THE UNIVERSITY
of EDINBURGH

Manipulation of Storage Polysaccharides in Microorganisms

Joseph White

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell Biology

The University of Edinburgh

September 2014

Contents

Declaration

I hereby declare that this thesis was completed by me, and the research presented is my own, except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Joseph White

Acknowledgements

I would like to thank Chris French for giving me the chance to undertake a PhD in his laboratory, and for his unfailing support and advice throughout the last four years. I would also like to thank the other members of the French lab, past and present, who have helped me over the course of this research. Especial thanks goes to David Radford, Maryia Trubitsyna, Jan Oltmanns, Craig Munns and Alejandro Salinas Vaccaro. I would also like to thank various members of surrounding laboratories for their help and advice along the way.

My thanks also goes to the Millar lab for their expert help in all matters to do with *Ostreococcus tauri*, and the Danos lab for their help and insight into polysaccharide synthesis and structure.

Finally, I would like to thank the BBSRC for funding my PhD.

Abstract

The rising demand for arable land, to meet the competing needs of food and energy for a growing population, will soon become unmanageable. There is therefore a pressing need to increase the efficiency with which these demands are met, or find alternative ways of meeting them. In this work, *E. coli* transformed with a copy of its own ADP-glucose pyrophosphorylase gene (*glgC*), under a lac promoter, was found to synthesise large inclusion bodies when grown in media supplemented with lactose and IPTG. Analysis of these inclusion bodies suggests that they are formed of polysaccharide, giving the cell a significantly higher total sugar content than a control grown under the same conditions. The inclusion bodies were found to react strongly with iodine, turning a blackish brown colour normally associated with iodine-starch reactions. Results also indicate that the bacteria are unable to digest these inclusion bodies once they have been formed. These findings suggest the presence of long α-helices, which would both bind with iodine and prohibit enzymatic digestion. It was therefore hypothesised that the extra GlgC enzymes were allowing glucan chains to extend at a faster rate during glycogen synthesis, leading to unbranched regions that were long enough to wind themselves into helices. The subsequent introduction of a copy of the *E. coli* branching enzyme gene (*glgB*), under the same lac promoter, was therefore expected to abolish the inclusion body phenotype, by allowing the branching of the polysaccharide to keep speed with the synthesis of its linear chains. This was indeed found to be the case. However, cells transformed with both additional gene copies were also found to accumulate a significantly higher total sugar content again: more than twice that of cells transformed with *glgC* alone and more than seven times that of a control, when grown under conditions designed to optimise polysaccharide synthesis. These transformants were also observed to grow to higher cell densities that a control, in various growth media. The results of both these transformations could be significant in meeting the demands of our growing society. In particular, the use of cyanobacterial glycogen as a carbon source for biofuels has recently been gaining interest, and the work presented here may well be applicable in this field, providing the possibility to significantly increase yields. Lastly, the effects of Isoamylases and Granule Bound Starch Synthase, taken from two starch producing organisms – *Zea mays* and *Ostreococcus tauri* – were investigated in an *E. coli* host. Results were inconclusive, but suggest many avenues for continuing the work.

Lay Summary

The population of the world is growing in both size and affluence. There are predicted to be over nine thousand million individuals on the planet by 2050, all consuming more food and energy than they have before. Concerns over anthropogenic global warming, brought about by fossil-fuel emissions, coupled with the growing energy demands from our ever-increasing population, have triggered a search for alternative sources of energy. One such alternative is biofuel, which has been successfully implemented in many countries. However, the large-scale use of current biofuels, made from existing plant energy stores such as corn starch, creates a demand for even more arable land. The existing biofuel industry is comparatively extremely small, but still the rise in food prices caused by this competing land use has probably been felt already. Meanwhile, one of the many negative effects of the increase in greenhouse gas emissions and the anthropogenic global warming that accompanies it, is a significant reduction of the land able to produce either food or biofuel crops. The result of all these factors is that, even allowing for the widescale destruction of natural ecosystems, demands for land will soon become unmanageable.

Polysaccharides such as starch and glycogen are complex sugars used by organisms as energy stores. They also account for around 70% of the food eaten by humans globally, and can be easily converted into biofuel. A better understanding of the biological systems that lead to the creation of polysaccharides may therefore lead to ways of producing food and energy with greater efficiency, or finding novel, more manageable means of production.

In this work, the polysaccharide stores of a model organism – the bacteria *Escherichia coli* – were altered by mutating the bacteria through the addition of extra copies of its own polysaccharide synthesis genes. In this way, bacteria were created that produced over twice the normal level of storage polysaccharide, in the form of a granule that may allow easy harvest. Another form of mutant *Escherichia coli* was created that produced over seven times the normal level of storage polysaccharide, and also seemed to grow to higher cell densities in a liquid culture. Such manipulations may therefore have potential for increasing the polysaccharide stores of organisms grown for biofuel production.

Chapter 1: Introduction

1.1 Storage Polysaccharides

Carbohydrates are a fundamental currency of life. They provide fuel for the Krebs cycle: a series of chemical reactions that is used across all life on Earth to produce energy as well as the precursors for the biosynthesis of many crucial molecules. This cycle – also known as the citric acid cycle - was almost certainly among the earliest of the life-defining reactions to evolve, when the biology we recognise today was just starting to cluster together, to self-regulate and resist the tug of entropy around 4 thousand million years ago. It has even been suggested that it evolved abiogenically, actually pre-dating the origin of life (Lane, 2009). The Krebs cycle is fed with acetyl-CoA derived from pyruvate, a three-carbon molecule which, on an ancient and barren planet (as well as many nooks and crannies of a proliferating one) can be formed from carbon dioxide and hydrogen, but which is now mostly formed in cells by glycolysis: the splitting of glucose.

The ability to store surplus carbohydrates can therefore confer significant evolutionary advantage. As such, the synthesis of large, complex polysaccharides such as glycogen and starch as a means of such storage is a phenomenon found across the spectrum of life on Earth, with the similarity of those polysaccharides hinting at an extremely ancient origin: it is thought that the universal common ancestor of every organism alive today was already using a glycogen-like polysaccharide to store energy when it lived and died back on the Archael Earth at least three and a half thousand million years ago (Zmasek et al., 2014).

1.2 Glycogen structure and function

Glycogen is composed exclusively of glucose residues: molecules of glucose that have been stripped of an oxygen atom (or occasionally two), allowing them to be linked together. The links are generally formed between the hydroxyl group at the α-1 carbon of one glucose residue and the carbon 4 of its neighbour, with the oxygen from one of them bridging this α-1,4 link. The occasional second link is formed within each chain, where side branches can occur, forming between the hydroxyl group of the α-1 carbon of the first glucose residue of the branch and the carbon 6 of the glucose residue within the pre-formed chain (figure 1.1). Although these branches are formed randomly, making the structure of glycogen a difficult one to characterise (Sullivan et al., 2010), they are thought to occur at an average ratio of around two branches per chain, giving the entire molecule an homogenous and roughly tiered growth.

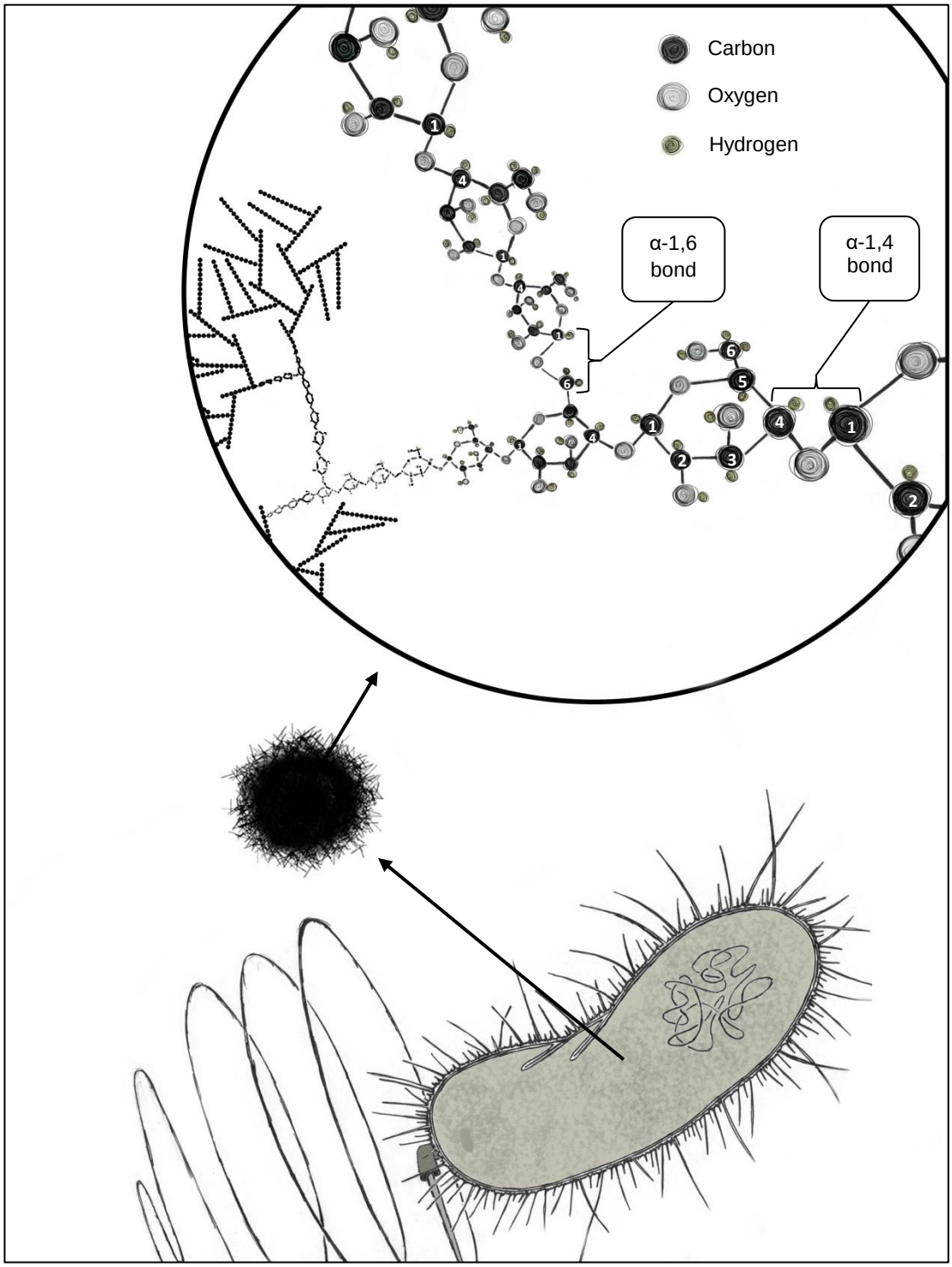


Figure 1.1: Glycogen structure (idealised). Glycogen molecules are small enough to remain soluble within the cell cytoplasm. Each molecule is made up of glucose units, bound into chains by α-1,4 bonds, with branch points forming from α-1,6 bonds.

A mathematical relationship between the structure of the glycogen molecule and its metabolic function has been suggested by Meléndez-Hevia et al. (1993), and thanks to this relationship it has been possible to derive a theoretical structure for the molecule based around defining two parameters: the amount of glucose directly available to phosphorolysis, and the compactness of the structure. The model suggests the following optimisation: a standard chain length of 13 glucose residues, with two branches per chain, forming an overall structure that is 12 tiers in size. By maintaining a regular degree of branching and limiting the size of the molecule in this way, an organism would maintain its glycogen's structural homogeneity, which is integral to its efficiency as a storage molecule, as it permits the highest level of total glucose as well as the highest level of glucose directly available to phosphorolysis and therefore use as an immediate energy supply. These models suggest that glycogen which continues to grow beyond 12 tiers would quickly loose homogeneity, and thereby reduce the amount of available glucose in the outer tier (Meléndez et al., 1998).

1.3 The glycogen synthesis pathway

Glycogen has been reported to exist in more than 50 different bacterial species, where the enzymology of its biosynthesis and degradation appears highly conserved (Preiss 2009). In both free-living and mildly parasitic bacteria, its presence is associated with a minimum of one of each of the following enzymes: ADP-glucose pyrophosphorylase, glycogen synthase, branching enzyme, debranching enzyme and glycogen phosphorylase (Ball & Morell, 2003). *E. coli* possess exactly this core pathway of five enzymes, which, in this organism, have been designated the following: GlgC for the ADP-glucose pyrophosphorylase, which creates the ADP-glucose units that the glycogen molecule is constructed from; GlgA for the glycogen synthase, which severs the bond between the ADP and the glucose of those units, while simultaneously linking the glucose to the end of the growing glucan chain with an α-1,4 bond; GlgB for the branching enzyme, which cuts off the end of a linear chain before re-attaching it to a mid-chain glucose residue via an α-1,6 bond; GlgX for the debranching enzyme, which cuts the α-1,6 bonds at the branch points, liberating the branch for further degradation and metabolism; and GlgP for the glycogen phosphorylase, which phosphorolyses the glucose residues from the ends of the chains inwards, plucking them off one at a time to be metabolised by the cell. With only this cast of workers, it would seem at face value that any specific parameters required to build an optimum molecule would be hard to regulate. However, such structural control is made possible by the substrate specificities of those enzymes.

1.3.1 ADP-Glucose pyrophosphorylase (GlgC)

GlgC, the ADP-glucose pyrophosphorylase (pyrophosphorylase being a diphosphate transferase), follows a biochemical cascade, where an enzyme from the phosphotransferase system first converts glucose taken in from the environment into glucose-6-phosphate (as it does for the first step of glycolysis). This in turn is converted by phosphoglucomutase into glucose-1-phosphate, which, in the presence of Mg^{2+} and ATP, is converted into ADP-glucose and inorganic phosphate by GlgC (Ballicora et al., 2003; Ball et al., 2010) (figure 1.2). *E. coli* mutants lacking in either the phosphotransferase or phosphoglucomutase genes have been shown to display glycogen-less phenotypes (Eydallin et al., 2007; Montero et al., 2009). This synthesis of ADP-glucose is both the first committed step and the rate-limiting step of glycogen synthesis (Ball & Morell, 2003). Through the regulation of this enzyme, the cell can control ADP-glucose production, and since these ADP-glucose molecules are the units that the glycogen chains are made from, limiting their formation limits the speed at which glycogen can be synthesised.

It was previously thought that GlgC was the sole supplier of ADP-glucose to glycogen synthesis. However, the absence of glycogen in *E. coli* lacking *glgC* has since been ascribed to the polar effect of this deletion on the downstream glycogen synthesis gene, *glgA* (Montero et al., 2009). More recent and comprehensive experiments have shown that, in fact, mutants lacking GlgC are still able to accumulate glycogen, indicating that it is not the sole source of the ADP-glucose used to build glycogen molecules. GalU, for example, is another nucleotide-sugar pyrophosphorylase. It is thought to produce UDP-glucose from glucose-1-phosphate and UTP, but reports have also shown that it is able to produce ADP-glucose from glucose-1-phosphate and ATP, and indeed *E. coli* lacking the *galU* gene show reduced glycogen content (Moran-Zorzano et al., 2007). However, although bacteria lacking *glgC* do still accumulate glycogen, it is to a far lesser degree than in wild-type cells. So while GlgC may not be the sole supplier of ADP-glucose, it is certainly the primary means by which the bacteria studied accumulate this substrate (Eydallin et al., 2007).

1.3.2 Glycogen synthase (GlgA)

GlgA, glycogen synthase, is responsible for taking the ADP-glucose molecules made by GlgC and attaching them one at a time to the non-reducing ends of the growing glucan chains of glycogen (figure 1.2). When grown on glucose, mutants lacking the *glgA* gene are unable to synthesise glycogen (Montero et al., 2009).

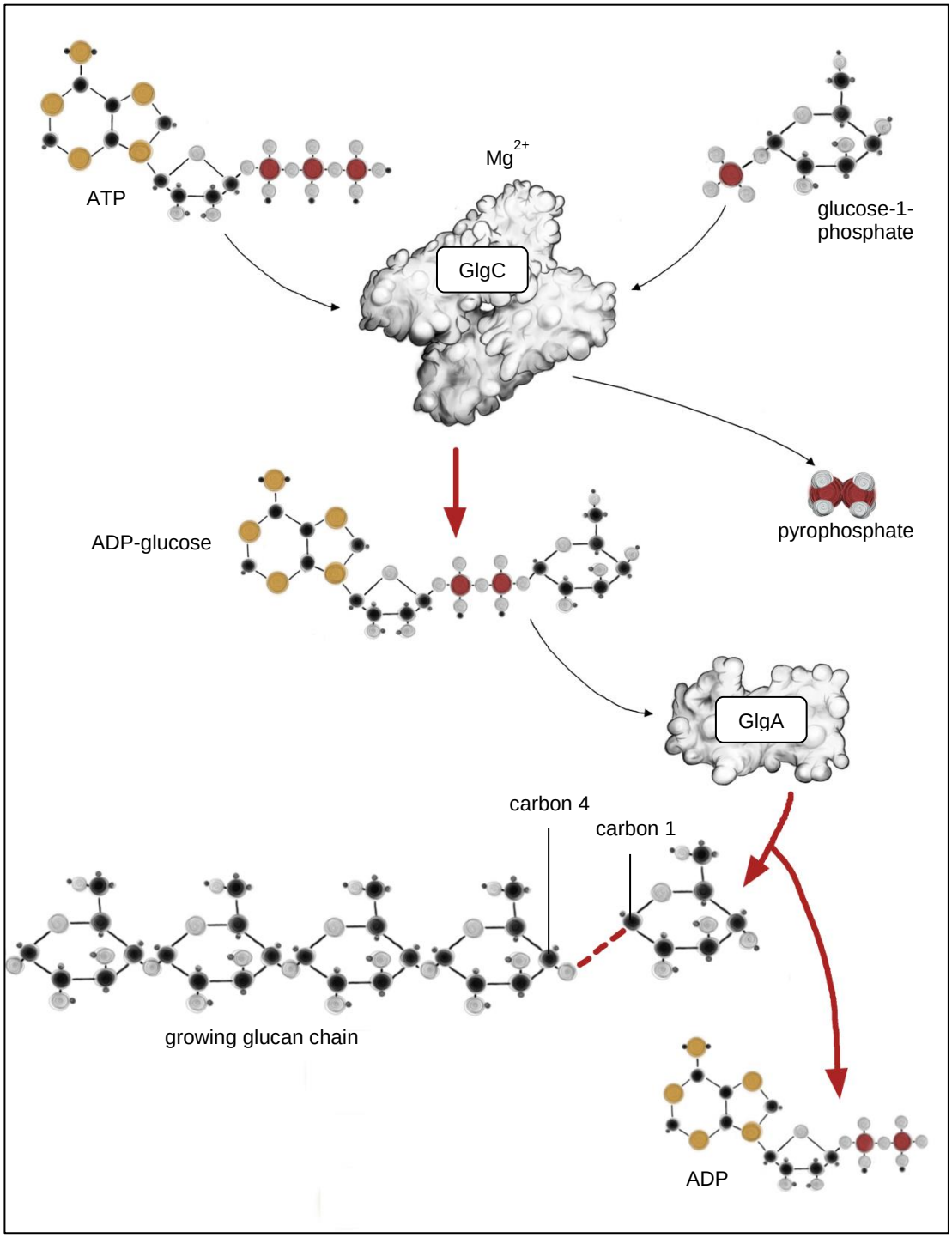


Figure 1.2: Activity of GlgC (ADP-glucose pyrophosphorylase) and GlgA (glycogen synthase). GlgC converts glucose-1-phosphate and ATP into ADP-glucose and pyrophosphate in the presence of Mg^{2+}. ADP-glucose is used as a 'building block' by GlgA, which severs the phosphate bond at the carbon 1 and uses the energy to create a covalent bond between the carbon 1 and the hydroxyl group of the carbon 4 on the glucose residue at the non-reducing end of a glucan chain.

The GlgA enzyme belongs to a glucosyltransferase family, whose members share a similar structure of two domains separated by a cleft in which the active site is located (Sheng et al., 2009), which severs the phosphate bond at the carbon 1 of the ADP-glucose and uses the energy released to create a covalent bond between that carbon 1 and the hydroxyl group of the carbon 4 on the glucose residue at the non-reducing end of a glucan chain. The two domains approach each other to generate a competent active site for this transfer of the glucose residue. The two domains then open again, to release the now elongated glucan chain as well as the freed ADP, in preparation for another round of catalysis (Sheng et al., 2009). Crystallisation of the enzyme in its bound state to ADP-glucose and a glucan chain suggest that the glucan chain binds only to the N-terminal domain, further suggesting that, *in vivo*, glycogen only binds to one side of the enzyme, allowing unencumbered interdomain movement (Sheng et al., 2009).

While opisthokonts (fungi, animals and a few other heterotrophic eukaryotes) require a primer – glycogenin – to initiate glycogen synthesis, it has been shown that bacteria lack any analogue for this protein (Deschamps et al., 2008a). Meanwhile, strong evidence has been provided by Ugalde et al. (2003) that GlgA from *Agrobacterium tumefaciens* (and by extension, perhaps most or all bacterial GlgA enzymes) can catalyse its own glucosylation, transferring glycogen residues from ADP-glucose onto its own amino acids, so that it then became the substrate for further GlgA catalysed glucan elongation, acting as its own primer.

