N, Nakamura Y, Yoshida M, et al** (2010) The primitive rhodophyte *Cyanidioschyzon merolae* contains a semiamylopectin-type, but not an amylose-type, alpha-glucan. *Plant Cell Physiol* **51**: 682–693

**Hirose T, Aoki N, Harada Y, Okamura M, Hashida Y, Ohsugi R, Akio M, Hirochika H, Terao T** (2013) Disruption of a rice gene for  $\alpha$ -glucan water dikinase, *OsGWD1*, leads to hyperaccumulation of starch in leaves but exhibits limited effects on growth. *Front Plant Sci* **4**: 147

**Hoffmann B, Charlot F** (2009) Transformation of the moss *Physcomitrella patens*. SGAP, INRA, Versailles

**Hovenkamp-Hermelink J, Jacobsen E, Ponstein A, Visser R, Vos-Scheperkeuter G, Bijmolt E, De Vries J, Witholt B, Feenstra W** (1987) Isolation of an amylose-free starch mutant of the potato (*Solanum tuberosum* L.). *Theor Appl Genet* **75**: 217

**Huang Y, Niu B, Gao Y, Fu L, Li W** (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680–682

**Hussain H, Mant A, Seale R, Zeeman S, Hinchliffe E, Edwards A, Hylton C, Bornemann S, Smith A, Martin C, et al** (2003) Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. *Plant Cell* **15**: 133–149

**Iglesias A, Charng Y, Ball S, Preiss J** (1994) Characterization of the kinetic, regulatory, and structural properties of ADP-glucose pyrophosphorylase from *Chlamydomonas reinhardtii*. *Plant Physiol* **104**: 1287–1294

- Imparl-Radosevich J, Keeling P, Guan H** (1999) Essential arginine residues in maize starch synthase IIa are involved in both ADP-glucose and primer binding. *FEBS Lett* **457**: 357–362
- Inoue H, Nojima H, Okayama H** (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **30**: 23–28
- James M, Robertson D, Myers A** (1995) Characterization of the maize gene sugary1, a determinant of starch composition in kernels. *Plant Cell* **7**: 417–429
- Jang S, Yamaoka Y, Ko D, Kurita T, Kim K, Song W, Hwang J, Kang B, Nishida I, Lee Y** (2015) Characterization of a *Chlamydomonas reinhardtii* mutant defective in a maltose transporter. *J Plant Biol* **58**: 344–351
- Jobling S, Schwall G, Westcott R, Sidebottom C, Debet M, Gidley M, Jeffcoat R, Safford R** (1999) A minor form of starch branching enzyme in potato (*Solanum tuberosum* L.) tubers has a major effect on starch structure: cloning and characterisation of multiple isoforms of SBE A. *Plant J* **18**: 163–171
- Kamisugi Y, Cuming A** (2009) Gene targeting. In C Knight, P Perroud, D Cove, eds, *Annu. Plant Rev. Moss Physcomitrella patens*. John Wiley & Sons, pp 76–112
- Kaneko T, Kihara M, Ito K, Takeda K** (2000) Molecular and chemical analysis of  $\beta$ -amylase-less mutant barley in Tibet. *Plant Breed* **119**: 383–387
- Kapustin Y, Souvorov A, Tatusova T** (2004) Splign - a hybrid approach to spliced alignments. In A Gramada, PE Bourne, eds, *Curr. Comput. Mol. Biol.* p 471
- Kapustin Y, Souvorov A, Tatusova T, Lipman D** (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* **3**: 20
- Katoh K, Toh H** (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* **9**: 286–298
- Keeling PJ** (2009) Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J Eukaryot Microbiol* **56**: 1–8
- Kenrick P, Crane P** (1997) The origin and early evolution of plants on land. *Nature* **389**: 33–39
- Kim W, Franceschi V, Okita T, Robinson N, Morell M, Preiss J** (1989) Immunocytochemical localization of ADPglucose pyrophosphorylase in developing potato tuber cells. *Plant Physiol* **91**: 217–220
- Kleczkowski L** (1994) Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. *Phytochemistry* **37**: 1507–1515
- Kleczkowski L** (1999) A phosphoglycerate to inorganic phosphate ratio is the major factor in controlling starch levels in chloroplasts via ADP-glucose pyrophosphorylase regulation. *FEBS Lett* **448**: 153–156
- Knight M, Harn C, Lilley C, Guan H, Singletary G, MuForster C, Wasserman B, Keeling P** (1998)

Molecular cloning of starch synthase I from maize (W64) endosperm and expression in *Escherichia coli*. *Plant J* **14**: 613–622

**Konishi Y, Nojima H, Okuno K, Asaoka M, Fuwa H** (1985) Characterisation of starch granules from waxy, nonwaxy, and hybrid seeds of *Amaranthus hypochondriacus* L. *Agric Biol Chem* **49**: 1965–1971

**Kotting O, Pusch K, Tiessen A, Geigenberger P, Steup M, Ritte G** (2005) Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. *Plant Physiol* **137**: 242–252

**Kotting O, Santelia D, Edner C, Eicke S, Marthaler T, Gentry M, Comparot-Moss S, Chen J, Smoth A, Steup M, et al** (2009) STARCH-EXCESS4 is a laforin-like phosphoglucan phosphatase required for starch degradation in *Arabidopsis thaliana*. *Plant Cell* **21**: 334–346

**Kräbs G, Watanabe M, Wiencke C** (2004) A monochromatic action spectrum for the photoinduction of the UV-absorbing mycosporine-like amino acid shinorine in the red alga *Chondrus crispus*. *Photochem Photobiol* **79**: 515–519

**Krogh A, Larsson B, Von Heijne G, Sonnhammer E** (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580

**Krueger-Hadfield S, Collén J, Daguin-Thiébaud C, Valero M** (2011) Genetic population structure and mating system in *Chondrus crispus* (Rhodophyta). *J Phycol* **47**: 440–450

**Kubo A, Colleoni C, Dinges J, Lin Q, Lappe R, Rivenbark J, Meyer A, Ball S, James M, Hennen-Bierwagen T, et al** (2010) Functions of heteromeric and homomeric isoamylase-type starch-debranching enzymes in developing maize endosperm. *Plant Physiol* **153**: 956–969

**Kumar S, Stecher G, Li M, Knyaz C, Tamura K** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549

**Leblanc C, Boyen C, Richard O, Bonnard G, Grienenberger J, Kloareg B** (1995) Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization. *J Mol Biol* **250**: 484–495

**Lemercier G, Bakalara N, Santarelli X** (2003) On-column refolding of an insoluble histidine tag recombinant exopolyphosphatase from *Trypanosoma brucei* overexpressed in *Escherichia coli*. *J Chromatogr* **786**: 305–309

**Leterrier M, Holappa L, Broglie K, Beckles D** (2008) Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications. *BMC Plant Biol* **8**: 98

**Li W, Godzik A** (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659

**Li W, Jaroszewski L, Godzik A** (2001) Clustering of highly homologous sequences to reduce the size of

large protein databases. *Bioinformatics* **17**: 282–283

**Li W, Jaroszewski L, Godzik A** (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* **18**: 77–82

**Lin Q, Facon M, Putaux J, Dinges J, Wattebled F, D’Hulst C, Hennen-Bierwagen T, Myers A** (2013) Function of isoamylase-type starch debranching enzymes ISA1 and ISA2 in the *Zea mays* leaf. *New Phytol* **200**: 1009–1021

**Lin T, Caspar T, Somerville C, Preiss J** (1988) Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADPglucose pyrophosphorylase activity. *Plant Physiol* **86**: 1131–1135

**Linka M, Jamai A, Weber A** (2008) Functional characterization of the plastidic phosphate translocator gene family from the thermo-acidophilic red alga *Galdieria sulphuraria* reveals specific adaptations of primary carbon partitioning in green plants and red algae. *Plant Physiol* **148**: 1487–1496

**Liu F, Zhao Q, Mano N, Ahmed Z, Nitschke F, Cai Y, Chapman K, Steup M, Tetlow I, Emes M** (2016) Modification of starch metabolism in transgenic *Arabidopsis thaliana* increases plant biomass and triples oilseed production. *Plant Biotechnol J* **14**: 976–985

**Liu H, Yu G, Wei B, Wang Y, Zhang J, Hu Y, Liu Y, Yu G, Zhang H, Huang Y** (2015) Identification and phylogenetic analysis of a novel starch synthase in maize. *Front Plant Sci* **6**: 1013

**Liu Y, Vidali L** (2011) Efficient polyethylene glycol (PEG) mediated transformation of the moss *Physcomitrella patens*. *J. Vis. Exp.*

**Lloyd J, Kossmann J** (2015) Transitory and storage starch metabolism: two sides of the same coin? *Curr Opin Biotechnol* **32**: 143–148

**Lorberth R, Ritte G, Willmitzer L, Kossmann J** (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nat Biotechnol* **16**: 473–477

**Lynch M, Conery J** (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155

**Macherey-Nagel** (2014) Purification of His-tag Proteins: User Manual, Revision 6.

**MacNeill G, Mehrpouyan S, Minow M, Patterson J, Tetlow I, Emes M** (2017) Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *J Exp Bot* **68**: 4433–4453

**Manners D, Sturgeon R** (1982) Reserve carbohydrate of red algae, fungi and lichens. *In* F Loewus, W Tanner, eds, *Encycl. Plant Physiol.* Springer, Verlag, Berlin, pp 472–514

**Manners D, Wright A** (1962) a-1,4-Glucosans. Part XIV. The interaction of concanavalin-A with glycogens. *J Chem Soc* 4592–4595