1.3.3 Glycogen branching enzyme (GlgB)

GlgB is the glycogen branching enzyme. It is responsible for creating the α-1,6 linked branching points, attaching the α-1 hydroxyl group at the start of the new branch to the carbon 6 hydroxyl group of a glucose residue within a pre-formed glucan chain. It is the only enzyme so far observed that can create α-1,6 linkages within glucans (Thiemann et al., 2006), so that mutants which lack *glgB* cannot accumulate any branched glycogen, and instead produce low yields of amylose-like, linear polysaccharide (Lares et al., 1974). GlgB operates by a two-step activity, where it first hydrolyses an existing α-1,4 link within a linear section of glucose residues at the non-reducing end of a chain. This hydrolysis is generally supposed occur between five and 14 glucose residues from the end of the chain, certainly always being over four glucose residues from the end, with the enzyme showing limited activity on longer chains (Wilson et al., 2010; Dauvillée et al., 2005; Wang & Wise 2011; Devillers et al., 2003). The enzyme then reattaches this severed chain-end to the carbon 6 hydroxyl group of a residue in the middle of a pre-existing chain (figure 1.3).

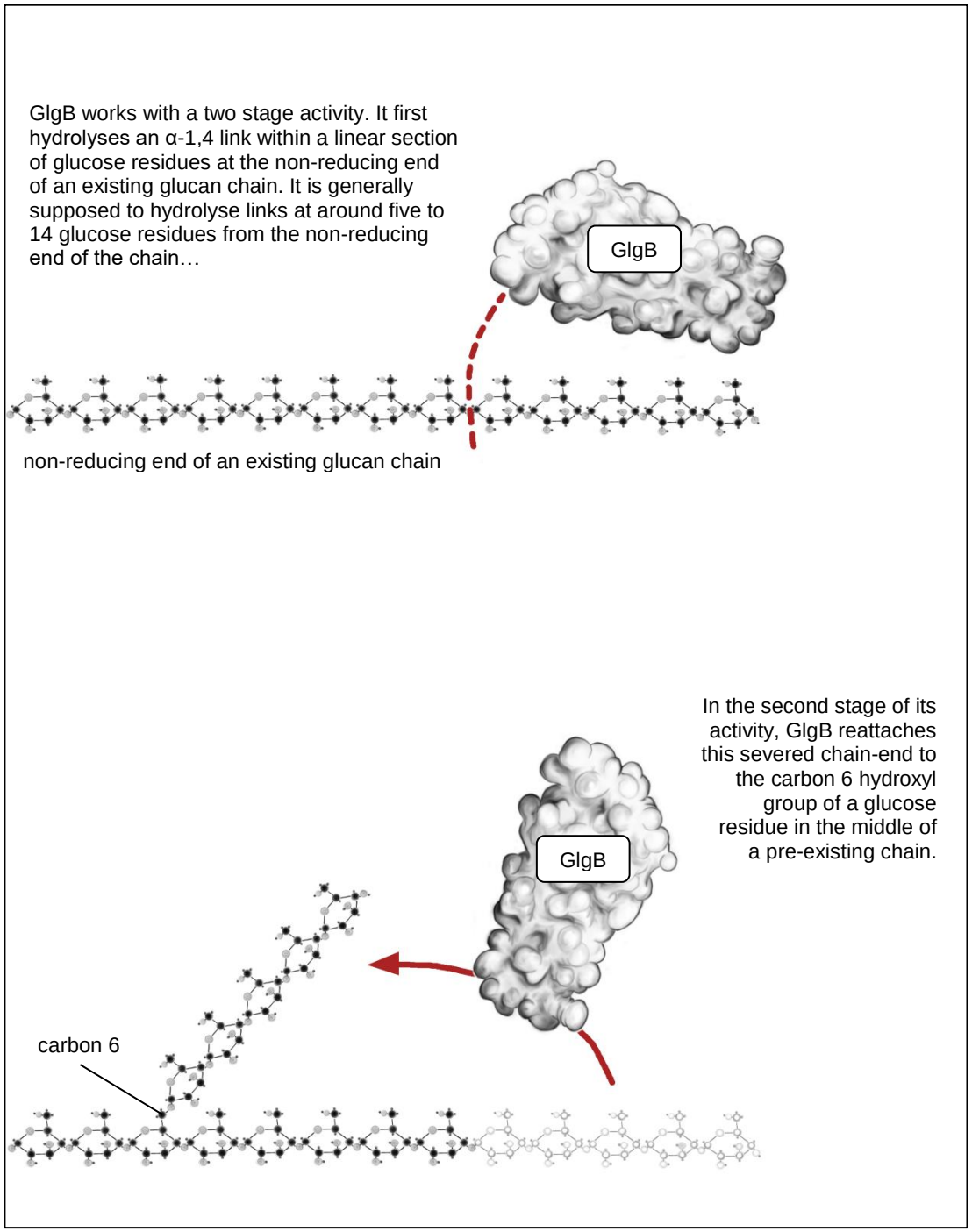


Figure 1.3: Activity of GlgB (branching enzyme). GlgB is the only enzyme known to create α-1,6 links within glucans. It first cuts a section off the non-reducing end of an existing glucan chain, then reattaches that chain section to the middle of an existing chain by an α-1,6 link. The chain ends that it severs and reattaches are always over four residues long, which avoids the problem of new branches being instantly pruned again by GlgX (see later).

The branching enzyme belongs to an α-amylase family of enzymes, all containing a central (α/β)-barrel which forms the catalytic site of the enzyme. The branching enzymes also have a C-terminal domain and an N-terminal domain; features which they share with the isoamylases of starch synthesis and degradation. There is strong evidence to suggest that the length of the N-terminal domain is key to determining the length of the cleaved chain. Mutated forms of the *E. coli* branching enzyme with shortened N-terminal domains showed no reduction in substrate specificity. However, the more the N-terminal domain was truncated, the longer were the chains that the enzyme cleaved and transferred, so that while the wild type enzyme mainly transfers oligosaccharides of around five to 14 glucose residues, an enzyme with 83 amino acids removed from its N-terminal domain preferentially transferred oligosaccharides of around 10 to 20 glucose residues, and removal of 112 amino acids led to an enzyme that cleaved and transferred a higher proportion of longer chains, of 15 to 20 glucose residues. This suggests that the N-terminal domain provides support for the glucan substrate during cleavage and transfer of the α-1,4 glucan chains, and aids in specifying which α-1,4 link is cleaved by the central catalytic domain of the (α/β)-barrel. Longer chains bind to the N-terminal domain and so avoid cleavage, while shorter chains would remain exposed and available for cleavage. The elongation of the N-terminus by domain duplication may therefore be an evolutionary mechanism by which the pattern of branching can be controlled. In support of this, the N-terminal domains from the branching enzymes of different bacteria (*E. coli*, *Haemophilus influenzae*, *Agrobacterium tumefaciens*, and *Pectobacterium chrysanthemi*) show a marked divergence, with a similarity of only 22%, while the similarity of their (α/β)-barrels is 51% (Binderup et al., 2000 & 2002; Devillers et al., 2003).

The breaking of an α-1,4 link and formation of an α-1,6 link is thermodynamically favoured, so that if α-1,4 links and α1-6 links were able to exchange freely and reversibly, the equilibrium mixture would be expected to contain around 70-90% of α-1,6 links (French, 1964). However, this is not the ratio found within living cells; rather, glycogen molecules tend to have an α-1,6 linkage ratio of around 9% (Theimann et al., 2006). The reason for this is likely to be due to the enzyme's substrate specificity. The enzyme will have a 'footprint' size, which is the number of glucose residues it needs to bind to during its operation, which will likely result in a gap of a certain number of residues it has to leave between each branch point when selecting a carbon 6 hydroxyl group within a glucan chain to attach its newly severed chain-end to.

Modelling by Meléndez et al. (1998) has suggested the size of this gap to be optimal at four glucose residues or more. If this is the case, a chain would have to be at least 11 glucose residues long before it could harbour two branches (while leaving a tail of one glucose residue after the second branch) or 16 glucose residues long before it could harbour three branches, and so on. Any chains of fewer than 16 glucose residues would therefore be limited to a maximum of two branches. Furthermore, as the glycogen molecule grows and more branches are added, they consequently become more tightly packed. This should limit the GlgB enzyme further, as it becomes unable to reach the relatively sparse inner regions of the glycogen molecule through the thicket of surrounding chains. This suggested footprint size is purely hypothetical. However, some minimum necessary distance between branch points is likely to exist, simply due to the physical restraints of the enzyme.

1.3.4 Glycogen debranching enzyme (GlgX)

GlgX is the debranching enzyme. It is a glucosidase, which can cleave the α-1,6 links at the base of the glycogen branches by hydrolysis, thereby releasing the stub of the branch as an oligosaccharide (Chiba 1997; Deschamps et al., 2008a) (figure 1.4). Importantly, GlgX cleaves branches of four glucose residues or fewer, so that the stub cleaved by GlgX is generally a maltotetraose. Dauvillée et al. (2005) have shown that the enzyme has strong substrate specificity, working with high preference on three or four residue chains and showing very little activity on longer chains. As such, it shows very little activity on the mature glycogen molecule in isolation (Song et al., 2010). This chain length specificity is of particular importance during synthesis, since when GlgB is creating branch points it does so almost exclusively by adhering branches of five glucose residues or more. GlgX can therefore only debranch chains that have already been shortened by phosphorolysis, which prevents the combined enzymic action from falling into a futile cycle of successive branching and debranching (Cenci et al., 2014; Chiba 1997; Dauvillée et al., 2005).

Three-dimensional models of the GlgX enzyme have suggested that this specificity is caused by the morphology of the substrate binding groove (Song et al., 2010). GlgX belongs to a family of glucosylhydrolases that are defined by a conserved domain, called domain A, which is composed of eight β-sheets and α-helices, arranged in a barrel shape. These β sheets and α helices are connected by loops (numbered according to their position within the polypeptide chain) and it has been suggested that loop 4 underpins the different substrate specificities of these enzymes. Enzymes with a longer glucan chain site specificity seem to have a longer loop 4 region, and previous studies have suggested that this loop may be required to accommodate the longer side chains (Dauvillée et al., 2005). However, a more

recent, detailed analysis of the structure of the *E. coli* GlgX enzyme by Song et al. (2010) concluded that this was probably not the case. They observed the binding groove within the enzyme, composed of several aromatic residues such as tryptophan and tyrosine, lined along its rim with polar residues such as glycine and histidine, which show strong alignment with residues involved in substrate binding in other members of the enzyme family. A key difference they observed with the *E. coli* GlgX structure, however, is that while its substrate groove spans the molecule, as it does in those other enzymes, the cleft of its groove is cut short by side chains of leucine and proline, creating a physical barrier half-way along its length. These side chains are found across a variety of GlgX enzymes, but are not present in enzymes of the same family which cleave longer branches, such as plant Isoamylase I. It seems likely, therefore, that it is these side chains which prevent the GlgX binding grooves from 'fitting' branches any longer than four residues. Loop 4, meanwhile, was observed to be located at the side of the enzyme, away from the active site of the binding groove, and is blocked from the active site by two other loops, suggesting that it does not play a direct role in the substrate specificity of the enzyme. However, the correlation between the size of loop 4 and the binding specificity of the enzymes it is in does still suggest that it might play some overall part in the enzyme specificity (Song et al., 2010).

As with GlgP, GlgX is expressed during glycogen synthesis, as well as its degradation. Mutants with disrupted *glgX* genes have been shown to overproduce glycogen, and also accumulate a surplus of short chains (Dauvillée et al., 2005). The overall molecule therefore shows a marked difference in chain length distribution from wild-type glycogen, towards a lower average chain length. The molecule also shows a severalfold higher proportion of external chains. This suggests that, as well as being vital to complete glycogen degradation, GlgX is important in maintaining the shape of the glycogen molecule during synthesis, by reducing the number of short external chains in each successive outer tier as the molecule is growing, so reducing the effects of crowding that would inhibit synthesis, and permitting the growth of larger, more efficient storage molecules (Dauvillée et al., 2005, Song et al, 2010).

1.3.5 Glycogen phosphorylase (GlgP)

GlgP is glycogen phosphorylase. It is the enzyme chiefly responsible for degrading the glycogen molecule, making the stored fuel once again available to the cell's metabolism. It acts through phosphorolysis of the non-reducing ends of the external glucan chains, catalysing the attack on the glucose residues by inorganic phosphate, which transforms them, one at a time, into glucose-1-phosphate and by doing so plucks them off the end of the chain (Deschamps et al., 2008a) (figure 1.4).

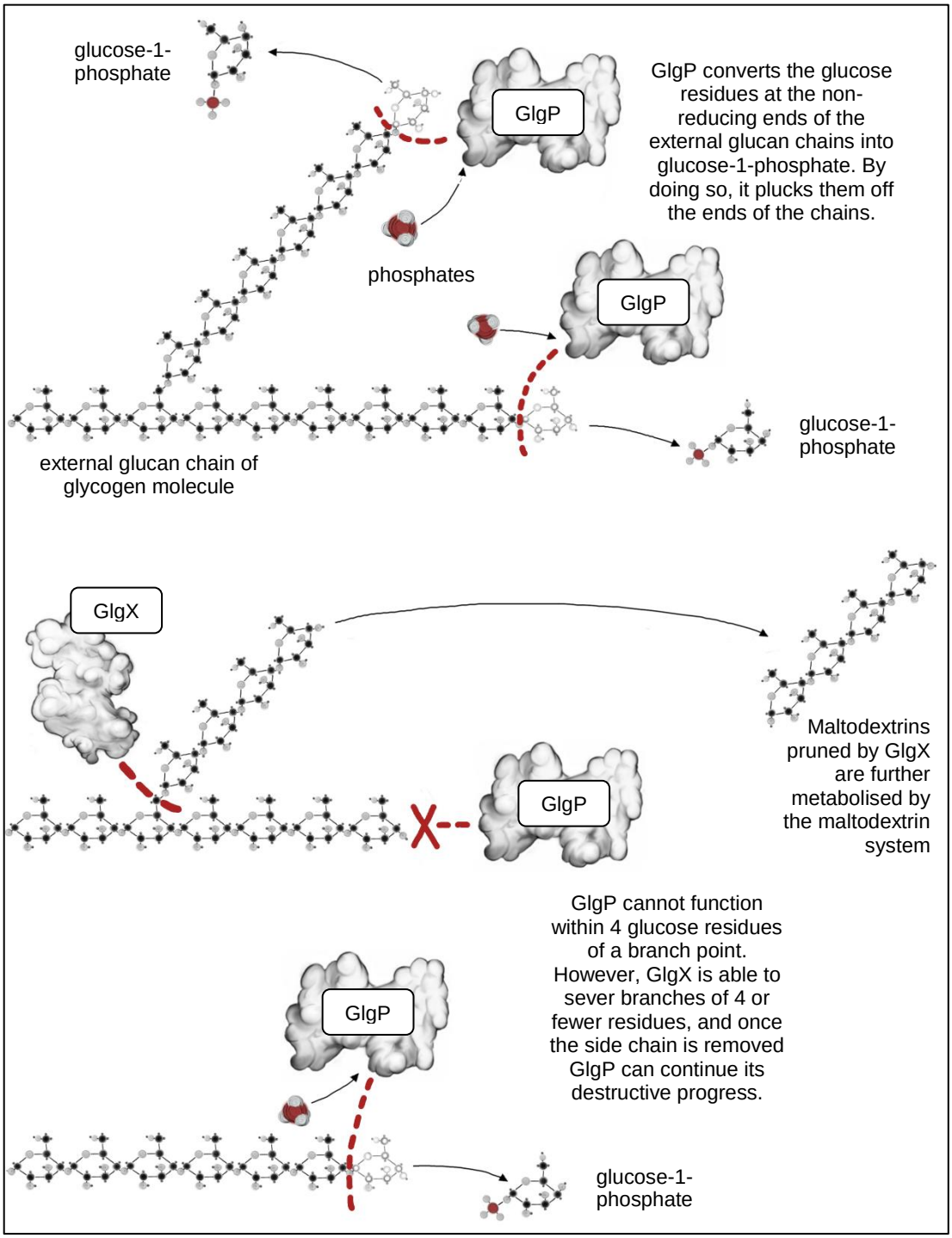


Figure 1.4: Activity of GlgP (glycogen phosphorylase) and GlgX (debranching enzyme). GlgP catalyses the attack on the glucose residues at the non-reducing ends of external glucan chains by inorganic phosphate, transforming them into glucose-1-phosphate and plucking them off the chain. However, GlgP's progress is cut short by proximity to the α-1,6 link of a branch point. These need to be cut off by GlgX before GlgP can continue down the chain.

Although this reaction is readily reversible (indeed the GlgP enzyme shows remarkable topological and structural similarity with GlgA (Buschiazzo et al., 2004)), the intracellular concentrations of the substrate and product favours phosphorolysis (Ball & Morell 2003). The enzyme therefore works its way down each chain, until its activity is cut short by its proximity to a branching point. GlgP is unable to digest α-1,6 links, and as a result cannot proceed beyond these branches. Furthermore, its binding specificity prevents it from phosphorolysing chains of four or fewer glucose residues. Its activity on α-1,4 links therefore stops when it gets to within four glucose residues upstream of an α-1,6 link (Alonso-Casajus et al., 2006). However, once it has reduced chains to this length, GlgX is able to cleave the α-1,6 links, thereby allowing GlgP to continue its deconstructive progress down the chain.

Deletion of the *glgP* gene in *E. coli* leads to a large increase (over 10-fold) in total glycogen content, and much longer glucan chains in the outer layer of the glycogen molecule. Those mutants also showed no glycogen phosphorylase activity, indicating that all the enzymatic activity is dependent on GlgP. Cells that lack GlgP activity also show an accumulation of glycogen when viewed under transmission electron microscopy, in the form of multiple scattered granules. Those same cells were also shown to stain dark brown with iodine, suggesting that the extended branches are forming double-helix structures, into which the iodine molecules collect (Alonso-Casajús et al., 2006. Also see later). Cells transformed to overexpress *glgP*, meanwhile, showed that just a moderate increase in *glgP* activity was accompanied by a proportionally much larger reduction of intracellular glycogen levels. Furthermore, even in untransformed, wild-type cells, the maximum catalytic activity of GlgP was higher than those of GlgC or GlgA during glycogen accumulation (Alonso-Casajús et al., 2006). This would suggest that GlgP must be highly regulated in order to allow for any glycogen accumulation whatsoever. Because GlgP has a binding specificity for linear branches of four glucose residues or more, it is unable to phosphorolyse the maltotetraose stubs that are pruned from the glycogen molecule by GlgX. These are dealt with by another group of carbohydrate-specific enzymes, referred to collectively as the maltodextrin system.

1.3.6 The maltodextrin system

The maltodextrin system in *E. coli* comprises 10 genes spread across five operons, most of which encode proteins involved in the uptake of maltose and short maltodextrins from the environment. However, four of them code for maltodextrin-utilising enzymes. They include one periplasmic amylase (MalS) and three cytoplasmic enzymes: MalQ (amylomaltase), MalP (maltodextrin phosphorylase) and MalZ (maltodextrin glucosidase). These three cytoplasmic maltodextrin-utilising enzymes are the ones brought into play during glycogen

degradation. MalP is able to depolymerise the maltotetraose pruned from the glycogen molecule to maltotriose – which is an inducer of the maltodextrin system – and glucose-1-phosphate (Dauvillée et al., 2005). The maltodextrins are also depolymerised by MalZ (a glucosidase), which is able to degrade chain lengths from maltotriose (3 glucose residues) to maltoheptaose (7 glucose residues) down to maltose and glucose. The glucose-1-phosphate and glucose produced by these depolymerisations can be transformed by phosphoglucomutase and glucokinase, respectively, to form glucose-6-phophate, which can then be fed into glycolysis. The remaining maltose molecules no longer induce MalT, which is the central activator of the maltodextrin system, and cannot be depolymerised directly. However, MalQ (which is an α-1,4 glucanotransferase) is employed to bind these remaining scraps of maltodextrins to larger polysaccharides, where they can continue to be degraded into smaller maltodextrins, including maltotriose. It preferentially removes glucose from the reducing end of these maltodextrin scaps before transferring the remaining residue(s) to the nonreducing ends of other maltodextrins, thus forming longer chains for MalP and MalZ to work on again. In this way, the system can continue ticking over (in conditions where maltodextrins are not a key environmental carbon source) under a constant but low level of induction (Dippel et al., 2005).

When *E. coli* cells are grown on a substrate of small maltodextrins, they use a binding protein-dependent ABC (ATP-Binding Cassette) transporter (ABC transporters are transmembrane protein with many different roles) to take up these oligosaccharides, which are then polymerised into longer dextrins by MalQ. The maltodextrin system as a whole operates independently of the GlgA glycogen synthesis enzyme, and neither sugar phosphorylation nor ADP-dependent sugar activation is needed for its polymerisation activity. The dextrins that result from MalQ polymerisation are even able to grow long enough to function as substrates for the branching enzyme GlgB, and so can eventually form glycogen molecules with a similar molecular weight and branch distribution to those of GlgC/GlgA initiated glycogen. Indeed, MalQ and GlgB have been shown to form glycogen from either maltose or linear maltodextrins *in vitro*. (Park et al., 2011).

When grown on maltose or maltodextrins, *E. coli* lacking *malP* were also found to produce more than 20 times the glycogen content of wild type cells (though they had a similar glycogen content to wild type when grown on glucose, which represses the maltodextrin system). Furthermore, when *E. coli* mutants totally lacking the *glgA* gene were grown on maltodextrins, they were still found to synthesise substantial amounts of glycogen. This was even more pronounced (resulting in around five times more total glycogen content) when

malP was also deleted. This not only showed that MalQ is able to form substantial linear glucan chains from maltodextrins, but also that MalP inhibits this pathway, by removing glucose residues from the nonreducing ends of the maltodextrin 'building blocks' and thereby reducing the amount of substrate available for conversion into glycogen by MalQ and GlgB (Park et al., 2011).