- Matsuzaki M, Misumi O, Shin T, Maruyama S, Takahara M, Miyagishima S, Nishida K, Yoshida Y, Nishimura Y, Nakao S, et al** (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**: 653–657
- McBride A, Ghilagaber S, Nikolaev A, Hardie DG** (2009) The glycogen-binding domain on the AMPK  $\beta$  subunit allows the kinase to act as a glycogen sensor. *Cell Metab* **9**: 23–34
- McCracken D, Cain J** (1981) Amylose in floridean starch. *New Phytol* **88**: 67–71
- Meeuse B, Andries M, Wood J** (1960) Floridean starch. *J Exp Bot* **11**: 129–140
- Mishra B, Kumar R, Mohan A, Gill K** (2017) Conservation and divergence of starch synthase III genes of monocots and dicots. *PLoS One* **12**: e0189303
- Momma M, Fujimoto Z** (2012) Interdomain disulfide bridge in the rice granule bound starch synthase I catalytic domain as elucidated by X-ray structure analysis. *Biosci Biotechnol Biochem* **76**: 1591–1595
- Moran-Zorzano M, Alonso-Casajus N, Munoz F, Viale A, Baroja-Fernandez E, Eydallin G, Pozueta-Romero J** (2007) Occurrence of more than one important source of ADP-glucose linked to glycogen biosynthesis in *Escherichia coli* and *Salmonella*. *FEBS Lett* **581**: 4423–4429
- Moreira D, Le Guyader H, Phillippe H** (2000) The origin of red algae and the evolution of chloroplasts. *Nature* **405**: 69–72
- Mouille G, Maddelein M, Libessart N, Talaga P, Decq A, Delrue B, Ball S** (1996) Preamylopectin processing: a mandatory step for starch biosynthesis in plants. *Plant Cell* **8**: 1353–1366
- Murata T, Sugiyama T, Akazawa T** (1965) Enzymatic mechanism of starch synthesis in glutinous rice grains. *Biochem Biophys Res Commun* **18**: 371–376
- Nagashima H, Nakamura S, Nisizawa K, Hori T** (1971) Enzymic synthesis of floridean starch in a red alga, *Serraticardia maxima*. *Plant Cell Physiol* **12**: 243–253
- Nakamura T, Yamamori M, Hirano H, Hidaka S, Nagamine T** (1995) Production of waxy (amylose-free) wheats. *Mol Gen Genet* **248**: 253–259
- Nakamura Y** (2002) Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant Cell Physiol* **43**: 718–725
- Nakamura Y, Aihara S, Crofts N, Sawada T, Fujita N** (2014) In vitro studies of enzymatic properties of starch synthases and interactions between starch synthase I and starch branching enzymes from rice. *Plant Sci* **224**: 1–8
- Nakamura Y, Kubo A, Shimamune T, Matsuda T, Harada K, Satoh H** (1997) Correlation between activities of starch debranching enzyme and  $\alpha$ -polyglucan structure in endosperms of sugary-1 mutants of rice. *Plant J* **12**: 143–153

- Nakamura Y, Takahashi J, Sakurai A, Inaba Y, Suzuki E, Nihei S, Fujiwara S, Tsuzuki M, Miyashita H, Ikemoto H, et al** (2005) Some cyanobacteria synthesize semi-amylopectin type a-polyglucans instead of glycogen. *Plant Cell Physiol* **46**: 539–545
- Nakosteen P, Hughes K** (1978) Sexual life cycle of three species of Funariaceae in culture. *Bryologist* **81**: 307–314
- Nashilevitz S, Melamed-Bessudo C, Aharoni A, Kossmann J, Wolf S, Levy A** (2008) The *legwd* mutant uncovers the role of starch phosphorylation in pollen development and germination in tomato. *Plant J* **57**: 1–13
- Nazarian-Firouzabadi F, Visser RGF** (2017) Potato starch synthases: functions and relationships. *Biochem Biophys Reports* **10**: 7–16
- Nelson O, Rines H** (1962) The enzyme deficiency in the waxy mutant of maize. *Biochem Biophys Res Commun* **9**: 297–300
- Niittylä T, Comparot-Moss S, Lue W, Messerli G, Trevisan M, Seymour M, Gatehouse J, Villadsen D, Smith S, Chen J, et al** (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. *J Biol Chem* **281**: 11815–11818
- Niittylä T, Messerli G, Trevisan M, Chen J, AM S, Zeeman S** (2004) A previously unknown maltose transporter essential for starch degradation in leaves. *Science* (80- ) **303**: 87–89
- Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T** (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol* **64**: 1694–1699
- Nofrarias M, Martinez-Puig D, Pujols J, Majo N, Perez J** (2007) Long-term intake of resistant starch improves colonic mucosal integrity and reduces gut apoptosis and blood immune cells. *Nutrition* **23**: 861–870
- Nougue O, Corbi J, Ball S, Manicacci D, Tenailon M** (2014) Molecular evolution accompanying functional divergence of duplicated genes along the plant starch biosynthesis pathway. *BMC Evol Biol* **14**: 103
- Nyvall P** (2000) Red algal starch metabolism: effects of endosymbionts and nutrients status on starch metabolising enzymes. University of Stockholm
- Nyvall P, Pelloux J, Davies H, Pedersén M, Viola R** (1999) Purification and characterisation of a novel starch synthase selective for uridine 5'-diphosphate glucose from the red alga *Gracilaria tenuistipitata*. *Planta* **209**: 143–152
- Okita T, Greenberg E, Kuhn D, Preiss J** (1979) Subcellular localization of the starch degradative and biosynthetic enzymes of spinach leaves. *Plant Physiol* **64**: 187–192
- Ono T, Suzuki H** (1957) Endosperm characters in hybrids between barley varieties with starchy and waxy

endosperm. *Seihen Ziho* **8**: 11–19

- Patron NJ, Keeling PJ** (2005) Common evolutionary origin of starch biosynthetic enzymes in green and red algae. *J Phycol* **41**: 1131–1141
- Peat S, Turvey J, Evans J** (1959) The structure of floridean starch. Part II. Enzymic hydrolysis and other studies. *J Chem Soc* 3341–3344
- Pires N, Dolan L** (2012) Morphological evolution in land plants: new designs with old genes. *Phil Trans R Soc B* **367**: 508–218
- Plancke C, Colleoni C, Deschamps P, Dauvillée D, Nakamura Y, Haebel S, Ritte G, Steup M, Buléon A, Putaux JL, et al** (2008) Pathway of cytosolic starch synthesis in the model glaucophyte *Cyanophora paradoxa*. *Eukaryot Cell* **7**: 247–257
- Preiss J** (1988) Biosynthesis of starch and its regulation. *In* J Preiss, ed, *Biochem. Starch its Regul.* Academic Press, New York, pp 181–254
- Qiu H, Yoon H, Bhattacharya D** (2016) Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. *PLOS Curr. Tree Life* December 2
- Raven J, Edwards D** (2001) Roots: evolutionary origins and biogeochemical significance. *J Exp Bot* **52**: 381–401
- Raynaud S, Ragel P, Rojas T, Mérida 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
- Reinhold H, Soyk S, Simková K, Hostettler C, Marafino J, Mainiero S, Vaughan C, Monroe J, Zeeman S** (2011)  $\beta$ -amylase-like proteins function as transcription factors in *Arabidopsis*, controlling shoot growth and development. *Plant Cell* **23**: 1391–1403
- Rensing S, Beike A, Lang D** (2013) Evolutionary importance of generative polyploidy for genome evolution of haploid-dominant land plants. *In* I Leitch, J Greilhuber, J Dolezel, J Wendel, eds, *Plant Genome Divers. Vol. 2 Phys. Struct. Behav. Evol. Plant Genomes.* Springer, Wien, pp 295–305
- Rensing S, Ick J, Fawcett J, Lang D, Zimmer A, Van de Peer Y, Reski R** (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* **7**: 130
- Rensing S, Lang D, Zimmer A, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P, Lindquist E, Kamisugi Y, et al** (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64–69
- Ritte G, Heydenreich M, Mahlow S, Haebel S, Kottling O, Steup M** (2006) Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. *FEBS Lett* **580**: 4872–4876



- Ritte G, Lloyd J, Eckermann N, Rottmann A, Kossmann J, Steup M** (2002) The starch-related R1 protein is an  $\alpha$ -glucan, water dikinase. *PNAS* **99**: 7166–7171
- Roach P** (2002) Glycogen and its metabolism. *Curr Mol Med* **2**: 101–120
- Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF** (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr Biol* **15**: 1325–1330
- Rösti S, Fahy B, Denyer K** (2007) A mutant of rice lacking the leaf large subunit of ADP-glucose pyrophosphorylase has drastically reduced leaf starch content but grows normally. *Funct Plant Biol* **34**: 480–489
- Rubinstein C, Gerrienne P, De la Puente G, Astini R, Steemans P** (2010) Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). *New Phytol* **188**: 365–369
- Safford R, Jobling S, Sidebottom C, Westcott R, Cooke D, Tober K, Strongitharm B, Russell A, Gidley M** (1998) Consequences of antisense RNA inhibition of starch branching enzyme activity on properties of potato starch. *Carbohydr Polym* **35**: 155–168
- Sambrook J, Fritsch EF, Maniatis T** (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. New York: Cold Spring Harbor Laboratory Press
- Sambrook J, Russel D** (2001) *Molecular Cloning*, 3rd ed. Cold Spring Harbor Laboratory Press, New York
- Samodien E, Jewell J, Loedolff B, Oberlander K, George G, Zeeman S, Damberger F, Van der Vyver C, Kossmann J, Lloyd J** (2018) Repression of Sex4 and Like Sex Four2 orthologs in potato increases tuber starch bound phosphate with concomitant alterations in starch physical properties. *Front Plant Sci* **9**: 1044
- San-Miguel T, Pérez-Bermúdez P, Gavidia I** (2013) Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *Springerplus* **2**: 89
- Santelia D, Kotting O, Seung D, Schubert M, Thalmann M, Bischof S, Meekins D, Lutz A, Patron N, Gentry M, et al** (2011) The phosphoglucan phosphatase Like Sex Four2 dephosphorylates starch at the C3-position in Arabidopsis. *Plant Cell* **23**: 4096–4111
- Santelia D, Zeeman S** (2010) Progress in Arabidopsis starch research and potential biotechnological applications. *Curr Opin Biotechnol* **22**: 271–280
- Schaefer D, Zrýd J** (1997) Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J* **11**: 1195–1206
- Scheidig A, Frohlich A, Schulze S, Lloyd J, Kossmann J** (2002) Downregulation of a chloroplast-targeted  $\beta$ -amylase leads to a starch-excess phenotype in leaves. *Plant J* **30**: 581–591
- Schonknecht G, Chen W, Ternes C, Barbier G, Shrestha R, Stanke M, Brautigam S, Baker B, Banfield**

- J, Garavito R, et al** (2013) Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**: 1207–1210
- Schwall G, Safford R, Westcott R, Jeffcoat R, Tayal A, Shi Y, Gidley M, Jobling S** (2000) Production of very-high-amylose potato starch by inhibition of SBE A and B. *Nat Biotechnol* **18**: 551–554
- Senoura T, Asao A, Takashima Y, Isono N, Hamada S, Ito H, Matsui H** (2007) Enzymatic characterization of starch synthase III from kidney bean (*Phaseolus vulgaris* L.). *FEBS J* **274**: 4550–4560
- Senoura T, Isono N, Yoshikawa M, Asao A, Hamada S, Watanabe K, Ito H, Matsui H** (2004) Characterization of starch synthase I and II expressed in early developing seeds of kidney bean (*Phaseolus vulgaris* L.). *Biosci Biotechnol Biochem* **68**: 1949–1960
- Sesma J, Iglesias A** (1998) Synthesis of floridean starch in the red alga *Gracilaria gracilis* occurs via ADP-glucose. In G Garab, ed, *Photosynth. Mech. Eff.*, V. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 3537–3540
- Shaw AJ, Schmutz J, Devos N, Shu S, Carrell AA, Weston DJ** (2016) The Sphagnum Genome Project: A New Model for Ecological and Evolutionary Genomics. In S Rensing, ed, *Adv. Bot. Res.* pp 167–187
- Sheath R, Hellebust J, Sawa T** (1979) Floridean starch metabolism of *Porphyridium purpureum* Rhodophyta I. Changes during ageing of batch culture. *Phycologia* **18**: 149–163
- Sheng F, Jia X, Yep A, Preiss J, Geiger J** (2009) The crystal structures of the open and catalytically competent closed conformation of *Escherichia coli* glycogen synthase. *J Biol Chem* **284**: 17796–17807
- Shimonaga T, Fujiwara S, Kaneko M, Izumo A, Nihei S, Francisco PB, Satoh A, Fujita N, Nakamura Y, Tsuzuki M** (2006) Variation in storage  $\alpha$ -polyglucans of red algae: Amylose and semi-amylopectin types in Porphyridium and glycogen type in Cyanidium. *Mar Biotechnol* **9**: 192–202
- Shimonaga T, Konishi M, Oyama Y, Fujiwara S, Satoh A, Fujita N, Colleoni C, Buleon A, Putaux JL, Ball SG, et al** (2008) Variation in storage  $\alpha$ -glucans of the Porphyridiales (Rhodophyta). *Plant Cell Physiol* **49**: 103–116
- Sievers F, Wilm A, Dineen D, Gibson T, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al** (2011) Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539
- Silhavy T** *Escherichia coli* strain TST6.
- Sonnhammer E, Von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In J Glasgow, ed, *Proceedings. Sixth Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, pp 175–182
- Stadnichuk IN, Semenova LR, Smirnova GP, Usov I** (2007) A highly branched storage polyglucan in the thermoacidophilic red microalga *Galdieria maxima* cells. *Appl Biochem Microbiol* **43**: 78–83

- Stander E** (2015) Roles of disproportionating enzymes in the moss *Physcomitrella patens*. Stellenbosch University
- Stanley D, Fitzgerald A, Farnden K, MacRae E** (2002) Characterisation of putative  $\alpha$ -amylases from apple (*Malus domestica*) and *Arabidopsis thaliana*. *Biologia (Bratisl)* **57**: 137–148
- Stanley D, Rejzek M, Naested H, Smedley M, Otero S, Fahy B, Thorpe F, Nash R, Harwood W, Svensson B, et al** (2011) The role of  $\alpha$ -glucosidase in germinating barley grains. *Plant Physiol* **155**: 932–943
- Stettler M, Eicke S, Mettler T, Messerli G, Hortensteiner S, Zeeman S** (2009) Blocking the metabolism of starch breakdown products in *Arabidopsis* leaves triggers chloroplast degradation. *Mol Plant* **2**: 1233–1246
- Stitt M, Zeeman S** (2012) Starch turnover: pathways, regulation and role in growth. *Curr Opin Plant Biol* **15**: 282–292
- Streb S, Eicke S, Zeeman S** (2012) The simultaneous abolition of three starch hydrolases blocks transient starch breakdown in *Arabidopsis*. *J Biol Chem* **287**: 41745–41756
- Streb S, Zeeman S** (2012) Starch metabolism in *Arabidopsis*. *Arab B* e0160
- Sulpice R, Pyl E, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques M, et al** (2009) Starch as a major integrator in the regulation of plant growth. *PNAS* **106**: 10348–10353
- Sun Z, Henson C** (1991) A quantitative assessment of the importance of barley seed  $\alpha$ -amylase,  $\beta$ -amylase, debranching enzyme, and  $\alpha$ -glucosidase in starch degradation. *Arch Biochem Biophys* **284**: 298–305
- Szydowski N, Ragel P, Raynaud S, Lucas M, Roldan I, Montero M, Munoz F, 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
- Takeda Y, Guan H, Preiss J** (1993) Branching of amylose by the branching isoenzymes of maize endosperm. *Carbohydr Res* **240**: 253–263
- Takeda Y, Hizukuri S** (1981) Studies on starch phosphate. Pt. 5. Re-examination of the action of sweet-potato  $\beta$ -amylase on phosphorylated (1-4)- $\alpha$ -D-glucan. *Carbohydr Res* **89**: 174–178
- Tetlow I, Beisel K, Cameron S, Makhmoudova A, Liu F, Bresolin N, Wait R, Morell M, Emes M** (2008) Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. *Plant Physiol* **146**: 1878–1891
- Tetlow I, Wait R, Lu Z, Akkasaeng R, Bowsher C, Esposito S, Kosar-Hashemi B, Morell M, Emes M** (2004) Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. *Plant Cell* **16**: 694–708

- Thalmann M, Santelia D** (2017) Starch as a determinant of plant fitness under abiotic stress. *New Phytol* **214**: 943–951
- Tsumoto K, Umetsu M, Kumagai I, Ejima D, Philo J, Arakawa T** (2004) Role of arginine in protein refolding, solubilization, and purification. *Biotechnol Prog* **20**: 1301–1308
- Umemoto T, Yano M, Satoh H, Shomura A, Nakamura Y** (2002) Mapping of a gene responsible for the difference in amylopectin structure between japonica-type and indica-type rice varieties. *Theor Appl Genet* **104**: 1–8
- Utsumi Y, Nakamura Y** (2006) Structural and enzymatic characterization of the isoamylase1 homo-oligomer and the isoamylase1-isoamylase2 hetro-oligomer from rice endosperm. *Planta* **225**: 75–87
- Valdez H, Busi M, Wayllace N, Parisi G, Ugalde R, Gomez-Casati D** (2008) Role of the N-terminal starch-binding domains in the kinetic properties of starch synthase III from *Arabidopsis thaliana*. *Biochemistry* **47**: 3026–3032
- Viola R, Nyvall P, Pedersén M** (2001) The unique features of starch metabolism in red algae. *Proc R Soc Lond* **268**: 1417–22
- Vriet C, Welham T, Brachmann A, Pike M, Pike J, Perry J, Parniske M, Sato S, Tabata S, Smith A, et al** (2010) A suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. *Plant Physiol* **154**: 643–655
- Wang Q, Monrow J, Sjölund R** (1995) Identification and characterization of a phloem-specific beta-amylase. *Plant Physiol* **109**: 743–750
- Wang X, Feng B, Xu Z, Sestili F, Zhao G, Xiang C, Lafiandra D, Wang T** (2014) Identification and characterization of granule bound starch synthase I (GBSSI) gene of tartary buckwheat (*Fagopyrum tataricum* Gaertn.). *Gene* **534**: 229–235
- Wang X, Zhao F, Hu Z, Critchley A, Morrell S, Duan D** (2008) Inter-simple sequence repeat (ISSR) analysis of genetic variation of *Chondrus crispus* populations from North Atlantic. *Aquat Bot* **88**: 154–159
- Wang Y, White P, Pollak L, Jane J** (1993) Characterization of starch structures of 17 maize endosperm mutant genotypes with Oh43 inbred line background. *Cereal Chem* **70**: 171–179
- Wattebled F, Dong Y, Dumez S, Delvalle D, Planchot V, Berbezy P, Vyas D, Colonna P, Chatterjee M, Ball S, et al** (2005) Mutants of *Arabidopsis* lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. *Plant Physiol* **138**: 184–195
- Wattebled F, Ral J, Dauvillee D, Myers A, James M, Schlichting R, Giersch C, Ball S, D’Hulst C** (2003) STA11, a *Chlamydomonas reinhardtii* locus required for normal starch granule biogenesis, encodes disproportionating enzyme. Further evidence for a function of  $\alpha$ -1,4 glucanotransferases during starch

granule biosynthesis in green algae. *Plant Physiol* **132**: 137-145

**Weber A, Servaites J, Geiger D, Kofler H, Hille D, Groner F, Hebbeker U, Flugge U** (2000) Identification, purification, and molecular cloning of a putative plastidic glucose translocator. *Plant Cell* **12**: 787–801

**Wellman C, Osterloff P, Mohiuddin U** (2003) Fragments of the earliest land plants. *Nature* **425**: 282–285

**Wettstein D** (1924) Morphologie und physiologie des formwechsels der moose auf genetischer grundlage. *Zeitschrift für Indukt Abstammungs- Vererbungslehre* **33**: 1–236

**Willis K, McElwain J** (2002) *The evolution of plants*. Oxford University Press, Oxford, UK

**Yamaguchi H, Miyazaki M** (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4**: 235–251

**Yang E, Boo S, Bhattacharya D, Saunders G, Knoll A, Fredericq S, Graf L, Yoon H** (2016) Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci. Rep.* **6**: e21361

**Yang Z, Zhang L, Zhang Y, Zhang T, Feng Y, Lu X, Lan W, Wang J, Wu H, Cao C, et al** (2011) Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PLoS One* **6**: e22981