1.4 Glycogen operon structure

Operons consist of a cluster of genes acting under the control of a single promoter. Bacterial genes with related functions are often found clustered onto a single operon within the genome, thereby ensuring the simultaneous expression of functionally related gene products. This seems to be the case with the core glycogen genes in many bacterial species (including but not limited to *Salmonella typhi*, *Yersina pestis*, *Shigella flexneri*, *Haemophilus influenzae*, *Pasteurella multocida*, *Shewanella oneidensis*, *Agrobacterium tumefaciens*, *Mesorhizobium loti*, *Sinorhizobium meliloti*, *Bacillus subtilis*, *Oceanobacillus iheyensis*, *Streptococcus mutans*, *Lactobacillus plantarum*, *Clostridium perfringens*, *Pectobacterium chrysanthemi* and *Rhizobium tropici*. Note that the *Bacillus* species have a *glgD* gene instead of the *glgX* gene; *glgD* is in fact partially homologous to *glgC* and its function remains unclear) (Cho et al., 2008; Wilson et al., 2010; Marroquí et al, 2001; Lepek et al., 2002). The observation of a single *glg* operon provides further evidence for the interdependence and role of the Glg enzymes as the core players in glycogen synthesis and degradation. However, there is an accumulation of evidence to suggest that in *E. coli* (as well as *Salmonella enterica*) the genes may be clustered in two tandemly arranged operons. The sequence of the genes, as found in the *E. coli* genome, is *glgB*, *glgX* (with an overlap between the stop codon of *glgB* and the start codon of *glgX*) *glgC*, *glgA*, *glgP* (with an overlap between the stop codon of *glgC* and the start codon of *glgA*). The first two genes in the sequence, *glgBX*, code for the branching and debranching enzymes, respectively, of the glycogen molecule, while the last three, *glgCAP*, code for genes involved in the synthesis and degradation of the glucan chains, and seem to be under strong regulatory control. Along with various experimental evidence, this has led to the supposition that these two groups also form two distinct operons (Preiss, 2009). Furthermore, a common side-effect of prokaryotic gene expression is the phenomenon of operon polarity, where when a single promoter controls the transcription of a number of genes in an operon, the genes furthest from that promoter show reduced expression as a result of the distance. The high transcription levels of *glgP* with respect to its upstream *glg* genes therefore further suggests a second promoter within the five-gene sequence. However, recent findings have subverted the two-operon model.

Dauvillée et al. (2005) showed that mutants lacking the complete *glgX* gene accumulate excess glycogen, indicating that *glgCAP* transcription can be initiated upstream of *glgX*. Meanwhile, work from Montero et al. (2011) shows that the expression levels of both *glgA* and *glgP* were higher than that of the upstream *glgC* gene, and also that the removal of the region upstream of *glgB*, resulting in the removal of the known, initial operonic promoter, led to an almost complete abolishment of *glgB*, *glgX* and *glgC* expression, but only a 50 to 60% reduction of *glgA* and *glgP*. They therefore suggest that the *E. coli glgBXCAP* genes are organised in a single operon, but *glgA* and *glgP* (not *glgC*) are also under the additional control of a suboperonic promoter located within the *glgC* gene.

As further confirmation of their theory, Montero et al. (2011) carried out computer searches for the primary *glgBXCAP* upstream promoter region and the suspected suboperonic promoter region within the *glgC* gene. The searches identified, in both cases, predicted -35 and -10 elements within the expected genomic domains (figures 1.5 and 1.6). Deletion of these specific regions led to the expected reduction of downstream gene expression, thereby providing strong support for their theory of a single *glgBXCAP* operon with a suboperonic promoter. The order of the core glycogen genes as found in *E. coli* and the position of their suboperonic promoter is different from that of many other bacterial species, and this may well be an evolutionary effect of different metabolic needs. There are two main functions of a suboperonic promoter: the first is to counteract the operon polarity effect described above, ensuring a high enough expression of downstream genes within an operon. However, the suboperonic promoter may also allow temporal expression of specific genes within the operon under conditions of reduced expression of their operonic neighbours, whose function may be redundantly fulfilled by genes in other loci. In the case of *E. coli*, this is exemplified by the presence of additional sources of ADP-glucose not produced by the action of GlgC, providing a motive for *glgA* expression even when *glgC* expression is reduced, resulting in continued polyglucan accumulation. An interesting additional feature of Montero et al.'s study is that the positions of the two promoter regions identified provide relatively long 5' untranslated regions in the case of both sets of transcripts (figures 1.5 and 1.6). Such 5' untranslated regions often serve as targets for regulatory factors, such as regulatory small RNAs, RNA-binding proteins and RNases, as well as allowing for the formation of riboswitch structures, all of which can control the fate and stability of the transcripts.

…**GCTTCGTCAACTGGCGTAA**TCTTTATTCATTAAATCTGGGGCGCGATGC
CGCCCCTGTTAGTGCGTAATACAGGAGTAAGCGCAGATGTTTCATGATTT
ACCGGGAGTTAAATAGAGCATTGGCTATTCTTTAAGGGTGGCTGAATACA
TGAGTATTCACAGCCTTACCTGAAGTGAGGACGACGCAGAGAGGATGCAC
AGAGTGCTGCGCCGTTCAGGTCAAAAAAATGTCACAACCAGAAGTCAAAA
ATCCAATTGGATGGGGTGACACAATAAAACAGGAAGACAAGC**ATGTCCG**…

Figure 1.5: Nucleotide sequence of the intergenic region upstream of *glgB*, showing the -35 (TTACCG) and -10 (GGCTATTCT) RNA polymerase binding elements (highlighted in dark grey) predicted by Montero et al. (2011). The end of the previous gene (*asd*, which codes for an enzyme involved in amino acid synthesis, and is not associated with glycogen metabolism) and the start of the *glgB* gene are in bold and underlined. The AGGA Shine-Dalgarno sequence of the *glgB* ribosome binding site is highlighted in pale grey.

…ATCGCAATTGGCCAATTCGCACCTACAATGAATCATTACCGCCAGCGAA
ATTCGTGCAGGATCGCTCCGGTAGCCACGGGATGACCCTTAACTCACTGG
TTTCCGGCGGTTGTGTGATCTCCGGTTCGGTGGTGGTGCAGTCCGTTCTG
TTCTCGCGCGTTCGCGTGAATTCATTCTGCAACATTGATTCCGCCGTATT
GTTACCGGAAGTATGGGTAGGTCGCTCGTGCCGTCTGCGCCGCTGCGTCA
TCGATCGTGCTTGTGTTATTCCGGAAGGCATGGTGATTGGTGAAAACGCA
GAGGAAGATGCACGTCGTTTCTATCGTTCAGAAGAAGGCATCGTGCTGGT
AACGCGCGAAATGCTACGGAAGTTAGGGCATAAACAGGAGCGA**TAATG**C…

Figure 1.6: Nucleotide sequence of the *glgC* region upstream of *glgA*, showing the -35 (TCGCAATT) and -10 (ACCTACAAT) RNA polymerase binding elements (highlighted in dark grey) predicted by Montero et al. (2011). The overlap between the stop codon of *glgC* and the start codon of *glgA* is in bold and underlined. The AGGA Shine-Dalgarno sequence of the *glgA* ribosome binding site is highlighted in pale grey.

1.5 A proposed optimal glycogen structure

The optimal glycogen structure proposed by Melendez et al. (1998) and described previously as 12 tiers consisting of 13-residue chains with two branches per chain becomes a more conceivable achievement within living cells once the substrate specificities of the enzymes described above are taken into account. Many of these enzymes have homologues with different specificities in different organisms, suggesting that their modes of action are plastic and easily manipulated by selective pressures. The GlgB branching enzyme's specificity for cutting glucan chains at around five to 14 residues from the non-reducing end appears amenable to evolutionary effects through the simple duplication of residues within its N-terminal domain. Evolution of the enzyme, at least in *E. coli* and many other organisms, therefore seems likely to have settled upon the length of chain GlgB cuts (and re-grafts as a side-branch) as the shortest it can get without that branch being instantly pruned again by GlgX. Meanwhile its footprint, disallowing it from grafting branches within a certain number of residues from each other, could well be used to maintain an even branching degree. Previous reports (Binderup et al., 2000; Binderup et al., 2002) have suggested that the mode of action of GlgB is one of the primary means by which glucose accumulating organisms control the structure of their glucose granules, and correspondingly it has been observed that the GlgB enzymes from different sources show considerable variation of specific activities and chain transfer patterns, which seems to relate to the differences in the final structure of the glycogen they accumulate.

The GlgX debranching enzyme has, at the same time, evolved a physical barrier within its binding groove, not present in related debranching enzymes such as Isoamylase I, which prevents it from being able to remove any branches longer than four glucose residues, allowing it, in turn, to work alongside GlgB without falling into a futile cycle of grafting and pruning. Rather it firsts needs the action of GlgP to trim the chains and branches down to four residues from the α-1,6 link before it can act; an action which GlgP is optimised to do.

The number of tiers of the complete glycogen molecule is expected to be self-limited by the interference between the branches as they get more and more crowded in the outer tier of the growing structure; at some point there will simply not be enough space for continued growth. Through mathematical modelling, Meléndez et al. (1998) showed it was possible for their idealised glycogen to form 13 tiers before its branching density restricted further growth. However, the 13^{th} tier of this model would also be incomplete, with several chains of the previous tier being unable to support their full quota of two branches. This reduced the

structure's homogeneity and limited the amount of glucose in the outer tier, which is the source of glucose directly available as an energy supply through phosphorolysis. They therefore concluded that 12 tiers is the optimal size for the molecule. This is supported by observation (see later) and might be achieved if the size of the GlgA synthesis enzyme is large enough to prevent growth beyond 12 tiers. At this point the enzyme can no longer fit amongst the tangle of branches, and so synthesis stops.

Despite being primarily involved in the metabolism of glycogen, both GlgP and GlgX are active during synthesis; an activity which, as described above, has been found to be integral in maintaining the optimum structure of the granule (Dauvillée et al., 2005). So rather than being an orderly process in which the different enzymes take it in turn to act their part, the synthesis of glycogen is instead a symphony of activity; of lengthening, cutting and sticking, trimming and pruning (figure 1.7). This high activity level will tend towards maximising the molecule's overall homogeneity. Since there are a large number of binding sites for the branches within each chain, if the chains and branches are being continuously trimmed, debranched and re-branched, the branches of each outer tier of the growing molecule will tend to move away from each other as much as possible as a result of probability. This homogeneity is key to maintaining the maximum amount of glucose in each tier; if any of the chains of a tier have more or fewer than 2 branches, the overall molecule becomes lop-sided, reducing the amount of glycogen available (figure 1.8).

The site-specificity and activity of the enzymes involved in glycogen synthesis and degradation are therefore theoretically able to control the branching degree and maximum size of the final glycogen structure. The only remaining factor is the speed at which the branches are synthesised and phosphorolysed, by which the average chain length of the structure might also be controlled (in addition to the chain-length specificity of GlgB). Investigations have indeed found that the expression of both GlgA and GlgP are tightly controlled by a complex web of regulatory systems, described in detail later, which are intimately linked to carbohydrate availability and the energy requirements of the cell.

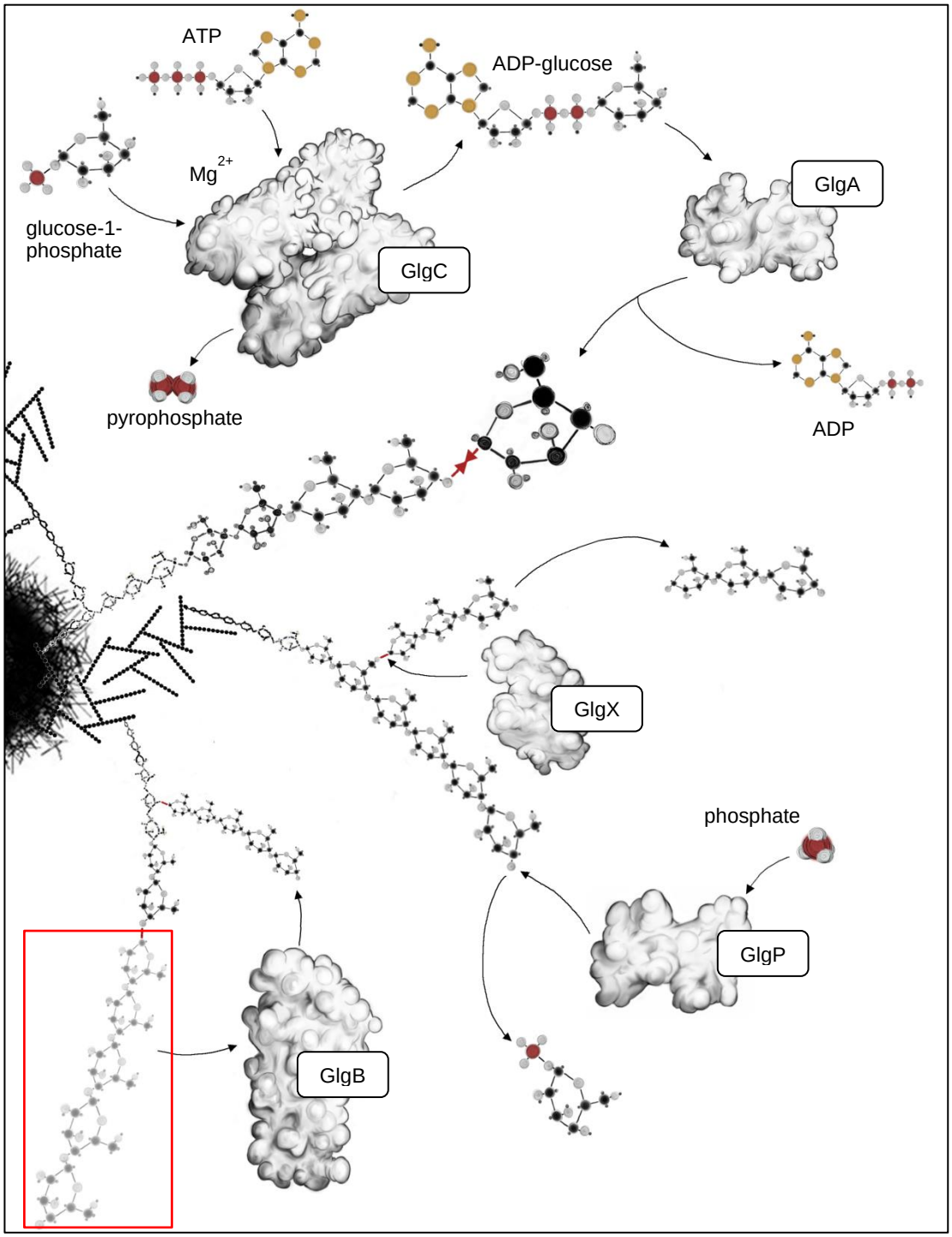
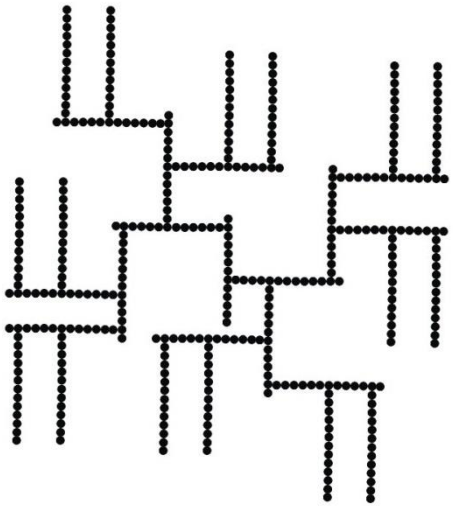
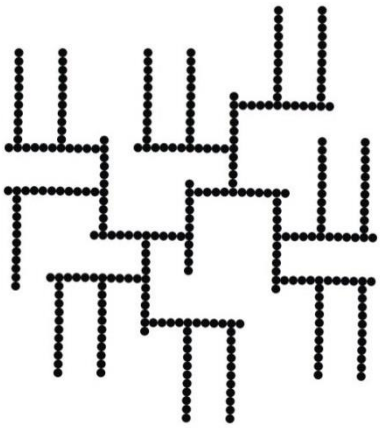


Figure 1.7: The full symphony of glycogen metabolism. GlgC and GlgA work in tandem to grow the non-reducing ends of external glucan chains one glucose unit at a time, while GlgP removes glucose residues from the non-reducing ends of the chains, until its activity is cut short by proximity to the branches created by GlgB. GlgX can sever the α-1,6 bonds at the fork of these branches, but only on branches shorter than those GlgB has originally attached. This continuous building, trimming, de-branching and re-branching maximises homogeneity, since the branches of the growing molecule will tend to spread out evenly as a result of probability.

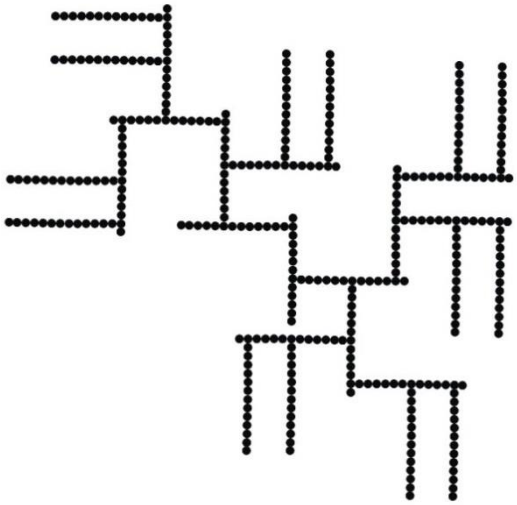
A)



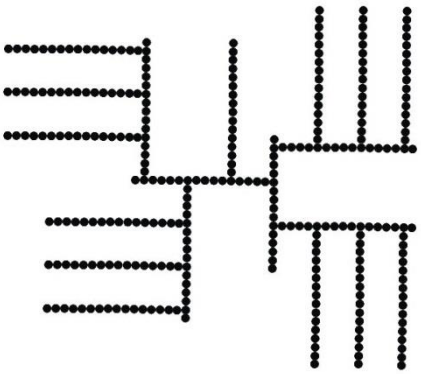
B)



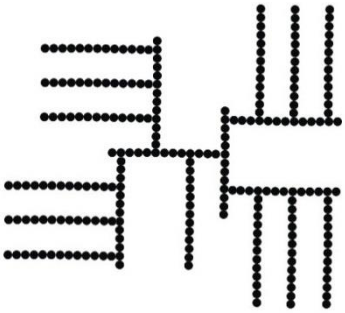
C)



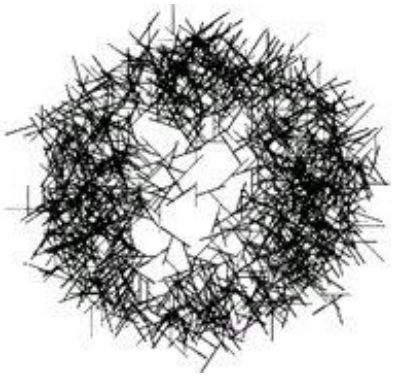
D)



E)



F)



Figures 1.8: Glycogen models to explore optimisation. A 2-dimensional model reduces the space available for branches, so the number of tiers in these is necessarily fewer than it would be for the true 3-dimensional molecule. All models also show a uniform chain length distribution and are not considered to represent true glycogen but only explore the relationships between branching and chain length. Branches are generally added until the point when overcrowding would result in a loss of homogeneity. In all these examples, 'optimum structure' is considered to be a combination of the total number of glucose residues stored by the individual molecule (compactness) and the degree by which they are available for immediate phosphorolysis, which is itself a combination of the number of glucose residues within the outer tier and the number of N-terminal ends available for phorsphorylases to metabolise simultaneously. Although these 'optimum' parameters may be considered arbitrary, they are reflected in the majority of glycogen molecules observed to date and therefore seem likely to provide some benefit to the host (Meléndez et al., 1998. See main text).
Figure A) shows the theoretically optimum structure, with chains of 13 residues in length, each with 2 branch points at least 4 residues apart. In the 2 dimensional model this allows for 5 tiers and a total of 31 chains, 16 of which are in the outer tier and therefore immediately available for phosphorolysis. Figure B) shows a similar structure but with 11 residues in each chain (the minimum required to support two branches) instead of 13. In the model this prevents the full formation of the fifth tier, resulting in 30 chains, 15 of which are in the outer tier. Figure C) shows the same structure to figure A), but with a lack of homogeneity: one inner chain only supports a single branch. This results in a lopsided molecule with 5/6 tiers, containing 28 chains, 14 of which are in the outer tier and available for phosphorolysis. Figure D) shows a model where chains are 16 residues long, permitting 3 branches per chain (with a 4 residue gap between branches). Again, this quickly leads to overcrowding and a reduction in homogeneity, so that the model has 3/4 tiers and 19 chains, 13 of which are in the outer tier and available for phosphorolysis. Figure E) shows a structure with chains of 13 residues in length, but where branches are permitted 3 residues apart. This also leads quickly to overcrowding, so that the model has 3/4 tiers and 19 chains, 13 of which are in the outer tier and available for phosphorolysis. Figure F) shows a 3-dimensional structure of the 'optimal' glycogen molecule described by figure A), having 13-residue chains with 2 branch points each, growing to 12 tiers in size. Some of the chains have been stripped from the image to reveal the relatively sparse arrangement of the inner tiers.
Figure F) is adapted from Meléndez et al., 1998.

1.6 The reality of glycogen structure

Glycogen is an extremely efficient storage form of glucose; one high-energy phosphate bond is spent in incorporating glucose-6-phosphate into its structure, and during metabolism about 91% of the residues are phosphorolytically cleaved to glucose-1-phosphate, which is converted at no extra cost to glucose-6-phosphate. The other roughly 9% are the α-1,6 branch residues, which are hydrolytically cleaved. One ATP is then used to phosphorylate each of these glucose molecules to glucose-6-phosphate. The complete oxidation of glucose-6-phosphate yields about 31 molecules of ATP and storage consumes slightly more than one ATP per glucose-6-phosphate. The overall storage efficiency of glycogen is therefore nearly 97% (Stryer, 1995). However, the inner 8 tiers, which contain only about 6% of the glucose, play mainly a skeletal role (Lomako et al., 1991). Only in extreme cases will their stored glycogen be used as fuel. Meanwhile the outer, unbranched tier contains 36% of the stored glycogen, all of which is readily accessible to catabolism without the need for debranching (Ball et al., 2010; Meléndez et al., 1998). The abundance of chains within this outer tier also aids in the fast response to metabolic needs, since degradation of the granule can only occur from the non-reducing ends of the linked glucose units. A more highly-branched structure has more non-reducing ends at its surface, allowing as many processes of phosphorolysis to occur concurrently. Additionally, as the glycogen granule increases its number of tiers it also increases in solubility (Shearer & Graham, 2002; Zmasek et al., 2014). As such, the fully-formed granule is entirely hydrosoluble, providing the added advantage that it has little effect on the internal osmotic pressure of the cell, while still being readily accessible to rapid metabolism (Ball et al., 2010; Wilson et al., 2010).

The model for optimal glycogen structure, put forward by Meléndez-Hevia et al. (1993) and described in previous sections of this work, was created by those authors to investigate the efficiency of the parameters already observed in glycogen molecules. From their modelling the authors concluded that the enzymes involved in shaping the growing glycogen must have evolved in order to create a granule optimised in compactness as well as in the number of glucose residues directly available to phosphorylase. These choices may seem arbitrary. For example, an organism synthesising many glycogen molecules with the highly-branched structure shown in figure 1.8 D) or E) would have far more immediately available glucose residues per total number of residues than one synthesising the apparently optimised structure shown by figure 1.8 A). However, figure 1.8 A) better describes those structures observed in the real world. Indeed, further investigations into the glycogen structure of different organisms have tended to find similar parameters. Across a large number of

organisms, including vertebrates, invertebrates, yeast, protozoa and bacteria, the molecular weight of fully-formed glycogen molecules is frequently found to be about 10^7-10^8 Da, accounting for around 55,000 glucose residues packed into 12 tiers, with a maximum diameter of 42nm, matching the original observations that inspired the model (Wilson et al., 2010; Meléndez et al., 1998; Shearer & Graham, 2002; Ball et al., 2010). It therefore seems plausible that 'compactness' is a parameter that has been selected for, perhaps to reduce the interference of the glycogen on other cellular systems, which may be an issue with more numerous, less soluble granules.

However, no glycogen analysed to date shows the uniform chain-length that the assumptions of the model are based on. The chain lengths of glycogen can be analysed by severing all of its α-1,6 branch points to completion with isoamylases (the starch debranching enzymes, which exclusively hydrolyse α-1,6 links), thereby producing a porridge of linear chains that can be measured by spectroscopy to reveal their length in glucose residues. This technique has shown that even in species such as *E. coli*, with its apparently optimum average chain length of 11-14 glucose residues, the glycogen granules are actually composed of a wide range of chain sizes, from around 5 to 50 glucose residues long (Sundberg et al., 2013).

A mathematical model underestimates the complexity observed in the real world by default. Whether or not real-life glycogen shows a less uniform structure than that of the model simply because of the variability always found in organic systems, or because the parameters set out by the model are not the only concerns of the organism, or because the parameters suggested by the model are better met by a higher level of complexity than was investigated, the model can perhaps only serve as a necessarily limited means to try to understand the driving forces behind the evolution of observed glycogen structure. For example, the model assumes granule compactness is a key parameter in order to fit with observations of glycogen structure, but as mentioned previously, the reasons for its importance have had to be applied retrospectively. There may in fact be no evolutionary advantage to maximising compactness, or different organisms may be better served by glycogen that is structured differently in this regard. Furthermore, while the model is based – for ease of the mathematics – on a uniform chain length distribution, there may in fact be an evolutionary advantage to the broad range of chain lengths revealed by spectroscopy. For example, longer chains in the outer tier of the granule might leave the proposed optimum structure undisturbed while also providing more glucose units immediately available for phosphorolysis (it certainly seems possible that chain-lengths could be manipulated during

granule growth, by altering the activity of the different enzymes). However, such suggestions can only be hypothetical at this point.

To return to firmer ground, the research described in section 1.2 into the substrate specificity of the glycogen enzymes does strongly suggest that their mode of action has been shaped by selective pressure to control the final glycogen structure. Furthermore, although the model proposed by Meléndez-Hevia et al. (1993) is heavily simplified, it does broadly reflect the structure observed in glycogen from a majority of organisms investigated, suggesting that, for whatever reason, the granule's compactness as well as the availability of glucose residues may well be integral to glycogen optimisation in some or all of these cases.

However, variability between the glycogen structures of different organisms certainly exists. It is assumed that this is primarily an adaptation to differences in lifestyle. At its most extreme, many bacteria have lost their ability to accumulate glycogen altogether. This loss of glycogen has so far been observed primarily in species that are parasitic, symbiotic or fastidious to one particular environment (Henrissat et al., 2002). In accordance with these findings, it has been suggested that the primary function of glycogen in bacteria (which is by no means agreed upon) may be to provide an energy supply during adaptation to fresh conditions, before initiating active proliferation (Rahimpour et al., 2013). In the case of *E. coli*, the need for this can easily be appreciated when considering the switch from its primary habitat, in the guts of warm-blooded animals wherein it tends to undergo slow and steady growth, to its secondary habitat, in the soil or water where it finds itself when shed outside of the host, which is generally characterised by low nutrient availability and lower temperatures. *E. coli* populations decay under these conditions, suggesting that such environments are stressful to the cells (Landini et al., 2013), so that any adaptations for survival under such conditions would be heavily selected for.

Even the exact function of glycogen probably varies across species, in accordance to lifestyle needs. Chain length in particular shows adaptation to different functions; although 13 residues might lead to the optimal structure in terms of compactness, a higher degree of branching has been shown to slow degradation, and the slow utilisation of glycogen is correlated with bacterial viability (Wang & Wise., 2011). A more branched form of glycogen, observed in some species, may therefore have evolved to purposefully slow down glycogen catabolism, eking out supplies in times of extended deprivation. As described earlier, spectroscopy performed on the linear chains of glycogens has revealed the average chain length in various species of bacteria (table 1.1) (Wang & Wise 2011).

Bacteria	**Average chain length**
Arthrobacter globiformis	6.6
Desulfurococcus	7
Sulfolobus	7
Thermococcus	7
Thermoproteus	7
Thermus thermophilus	7
Arthrobacter spp.	7~9
Synechocystis sp. PCC6803	7.5~10.4
Prevotella ruminicola	8
Pseudomonas V-19	8
Mycobacterium tuberculosis	7~9, 10~12
Bacillus megaterium	10~11
Chromatium strain D	11.1~11.4
Klebsiella pneumoniae	11.6
Neisseria perflava	11~12
Streptococcus mitis	12
Escherichia coli	12, 10.8, 13, 14
Selenomonas ruminantium	12, 23.5
Aerobacter aerogenes	13
Agrobacterium tumefaciens	13
Nostoc muscorum	13
Sphaerotilus natans	13
Clostridium botulinum	17
Geobacillus stearothermophilus	21

Table 1.1: Glycogen average chain length (number of glucose residues/number of branching points) across a spectrum of bacterial species. Adapted from Wang & Wise, 2011.

From table 1.1 it can be seen that *Geobacillus stearothermophilus* has the longest average chain length reported so far in wild type bacteria, at 21 glucose residues. Its branching enzyme also has the shortest N-terminal region of any known branching enzyme, at 149 residues (Binderup et al., 2000), thereby supporting the theory that these two features are linked. Correspondingly, this bacterium also shows the fastest degradation of glycogen (Takata et al., 1998). A direct comparison of GlgP affinity for the glycogen of bacterial species with different average chain lengths show that, for *E. coli* (with an average chain length of 11-14), the affinity is lower than for *G. stearothermophilus* (average chain length of 21) but higher than in *Arthrobacter* spp (average chain length 7-9) (Wang & Wise, 2011).

The relationship between average chain length and lifestyle of the various bacteria in table 1.1 has not been fully elucidated and is almost certainly complex in each example. Whether the longer chains found in the glycogen granules of *G. stearothermophilus* and *Clostridium botulinum* are the result of selective pressure or a lack of it is beyond the scope of this report. However, the majority of glucose-accumulating organisms studied so far (including eukaryotes) accumulate glycogen with a structure that roughly fits the proposed model, suggesting compactness and the immediacy of available residues may well be selective forces that have shaped the substrate specificities of glycogen enzymes. Meanwhile the majority of those that don't fit the model have glycogen with a shorter average chain length, which in most if not all cases is presumably a means to slow glycogen catabolism in times of nutrient deprivation. While many cases remain obscure, a general correlation between glycogen accumulation and lifestyle has been observed, and the function of glycogen within bacteria seems most likely to be a means to persist through changeable conditions.

1.7 Regulation of glycogen

Regulation of glycogen accumulation and metabolism has been found to be closely tied to the different growth phases of *E. coli*. As the bacterial cells shift from exponential to stationary phase, a cascade of genetic regulation is initiated to increase cell viability under these reduced nutrient conditions, tuning the cells for maintenance, survival and biosynthesis rather than growth. Glycogen biosynthesis occurs primarily during this phase, when nutrients other than carbon are limiting. It makes metabolic sense for the bacteria, if they are unable to metabolise environmental glucose directly due to the limitation of other essential nutrients, to sequester that glucose until a time when more optimal cellular function may recommence.

GlgC, which controls the first committed and rate-limiting step of glycogen biosynthesis, is activated by metabolites that signal a high carbon and energy content within the cell. In

E. coli this is chiefly an excess of fructose-1,6-bisphosphate (a product of glycolysis). It is inhibited, meanwhile, by intermediates of low metabolic energy levels (Wilson et al., 2010), which in *E. coli* is chiefly an excess of AMP (Deschamps et al., 2008a). Supporting this, *E. coli* mutants lacking the *purA* gene for adenylosuccinate synthetase, needed for AMP biosynthesis, show a glycogen excess phenotype (Montero et al., 2009). Meanwhile, as far back as 1973, Govons et al. identified a mutant form of GlgC showing a reduced affinity to its AMP inhibitor, resulting in approximately double the rate of glycogen accumulation and two to three times the normal glycogen content, along with three other mutant forms of GlgC which also showed altered regulatory properties, affecting the host cell's ability to accumulate glycogen. More recently, in 2011, Figueroa et al. performed a site-directed mutagenesis of amino acid residues within the GlgC enzyme that were found to be conserved across all ADP-glucose pyrophosphorylases (even those with different allosteric regulators and low sequence identity), which resulted in a form defective to activation by fructose-1,6-bisphosphate, despite those residues being distant from the activator binding domain. Overall this presents strong evidence that glycogen synthesis is controlled by allosteric regulators (effector molecules that bind to the enzyme's allosteric site, as opposed to its active site). Figueroa et al. (2011) therefore proposed that the allosteric regulator communicates its regulatory signal through conformational changes of loops containing those residues they mutated. The ADP-glucose pyrophosphorylases of different organisms are finely tuned to allosteric activation and inhibition, and as such have so far been classified into 9 distinct groups, based on the activators and inhibitors found to be specific to the metabolic specialisations of the organisms involved (Figueroa et al., 2011). Cyanobacterial ADP-glucose pyrophosphorylases, for example, are found to be most effectively activated by orthophosphate and inhibited by 3-phosphoglycerate, in the presence of ATP and glucose-1-phosphate substrates, thereby coupling glycogen synthesis to the light and dark reactions of photosynthesis. This function has also been conserved in the ADP-glucose pyrophosphorylases of starch biosynthesis (Ball & Morell, 2003).

GlgP, which controls the catabolism of the glycogen granule, has been suggested by most studies to be regulated at the level of gene expression. It seems that although GlgP (and GlgX) are present during synthesis of the glycogen molecule, their activity is low and it is only when levels of ADP-glucose become limited that there are significant rates of net glycogen degradation (Dauvillée et al., 2005). However, several recent reports now indicate that post-transcriptional regulation may occur after all. Alonso-Casajus et al. (2006) showed that, unlike GlgC regulation, the regulation of GlgP is not caused by substrates of

physiological relevance, such as ADP-glucose, glucose or AMP. However, glycogen degradation is also found to accelerate under conditions of reduced glucose-1-phosphate and accumulated inorganic phosphate (Dauvillée et al., 2005), suggesting that GlgP may be allosterically modulated as a result of its interaction with the phosphotransferase system: the system by which environmental sugars are taken up by the cell in the first place.

1.7.1 The phosphotransferase system (PTS)

The Phosphotranspherase system (PTS) is the primary means by which carbohydrates are taken up by the cell and phosphorylated. It entails the shuttling of phosphate along a chain of enzymes before it is finally donated to a newly-acquired sugar molecule as that sugar passes into the cell, which gives the molecule an overall negative charge, thereby prohibiting it from freely leaving the cell through the hydrophobic cytoplasmic membrane, while also priming it for entry into glycolysis or conversion into ADP-glucose for further storage as glycogen, as described previously. The phosphate used by the phosphotransferase system is originally donated by phosphoenolpyruvate (PEP); not an enzyme, but a phosphorylated three-carbon hydrocarbon that exists in the cell as a precursor of many biosynthetic pathways, and in particular is an intermediate molecule in glycolysis (and by donating its phosphate group, PEP transforms into pyruvate, which can then be converted into acetyl-CoA and fed into the Krebs cycle). Thus the phosphotransferase system is self-regulating, taking phosphate from glycolysis to import sugars that are fed back into glycolysis.

The phosphate attached to PEP is snatched by the first enzyme of the phosphotransferase enzyme chain, known as Enzyme I, in the presence of Mg^{2+}. From Enzyme I, the phosphate is passed to the histidine phosphocarrier protein HPr. HPr then passes the phosphate to the sugar-specific Enzyme IIA, which passes it to the sugar-specific Enzyme IIB. Enzyme IIB forms a dimer with the cytoplasmic-membrane-bound Enzyme IIC, which is an integral membrane protein permease via which the sugar molecule is recognised and transported into the cell (the sugar first enters the periplasmic space of gram-negative bacteria through a series of abundant and non-specific porins in the outer membrane. However, because the cytoplasmic membrane is a hydrophobic barrier, active carbohydrate transport systems are required to transfer those sugars into the cytoplasm). As the sugar crosses Enzyme IIC, the phosphate held by Enzyme IIB is donated to it, thereby ending the chain (Escalante et al., 2012) (figure 1.9).

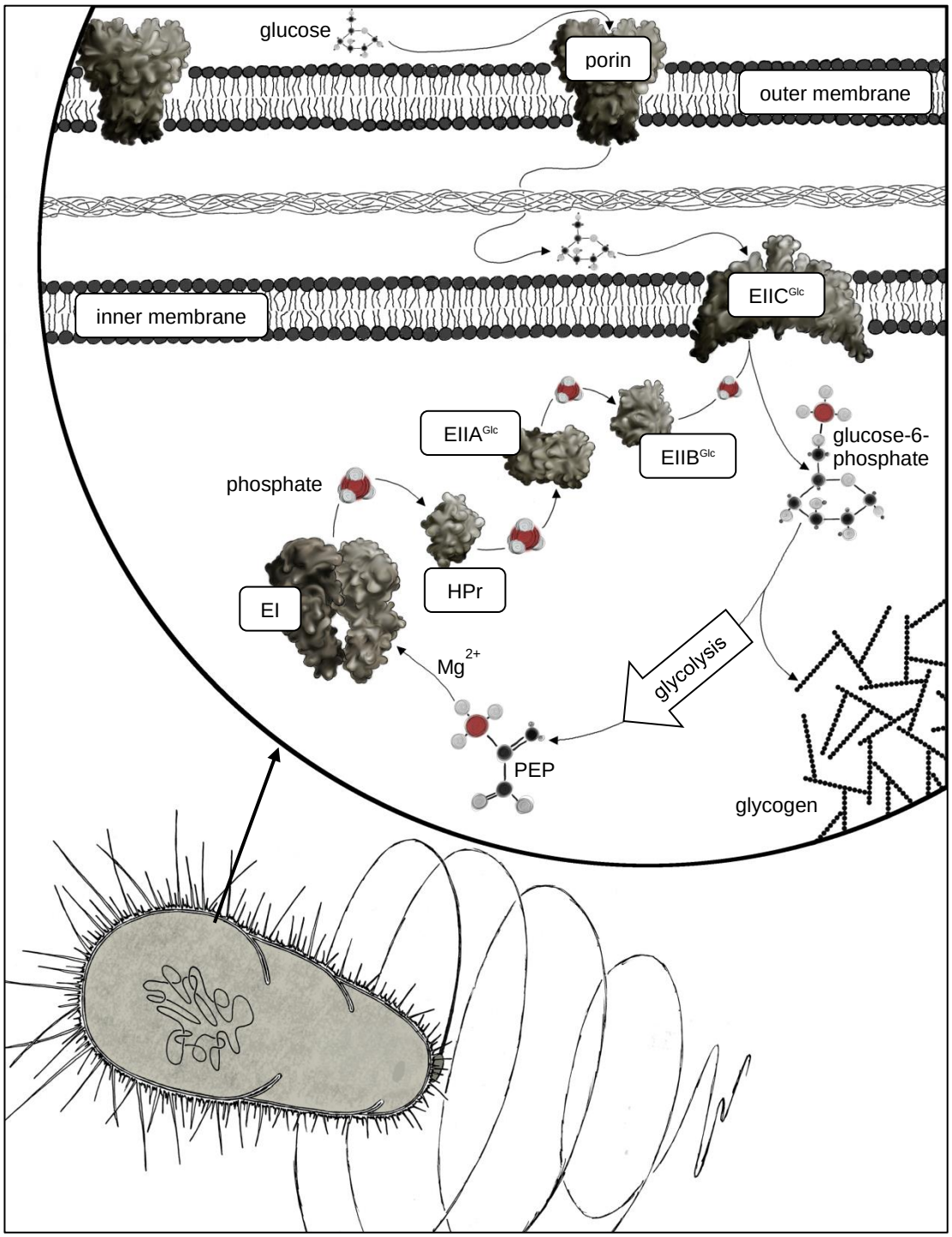


Figure 1.9: The phosphotransferase system. A phosphate is donated by PEP, a hydrocarbon intermediate in glycolysis. It is shuttled along a chain of enzymes until it reaches the transmembrane Enzyme IIC, which donates the phosphate to an incoming glucose molecule, preventing it from exiting the cell while also priming it for glycolysis itself.

GlgP has been shown to strongly interact with HPr, the second protein in the phosphotrasferase system. This interaction was found to be highly specific; the HPrs of different bacteria do not seem to bind to *E. coli* GlgP, and nor do HPr-related *E. coli* enzymes such as NPr (a homologue of HPr used in the nitrogen regulatory phosphoryl transfer chain) or FPr (a diphosphoryl transfer protein). Equally, *E. coli* MalP (maltodextrin phosphorylase), which shows high similarity with GlgP, does not bind to *E. coli* HPr (Deutscher et al., 2006). Conversely, overproduction of GlgP was found to result in the inhibition of the phosphotransferase system, through sequestering of more HPr (Koo & Seok, 2001). Whether this contributes to the glycogen reduction of cells overexpressing *glgP* is unconfirmed, but seems entirely possible. The binding of GlgP to HPr was found to be strongest (with four times greater affinity) when HPr is totally phosphorylated, but GlgP activity was increased (by 2.5 times) when it was bound to dephosphorylated HPr, under which conditions GlgP is also stimulated to form dimers and tetramers (Seok et al., 2001; Ball & Morell, 2003). The cellular concentration of HPr is far greater than that of GlgP, so that at any one time a large proportion of the cellular GlgP enzymes will form complexes with the HPr enzymes. It has therefore been proposed that GlgP activity is directly regulated by the phosphorylation state of HPr. When cells are grown on glucose, this would result in an accumulation of glycogen at the onset of stationary phase, when glycolysis is still in full flow but sugar entry into the cell is dwindling, leading to a relatively high proportion of phosphorylated HPr. As stationary phase continues, the reduction of sugar entering the cell slows glycolysis, leading to a reduction of PEP. Because the first phosphoryl transfer steps of the phosphotransferase system are reversible (Meadow & Roseman., 1996) this leads to a higher proportion of dephosphorylated HPr, which increases GlgP activity, so that the glycogen stores are metabolised (Deutscher et al., 2006) (figure 1.10).

The activity of the phosphotransferase chain also has a primary role in the signalling systems that control the hierarchy of preference for certain carbohydrate substances over others being transported into the cell and metabolised; in particular, for the preferential consumption of glucose over other forms of carbohydrate (Escalante et al., 2012). So far, 21 different Enzyme II complexes have been identified in *E. coli*, involved in the transport and phosphorylation of about 20 different carbohydrates (Escalante et al., 2012). When the bacterial cell is exposed to a mixture of extracellular carbohydrate sources, the phosphotransferase system prevents the expression of several catabolic genes and the activity of ‘non-PTS sugars’ transport systems, through the activity of its enzymes in their different states of phosphorylation.