**Yep A, Ballicora M, Preiss J** (2006) The ADP-glucose binding site of the *Escherichia coli* glycogen synthase. *Arch Biochem Biophys* **453**: 188–196

**Yep A, Ballicora MA, Sivak MN, Preiss J** (2004) Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli*. *J Biol Chem* **279**: 8359–8367

**Yu S, Ahamad T, Kenne L, Pedersén M** (1995)  $\alpha$ -1,4-Glucan lyase, a new class of starch/glycogen degrading enzyme. III. Substrate specificity, mode of action and cleavage mechanism. *Biochim Biophys Acta* **1244**: 1–9

**Yu S, Bojko M, Madsen F, Olsen CE** (2002) Physico-chemical characterization of floridean starch of red algae. *Starch/Starke* **54**: 66–74

**Yu S, Kenne L, Pedersén M** (1993)  $\alpha$ -1,4-Glucan lyase, a new class of starch/glycogen degrading enzyme. I. Efficient purification and characterization from red seaweeds. *Biochim Biophys Acta* **1156**: 313–320

**Yu S, Pedersén M** (1993)  $\alpha$ -1,4-Glucan lyase, a new class of starch/glycogen degrading enzyme. II. Subcellular localization and partial amino-acid sequence. *Planta* **191**: 137–142

**Yu T, Kofler H, Häusler R, Hille D, Flüge U, Zeeman S, Smith A, Kossmann J, Lloyd J, Ritte G, et al** (2001) The *Arabidopsis* *sex1* mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. *Plant Cell* **13**: 1907–1918

**Yu T, Zeeman S, Thorneycroft D, Fulton D, Dunstan H, Lue W, Hegemann B, Tung S, Umemoto T, Chapple A, et al** (2005)  $\alpha$ -Amylase is not required for breakdown of transitory leaf starch in *Arabidopsis*

leaves. *J Biol Chem* **280**: 9773–9779

**Zabawinski C, Van Den Koornhuise N, D’Hulst C, Schlichting R, Giersch C, Delrue B, Lacroix J, Preiss J, Ball S** (2001) Starchless mutants of *Chlamydomonas reinhardtii* lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase. *J Bacteriol* **183**: 1069–1077

**Zeeman S, Delatte T, Messerli G, Umhang M, Stettler M, Mettler-Altmann T, Streb S, Reinhold H, Kötting O** (2007) Starch breakdown: recent discoveries suggest distinct pathways and novel mechanisms. *Funct Plant Biol* **34**: 465–473

**Zeeman S, Northrop F, Smith A, Ap Rees T** (1998a) A starch-accumulating mutant of *Arabidopsis thaliana* deficient in a starch-hydrolyzing enzyme. *Plant J* **15**: 357–365

**Zeeman S, Umemoto T, Lue W, Au-Yeung P, Martin C, Smith A, Chen J** (1998b) A mutant of *Arabidopsis* lacking a chloroplastic isoamylase accumulates both starch and phytyglycogen. *Plant Cell* **10**: 1699–1711

**Zeeman SC, Kossmann J, Smith AM** (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* **61**: 209–234

**Zhu L, Gu M, Meng X, Cheung S, Yu H, Huang J, Sun Y, Shi Y, Liu Q** (2012) High-amylose rice improves indices of animal health in normal and diabetic rats. *Plant Biotechnol J* **10**: 353–362

## Appendix A: List of primers

Primer name	Sequence (5' → 3')	
CcSS Fw	ATGGTGTCTTCTGACTGCCGAACCTGGGT	<i>C. crispus</i> starch synthase
CcSS Rv	TTACTGAACTTGCAGAACGTTGTTTCGT	
E1200C Fw	AAGTCTGTTCTGCCAGGTGGCATTGTTC	Site-directed mutagenesis
E1200C Rv	GGCATCAGCCCAAAGTCA	
T1181Y Fw	GTGGTGCTTTTTTCGCACAGCCTGACG	
T1181Y Rv	GGATGGCGACGACGCAGG	
TA829-30KT Fw	TCGCTTCAGTAAGACGGGCGGTCTGG	
TA829-30KT Rv	CCCAGTTCGGCAGTCAGAAA	
BBH CcSS Fw	TTATCAACACACAAACACTAAATCAAAGA- ATTCATGCGGGGTTCTCATCATCATCATCAT	Yeast-mediated ligation
BBH CcSS Rv	TTGGACTAGAAGGCTTAATCAAAAGCTCTC- GAGTTACTGAACTTGCAGAACGTTGTTTCGT	
PpMEX1a gene Fw	AGTACCAATGTTGCCAGCGA	<i>P. patens</i> MEX gene sequences
PpMEX1a gene Rv	CTGGAGGAGGTGTGCTGTTT	
PpMEX1b gene Fw	GCATTCCATGGCCTCCTCTT	
PpMEX1b gene Rv	TTGCGGACAGACAAGGGTAC	
PpMEX1c gene Fw	TCAACATTCACTTTGCACCA	
PpMEX1c gene Rv	TCATAGCAAAGATAAAGTCACTCAAA	
PpMEX1d gene Fw	ATGGCCTCCATTGCAGTG	
PpMEX1d gene Rv	GGCATTGGGTGGATGGAAGA	
PpMEX1a cds Fw	CAGATCTATATGGCCTCCATTGCAGT	<i>P. patens</i> MEX coding sequences
PpMEX1a cds Rv	ACATTACGCAGACCTACTTCC	
PpMEX1b cds Fw	CAGATCTATATGTCGTCTGAGGGTAGGT	
PpMEX1b cds Rv	TTGGTTTTAGAATGTCCACAGAGT	
PpMEX1c cds Fw	CAGATCTATATGGCCTCCATTGCAGT	
PpMEX1c cds Rv	TCAGACGTAGTATCGCCTTGA	
PpMEX1d cds Fw	CAGATCTATATGGCCTCCATTGCAGT	
PpMEX1d cds Rv	CTTTCAGCCGTGATATTGCCT	
PpMEX1a sq Fw	AGCTTAGCCCTAGCCTCTGG	<i>P. patens</i> MEX sqRT-PCR
PpMEX1a sq Rv	ATGGAGGCCGGAATGAATGG	
PpMEX1b sq Fw	GCATACTTGTTGCTGTTGGCA	
PpMEX1b sq Rv	TCCTGAGACTGAGTCATATCTCCAT	
PpMEX1c sq Fw	CTACCTCATTTCGTTGCTGCCT	
PpMEX1c sq Rv	GCACCACCCACATGTAATAGCA	

<b>PpMEX1d sq Fw</b>	CCAGTCTCTGGTGATTGGCG	
<b>PpMEX1d sq Rv</b>	CGTCAGCTGAGTTCTTCCAGT	
<b>PpACT sq Fw</b>	ACCGAGTCCAACATTCTACC	<i>P. patens</i> reference genes sqRT-PCR
<b>PpACT sq Rv</b>	GTCCACATTAGATTCTCGCA	
<b>PpEF1a sq Fw</b>	AATCATACATTTTCACCTCGCC	
<b>PpEF1a sq Rv</b>	GATCAGTGGGTAGAAAGTGAC	



## Appendix B: Codon-optimised *Chondrus crispus* starch synthase sequence

ATGGCTGCTCCGGCTTGGACTGACTTTCCTCTGCGTAGTGGCGAAACCCACCCAGCCGTGCTGG  
 ACGGCTACCCTGCCGACGTAGCTCGTCGTCTGCGTGGTCGCTTCATCGTCCAGCCAGCAGGCAT  
 GCGCGAAGCAAGCCTGATCGAAGTCCCTATCGATGAAGTTGGTGGGAGTGGGACGCAGCCAC  
 TGGCGACCTGAACAGTCGTGGCACCTTCGAGAGTGTTCTGACGATCATTCTGGGCTGAAAGTT  
 CAGAACTTTACTGGTGTTCATGATGGGTGCCCTGGAACGTCCGTCTGACGACCCTAACCTT  
 CTCCATTCAGCGTTGCCGATCGTGAAATCCCTGCAAACATCCTGGGTGGCAAGAGTGGTTCGG  
 TAATCTGTGTGCCGAGATGAAACGTCTGGGGCTGTGCCCTATCATTGATGGGATCGATCGTGTA  
 TCTCGTACGCGCATGAGCCGCAAGTATCGCCATCTGACGGTTGAAACGCTGAGTAACAAAGGC  
 ATCCCACTGCGCCATCCTGGGACCGATGGTCGCGAGAACCAGTGGGAGGATACGGCACTGCTG  
 AACCATCGTCGTGTGGAGACCTGGAACCTGATGGTTGCAGAAGTGAAAGCGATGGCAGAGGAA  
 TACGGCGTAGGTGGCATCCGTCTGGACAACGCTCAAAGCCTGCCACCGATCATGGCTCCGAAC  
 ATGGACGAGCTGCTGCGCCACGATCCTGATGGTGAGCCTCACTATTCTCTGAGCGAGGTGTTCT  
 ATGGGGCAGTCGTGAAGGCCAATGAAGAGTACGGCTATTGGACGTCTTCTGCTGGGATTGAAC  
 GTGGTTACCCAAATCCATTTCTGGTGAAGTTTTGCCGTGAGATGTGGAATGCATTTCTGACTTC  
 ATTGTCATGGCGGAAGCACACTTTCACCGCGAGGCACAACCTGCTGACTTCTGGTCCGGTAGTTC  
 ATACTGTACGCATTCCACAGATTCTGGCCAGCATTAGTGGCAAGTCTCTGCGTAAAGATGGCAC  
 TGTAGGTCGCGTTCCTGGCAAGAATCGCTCTACCGCTCGCACTCTGAGTCGTCTGTACCGTAAC  
 GATAGCGATTGGCTGCCGAAGAATGCCATCATGGTGAATTGCACGTGTACCCATTCTTCTCCGT  
 ACCCTGGGCTGCTGTATGGGCGTCGTTCTTGGATTGCGGTTGATCTGCTGAACTTTCTGCCTGAA  
 GTACCTATGACTGTCTATGGGGAAGAACGTGGTTCGTGCGTATCGTATGAACATGAACGGCGTG  
 AGCAATACCGAGGAAATGACTGAGTATGATGTGAACTTTGATGCAGTCCTGCCTAAGAGTCCTC  
 CGCTGCGCACTGGTCAGACGATGCCATCTAAACCTAAAGGGCTGCCACCACTGACTCCGCCACA  
 TTCTGTAGAACGTAAGCTGAAGATGAAGCGCAAGGGTTCTCTGGCGGACCTGCGTCGTGTCCCT  
 AGCAATTCTAATCTGGTTCGTTCTCGTTCTCGTGACGATATGAATGGTATGAGCGTCCGCAGTGT  
 ATCTGCTGCCGATTTCCGTAAGATGAGCGCGATGGAGGAACAACTCGCCAAGAGATTGGTCC  
 GGCAAGTGGGTACGACATCGCGCAGATTGAGGGGCACTATAGTCATCGTCGCATGCTGCGCCA  
 GGAGCTGGCTGCGCTGCACTCTGGCTTCATGTGCGTGCTGACGATCGAACCTCAGCTGAAAGAG  
 CAAGTCTTTGCCTTTGCTCGCTATAACGGAGGACCAAGTCGTGATTGTGGCTGCGAACTTCAAGG  
 ATAACCGCGATGGCCACAGTATAGCAACGGGTGCGATGTCGAACTGGACTTTCAGACCCTGT  
 GGGACGTGCTGCCGGATACCTTTACGACTGGTGTGCTGCGCCATGCGCCTTCTATTCTGTGGTCAA  
 CACGTTTACCGGCAAAGAGCACAGCACGGATGTCCAGACCCTGGAGGAACTGGTCTTTCGCAA  
 GTACAAGACGCATCTGGACCCGCTGGGTATCAGCCTGCTGACTTTCAGCCGGTTCAGGACACG  
 CCTGAACGTCGTGGTGGCAGCTTTAGCGAATGCATCAACCGTCTGCGCTCTCAGGAAGCTAATG  
 ACATCAAGGATGCGCGTGAGAACGACATCATTGCTCGCCTGGCACGTGGTGCAGCAGCCAGCG

CGAGTGACTTTGTTGGGGCTATGGAAAGTCTGCGCAATGGGCTGCGCAATGAAGGCTGCGAGA  
AAGCGGAGATGGAGCGCATCATGCAGCTGTGCATGCAGCGTGCAGCCAACTGCGCTTCATGG  
TCGCTTATGAAGGCGTCCCTCGTCCGAAAGACTTCGAACCACCTGCCGCAGAGCATATCGTAGC  
GTATCTGACCCACATGAGTACGTGTGCGAAAGACCCGGATCTGATGACCCTGGCTCGTAGCGTA  
GTAGCCAAAACCACCAAGCTGGGTCCACTGGTGTCTGACTGCCGAACTGGGTCGCTTCAGTA  
CGGCAGGCGGTCTGGGGTTATGGTGGACGAACTGACCAAAGGGCTGGCAGGTCTGGGTCTGG  
AAGTCTATGTAGTTTCTCCATACTACACTGTGAATCGCAAGAATCGTTCTGGTTATCTGGGGGA  
CAACATTCAGTGGACTCGCAACATTAGTGTCAACATCGGGACTCACATTGTCGAAGCTGGCGTA  
TTCGAGGGCGTAGAAAACGATGTGAATCTGATCTTTCTGGAACGTGGCGATTTCTTTCTAAAG  
TCTACGCTGATCCTGGTGGGAGCGTACGTCATCTGCAAACGGTAGTCCTGATGTCTCTGGGGAG  
CCTGGAAGTGTGTTGTCAAAGCAGCTGTATCCGAGCGTCATTGTAACCTAACGACTGGCTGCCA  
TCTATGGCTGCTGGTTACCGGATTTCTTTGGCGACTACTTCAAGAATACCTCTTTCTTCCATCT  
GATTCACAACCTGGGTGAAGGTGCGTATGAGGGTTCGTGTGTATCCTAATCCAGGCGAAGGCAC  
CCTGGACCATATCCACCGCCTGCCGACGCACGTCATGGTAAACCCGTGGTGGAGCACGCTGGTA  
GTGAATCCATCTCGCTGTGCGATCATGCGTAGTGAAAGCTGGGGCACTGTGAGTCCTTCTTATC  
TGCGCGAGCTGCGTGCCGGTCACCCTCTGAGTGACCTGCTGCAACAAGCCAAAAGCCTTTTGC  
ATACCCTAATGGCATTTCGTAAGGCTGAACGTGAAGAAGCCCTGCGCGTCAAAGGGGGCAGAGAG  
TCATGCGGCAGCTAAGGAAATCCTGCAAAAACGCTACTTTGGGTTCCAGAAAGGTGATCCTAC  
GATCCCTCTGTTTGCCTTCGTAGGTGCGATTACGAGTCAAAGGGGCGTTCACCTGATTCTGAAC  
GCAGTTGACGAACTGATCGGTCATACGAACGGGAAGATCCAGATCCTGGTAGGTGGCCCAGCC  
AACTACAGTGACGAGTATTCTAGCGGGTGTGCTCGTCACATGGTAGACCTGCGTCGTCGCCATC  
CGTGGTGTCTTTTGGGCACAGCCTGACGAGTTCTTCACGGATGGTCCAATGTGTAATCTGGGTGC  
TGACTTTGGGCTGATGCCAAGTCTGTTTCGAGCCAGGTGGCATTGTTCAACAGGAGTTCTTCGTG  
GCTGGCACTCCAGTGATTGCCTACAAGACTGGTGGCCTGAAAGACACTGTGCACGAATGGAAA  
AGTAGCCAAGGGGAAGGTAACGGCTTTACGTTTGGAGGAGTATTCTCATGCTGACTTCGTGTGGG  
CTGTAAAGCGTGCCCTGCGCGTATTTCGCCAACCTCACGAGTACGAAGAAATGCGCGTTCGCAG  
CAGCGGAGACGACTATTGACGTTTCTCAGGTGGCATGGGCCTGGAGTAGCGAGTTCCACCGTAT  
TCGTAACGCTATGTATACTCGCGGTGACGTTGTTGCCAGCATCATTAGTAGCACCGTAGATGAA  
GAGACGGACCTGTACGATCGTAGCGGAACCCAGTCCTGATTCAGTGGACGGGGAGCGGCAAC  
TCTGTGGTGCTGAAAGGGAGCTTCGACAACCTGGACCGCTGAATGGCCACTGAGCCAAGCGGTC  
GGGGATCATGGCGCATTTGGCCTGAAACTGCTGCTGCGTCCTGGGGAGTATGTTTGCAAGTTCA  
AAGTAAACCAGGAGTGGACGGTAGCTGACGACCTGCCGCAGAAACAAGACGAGGCTGGGTTTA  
CGAACAACGTTCTGCAAGTTCAGTAA

# Appendix C: Multiple sequence alignment of starch and glycogen synthases from bacteria, algae and plants

<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	1 ----- MK-----	2
<i>Creinhardtii_GBSS</i>	1 ----- MA-----	2
<i>Athaliana_SSI</i>	1 ----- MA-----	2
<i>Athaliana_SSII</i>	1 ----- MA-----	2
<i>Athaliana_SSIII</i>	1 M I S Y - - - - F L N Q D F S - - - - - R K - - - - K Q G - - - - -	16
<i>Athaliana_SSIV</i>	1 M T T K L S S F C F L T H G L A G I S C E R E H G S S - - - - - R R - - - - -	29
<i>Athaliana_GBSSI</i>	1 ----- MA-----	2
<i>Gsulphuraria_SS</i>	1 ----- M S S S P S E L E S S A T W G A S G A D E D W E N S S S S D L E G I A S S K L R K Q T T Q K Q D Y Q N A V L S	56
<i>Cmerolae_SS</i>	1 ----- M P G S T Q R G G - - - - - W R E Q - - - - -	13
<i>Ccrispus_SS</i>	1 ----- M A A P A - - - - - W T D - - - - -	8
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	-----	
<i>Creinhardtii_GBSS</i>	3 ----- V A S T S R P S S A R P I V - - - - -	16
<i>Athaliana_SSI</i>	3 ----- S L Q I S G - - - - -	8
<i>Athaliana_SSII</i>	3 ----- S V A E S S F P L L - - - - - C - - - - -	13
<i>Athaliana_SSIII</i>	17 ----- R M A A S G P K - - - - -	24
<i>Athaliana_SSIV</i>	30 F F Y - - - - L P S R R L V S T S C K M R - - - - -	46
<i>Athaliana_GBSSI</i>	3 ----- T V T A S S - - - - -	8
<i>Gsulphuraria_SS</i>	57 F F S G K D L G I P - - S L G R S S E T L T M A P G G W E R S C E V L Y E Y F L K Q F A S M - D E Q P V V V A F N E L Q T L L I R L Q E D Y	123
<i>Cmerolae_SS</i>	14 ----- R P - - A L S R G T P V N Q R R W S A W - - - - - L P R L S S L V D E Q P V Q D V L D E V Q Q L L L - - - E R S	59
<i>Ccrispus_SS</i>	9 ----- F P - - L - - R S G E T - - - - -	16
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	3 ----- - - - - - S F M R R D A L G A G L - - - - -	14
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	9 ----- - - - - - S V K F E P F V G F N R I R H F R - - - - -	25
<i>Athaliana_SSII</i>	14 ----- - - - - - Q I K T Q R R I N S S T L R H S R - - - - -	30
<i>Athaliana_SSIII</i>	25 ----- - - - - - S S G P R G F G R R T T V G S A Q - - - - -	41
<i>Athaliana_SSIV</i>	47 ----- - - - - - Q Q R G F D S S K R Q E V K K G S - - - - -	63
<i>Athaliana_GBSSI</i>	9 ----- - - - - - N F V S R T S L F N N H - - - - -	20
<i>Gsulphuraria_SS</i>	124 R T V C A Q K T S P G D V A G I L G A F K - - - - - T L E R V L I P L L G H P V P E V R E R T V I L L N V L Y D G H	176
<i>Cmerolae_SS</i>	60 T W V T Y E D G A R G G A S G - R G A N K P P A P F S S A G W E R L Q T L L E V E K A L V S A L G H H C E D V R D R A V V L L S V L Y D G H	128
<i>Ccrispus_SS</i>	17 ----- - - - - - H P - - - - - A V L D G Y - - - - -	24
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	15 R - - - - -	15
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	-----	
<i>Athaliana_SSII</i>	-----	
<i>Athaliana_SSIII</i>	42 K R - - - - - T Q - - - - -	45
<i>Athaliana_SSIV</i>	64 P K - - - - -	65
<i>Athaliana_GBSSI</i>	21 G - - - - -	21
<i>Gsulphuraria_SS</i>	177 E L Q L T - E S L N V S V Q C V G E T A D L E I P V H G L T D I R E I - - - - - D N Y - - - V L C L S E T N - -	221
<i>Cmerolae_SS</i>	129 P L Q L A G D A L P V A V T C V G E P V T V C I P F A S E A D - R E A F Q P Q R A V L R L F R P W P N K N K N G Y T R E V L S K Q A T N S L	197
<i>Ccrispus_SS</i>	-----	
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	-----	
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	-----	
<i>Athaliana_SSII</i>	-----	
<i>Athaliana_SSIII</i>	46 ----- K K N G E K D S N A - - - - -	55
<i>Athaliana_SSIV</i>	66 ----- P I L S I N S G L - - - - - Q S N N D E E S D L - - - - -	84
<i>Athaliana_GBSSI</i>	-----	
<i>Gsulphuraria_SS</i>	222 ----- R E S Y L A - - - - - P C R - - - - -	230
<i>Cmerolae_SS</i>	198 Q L G G S R P P R S P S S D G P Q Q R K P S L A T E T G P N A T S T S P S R E Q T A Q D W T N A E E E A E Q E L P V G T H A L W A A C H G H	267
<i>Ccrispus_SS</i>	-----	
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	-----	
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	-----	
<i>Athaliana_SSII</i>	-----	
<i>Athaliana_SSIII</i>	-----	
<i>Athaliana_SSIV</i>	-----	
<i>Athaliana_GBSSI</i>	-----	
<i>Gsulphuraria_SS</i>	231 W L R Y S I L Y K N G Y L K V K G - - - - - L P G F S R S G F Y D W Y F A E K P K D V K P E S S F V F L P P V - - - S L P F C E A R L Q K	291
<i>Cmerolae_SS</i>	268 W Q S Y P I - - - - - S V Q G E S I L V H L G S F E Q P G F Y D W Y I A R A S D G V - - A S P L V I A H P V V Y A D V P G M D F R R L R	328
<i>Ccrispus_SS</i>	25 ----- - - - - - P A D V - - - - - A R R L R - - - - -	33