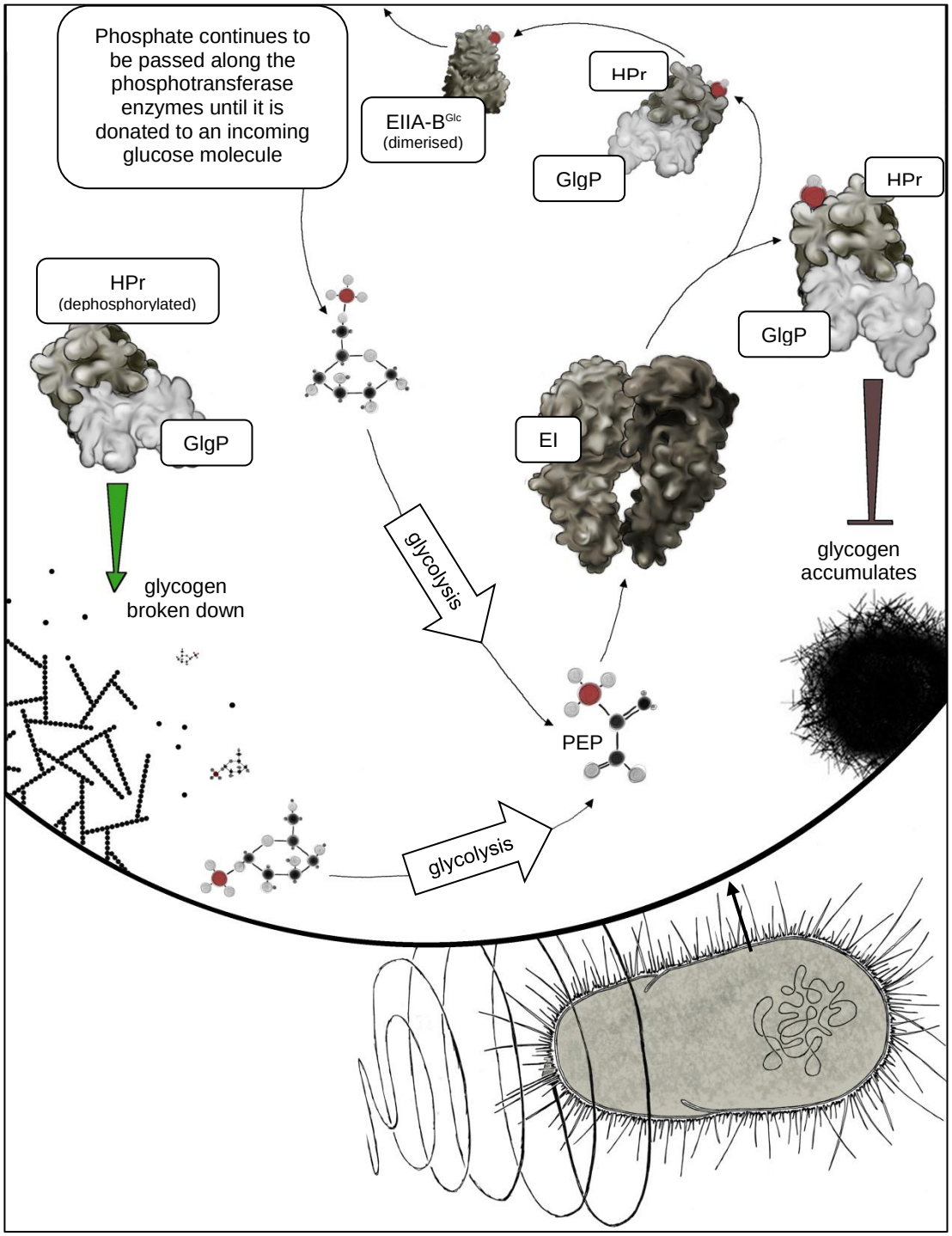


Figure 1.10: The effect of the phosphotransferase system on GlgP. When HPr is phosphorylated, it binds strongly to GlgP. However, dephosphorylated HPr still binds weakly to GlgP and increases its activity, causing it to break down glycogen (which in turn releases glucose-1-phosphate for glycolysis, leading to an increase of PEP and a proportional increase of phosphorylated HPr, so that the system is self-regulating).

For example, during growth in media containing mixed carbohydrate sources of glucose and the 'non-PTS sugar' lactose, the bacteria will preferentially take up the easily-metabolised glucose first, and only start to acquire and metabolise lactose once the glucose has been exhausted. It achieves this through the activity of the glucose-specific Enzyme IIA^{Glc} in its different phosphorylation states. While the extracellular glucose source is high, the concentration of PEP (a product of glycolysis) is high, but the phosphates that they donate to the enzymes of the phosphotrasferase system are rapidly shuttled to the incoming glucose molecules, so that they remain preferentially dephosphorylated (Escalante et al., 2012).

Meanwhile, the genes expressing enzymes needed for lactose acquisition and metabolism, which make up the *lac* operon and include *lacY* (coding for lactose permease, the integral membrane protein enzyme needed to transport lactose from the periplasmic space into the cytoplasm) as well as *lacZ* and *lacA* (coding, respectively, for the enzymes β-galactosidase, needed for lactose metabolism, and thiogalactoside transacetylase, whose function remains unclear but seems to be involved in detoxifying the cell of certain compounds, including, interestingly, IPTG (Roderick, 2005; Marbach & Battenbrock, 2012)) are repressed by the activity of LacI, the lactose inhibitor. The *lacI* gene for this repressor enzyme sits just upstream of the *lac* operon and is constitutively expressed, so that under growth on glucose it stops the transcription of the *lac* operon genes. Its repression is not absolute, however, so some transcription of the *lac* operon occurs, leading to a low level of cellular Lac enzymes, including the LacY lactose-permease. However, at the same time, the dephosphorylated version of Enzyme IIA^{Glc} binds to any LacY enzymes embedded in the cellular membrane (though only in the presence of lactose, so that Enzyme IIA^{Glc} isn't wasted if no lactose is present in the growth medium. Enzyme IIA^{Glc} also carries out similar forms of repression on several other 'non-PTS sugars' transfer or metabolism enzymes, such as the melibiose permease MelB, the MalK component of the maltose transport system and the glycerol kinase GlpK). The binding of Enzyme IIA^{Glc} inhibits LacY from transferring lactose into the cytoplasm. Glucose is therefore taken up and phosphorylated by Enzymes $IICB^{Glc}$, while lactose is excluded from the cell and little energy is wasted on transcribing and translating the genes that would be needed for its metabolism, since these are largely repressed (Alacante et al., 2012; Deutscher et al., 2006; Wilson et al., 2007) (figure 1.11).

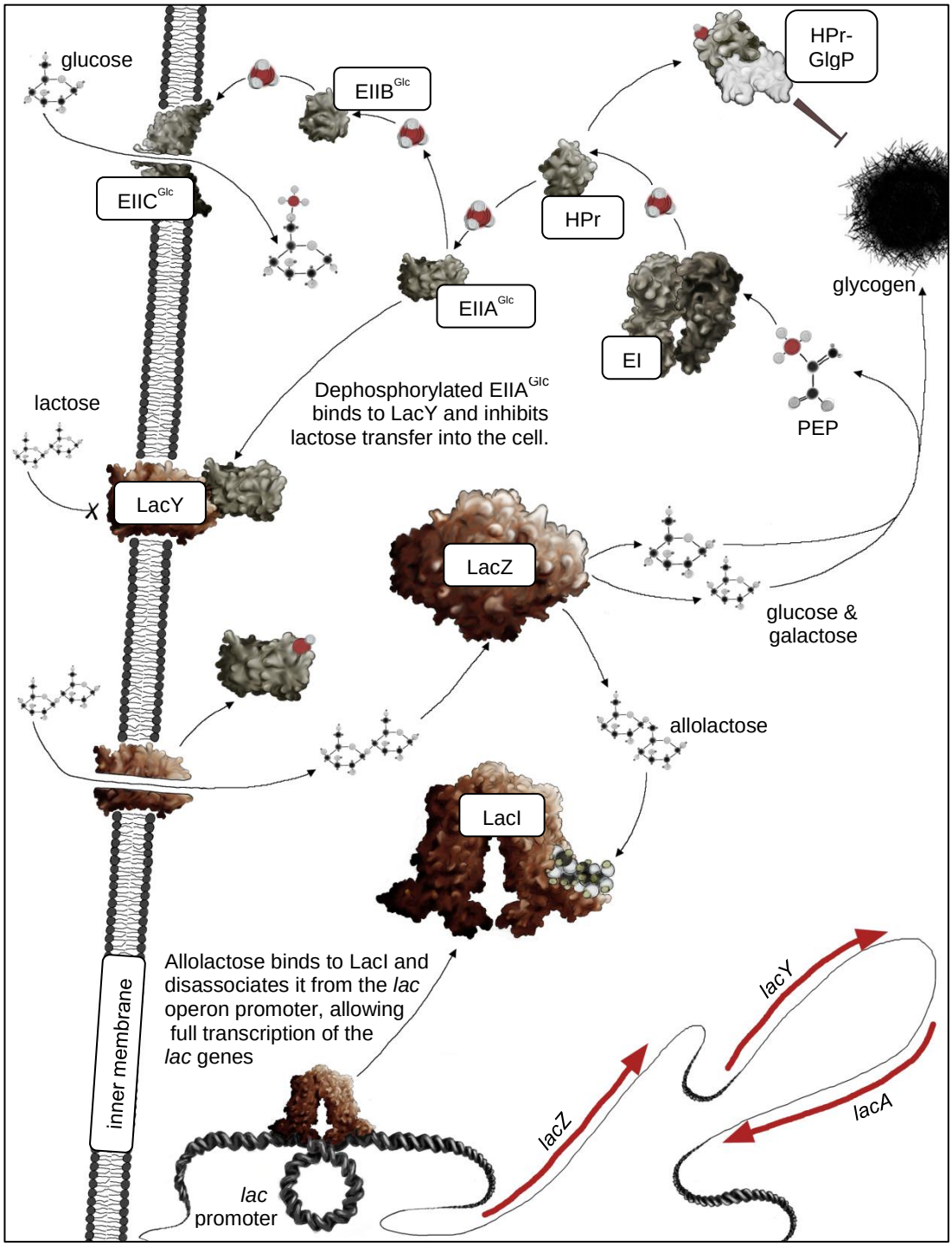


Figure 1.11: The effect of the phosphotransferase system on lactose metabolism. When EIIA^{Glc} (a middle enzyme in the phosphotransferase system) is in a dephosphorylated state it is able to bind to LacY, the lactose permease, and inhibit lactose from entering the cell. This therefore prevents the cell from taking in lactose – which is more costly to metabolise – when there is a surplus of glucose in the environment. However, when glucose is scarce this inhibition is switched off and lactose enters the cell. It is metabolised by LacZ, which turns it either into its constituent glucose and galactose parts, or converts it into allolactose. Allolactose binds to LacI, the lac inhibitor, thereby disassociating it from the *lac* promoter and allowing full transcription of all *lac* genes.

However, as the extracellular glucose concentration starts to drop, its entry into the cytoplasm via Enzyme IIC^{Glc} slows, so that the enzymes of its phosphotransferase system are unable to pass on their phosphate as quickly. They therefore start to become preferentially phosphorylated. Phosphorylated Enzyme IIA^{Glc} does not bind to LacY, so with this shift in phosphorylation states, the LacY enzymes embedded in the cell membrane start to lose their inhibition and actively transport lactose into the cytoplasm. Due to the slight leakage of the repression of the *lac* operon by LacI, *lacZ* is also transcribed in low quantities. LacZ cleaves lactose into glucose and galactose, but can also lead to the formation of allolactose by transglycosylation, basically the switching of the linkage between the glucose and galactose monomers from a β-1,4 glycosidic linkage to a β-1,6 glycosidic linkage. Allolactose binds to the LacI repressor enzyme, causing a conformational change that removes it from the upstream sequence of the *lac* operon, so that full LacZYA expression can proceed. At the same time, the phosphorylated version of Enzyme II^{Glc} may no longer bind to LacY, but it does now stimulate expression of another gene, *cya*, which codes for adenylate cyclase, the enzyme responsible for turning ATP into cyclic AMP (cAMP). cAMP is the cytoplasmic signal metabolite of carbon and energy sufficiency. It binds to another protein, the cAMP regulator protein (CRP), activating it to bind to DNA as a positive transcriptional regulator of a number of catabolic operons. cAMP/CRP is thought to regulate the transcription of more than 150 genes, mostly involved in energy metabolism, such as galactose, citrate or, in the instance of this example, lactose metabolism. Indeed, mutant cells lacking either the *cya* or *crp* gene show a total repression of the *lac* operon, revealing that its expression requires at least a basal level of cAMP/CRP (Kuo et al., 2003). The cAMP/CRP complex achieves this positive regulation by binding to a specific site upstream of the *lac* promoter, which enhances binding of RNA polymerase and thereby increases transcription of the *lacZYA* genes by nearly 50 times, enabling the cells to readily metabolise lactose (Kolb et al., 1993; Tutar, 2008; Wilson et al., 2007). A positive-feedback loop is therefore set up, where the more lactose is transported into the cell and metabolised, the more the expression of those genes is increased (Alacante et al., 2012; Deutscher et al., 2006) (figure 1.12).

Relief of the LacI repressor on transcription of the *lac* genes can also be achieved by the addition of the non-metabolisable inducer IPTG (Isopropyl β-D-1-thiogalactopyranoside). This binds to the LacI inhibitor enzyme in the same way that allolactose does, changing its structural conformation so that it disengages with the upstream region of the *lac* operon, allowing transcription of the *lac* genes to commence. However, while glucose is still available in the growth medium the transcription of the *lac* operon remains relatively low,

due to the inhibition of cAMP expression while Enzyme IIA^{Glc} is still largely dephosphorylated. It is only when glucose runs out that phosphorylated Enzyme IIA^{Glc} builds up and activates the conversion of AMP into cAMP, leading to an increase in the cAMP/CRP complex which in turn binds upstream of the *lac* operon and increases its expression nearly 50-fold (Wilson et al., 2007). This in turn leads to an upregulation of the *lacY* gene, creating an abundance of LacY permease proteins in the inner membrane. While the general model outlined above suggests that dephosphorylated Enzyme IIA^{Glc} inhibits LacY activity, the same can be said the other way around, since binding to LacY blocks the Enzyme IIA^{Glc} phosphorylation site. A large proportion of the remaining dephosphorylated Enzyme IIA^{Glc} is therefore removed from the phosphotransferase system, lowering the uptake of glucose (Deutscher et al., 2006).

The dephosphorylated form of Enzyme IIB^{Glc} (which is the predominant form under high glucose concentrations) also binds to a protein: Mlc (named for the phenotype of its mutation 'makes large colonies'). When glucose runs dry and Enzyme IIB^{Glc} becomes predominantly phosphorylated, the Mlc protein detaches and is free to bind to DNA. The DNA that it binds to is a site either close to or overlapping the promoter regions of its own *mlc* gene, as well as the genes that code for Enzyme IIA^{Glc}, Enzyme IIB^{Glc} and Enzyme IIC^{Glc} (and a handful of other glucose-specific genes). Therefore when glucose is no longer available in the growth medium, the cell represses expression of its glucose transport and metabolism genes. Expression of *mlc*, along with genes that code for Enzyme IIA^{Glc}, Enzyme IIB^{Glc} and Enzyme IIC^{Glc}, is also stimulated by the cAMP/CRP complex, so that these genes are actually transcribed and translated at lower levels while glucose is available. Although this results in lower induction of the genes for Enzyme IIA^{Glc}, Enzyme IIB^{Glc} and Enzyme IIC^{Glc}, it also causes relatively low *mlc* expression, which means those glucose genes are not repressed by Mlc (Plumbridge, 2002; Escalante et al., 2012; Deutscher et al., 2012). Mlc expression was also shown to decay rapidly in the presence of glucose, suggesting that it is also subject to post-transcriptional control. Taken together, this network of interconnecting regulation perhaps acts as a means to fine-tune the expression of glucose transport and metabolism genes, to avoid the stress of metabolite overload: under high glucose there is little stimulation of the genes by cAMP/CRP, but nor is there inhibition by Mlc, so levels are relatively stable. As glucose dwindles, cAMP/CRP expression goes up, stimulating the expression of glucose transport and metabolism genes to take advantage of the little glucose left, while also switching on genes to metabolise other sugars. Meanwhile, Mlc starts to disassociate from Enzyme IIB^{Glc} and repress expression of the glucose genes (figure 1.13).

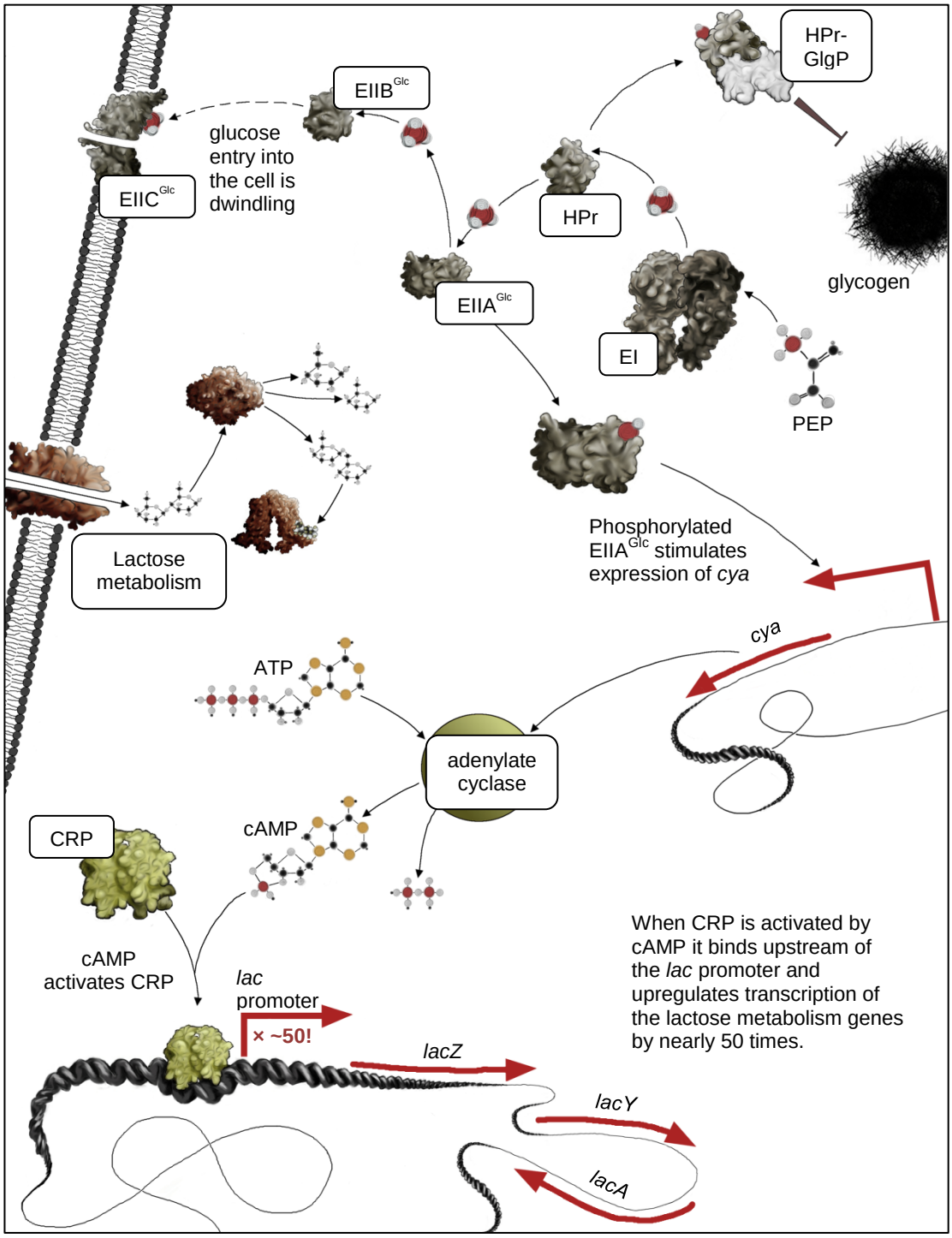


Figure 1.12: The further effect of the phosphotransferase system on lactose metabolism. When $EIIA^{Glc}$ is phosphorylated it no longer bind to LacY, the lactose permease, thereby allowing lactose to enter the cell and switch off the inhibition of the *lac* promoter, as described previously. Additionally, the phosphorylated $EIIA^{Glc}$ stimulates expression of another gene – *cya* – which codes for adenylate cyclase (structure unknown). This turns ATP into cAMP (cyclic AMP). In turn, cAMP binds to its Receptor Protein (CRP) and activates it to bind to DNA as a positive transcriptional regulator. One of the many operons it regulates is the *lac* operon, for which it increases transcription by nearly 50 times. Lactose metabolism is therefore now in full swing!