<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	16 ----- GAASTKPVSRVASVRPAP-----	33
<i>Creinhardtii_GBSS</i>	17 ----- INAASFGVKKTA-----	28
<i>Athaliana_SSI</i>	26 ----- P IASLGFRFR-----	36
<i>Athaliana_SSII</i>	31 ----- VSYHDLPSGSLFRSRFV-----	49
<i>Athaliana_SSIII</i>	56 ---- TSTATNEVSG I SKLPAAKVDVQKQSSVVLNERNVLDKSD I EDGSDRLDKKTTDD-----	109
<i>Athaliana_SSIV</i>	85 ----- ENGSADSVPSLKSDAEKGS I ---- HGS I ---- DMNHADENLEKK-----	121
<i>Athaliana_GBSSI</i>	22 ----- ASSCSDVAQ I TLKGQSL-----	38
<i>Gsulphuraria_SS</i>	292 GR I I VHPSN I RDSLLYELPVDQVDATWDSKT---- GELKKRGSFDLVAQRPLDLRLDG I THVYLMGALAR	357
<i>Cmerolae_SS</i>	329 GR F I VQPRGARAHRLYE I PVDQVGARWDATT---- GALLSRGSFDAQVELLPRLAASG I TG I YVSGCLER	394
<i>Ccrispus_SS</i>	34 GR F I VQPMREASL I EVP I DEVGAEWDAAT---- GDLNSRGTTFESVLT I I PGLKVQNF TGVYMMGALER	99
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	34 ----- TAYR----- TACQVAKVDEM--	48
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	37 ----- RRFS I G-----	42
<i>Athaliana_SSII</i>	50 ----- LGHR----- CKCVSRVEAS-----	63
<i>Athaliana_SSIII</i>	110 -- DDLLEQKLLERE----- NLRKE I E----- TLAAENLARGDRMFV	145
<i>Athaliana_SSIV</i>	122 -- DD I QTT- - EVTRR----- KSKTAKKKGES I HAT I D I GHDDGKNLDN I TVPEVAKALS LN-	173
<i>Athaliana_GBSSI</i>	39 ----- THCGLRSFNMV-----	49
<i>Gsulphuraria_SS</i>	358 PTDDEAPPGE I ADRSQPA A I LGGA I AFKNLVSEANRLGVGT I V----- D- - GFCRVSRNAHHRK	415
<i>Cmerolae_SS</i>	395 RLDEHEPTPTHTVVDAMPAT I LGGVHKFRMCAEARHQLAT I V----- D- - CLDRVSLARAHRR	452
<i>Ccrispus_SS</i>	100 PSDDPNPSPFSVADRE I PAN I LGGKSAFRNLCAEMKRLGLCP I I----- D- - G I DRVSRTRMSRK	157
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	49 ----- VSVDEELT----- RLRK-----	60
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	-----	
<i>Athaliana_SSII</i>	64 ----- GSDDDEPED----- ALQA-----	76
<i>Athaliana_SSIII</i>	146 YPVI --- VKPDEDIEVF LNRNLSTLNNEPDVL I MGAFNEWRWKSF T- - RRLEK- TWIHEDWLSCLLH I PK	209
<i>Athaliana_SSIV</i>	174 ----- KSEGEQ I SDGQFGE L----- MTM I RSAEK-----	197
<i>Athaliana_GBSSI</i>	-----	
<i>Gsulphuraria_SS</i>	416 YNPL- - - VVYTKNSEGLL I PHAGTDGRE LQ- - - - - WDNTCLLN YRKFEAWELFYQD I YRL I HEFG I	473
<i>Cmerolae_SS</i>	453 YRR I GYVR I VDAKRQVPALPHPGTDCHVQ- - - - - WEDTALLNYRRLESWYSL I ED I SQMATEHGA	513
<i>Ccrispus_SS</i>	158 YRHL- - - TVETLSNKG I PLRHPGTDGRENQ- - - - - WEDTALLNHRRVE TWNLMVAEVKAMAEYGV	215
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	-----	
<i>Creinhardtii_GBSS</i>	-----	
<i>Athaliana_SSI</i>	-----	
<i>Athaliana_SSII</i>	77 --- T I D- - - - KSKKV-----	84
<i>Athaliana_SSIII</i>	210 EAYKMD FVFFNGQSV----- YDNND S-----	230
<i>Athaliana_SSIV</i>	198 N I LRLD- - - - EARATA-----	209
<i>Athaliana_GBSSI</i>	-----	
<i>Gsulphuraria_SS</i>	474 QGVRLD- - - - NAQSYPL I MKADLEELFRVDVDELHYSLDD I LHAKVVKTNEDCGYWLTEAALDFGYPNP	539
<i>Cmerolae_SS</i>	514 GQVRLD- - - - NAQCVPCLAPDVTELAR I DNDG I AHYDDEEKAFGDVVLANVEGGYWRDAA I D- GYPNP	578
<i>Ccrispus_SS</i>	216 GG I RLD- - - - NAQSLPP I MAPNMD ELLRHDPDGEPHYSLSEV FYGAVVKANEE YGYWTSSAG I ERGYPNP	281
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	61 ----- ENE LLR----- AQLALYQQNQQPSVG	81
<i>Creinhardtii_GBSS</i>	29 ----- NQLLRE----- LARGSARKS TSRSAVT	50
<i>Athaliana_SSI</i>	43 ----- RSLL----- LRRSSSFSGDSRESDEERF I -	66
<i>Athaliana_SSII</i>	85 ----- LAMQRNLLHQ I AERRKLVSS----- I KESTPDLD DAKASSKQESAS	125
<i>Athaliana_SSIII</i>	231 ----- KDFCVE I KGGMDKVD FENF LLEE----- KL----- REQEK LAKEEAEREROKEEKR	276
<i>Athaliana_SSIV</i>	210 ----- LDLLNK I LSDKE----- ALQGE I NVLEMK LSETDER I K	242
<i>Athaliana_GBSSI</i>	50 ----- DN LQ----- RRSQAKPVS AKSSKRSSKVK	73
<i>Gsulphuraria_SS</i>	540 FLVRLTKR I WNDFPNF I I LAE AHFQREPOLA----- FSGV I PHT I RVAQ I LAS I CGQS LRRDGSVS	600
<i>Cmerolae_SS</i>	579 LLLKLTREMMHWNPSFLVLGESHFHRE RNLI----- VSGL I PHTLRVAT I LAG I SGKSLRRDGSVR	639
<i>Ccrispus_SS</i>	282 FLVKFCREMMNAFPDF I VMAE AHFHREAQLL----- TSGPVVHTVRI PQ I LAS I SGKSLRKLDGTVG	342
<i>Ecoli_GS</i>	-----	
<i>Atumefaciens_GS</i>	-----	
<i>Creinhardtii_SS</i>	82 ----- AA AVAP-----	87
<i>Creinhardtii_GBSS</i>	51 ----- GATGATCA-----	58
<i>Athaliana_SSI</i>	67 ----- TDAER----- DGSGSVL-----	78
<i>Athaliana_SSII</i>	126 SVN- - - - - ANTDATKK- - - - - E I MDGDANGSVSP-----	149
<i>Athaliana_SSIII</i>	277 R I EAQKAA I EADRAQAKAETQKRRELLQPA I KKAVVS- - - - - AENWY I E----- P	322
<i>Athaliana_SSIV</i>	243 ----- TAAQEKAHVELLEEQLERHEM I SP----- IESDGY-----	274
<i>Athaliana_GBSSI</i>	74 ----- TAGK I VC-----	80
<i>Gsulphuraria_SS</i>	601 KLP- - - - - ESRKSTARTLSRLYRSCKYTMPKGA I QLGCTCTHNSP YPGVLYGRRAWLAVDLLYFLPEVP	664
<i>Cmerolae_SS</i>	640 ALG- - - - - ENKHTVDF I ARLYRNDAAH I PSGA I MVGGTCSDTSP YPSVLYGRRAWTAVDMLCFLPDIP	703
<i>Ccrispus_SS</i>	343 RVP- - - - - GKNRSTARTLSRLYRNDSDWLPKNA I MVNCTCTHSSP YPGLLYGRRSWI AVDLLNFLPEVP	406

Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS	88	-----PAAATK-----VLEKP-----	98
Creinhardtii_GBSS			
Athaliana_SSI	79	-----GFQLTPPGD-----QQT-----	90
Athaliana_SSII	150	-----STYGKSSLSKEPE-----AKT-----	165
Athaliana_SSIII	323	SDFKAEDTVKLYYNKRSGPLTNSKELWLHG----GFNMWWDGL-----SIVVKL-----	367
Athaliana_SSIV	275	-----VLALSKELETLKLENLSLRNDIEMLKSELDVSKDTGERVVVLEKE-----	319
Athaliana_GBSSI			
Gsulphuraria_SS	665	ILFYGEENGRMY---RFNMTVVSQS IETHP----FYDVNYEN-----VLPKS-PRSTD	710
Cmerolae_SS	704	LLLLGEEDGRAY---RINMASVSREVDP-----DTNLEL-----EVPKS-PRLGGH	745
Ccrispus_SS	407	MTVYGEERGRAY---RMNMNGVSNTEEMT----EYDVNFDA-----VLPKSPPLRTGQ	452
Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS	99	-----APAKQASVD-----	107
Creinhardtii_GBSS			
Athaliana_SSI	91	-----VSTSTGEIT-----	99
Athaliana_SSII	166	-----FSPSTESLK-----	174
Athaliana_SSIII	368	-----VNAELKDVPKSG-----NWWFAEVVPPGGALVIDWV	399
Athaliana_SSIV	320	-----CSG-----LESSVKDLESKL-----	334
Athaliana_GBSSI			
Gsulphuraria_SS	711	SSPA----DGISTLVLDELSHLDTDSDPPSKALTTSPSSAKKK---KKKSS----LFSLE-----L	758
Cmerolae_SS	746	GLPRGSSVASGMSML-----HLQ-----QTSMDVRRLL---KRSRSGRETGLARLSS-TNLLGRNTT	797
Ccrispus_SS	453	TMPK---PKGL-----PP---LTPPHSVERKLMKMKRKS----LADL-----	485
Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS	108	-----G-----GIWPK-----PGEAFW-----	120
Creinhardtii_GBSS			
Athaliana_SSI	100	-HHEEKKE-----AI-----	108
Athaliana_SSII	175	-NRKQSSA-----SVISSS-----PVTSPQKPSDVATNGKPW-----	205
Athaliana_SSIII	400	FADGPPKG-----AFLYDNNGYQDFHALVLPQKLPEELYWLEEEENMIFRKLQEDRRLK-----	451
Athaliana_SSIV	335	-----SV-----SQEDVSQLSTLKI ECTDLWA---KVETLQLLLDRATK-----	370
Athaliana_GBSSI			
Gsulphuraria_SS	759	KRSSGSSQ-----SLVRSQ-SIDDVVRGMSIRSVSVDDIGS-----LSRLEEDTRLKIGPGVGYDIRMI	815
Cmerolae_SS	798	NRADSGSMGAAAMVRSR-STEDMLKLSIRSVSAEDIRQ-----LDLAEDIRREIGPQEGYDIRQI	860
Ccrispus_SS	486	RRVPSNS-----NLVRSR-SRDDMNGMSVRSVSAADFRK-----MSAMEEQTRQEIGPASGYDIAQI	541
Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS			
Creinhardtii_GBSS			
Athaliana_SSI	109	-----DQIVMADFGVP-----	119
Athaliana_SSII	206	-----SSVVASSVDPYKPS--SVMTSPEKT-----	229
Athaliana_SSIII	452	-----EEVMRAKMEKTARLK---AETKERTLKKFLL---SQKDVVYTEPLEIQ	494
Athaliana_SSIV	371	-----QAEQAVIVLQQNQDLRNLK---VDKIEESLKEANVYKESSEKIQQYNELMQHKV	420
Athaliana_GBSSI			
Gsulphuraria_SS	816	RGHYEHRLSIRMFHPAFRHGSLTLDVSLHLKEQVFAFVRSDDSLIVVAMNMKCDLDGNEFREPCVEVL	885
Cmerolae_SS	861	RGHYEHRLLRLRASYEALREGHMVALVVREQAKYQVLAFAFRTRQQIVLLFIVNRGGHDGAPFAVPVDTIV	930
Ccrispus_SS	542	EGHYSHRRMLRQELAAALHSGFMCVLTIEPQLKEQVFAFARYTEDQVVIVAANFKDNRDGPQYSNGCDVEL	611
Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS			
Creinhardtii_GBSS			
Athaliana_SSI			
Athaliana_SSII	230	-----SDPVTSPGKPS-----KSRAGAFWS-----DPLPSY	255
Athaliana_SSIII	495	-----GNPVTVLYNPANTV-----LNGKPEWV----FRGSFNRWTHRIGPLP--	532
Athaliana_SSIV	421	TL--LEERL-----EKSDAEIFSYVQLYQESIKEFQETLES LK--	456
Athaliana_GBSSI			
Gsulphuraria_SS	886	DFKPLSETLRNEIYQKNEDKLF-Y-----FLECF-T-KESYPELLTFHELL---FRKYPVKLPLGTI	942
Cmerolae_SS	931	DLRPLAEALVAAAFRNG--CTY-FGPWDRIEELRDCF-TGAREPSRYAFEEFL---FRRLQITLKPLETE	993
Ccrispus_SS	612	DFQTLWDVLPDFTTGAAPCAF-Y-----SVVNTFTGKEHSTDVQTLLELV---FRKYKTHLDPLGIS	670
Ecoli_GS			
Atumefaciens_GS			
Creinhardtii_SS	121	-----ERSPRASP-----MP-----LQ-----	132
Creinhardtii_GBSS			
Athaliana_SSI			
Athaliana_SSII	256	L-----TKAPQTSTMKTEKYVEKTP-----	275
Athaliana_SSIII	533	----PQKMEATDDESSHV-KTTAKVP-----LDAYMDFV-----	562
Athaliana_SSIV	457	----EESKKKSRDEP-----VDDMP-----WDYWSRLLLT-----V	483
Athaliana_GBSSI			
Gsulphuraria_SS	943	VLSPERSTQSRNEHLHLEQCLARLQ-MDGVDMKDPRENALAWKLTTRAAADS LQQFSNVFYDIYSLMS--	1009
Cmerolae_SS	994	I LELCPSSAAESSTDLHQQTQVRLD-EDDEMLLDARANWLAGYLARHCYD-LKAFARALGFLQGLLAPG	1061
Ccrispus_SS	671	LLTFKPVQDTPERRGAHFSECI NRLRSQEANDIKDARENDI IARLARGAAASASDFVGAMESLRNGLR--	738