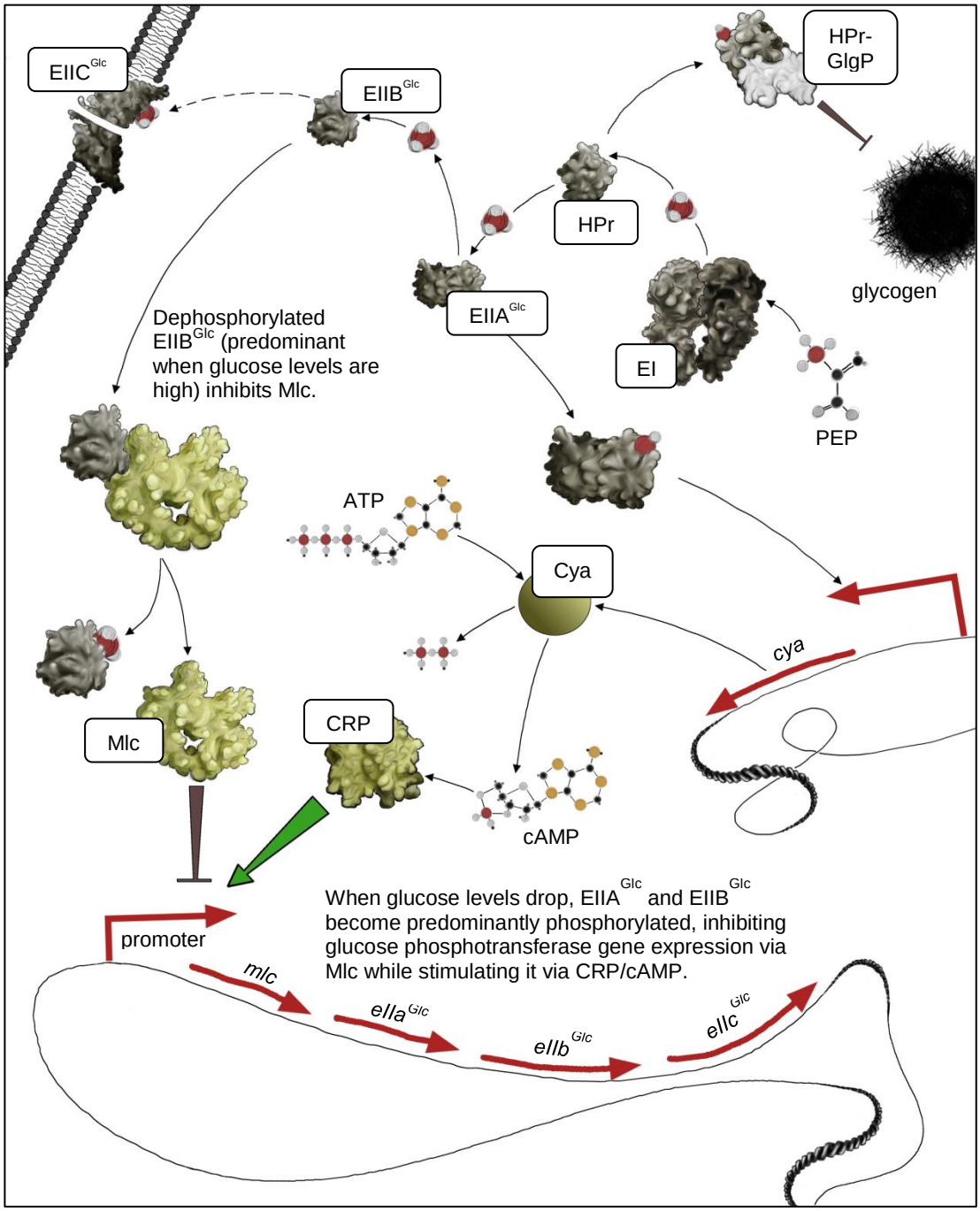


Figure 1.13: Fine-tuning the phosphotransferase system. The Mlc protein is inhibited by dephosphorylated EIIB^{Glc}, but not phosphorylated EIIB^{Glc}. It represses expression of its own *mlc* gene, as well as the genes that code for EIIA^{Glc}, EIIB^{Glc} and EIIC^{Glc}, so that glucose transport and metabolism is reduced when glucose levels are low. However, the same genes are stimulated by cAMP/CRP. So under high glucose there is little stimulation of the glucose phosphotranspherase genes (perhaps to avoid metabolite overload) but nor is there inhibition by Mlc, keeping levels stable. As glucose levels dwindle, cAMP/CRP increases, stimulating glucose phosphotranspherase gene expression to take advantage of the little glucose left, while also switching on the genes to metabolise other sugars. Meanwhile, Mlc starts to disassociate from EIIB^{Glc} and repress expression of the glucose phosphotranspherase genes once the glucose has run dry.

1.7.2 Cyclic AMP (cAMP)

Cyclic AMP (cAMP), which is synthesised by the product of the *cya* (*adenylate cyclase*) gene acting on ATP, forms a complex with the cAMP receptor protein (CRP) that is an important cellular mediator of glucose effects in *E. coli*. As previously described, cAMP expression is low when glucose is the primary carbon source, due to the inhibition of the *cya* gene as a side-effect of glucose transport into the cell and the predominantly dephosphorylated state of Enzyme IIA^{Glc}. However, under conditions of low glucose or alternative carbon sources, cAMP binds to CRP, forming a complex which is activated to bind to DNA as a positive transcriptional regulator of more than 150 genes. Among these are the operons encoding the ABC transporter of the maltodextrin system, as is the expression of *malT*, which encodes the activator of all the *mal* genes. In this way, the use of glucose as a substrate only permits glycogen synthesis via GlgA activity.

In 1989, Romeo & Preiss suggested that the expression of *glgC* and *glgA* (though not *glgB*) were positively regulated by increased levels of cAMP/CRP, while mutations in the *cya* and *crp* genes lead to a decrease in glycogen accumulation. Addition of cAMP to the cells lacking *cya* restored normal glycogen levels, strongly suggesting that expression of the *glg* genes are controlled by the cAMP/CRP complex. However, subsequent research has failed to show a change of *glgC* expression in *E. coli* mutants lacking the *cya* gene, or of *glgBXCAP* expression in *E. coli* lacking *cya* function, or *glgA* expression in *E. coli* mutants lacking the *glgBX* genes and *cya* function (Montero et al., 2011). Furthermore, transcriptome analysis failed to include *glg* genes in the CRP regulon (Gosset et al., 2004; Hollands et al., 2007).

However, experimental evidence does suggest that cAMP/CRP directly upregulates *glgS*; a sixth, rather enigmatic *glg* gene which has not been discussed so far (Wilson et al., 2010). *glgS* expression has been known for some time to positively affect glycogen accumulation, so that its upregulation, through cAMP/CRP interaction, results in a higher glycogen content. However, it is not found on the same operon as the five core *glg* genes, knockout mutation experiments have suggested that it is not essential for glycogen synthesis or degradation in most glycogen producing bacteria (Wang & Wise, 2011), and its function remains obscure. It is a hydrophilic, highly-charged protein that has no significant sequence similarity to any other protein currently found in databases outside those of the enterobacteria (Wilson et al., 2010). Its structure is highly conserved, except for a variable N-terminal portion, and overall it has a similar architecture to the peripheral subunit binding domain of dihydrolipoyl acyltransferase, an enzyme from the pyruvate dehydrogenase multienzyme complex that is involved in glucose metabolism. It has been shown to stimulate glycogen production when

overexpressed (Kozlov et al., 2004), while in mutant *E. coli* cells lacking the *glgS* gene, high levels of AMP were accumulated, which negatively regulate GlgC and so lead to a reduction of glycogen content (Rahimpour et al., 2013). GlgS has also been found to be negatively regulated by the global post-transcriptional carbon storage regulator CsrA, and positively regulated by the stringent response regulators (p)ppGpp, as well as RpoS and the RNA chaperone Hfq (whose translation is in turn inhibited by CsrA) (Rahimpour et al., 2013). All of these factors would suggest an interaction with proteins involved in glycogen biosynthesis. However, mutants lacking the *glgS* gene have recently also been found to be hypermobile, with increased flagella and fimbriae production and an increase in biofilm production. Along with predictions of genomic interaction between *glgS* and the non-essential genes of *E. coli*, these results indicate that GlgS is in fact a major negative regulator of enzymes involved in *E. coli* propulsion, adhesion and synthesis of biofilm exopolysaccharides, processes which would otherwise consume large amounts of ATP and glucose-1-phosphate, diverting them away from glycogen synthesis. As such, it has even been suggested that the enzyme is renamed ScoR (surface composition regulator) (Rahimpour et al., 2013). The glycogen excess phenotype found in mutant cells overexpressing GlgS is therefore most likely an indirect effect of those cells having more ATP and glucose available because they need less for their reduced surface composition. It seems likely then that it is mainly through this indirect route – through the upregulation of *glgS* and therefore the repression of surface composition genes and a subsequent increase in cellular ATP and glucose – that cAMP/CRP expression indirectly increases glycogen accumulation. Mutant cells lacking either the *cya* or the *crp* genes therefore experience a downregulation of *glgS* which leads to increased expression of surface composition proteins, and as a result a hypermobile phenotype, with increased flagella and fimbriae production and an increase in biofilm production, and a subsequent reduction of ATP and glucose available for glycogen biosynthesis, leading to the glycogen-deficient phenotype. Mutant *E. coli* lacking either *cya* or *crp* also show a high accumulation of AMP, though the reason for this is, as yet, unknown (Montero et al., 2009). Since AMP is an important allosteric inhibitor of GlgC, this is again likely to contribute indirectly to the reduction of glycogen accumulation.

1.7.3 (p)ppGpp, the product of RelA

(p)ppGpp, the product of RelA, is an alarmone that directly affects the transcription of many genes in response to changing levels of amino acids taken up by the cell. In exponential growth phase, when the level of amino acid intake is high, around 85% of a bacteria's tRNA will be bound to amino acids and available for protein synthesis at any one time, while the majority of the remaining 15% will be bound to synthetases, in preparation for receiving an amino acid, or to ribosomes after transferring their amino acid to a growing peptide chain. However, when amino acid levels drop, the proportion of tRNAs carrying amino acids can reduce to less than 20%. Amino acid carrying (aminoacylated) tRNAs bind to ribosomes with a much higher affinity (about 50 times higher) that deacylated 'empty' tRNAs. However, under conditions of a high proportion of deacylated tRNAs in comparison to aminoacylated tRNAs, many of the deacylated tRNAs are given the opportunity to bind weakly to ribosomes, if those ribosomes contain mRNA with the 'empty' tRNA's corresponding codon. This effectively blocks the ribosome from carrying out translation. When this happens, the RelA protein, swiftly followed by the RelC protein (which until this point have been inactive) also bind to the ribosome and, in this bound state, catalyse the reaction of ATP and GTP or GDP to synthesise AMP and the alarmones pppGpp (guanosine pentaphosphate) or ppGpp (guanosine tetraphosphate) (collectively referred to as (p)ppGpp). As this reaction occurs it causes a conformational change of the RelA protein which releases it from the ribosome, whereupon it is able to be recycled by hopping to another bound ribosome and repeating the process (Wendrich et al., 2002). (p)ppGpp synthesis is at its most efficient when there is a vast excess of ribosomes bearing deacylated tRNAs over RelA proteins (and indeed there is around one RelA protein per 200 ribosomes in the average *E. coli* cell (Srivatson & Wang, 2008)). Under these conditions, the RelA proteins hop continuously from one blocked ribosome to another, producing ppGpp and pppGpp. This allows for regulatory control, providing a guarantee that the level of (p)ppGpp production reflects, at any one time, the number of blocked ribosomes in the cell (Wedrich et al., 2002). The (p)ppGpp (produced as a result of this starvation-induced inhibition of translation) then almost immediately go on to inhibit the transcription of genes associated with the translation apparatus, such as genes encoding ribosomal proteins, rRNAs, tRNAs, translation factors and synthetases, while upregulating the transcription of genes encoding metabolic enzymes, especially those involved in amino acid synthesis, in a process called the stringent response (Wendrich et al., 2002). This discrepancy is thought to be due to a difference in the promoters of the genes involved. The genes for translation factors such as stable RNAs (rRNA and tRNA) have a GC-rich 'discriminator' sequence between their transcriptional

start site and the -10 site, which has suboptimal interaction with the corresponding region of RNA polymerase. They therefore normally form open complexes with RNA polymerase during transcription initiation, allowing for optimal transcription. However, this instability is increased by the association of (p)ppGpp with RNA polymerase, to the point where transcription is inhibited (Magnusson et al., 2005). Expression of ribosomal protein genes, meanwhile, is controlled by cellular rRNA levels, so that a downstream effect of the inhibition of stable RNA synthesis is a large-scale downregulation of the entire translation apparatus (Paul et al., 2004). At the same time, promoters for genes that are upregulated during the stringent response, such as those for amino acid synthesis, have a different 'discriminator' sequence between their transcriptional start site and -10 site, which is rich in ATs and allows for optimal binding with the RNA polymerase. (p)ppGpp association during transcription of these genes weakens this binding enough to increase the rate of formation of open complexes, and thereby stimulates transcription. At the same time, the transcriptional inhibition of translation apparatus genes caused by (p)ppGpp association leads to the release of RNA polymerases from their promoters, leaving those RNA polymerases available for transcription of genes without a CG-rich 'discriminator' region, such as those for amino acid biosynthesis. Since transcription of the genes of the translation apparatus can account for around 60-80% of total cellular transcription, this passive upregulation is thought to lead to a significant increase in the transcription of genes that are favoured at the onset of stationary phase (Barker et al., 2001). Furthermore, (p)ppGpp is thought to inhibit the initiation of DNA replication in *E. coli* through interaction with primase, and may also inhibit the activity of translation elongation and initiation factors (Srivatson & Wang, 2008).

Deletion of the *relA* or *relC* genes leads to a relaxed phenotype, which gives these proteins their names, where RNA synthesis can continue for up to an hour after translation has been inhibited by starvation (Yang & Ishiguro, 2001). Transcriptomic analysis of the *glg* upstream region has provided conflicting data, suggesting that either (p)ppGpp is a negative allosteric regulator of the *glg* genes or that it enhances *glgC* and *glgA* transcription, but not that of *glgB* (Romeo & Priess, 1989; Montero et al., 2011). Analysis of mutant *E. coli* lacking the *relA* gene show that the expression of the five core *glg* genes was almost entirely abolished in these cells, strongly indicating that the entire *glg* operon belongs to the RelA regulon, while also reinforcing the notion that *glgBXCAP* operates as a single transcriptional unit (Romeo & Preiss, 1989; Montero et al., 2011). Further experiments by Motero et al. (2011), which show lower *glgA* expression in mutants lacking *relA* and *glgBX* than those only lacking *glgBX*, further suggest that the *glgAP* suboperonic promoter is also part of the RelA

regulon. Due to the dramatic shift away from the synthesis of translation apparatus, which under normal growth conditions accounts for the majority of cellular transcription, and without which growth and division are arrested, (p)ppGpp regulation during the stringent response also leads to a significant increase in the cell's available pool of ATP. Under conditions of carbon source excess, this ATP can then be diverted towards glycogen biosynthesis as a means of energy storage (Montero et al., 2009).

Another protein, SpoT, is also able to synthesise (p)ppGpp, but rather than responding to amino acid starvation it is activated under conditions of limited phosphorus, iron, fatty acids and carbon sources. SpoT is also able to shift its activity to the hydrolysis of (p)ppGpp to GDP and inorganic phosphate under conditions of nutrient availability (Srivatson & Wang, 2008; Magnusson et al., 2005), and as such is an important mechanism for controlling the concentration of intracellular (p)ppGpp. The deletion of both *relA* and *spoT* genes leads to the total inability of cells to produce (p)ppGpp (Traxler et al., 2008; Lange et al., 1995). This would suggest that under conditions of limited phosphorus, iron or fatty acids the initiation of the stringent response would again induce glycogen accumulation (Montero et al., 2011).

Another suspected impact of (p)ppGpp expression during the stringent response is that it potentially inhibits another enzyme, PurA, which catalysis the first committed step in AMP biosynthesis (Hou et al., 1999; Ballorica et al., 2003). So although the production of (p)ppGpp itself produces AMP as a biproduct, the overall cellular content of AMP is still reduced. Since AMP is the main allosteric inhibitor of GlgC, a knock-on effect of (p)ppGpp activity is the post-transcriptional upregulation of GlgC and therefore glycogen biosynthesis, due to the reduction of cellular AMP. Certainly mutants lacking the *relA* gene have been shown to accumulate high levels of AMP while also showing reduced glycogen content. Meanwhile, as already mentioned, mutants lacking *purA* show reduced AMP accumulation and high glycogen levels (Montero et al., 2011), while cells bearing a mutant form of the GlgC enzyme that is insensitive to AMP allosteric regulation also accumulate an excess of glycogen (Govons et al., 1973).

Lastly, (p)ppGpp also alters the utilisation of sigma factors – proteins which enable specific binding of RNA polymerase to promoters, and are therefore needed for the initiation of RNA synthesis – by shifting RNA polymerase away from genes dependent on the σ^{70} 'housekeeping' sigma factor towards genes dependent on alternative sigma factors, such as the stationary phase sigma factor σ^{S} (RpoS), as well as the extracytoplasmic stress sigma factor σ^{E} (RpoE) while also inducing transcription and increasing stability of those sigma

factors. It does this by one of two mechanisms. The first is through direct upregulation of the sigma factor transcription by binding to RNA polymerase and aiding formation of an open complex, as described above, which seems to be the case for RpoS. However, further action of upregulation appears to occur through the second, more complex route of the translational upregulation of genes with AT-rich discrimination factors which then either directly upregulate the expression of those sigma factors or block the action of other proteins which aid in their degradation. The RpoE-mediated envelope stress response actually seems to negatively affect glycogen accumulation, since mutants lacking the gene for the enzyme that degrades it (RseA) show a reduction in glycogen accumulation (Montero et al 2009).

1.7.4 The RpoS sigma factor

The RpoS Sigma factor is considered the master regulator of the stress response, and is thought to upregulate the transcription of at least 500 genes, which accounts for around 10% of the *E.* coli genome (although only about 140 of these are considered the 'core *rpoS* regulon', being induced by an increase of RpoS regardless of other factors). Expression of RpoS seems to be induced through sensing of the cell's growth rate, and therefore also its energy state, and as such is intimately linked not only with (p)ppGpp expression, but also the phosphotransferase system and cAMP/CRP. Indeed, there are a great many factors controlling RpoS expression, as there are for the expression of the other five alternative sigma factors identified in *E. coli*. The 5' untranslated region of *rpoS* is one of the longest identified in *E. coli*, spanning 567 nucleotides, and is known to bind at least three regulatory RNAs, all of which promote *rpoS* expression through either mRNA stabilisation or enhanced translation. This positive control of *rpoS* stability by regulatory RNAs also requires the RNA helicase Hfq, which is a chaperone that stabilises many regulatory sRNAs and aids their annealing to target mRNAs. (Landini et al., 2013; Wilson et al., 2010).

The expression of RpoS has been linked to glycogen accumulation (Wilson et al., 2010). However, the *glgBXCAP* operon has not been found to be part of the core *rpoS* regulon. Instead, it seems that the upregulation of glycogen through RpoS expression is indirect. A means by which this might occur is made apparent by the fact that one of the genes that does belong to the *rpoS* regulon is *glgS* (Kozlov et al., 2004; Wilson et al., 2010). Experimental evidence suggests, then, that RpoS – like cAMP/CRP – upregulates *glgS*, and it is therefore very possible that by the same mechanism – of GlgS inhibition of surface composition proteins – RpoS expression indirectly increases glycogen accumulation by increasing the pool of available ATP and glucose.

1.7.5 CsrA

CsrA is the global post-transcriptional carbon storage regulator. It is an RNA-binding protein that controls the expression of genes involved in carbohydrate metabolism, by repressing gluconeogenesis and glycogen biosynthesis and catabolism while activating glycolysis, as well as cell motility and acetate metabolism, by acting on the translation of target genes and / or the stability of target mRNA transcripts (Baker et al., 2002). It also seems to influence biofilm formation, secondary metabolite production, environmental stress resistance and cytotoxic factor production. Its possible mode of action to prevent translation is through binding to a site near the Shine-Dalgarno sequence, thereby blocking ribosome binding and facilitating mRNA decay. Its positive regulation seems to occur through an increase of target transcript levels rather than translational efficiency, again through binding to mRNA, but in this case stabilising it (Lucchetti-Miganeh et al., 2008). Although not many direct targets of CsrA have so far been identified, among those that have are the *E. coli glg* genes (Baker et al., 2001). CsrA has been found to directly negatively control *glgC* , *glgA*, *glgB* and *glgP* expression, though interestingly not *glgX* expression (Yang et al., 1996). Since the debranching activity of GlgX is dependent on those branches first being shortened by the activity of GlgP, it may simply be that inhibition of *glgP* translation makes the inhibition of *glgX* redundant. In particular, CsrA has been found to bind upstream of *glgC*, overlapping the ribosome binding site and thereby preventing *glgC* translation and promoting its mRNA decay (Lucchetti-Miganeh et al., 2008). Two other direct targets of CsrA that have been identified are *glgS* and *hfq*, both of which it negatively regulates (as mentioned above), thereby increasing its overall negative effect on glycogen accumulation (Montero et al., 2013; Wilson et al., 2010).

1.7.6 The PhoP-PhoQ system

The PhoP-PhoQ system is another regulatory system, occurring in *E. coli* and other bacteria under conditions of nutrient limitation, which responds to extracellular levels of Mg^{2+}. Mg^{2+} is the most abundant divalent cation within living cells, which strongly determines the cell's metabolic and energy status. It is used, among other things, as a stabilising factor for membranes, tRNA and ribosomes as well as being an essential cofactor in a variety of enzymic reactions including all phosphoryl transfer reactions that involve ATP (Reinhart, 1988; Groisman et al., 2013). This includes the conversion of glucose-1-phosphate and ATP to ADP-glucose and inorganic phosphate by GlgC, as well as the transfer of phosphate from PEP to Enzyme I in the phosphotransferase system.