<i>Ecoli_GS</i>	-----		
<i>Atumefaciens_GS</i>	-----		
<i>Creinhardtii_SS</i>	133	-----GGAAEAPPV-----ERD-----	144
<i>Creinhardtii_GBSS</i>	-----		
<i>Athaliana_SSI</i>	120	-----GNRAVEE-----GAAEVGIP-----	134
<i>Athaliana_SSII</i>	276	-----DVASSETNEP-----GKDEEKPPPLAG-----	297
<i>Athaliana_SSIII</i>	563	---FSEKEDG---GIFDNKGLDYHLPVVGGISKEPP-----	593
<i>Athaliana_SSIV</i>	484	DGWLLEK-----KIASNDADLLRDM--WVKDDRRIHD-----TYIDVKDKNERD-AISAFLLKV	534
<i>Athaliana_GBSSI</i>	-----		
<i>Gsulphuraria_SS</i>	1010	ASGMLESTIVHLCVSVCLQRTATVPR-----YEAFYAPKNLIPPRGERVLSLLIHLSCCAKDFQFRKTCQKI	1074
<i>Cmerolae_SS</i>	1062	KQQLTEARARHLMCVCLQRTALLA-----GEQEYPPHGFRAVSGERLLAYLMLVLSSCGID-ALRTVARRI	1125
<i>Ccrispus_SS</i>	739	NEGCEKAEMERIMQLCMQRASQLRFM-VAYEGVPRPKDFEPPAAEHI VAYLTHMSTCAKDPDLMTLARSV	807
<i>Ecoli_GS</i>	1	-----MQVLHVCSMFPLKKTGGLADVIGALPAAQIADGVDAVLLPFAF-----PDIRRGVT-----	52
<i>Atumefaciens_GS</i>	1	-----MNVLSVSSIEIYPLIKTGGGLADVVGALPIALEAHGVRTRTLIPGY-----PAVKAAVT-----	52
<i>Creinhardtii_SS</i>	145	---GNPMHIHITAEAMAPIAKVGGLGDVVTVGLAKAALARGHFVTVMLPFY-----ECLPKDQIEGLKHE	205
<i>Creinhardtii_GBSS</i>	59	---LDIVMVAEEVAPWSKTGGGLGDVVTVGLPIELVKRGHRVMTIAPRY-----DQYADAW-----	109
<i>Athaliana_SSI</i>	135	SGKAEVNNLVFVTSEAAPYSKTGGGLGDVCGSLPIALAGRGRHVMVISPRLNGTAADKNYARAK-----	199
<i>Athaliana_SSII</i>	298	---ANVMNVILVAAECAPFSKTGGGLGDVAGALPKSLARRGHRVMVVPVRY-----AEYAEAK-----	351
<i>Athaliana_SSIII</i>	594	---LHVHIAVEMAPIAKVGGLGDVVTVLSRAVQELNHNVDIVFPKY-----DCIKHNFVK-----	646
<i>Athaliana_SSIV</i>	535	SSPTSSGLYVHI AEMAPVAKVGGLGDVVTVAGLGKALQRKQHLVEIILPKY-----DCMQYDRVRDLR-A	598
<i>Athaliana_GBSSI</i>	81	---EKGMSVIFIGAEVGPWSKTGGGLGDVVGGLPPALAARGHRVMTICPRY-----DQYKDAW-----	134
<i>Gsulphuraria_SS</i>	1075	LKNISEIGPLVFAPELGRFSTAGGLGVMDVDELTKGLAALGSEVYVISPYY-----SVNRKGQWKYL--E	1137
<i>Cmerolae_SS</i>	1126	LSG-NRMGPIVLFSPLELGRFSSIGGLGTMVDELSKGLADLGLDYYVISPAY-----TFNRRGETRYL--E	1187
<i>Ccrispus_SS</i>	808	VAKTTKLGPLVFLTAELGRFSTAGGLGVMDVDELTKGLAGLGLLEVYVSPYY-----TVNRKNRSGYL--	869
<i>Ecoli_GS</i>	53	-D-AQVVSRRDTFAG--HI-----TLLFGHYNGVGIYLIDAPHL-----YDRPGSPYHDTNLFAYTDNVL	108
<i>Atumefaciens_GS</i>	53	-DPVKCFEFTDLLGE-----KADLLEVQHERLDLLILDAPAY-----YERSGGPYLGQTGKDYPDNMK	109
<i>Creinhardtii_SS</i>	206	CD-IEVPKGYRWDGE-IRVGPLKTSVFWGRVGGCPVYLIPKADDTN-CNIFRGGRIYGGG-----YNEME	267
<i>Creinhardtii_GBSS</i>	110	-D-TSVV--VDIMGE--KV-----RYFHSIKKGVHRWIDHPWF LAKWVGKTGSKLYGPRSGADYLDNHK	168
<i>Athaliana_SSI</i>	200	-D-LGIRVTVNCFFGGSQEV-----SFYHEYRDGVDWVFDHKS-----HRPGNPGDSKGA-FGDNQF	255
<i>Athaliana_SSII</i>	352	-D-LGVRKRYKVAQGDMEV-----MYFHAFIDGVDFVIFDSPEF-----RHLSNNIYGGN---RLDILK	405
<i>Athaliana_SSIII</i>	647	-D-LQFNRSYHMGGT--E-----KVMHGKVEGLSVYFLDPQNG-----LFQRGCVYG-----CADDAG	696
<i>Athaliana_SSIV</i>	599	LD-TVVESYFD--GKL--YKNKIWI GTVEGLPVHFIEPQHPSK--F-FWRGQFYG-----EQDDFR	651
<i>Athaliana_GBSSI</i>	135	-D-TCYVVQIKVGDKVENV-----RFHCHYKRGVDRVFDHPHIFLAKVVGKTGSKIYGPITGVYDNDNQL	197
<i>Gsulphuraria_SS</i>	1138	PDGIIWTRNIQVRVGRSDV--TVGVYEGVEEGVHLIFLEQGEYYP-----KVYADMANQ--RRQLE	1194
<i>Cmerolae_SS</i>	1188	RDGIRWTRNIDVRI GNVGV--ATLGI FEG IENSURLIF FENHSYFP-----RVYQDLGSQ--ARMME	1245
<i>Ccrispus_SS</i>	870	GDNIQWTRNISVNI GTHIV--EAGVFEGVENDVNLIFLERGDFFP-----KVYADPPGGS--VRHLQ	926
<i>Ecoli_GS</i>	109	RFALLGWGAE MASGL-----DPFWRP DVVHAHDWHAGLAPAYLAARGRP-----AKSVFTVHNL	163
<i>Atumefaciens_GS</i>	110	RFAALS LAAARIGAGV-----LPGWRPDMVHAHDWQAAMTPVYMRYAETP-----EIPSLTITHNI	165
<i>Creinhardtii_SS</i>	268	AYLYFCRACLE-----YLNVSQQNPHVLQLHDWHAHAASMLYWDVYVNPNG-FSRTRLMLTITHNL	325
<i>Creinhardtii_GBSS</i>	169	RFALFCKAAIEAARVLPFGP-----GEDCVFVANDWHSALVPVLLKDEYQPKGQFTKAKSVLAITHNI	230
<i>Athaliana_SSI</i>	256	RFTLLCHAACEAPLVLP LGGFTY-----GEKSLFVNVDWHAGLVPILLAAKYRPGVYKIDARSILITHNL	320
<i>Athaliana_SSII</i>	406	RMVLFCKAAVEVPWYVPCGGVCY---GDGNLAFIANDWHTALLPVYLKAYYRDHGMKYTRSVLVIITHNI	471
<i>Athaliana_SSIII</i>	697	RFGF FCHAAL-----FLLQGGFHPDILHCHDWSSAPVSWLFDKHYTQYG-LIKTRIVFTITHNL	754
<i>Athaliana_SSIV</i>	652	RFSYFSRAALE-----LLQSGKKPDIHCHDWQTAFAVAPLYWDLYAPKG-LDSARICFTCHNF	709
<i>Athaliana_GBSSI</i>	198	RFSLLCQAALAPQVLNLSKSYFSGYGEDVVFVANDWHTALLPCYLKSMYQSRGVYMNKAVVFCITHNI	267
<i>Gsulphuraria_SS</i>	1195	LIVLTSLASLE-----VLCQKSLPPALFITNDWI AALSAGYAKQGF FGSF-FENTTF FHI IHNL	1252
<i>Cmerolae_SS</i>	1246	MLVLANRGVLE-----ICCHKQLRPSLLVTNDWAMGGLVPAYGRLGYFGGY-FDDTCF FHI VHNL	1303
<i>Ccrispus_SS</i>	927	TVVLMSLGSL-----VCCQKQLYPSVIVTNDWLPSMAAGY--RDFFGDY-FKNTSFFH IHNL	982
<i>Ecoli_GS</i>	164	---AYQGMF---YAHHMNDI-QLPWSF---FN-----IHGLEF	191
<i>Atumefaciens_GS</i>	166	---AFQGF---GANIFSKL-ALPAHA---FG-----MGEIEY	193
<i>Creinhardtii_SS</i>	326	---DNTGET---RQDEFFFT-GVPGENFATIDK-----ALDE---R-TIGHNP	362
<i>Creinhardtii_GBSS</i>	231	---AFQGRM---WEEAFKDM-KLPQAA---FDKLAFSGDYAKV-----YTEATPMEEDEKPLLTGKTY	283
<i>Athaliana_SSI</i>	321	---AHQGE---PAATYTNL-GLPSEW-----YGAVGWVFTWARTHALD-----TG---	360
<i>Athaliana_SSII</i>	472	---AHQGRG---PVDDFSYV-DLPSHYLDSFKL-----YD-----PVGG---	503
<i>Athaliana_SSIII</i>	755	-----EFG---	757
<i>Athaliana_SSIV</i>	710	---EYQGT---SASELGSC-GLDVNQLNRPDR-----MQDH-----SSG---	742
<i>Athaliana_GBSSI</i>	268	---AYQGRF---AFDDYSLL-NLPI SFKSSFD---FMDGYEK-----PVKG---	303
<i>Gsulphuraria_SS</i>	1253	GDAVYEGRVYPNESEGLFEDVHRLPVLSL--VFD-----PSWSQ-----1288	
<i>Cmerolae_SS</i>	1304	GDAAYEGRLYSPAEAGFEA I HQLPRDI--VID-----PTWNR-----1339	
<i>Ccrispus_SS</i>	983	GEGAYEGRVYNNPGEGLDHIHRLP THV--MVN-----PWWST-----1018	
<i>Ecoli_GS</i>	192	NGQISFLKAGLYADHITAVSPTYAREITE-PQFAYGMEGLLQQRHREGRLSGVLNGVDEKIWSPETDLL	260
<i>Atumefaciens_GS</i>	194	YNDVSFLKGGQLTATALSTVSPSYAEILT-AEFGMLEGVI GSRAH--VLHGIVNGIDADVWNPATDHL	260
<i>Creinhardtii_SS</i>	363	-ERLNLKMGKGIYYCNAVTTVSPTYANVNLNGGAAGWLRSTFARPELR-SKFHGLINGIDCEEVNPATDAL	430
<i>Creinhardtii_GBSS</i>	284	-KKINWLKGGIIAADKLVTVSPNYATEIAADAAGGVELDVTIRAKG---IEGIVNGMDIEEWNPKTDKF	348
<i>Athaliana_SSI</i>	361	-EAVNVLKGAIVTSDRIITVSQGYAWEITT-VEGGYQLDQLSSRKS--VINGITNGINVDENWNPSTDEH	426
<i>Athaliana_SSII</i>	504	-EHNIFIAAGLKAADRVLTVSHGYSWEVKT-LEGGWGLHNIINENDW--KFRGIVNGIDTQEWNPSTFDY	569
<i>Athaliana_SSIII</i>	758	---ANAI GKAMTFADKATTVSPTYAKEVAG-----NSVISAHLY--KFHGLINGIDPDIDWPYNDNF	814
<i>Athaliana_SSIV</i>	743	-DRVNPVKGAIIFSNIVTVSPTYAQVVRT-AEGGKGLHSTLNFHSK--KFIGILINGIDTDSWNPATDPF	808
<i>Athaliana_GBSSI</i>	304	-RKINMMKAAILEAHRVLTVSPYAAQELISGVDRGVELHKYLRMKT---VSGIINGMDVQEWNPSTDKY	368
<i>Gsulphuraria_SS</i>	1289	-PIVFNPSRCALLCSDTWGTVSNLSYLELVTT---FHP LKHILRLAR---CPFGYPNGIRKKE-----1342	
<i>Cmerolae_SS</i>	1340	-VVVNPSRAAFKCAHSWGTVSPSYLELLT---NHLRPMVRLCR---APFGTSNGIRVQD-----1393	
<i>Ccrispus_SS</i>	1019	-LVVNPSRCAIMRSESWGTVSPSYLRELR---GHP LSDL LQQAQ---SPFAYPNGIRKAE-----1072	