This two-component regulatory system consists of a sensor (PhoQ) and a response regulator (PhoP). The PhoQ sensor is an integral cytoplasmic membrane protein that becomes phosphorylated in response to low environmental Mg^{2+}. The phosphate is then transferred to a conserved residue on the PhoP response regulator protein, whose affinity to DNA is modulated by phosphorylation. Low Mg^{2+} therefore promotes the transcription of PhoP-activated genes, such as those coding for enhanced Mg^{2+} transporter proteins, and decreases expression of PhoP-repressed genes, such as those involved in cellular respiration (Groisman et al., 2013). The expression of over 40 different proteins with widely divergent properties is modulated by PhoP (Véscovi et al., 1996), among which, it was thought, were the *glg* genes of glycogen synthesis and metabolism. Indeed, expression of the five core *glg* genes is lower in mutant cells lacking either the *phoP or phoQ* gene, while the comparisons of *E. coli* lacking the *glgBX* genes with those lacking both *glgBX* and *phoP* showed a lower activity of *glgA* in those with the double deletion, suggesting that both the *glgBXCAP* operon and the suboperonic promoter driving *glgAP* expression both belonged to the PhoP-PhoQ regulon. This control only occurs under conditions of low extracellular Mg^{2+}, since in culture supplemented with Mg^{2+}, normal glycogen levels were restored (Montero et al., 2009; Montero et al., 2011). Furthermore, the ability of *E. coli* cells to accumulate glycogen was found to be profoundly affected by submillimolar changes in external Mg^{2+} concentrations, whereby supplementation of culture media with Mg^{2+} led to an almost doubling of glycogen content at the onset of stationary phase in both wild type cells and mutant cells lacking the *phoP* or *phoQ* genes (Montero et al., 2009). However, analyses of the 1000 nucleotide promoter region upstream from the start codon of *glgC* has not revealed the presence of a putative PhoP box (Montero et al., 2009; Wilson et al., 2010), suggesting that in fact the mediation of *glg* genes by the PhoP-PhoQ system is indirect. Since the ATP content of cells cultured under non-limiting conditions of Mg^{2+} was over twice that of cells cultured in limited Mg^{2+}, the effect of Mg^{2+} concentration on glycogen content may be an indirect consequence of its effect on cellular ATP concentration. However, the downregulation of *glg* genes observed under Mg^{2+} limited conditions in mutant cells lacking either *phoP* or *phoQ* does still suggest that the *glg* genes are under control of the PhoP-PhoQ system, although it now seems likely that this is achieved in an indirect and so far unresolved manner.

1.8 The origins of starch synthesis

The limited size and hydrosoluble nature of glycogen is thought to have been problematic for an early member of the cyanobacteria, whose descendant – a recently isolated member of the subgroup V diazotrophic cyanobacteria (Clg1) (Falcon et al., 2004) – was found to be unique among prokaryotes so far studied in that, instead of glycogen, it accumulates starch (Nakamura et al., 2005; Deschamps et al., 2008b). As a diazotroph, this cyanobacterium fixes nitrogen using the primitive nitrogenase enzyme, which evolved on an early Earth devoid of atmospheric oxygen, and as such remains easily inactivated by the presence of O_2. This creates a serious complication for the cyanobacterium, which produces oxygen as a by-product of its photosynthetic activity. To bypass this complication, it controls nitrogen fixation under circadian regulation so that it only occurs at night. Since photosynthesis by definition only occurs during the day, a temporal boundary is set up between the oxygen-producing photosynthesis pathway and the oxygen intolerant nitrogen-fixation pathway. Carbohydrates stored during the day are used in nocturnal respiration, thereby diminishing any O_2 still present, and supplying energy and reducing power for the nitrogenase. It was this need to supply large amounts of energy in the form of carbohydrates that may have been the selective force that led to the evolution of starch – a large and osmotically-inert granule that would undergo relatively little turnover during the diurnal phase (Ball et al., 2011).

At the molecular level, starch is very similar to glycogen, being made up of glucose residues connected by α-1,4 linkages with side-branches forming from α-1,6 linkages (Zeeman et al., 2010). However, the granule is endowed with distinct properties due largely to its branching pattern. True starch is composed of two distinct fractions: amylopectin and amylose. Due to alterations in the branching pattern of the amylopectin fraction, there is no increasing interference between the branches as the granule grows in size, and therefore practically no restriction to the size each crystal can reach (figure 1.14). As such, starch granules range from <0.5 μm in the smallest free-living eukaryote yet identified, the green alga *Ostreococcus tauri*, to several tens of μm in potatoes (D'Hulst & Merida, 2010). The amylose, meanwhile, is not in fact requisite for starch formation, and requires a pre-existing amylopectin granule for its synthesis. However, starch that lacks an amylose fraction is termed 'waxy starch' as opposed to 'true starch'. It is thought to exist largely to increase the density and therefore the overall storage efficiency of the starch granule (Zeeman et al., 2010).

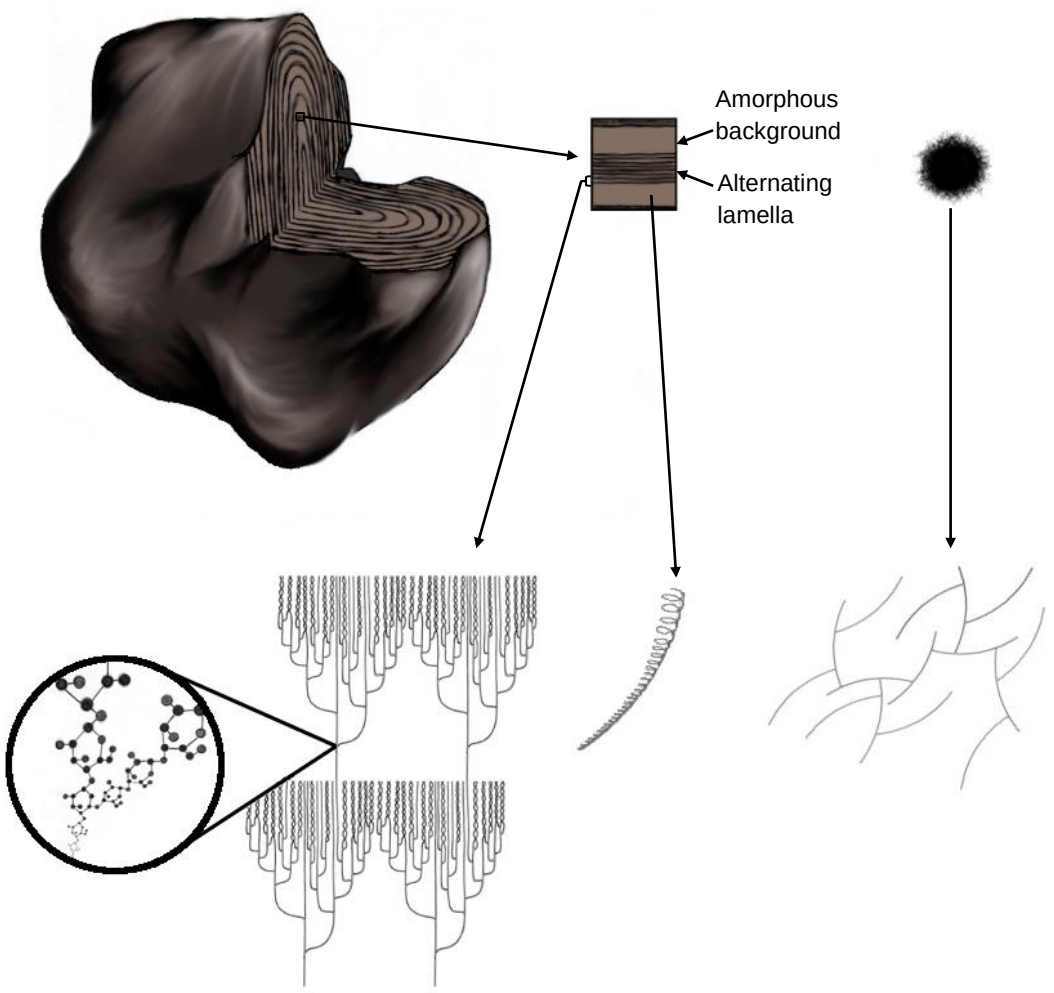


Figure 1.14: Structures of the three major storage polysaccharides: amylopectin and amylose, which together make up true starch, and glycogen.

1.9 Glucan structures

Starch is mainly composed of amylopectin (Ball & Morell, 2003). Its branching level is below 6% (compared to around 9% in glycogen) and occurs discontinuously, in tiers, thereby creating alternating amorphous and crystalline lamella (D'Hulst & Merida, 2010). The crystalline lamella are likely composed of unbranched glucan chains around 12 to 20 glucosyl residues in length, that pair together forming double helices with their neighbours, while the amorphous lamella are thought to be the recurrent regions of the molecule that contain the branch points, and as such are prevented from forming helices by the interference of the α-1,6-bonds. The lamellae repeat with a 9 - 10 nm periodicity and are joined by relatively sparse, long chains of around 35 to 80 glucosyl residues in length. It is further supposed that the double-helices of the crystalline lamella twist into superhelices, although the details of this have yet to be elucidated (Smith, 1999; Smith 2001; Zeeman et al., 2010).

The amylose fraction generally makes up around 15% of the granule in the Chloroplastida and is thought to be interspersed within the amorphous lamella of the amylopectin (Tetlow, 2006). It is far simpler than amylopectin, being mostly unbranched chains of α-1,4 linked glucosyl residues. It is also far smaller, with molecular weight estimates varying between 10^5 and 10^6 Daltons (Zeeman et al., 2010). Because of its sparsity of branches, amylose twists itself into single helices as shown in figure 1.14, where each coil of the helix is thought to be 6 glucosyl residues long (Yu et al., 1996).

Iodine is often used in the colourimetric assaying of starch, and it is the single helices of the amylose fraction that react to form the intense blue-black colour typical of this assay. X-ray diffraction analyses of amylose-iodine complexes have shown inclusion complexes, where atoms of iodide, rather than I_2, are encircled by the amylose helix, most likely as either triiodide or pentaiodide chains. The resulting colour and wavelength of maximum absorbance (λ_{max}) then depends on the length of the amylose-iodine complex, giving a brown colour, with a λ_{max} of around 550 nm, for helices of 21-24 residues, red for 25-29 residues, red-violet for 30-38 residues, blue-violet for 39-46 residues, and finally blue for 47 residues or more, with a λ_{max} of around 630 nm (John et al,. 1983; McGance et al., 1998; Saibene et al., 2008). Amylopectin also stains with iodine solution, forming a brown colour suggestive of short single helices within the molecule (iodine does not complex with the double-helices of the crystalline lamella). Glycogen, meanwhile, tends to form a reddish complex with iodine, suggesting the presence of slightly longer single helices, although this tends to vary considerably due to glycogen's lower structural organisation.

1.10 Endosymbiosis and a return to the plastid

At some point between 1.5 and 0.7 billion years ago (Ball et al., 2011) a heterotrophic, eukaryotic cell internalised a member of this early cyanobacterial lineage, and this single event led to a symbiosis that is likely to have given rise to all three lineages of photosynthetic eukaryotes: the Glaucophyta, the Rhodophyceae (red algae) and the Chloroplastida (green algae and land plants), that are collectively called the Archaeplastida (Rodriguez-Ezpeleta et al., 2005). Not only did this event enable the eukaryotic kingdom to start harvesting energy by oxygenic photosynthesis, it also led to the widespread use of starch as the chief storage polysaccharide among plants and algae.

After the cyanobacterial ancestor was engulfed by its eukaryotic host on the ancient Earth, a symbiotic relationship evolved between the two organisms, leading to a simplification of the cyanobacterium that has resulted in its reduced form: the plastids we find within Archaeplastida today. Many of the cyanobacterial genes became surplus to requirement, and so were lost, while others were transferred to the host genome. Among these were the genes for starch synthesis, which, since homologues for many of them already existed within the host's glycogen synthesis pathway, were able to replace or function alongside the genes of the host.

Starch synthesis is a relatively large and complex pathway, but is needed in its entirety if it is to function and provide an evolutionary advantage. The transfer of such a complex pathway in its entirety is extremely unlikely. However, since there is no mechanism by which the protein products of such relocated genes are automatically directed back to the organelle they came from, they are instead free to explore all targeting possibilities within the cell, and as such are able to replace host genes (Martin, 2010). The transfer could therefore have occurred one gene at a time, replacing or working alongside glycogen genes, until the whole pathway had relocated to the nucleus without the need for any gene duplication or subfunctionalisation (Deschamps et al., 2008c).

It has recently been shown that the genes from a third organism, a member of the bacterial genus *Chlamydia*, also became integrated into this mix (Horn, 2008). It is postulated that if the eukaryotic ancestor of the Archaeplastida was routinely infected by an ancestor of the parasitic chlamydia, these repeated infections would have made the chlamydia's genes available via lateral gene transfer during establishment of the endosymbiont (Huang & Gogarten, 2007) (figure 1.15). Among the donated genes of the chlamydial parasite are those for ATP/ADP Translocase, which allowed the cyanobacterial endosymbiont to import ATP

from its host and therefore survive the loss of its storage polysaccharide to the host cytosol. Also donated were at least one and perhaps all of the *isoamylase* genes used in the starch pathway of the Archaeplastida (see later), despite the presence of its analogue within the original cyanobacteria (Ball et al., 2011). Semi-crystalline storage polysaccharides may therefore have actually evolved independently from the cyanobacteria within the Archaeplastida. Still, the cyanobacterium was certainly responsible for some enzymes of the Archaeplastidial starch pathway. Among the genes that it passed on are those for Granule Bound Starch Synthases (Ball et al., 2011). These are responsible for the synthesis of amylose and therefore requisite for 'true' starch, and may also have had a more significant function in the past, making the cyanobacterial contribution to the pathway important for the present-day success of starch-accumulating eukaryotes.

This relocation of starch biosynthesis to the cytosol of the host cell can still be seen today within the relatively simple pathways of the Glaucophyta and Rhodophyceae. The cytosolic storage starch they accumulate has been collectively termed "floridean starch", as it was first described in the red algae Florideophyceae (Dauvillee et al., 2009). The Chloroplastida, however, accumulate starch within their plastids. The starch pathway moved back to the plastid at some point early in the evolution of the Chloroplastida, possibly to reduce the increased oxidative stress brought about by the evolution of chlorophyll b-containing light harvesting complexes (LHCs) (Deschamps et al., 2008b) (Figure 1.15). It is suggested that this switch from phycobilins to chlorophyll b, and the evolution of the chlorophyll-rich light harvesting complexes (LHCs) that coincided with this in the Chloroplastida, was the causative agent for the movement of Chloroplastidal starch back to the plastid because the need for a plastidial source of ATP at night would have increased under these circumstances (Deschamps et al., 2008b). Supporting this theory, *Arabidopsis* defective for plastidial ATP import have been shown to undergo oxidative stress, where the severity of this stress depends on the size of the intraplastidial starch pool (Deschamps et al., 2008c).

Of the three Archaeplastidal clades, the Glaucophyta are considered to most resemble the ancestral Archaeplastida (Ball et al., 2011), and its photosynthetic plastid (called a cyanelle as opposed to a chloroplast) still has a peptidoglycan layer; a relic of the ancestral cyanobacterial cell wall. Like all photosynthesis plastids, they contain chlorophyll a. However, along with the plastids of the Rhodophyceae (and the cyanobacteria), they also contain the photosynthetic pigment phycobilin, which in the Chloroplastidal plastids has been replaced with chlorophyll b.

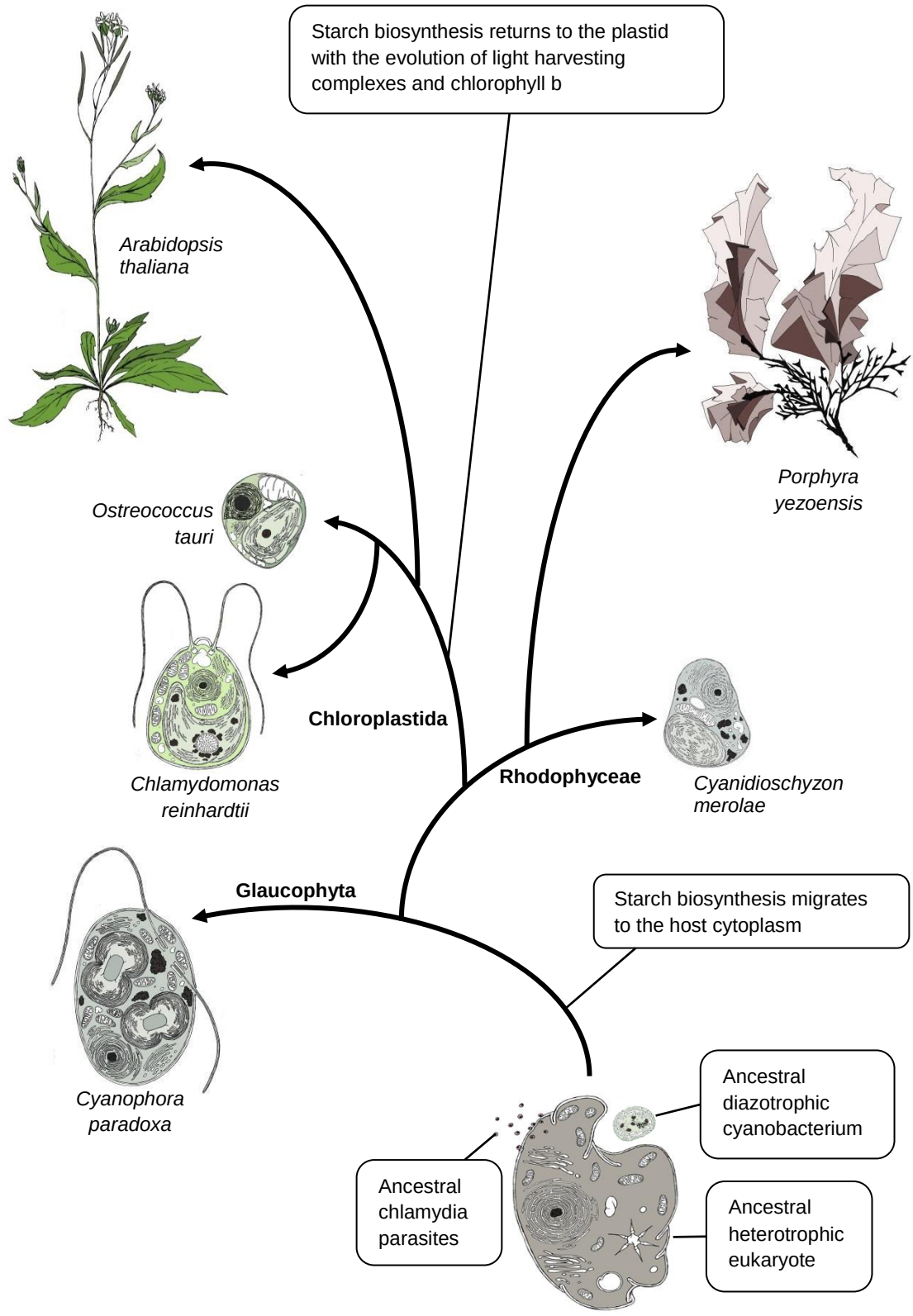


Figure 1.15: Simplified phylogeny of Archaeplastida, showing the position of model organisms referred to in the text. The mixing of synthesis enzymes from a cyanobacterium, a heterotrophic eukaryote and a chlamydial bacteria lead to the starch pathway found in all modern species.

The second relocation of the starch pathway, back to its original home in the plastid, could not occur as easily as its initial move to the cytoplasm, since no homologous genes were left within the plastid, and until the whole pathway was established any newly redirected individual enzymes would have had no function. 'Minor mistargeting', whereby the imperfections in the eukaryotic protein targeting system allow small amounts of pathways to leak into the 'wrong' compartment, such as a plastid, may have been enough to provide a small selective advantage for organisms containing such redirected enzymes, allowing them to endure until enough enzymes joined them to form a new pathway (Martin, 2010). This rewiring may also have involved steps of increasing biochemical complexity, co-evolving with the increasing complexity of the LHC antennae, where first the enzymes necessary for the accumulation of malto-oligosaccharides moved back, followed by the enzymes necessary to convert those malto-oligosaccharides into glycerol, before finally the remaining proteins followed suit and the plastid could again accumulate starch (Ball et al., 2011). This stepwise return to the plastid would mean that the genes experienced at least two rounds of duplication and subfunctionalisation, which fits with the presence of the relatively higher number of genes in the Chloroplastidal starch pathway (where even the smallest number seen so far is 33 genes, in *Ostreococcus tauri*) composed of multiple isoforms catalysing each step, playing a partly redundant function with one another.

The pathways of the different clades of starch-producing organisms therefore differ significantly. Most of this is accounted for by gene duplication, though some is also due to specialised functions that have evolved in various starches. As such, the number of genes responsible for starch metabolism ranges from 12 in the cyanobacteria, to over 40 in some members of the Chloroplastida. Furthermore, although the starch pathways of the Rhodophyceae and the Glaucophyta can consist of as few as 12 and in one species perhaps even fewer than 10 genes (see later), many are distinct from those found in the cyanobacteria (Deschamps et al., 2008b).

A surprising side-effect of all this complexity is the fact that it has clearly been a highly successful evolutionary strategy. Both Chloroplastidal and Floridean starch seem to have the same basic structure, where even the length of the amorphous and crystalline regions are the same (approximately 9-10nm) (Tetlow, 2006), so why Chloroplastidal starch is advantageous remains a mystery. The Chloroplastidal pathway does seem to provide a greater degree of plasticity, so it may be thanks to the greater potential for specialisation of starches between different species and different tissues of the same organism.

1.11 The starch synthesis pathway

Despite these overarching differences, all of the pathways investigated to date seem to include at least one gene for each of the following enzymes: Glucose Pyrophosphorylase, Granule Bound Starch Synthase, Soluble Starch Synthase, Branching Enzyme and debranching enzyme (Isoamylase). Additionally, there are around eight further classes of genes identified in the Archaeplastida whose function seems predominantly to do with starch hydrolysis, but which might also have some role in starch synthesis. But where the Rhodophyceae and the Glaucophyta tend to possess only one or two copies of each gene, the Chloroplastida can contain up to six isoforms for each, sometimes with more than one copy of each isoform.

1.11.1 Glucose Pyrophosphorylase

Glucose Pyrophosphorylase activates the polymerisation of glucose molecules by catalysing the synthesis of their nucleotide sugar substrate, the metabolite NDP-glucose. The enzyme has a homologue within the glycogen synthesis pathway, but where the ancient cyanobacterial symbiont (and its modern free-living descendent) used Adenosine diphosphate-glucose, its host (like all heterotrophic eukaryotes) would have begun its glycogen metabolism with the synthesis of Uracil diphosphate-glucose.

Of the modern clades of Archaeplastida, the Glaucophyta and the Rhodophyceae assimilated the starch synthesis pathway into their own genome and today synthesise starch within the cytosol of their cells using the host-derived UDP-Glucose Pyrophosphorylase (UGPase) to initiate polymerisation. Within the Chloroplastida, meanwhile, the genes for starch synthesis are found in the plastidial genome, and starch is produced and stored in the plastid. These plants and algae therefore use the cyanobacteria-derived ADP-Glucose Pyrophosphorylase (AGPase) to initiate the pathway (Plancke et al., 2008; Deschamps et al., 2008c).

Furthermore, within the Glaucophyta and the Rhodophyceae there appears to only be one gene responsible for UDP-Glucose Pyrophosphorylase. However, within the Chloroplastida the AGPase consists of at least two and sometimes as many as six genes, coding for enzyme parts that are divided into large and small subunits (Tetlow, 2006; Ball et al., 2011), and the composition of these subunits may vary even within different tissues of the same plant (Tetlow, 2006). The large subunit seems to have little to do with the catalytic activity of the enzyme, but is important for determining its regulatory properties (Smith, 2008). Meanwhile, mutants of both *Chlamydomonas* and *Arabidopsis* that lack the small subunit are not able to

accumulate starch at all (Deschamps et al., 2008c). The expression and function of the subsequent subunits present in some species remains unknown (Deschamps et al., 2008c).

It is widely thought that as the initial specific step in the starch synthesis pathway, synthesis of NDP-Glucose Pyrophosphorylase is also the rate-limiting step of starch biosynthesis. Much experimental work exists to back this up: for example, in 2004, Sakulsingharoj et al. increased starch production in rice seeds by transforming them with an allosterically insensitive bacterial mutant AGPase (the *E. coli glgC* mutant) (Tetlow, 2006). Quantitative analyses, however, have revealed problems with the theory, as there is often little correlation between the level of expression of the AGPase and the level of starch produced, even after more than two decades of attempts to enhance starch production through the manipulation of these genes (Smith, 2008).

1.11.2 Granule Bound Starch Synthase (GBSS)

Unlike NGPase, Granule Bound Starch Synthase (GBSS) would seem to be of cyanobacterial origin in all cases, as it is found only within the cyanobacteria, the Archaeplastida and some of the Archaeplastida's secondary endosymbionts (Ball et al., 2011). This single common ancestry can also be glimpsed in the fact that, although the Floridean starch Granule Bound Starch Synthase shows a strong preference for the UDP-glucose of the eukaryotic glycogen pathway it combined with, it is still capable of using ADP-glucose (although at efficiencies around six times lower, at least in *Cyanophora paradoxa*) (Plancke et al., 2008). Equally, the Chloroplastidal Granule Bound Starch Synthase has proved to prefer ADP-glucose but is again capable of using either, while the starch-accumulating cyanobacterium Clg1 contains a Granule Bound Starch Synthase enzyme that is far more strongly selective for ADP-glucose (Ball et al., 2011).

Granule Bound Starch Synthase is the sole enzyme within the starch synthesis pathway that is responsible for the biosynthesis of amylose and in most of the starch-accumulating organisms studied to date only one copy of the gene has been found. Where two copies are found (designated *gbssI* and *gbssII*) the amino acid sequence remains very similar (generally higher than 72% identical). In these cases the host tends to be a member of the higher plants and null-mutations studies have confirmed that GBSSI is confined to storage tissues, such as tubers, endosperm tissue and pollen-grains, while GBSSII is used for transitory starch accumulation in non-storage tissues like leaves (Vrinten & Nakamura, 2000). An exception to this pattern is the single-celled green alga *Chlamydomonas reinhardtii*, which has also been found to contain two *gbss* genes, presumably through a duplication event. Deschamps

et al. (2008b) were unable to find corresponding sequences in any other *Chlamydomonas* genomes, and the evolutionary significance of this anomaly remains elusive.

Granule Bound Starch Synthase is the only enzyme to be active within the insoluble amylopectin matrix of a pre-existing starch granule, hence its name. Within this matrix it is protected from the action of hydrosoluble branching enzymes, allowing it to synthesise the low branch-density glucan chain, amylose, which is thought to be dispersed in the amorphous, non-crystalline regions as random-coils or single-helices (Tetlow, 2006). Granule Bound Starch Synthase has also been shown to have some effect on the extension of amylopectin glucans and can therefore affect the storage-organ morphology (Tetlow, 2006). Experiments in *gbssI*-defective mutants of *Chlamydomonas reinhardtii* by Ral et al. (2006) show that, in this organism, Granule Bound Starch Synthase is actually integral to the synthesis of the long glucan fractions that interconnect amylopectin clusters. The authors suggest that it was probably an important factor in the evolutionary transition from the synthesis of glycogen-like polymers to amylopectin-like polymers, but that this primary function has been rendered redundant in most Archaeplastida. An exploration of Granule Bound Starch Synthase might therefore be illuminating from an evolutionary perspective.

In all other members of the Archaeplastida studied to date, an absence of Granule Bound Starch Synthase leads to the absence of amylose but the starch granule is still synthesised, although with a different morphology termed ‘waxy starch’, which is less dense, perhaps with lower storage efficiency, explaining the conservation of Granule Bound Starch Synthase (Zeeman et al., 2010). However, there are some red algae, such as *Cyanidium* species, that lack amylose altogether (Izumo et al., 2011).

Different amylose contents give starch different commercial properties. For example, the waxy starch produced by *gbss* mutants actually has improved paste-clarity, freeze-thaw stability and rapid gelatinisation, making it desirable in processed foods, while starch with a high amylose content has a higher gelatinisation temperature, providing better gel texture and adhesion capacity, thereby making it useful for the production of paper and adhesive products (Santelia & Zeeman, 2011). High amylose starches (referred to as resistant starches) are also less digestible within the small intestine, meaning that a higher proportion is passed into the large intestine where it is fermented by gut bacteria. This fermentation produces short chain fatty acids such as butyrate and acetate that are beneficial for colon health, so these resistant starches have recently become a high-value health food. For less

worthy diets, a high gelling strength also makes high amylose starches useful in the production of boiled sweets (Tetlow, 2006).

Granule Bound Starch Synthase therefore provides an interesting target for manipulation when considering the synthesis of starch for commercial purposes. The accumulation of amylose seems naturally to vary dramatically between species and within individuals under different conditions, from 30 to 60% of the total polysaccharide in *Cyanophora paradoxa* when grown in alternating or continuous light, respectively, to as little as between 10 and 2% in *Chlamydomonas reinhardtii* when starved or supplied with nitrogen, respectively (Plancke et al., 2008), to 0% in *Cyanidium*. It should presumably be malleable in an engineered pathway through the simple up- or down- regulation of the one gene, and indeed Jobling et al. (2002) have already produced a 'waxy' potato through the downregulation of Granule Bound Starch Synthase, as well as SSII and SSIII (see later) so that starch composed of short-chain amylopectin is synthesised.

1.11.3 Soluble Starch Synthase (SS)

The Soluble Starch Synthase (SS) is the most diverse of the starch synthesis enzymes, with four or five different isoforms within the Chloroplastida (SSI, SSII, SSIII, SSIV and sometimes SSV), all of which transfer ADP-glucose to the growing α-1,4-linked glucan (Ball & Morell, 2003). Three of these (SSII, SSIV and SSV) are clearly related to one of the two cyanobacterial synthases, though the relationship is less clear in the other two (SSI and SSIII) (Deschamps et al., 2008c). A marked difference to this is again seen within the Rhodophyceae and Glaucophyta, where there is only one *ss* gene, whose product does the work of all four (or five) Chloroplastidial Soluble Starch Synthase isoforms, although it extends the α-1,4-linked glucan with UDP-glucans rather than ADP-glucose, in a manner analogous to the eukaryotic glycogen synthesis pathway.

The function of the SSV enzyme, when it is present within an organism, has yet to be deduced (Deschamps et al., 2008c). However, simple models have been suggested for the functions of the other four isotopes in the Chlorophyta – supported by biochemical evidence – to the general impression that SSI is responsible for synthesising the shortest of the glucan chains, with ten units or less, whereupon SSII and SSIII take over to make them progressively longer (Tetlow, 2006). Further reports have suggested that SSI might also decorate amylopectin outer chains with glucose residues and possibly terminate synthesis, and that SSIII conditions the synthesis of the long glucan fractions that interconnect clusters of amylopectin chains (sometimes in tandem with Granule Bound Starch Synthase, as in

Chlamydomonas reinhardtii, see above), since the starch granules of mutants defective for SSIII show an increase in amylose content (Deschamps et al., 2008c). Although the exact mechanisms that determine granule size, number and morphology are still unclear (Zeeman et al., 2010), mutations in *ssIV* have suggested that it controls the number of starch granules, at least in higher plants – *Arabadopisis* mutants lacking the gene displayed only one large starch granule per plastid (Zeeman et al., 2010). That this enzyme is missing from the normal starch synthesis pathway of *Ostreococcus tauri* supports this, as this alga only ever has one granule in its plastid. SSIV may also be involved in the priming of the starch granule, or of polysaccharide biosynthesis, or both (Deschamps et al., 2008c). Although sequences in the *Arabidopsis* genome similar to those for the animal glycogen primer glycogenin provide other candidates for the starch primer (Ball & Morell, 2003), evidence for their use in this role is extremely limited (Zeeman et al., 2010). Structural characteristics of SSIV also hint at its role as a primer: the N-terminal half of the protein is largely composed of long coiled-coil domains, which have often been found to act as protein-protein interaction motifs. These motifs are highly conserved in the SSIV of all species studied to date, suggesting an essential nature (D'Hulst & Merida, 2010). However, if SSIV is the starch granule primer it seems odd that such a gene would be lacking from *O. tauri*. With this is mind, it might be illuminating to transform *O. tauri* with *ssIV* and observe the effects. If possible, it might also be worthwhile to remove the starch granule from an *O. tauri* cell, to see if it can grow a fresh one. Ral et al. (2004) note that they have never been able to make *O. tauri* degrade its starch to completion, so perhaps the complicated fission of the starch granule and the partitioning of the polysaccharide into the two daughter cells when *O. tauri* divides (Deschamps et al., 2008c) has removed the need for a primer, as instead a pre-formed granule is always passed on from cell to cell like an Olympic flame.

The Soluble Starch Synthase enzymes together also show a maximum catalytic capacity close to the actual flux through the biosynthesis system, indicative of a rate-limiting-step. This would explain the inconsistencies seen in earlier attempts to increase starch content by upregulating AGPase (see above). Unfortunately, the multiple isoforms of Soluble Starch Synthase have complicated attempts to enhance its activity (Smith, 2008), as it transpires that the relatively simple model whereby different Soluble Starch Synthase enzymes fulfil discrete roles is not the whole story. Instead it seems the enzymes are working synergistically: simultaneous mutations of multiple *ss* genes result in a starch granule phenotype that is often more severe than just the sum of its individual mutations (Table 1.2).

Mutation	**Starch content**	**Amylose content**	**Granule morphology**	**Amylopectin chain length distribution**
ssI	slightly reduced	normal	smaller	more intermediate chains, fewer short ones
ssII	normal	increased	larger	more short chains, fewer intermediate ones
ssIII	slightly increased	normal	normal	minor changes
ssIV	slightly reduced	normal	single granule	minor changes
ssII / *ssIII*	much reduced	increased	larger	many more short chains and fewer longer ones
ssI / *ssIV*	much reduced	not detected (abolished?)	single granule	more intermediate chains and fewer short ones
ssII / *ssIV*	much reduced	not detected (abolished?)	single granule	more short chains, fewer intermediate ones
ssIII / *ssIV*	abolished	abolished	none	None
ssI / *ssII* / *ssIII*	much reduced	increased	smaller	many more short chains, fewer longer ones
ssI / *ssII* / *ssIV*	much reduced	increased	single granule	slightly more intermediate chains and fewer short ones

Table 1.2: Summary of the effects of single, double and triple mutations in different Soluble Starch Synthase isoforms of *Arabidopsis thaliana*. Adapted from Santelia and Zeeman, 2011.

If different Soluble Starch Synthase isoforms are all responsible for extending the linear amylopectin chains, they will jostle over the growing ends, influencing their structure (Santelia & Zeeman, 2011), providing one possible mechanism for this phenomenon. Furthermore, multi-enzyme complexes of Soluble Starch Synthases have been identified and many starch synthesis enzymes are predicted to have domains for protein-protein interactions like those found in SSIV (Koetting et al., 2010), although there is still no direct evidence that complex formation influences either the rate of starch accumulation or the structure of glucans synthesised *in vivo* (Zeeman et al., 2010).

Plants with reduced expression of *ssI* show little (such as in *Arabidopsis*) to no (such as in potatoes) phenotypic change to the starch granule, and other Soluble Starch Synthase isoforms may therefore be capable of partly compensating its function (Tetlow, 2006). Reduction in the expression of *ssI* and *ssIII* seems to only result in the equivalent sum of *ssI* and *ssIII* single mutation phenotypes. However, simultaneous deficiencies in *ssII* and *ssIII* result in a much more severe phenotypic change, indicative of synergistic action between these enzymes. Furthermore, while the reduced expression of either *ssIII* or *ssIV* again results in a relatively mild phenotypic change, a simultaneous deficiency of both the enzymes coded for by both these genes has been shown to abolish starch accumulation altogether (Santelia & Zeeman, 2011) (Table 1.2). Loss of *ssIII* therefore seems to differ strongly depending on the genetic background.

Starch synthase genes have also been shown to influence phosphate levels. Potatoes with reduced expression of *ssII*, for example, show a 50% reduction in phosphate. Meanwhile, potatoes with reduced expression of *ssIII* show a 70% higher starch phosphate level (Tetlow, 2006). The covalently bound-phosphate level within the starch granule is negatively correlated to its crystallinity, and high-phosphate starches have improved viscosity and freeze-thaw stability, as well as being more transparent. Because they are more highly charged, they have also proved useful as surface coatings for paper manufacture (Santelia & Zeeman, 2011).

Overall, *soluble starch synthase* mutations in both vascular plants and green algae indicate that these genes are orthologous within all members of the Chloroplastida. However, a further layer of complication arises from the fact that some species contain more than one copy of some of the Soluble Starch Synthase isoforms. Various species within the algal classes Chlorophyceae and Prasinophyceae have been shown to contain multiple copies of SSIII, and the Chlorophytes *Chlamydomonas* and *Volvox* also contain two SSI-like

sequences, while the Prasinophyte *Micromonas* contains two copies of SSI and of SSII (Deschamps et al., 2008c). Two copies of *ssII* genes (*ssIIa* and *ssIIb*) are also found in monocots. *ssIIa* is expressed mainly within the cereal endosperm, and while it does not contribute much to the total measurable Soluble Starch Synthase activity in this tissue, its disruption results in reduced starch content, reduced amylopectin chain length, altered granule morphology and reduced crystallinity, all of which is far more severe than the effects of disrupting the single SSII enzyme in dicots. SSIIb, meanwhile, is mostly found within photosynthetic tissue and its role is still unknown (Tetlow, 2006). No functional meaning for the duplication found in these species has yet been uncovered and, at either end of the Chloroplastidal phylogenetic spectrum, both *Ostreococcus* and *Arabidopsis* only contain single copies of the *soluble starch synthase* isoform genes (Deschamps et al., 2008c).

In stark contrast to all this complexity, the single Soluble Starch Synthase enzyme found in the Glaucophyta and Rhodophyceae is very well conserved (Ball et al., 2011). Furthermore, whereas the Chloroplastidal starch synthases seem to be of cyanobacterial origin, the Floridean Soluble Starch Synthase represents one of the two major forms of glycogen synthase found in heterotrophic eukaryotes (Ball et al., 2011). After symbiosis, ADP-glucose entering the host cytosol from the cyanobacterial symbiont would have barely impacted on the host's metabolism, as it is neither used nor recognised by eukaryotes. It could therefore only be metabolised by bacterial enzymes that had transferred to the host's genome (Ball et al., 2011). The existence of both UDP and ADP utilising pathways in the Archaeplastida today therefore suggests that an ancestral Archaeplastida used both ADP-glucose and UDP-glucose pathways to generate its cytosolic (Floridean) starch (Dauvillee et al., 2009) (figure 1.16). This would imply that mixing genes from the different pathways might yield some product, although work on the UDP-glucose metabolising Glaucophyte *Cyanophora paradoxa* (Plancke et al., 2008) revealed that, distinct from the seemingly unfussy Granule Bound Starch Synthase, their Soluble Starch Synthases could not use ADP-glucose at significant rates.

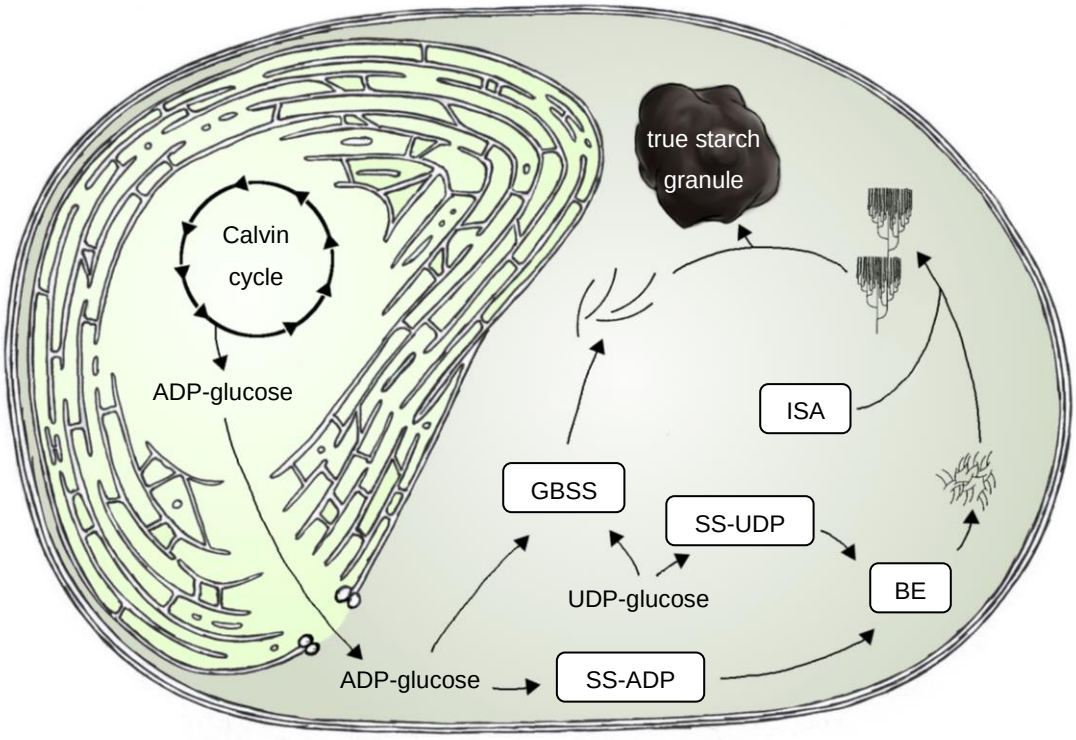


Figure 1.16: Summary of the actions of the combined endosymbiont and host starch biosynthesis enzymes in a hypothetical Archaeplastidal ancestor.

1.11.4 Branching Enzyme (BE)

The branching enzymes hydrolyse some of the α-1,4-linkages of the amylopectin chain and transfer the released reducing ends to α-1,6 positions, generating branches. The selective use of either ADP-glucose or UDP-glucose is not relevant for this enzyme or indeed any of the remaining classes of genes in the pathway. As such, the differentiation between the Chloroplastida and the Rhodophyceae and Glaucophyta does not occur beyond this point and their corresponding genes within these classes are thought to be homologous. Furthermore, there are genes homologous to those for the branching enzymes within the glycogen synthesis pathway of the opisthokonts and bacteria, much as there are for the gene classes of the glucose pyrophosphorylases and soluble starch synthases.

There are two isoforms of this enzyme within the Chloroplastida, designated BEI and BEII, which combine to do the work of the single Branching Enzyme found in the Floridean starch synthesis pathway. Furthermore, within the monocots a second duplication event seems to have occurred and BEII is made up of BEIIa and BEIIb. BEII has also undergone duplication in the Volvocales (Deschamps et al., 2008c). BEI and BEII differ in their substrate specificities and the length of glucan chain they transfer, at least *in vitro*. BEII enzymes show a higher affinity towards amylopectin and transfer relatively short chains, while BEI enzymes show a higher affinity towards amylose and transfer longer chains (Tetlow, 2006). As such, it is only the BEII enzymes that are essential for starch synthesis, and mutants defective for *beII* display similar severe phenotypes across different plant species (Deschamps et al., 2008c). Meanwhile, the disruption of BEI activity displays little phenotypic effect in either monocots or dicots, although when *beI* was repressed in potatoes the phosphate content of the starch produced was clearly affected (Tetlow, 2006).

As with the soluble starch synthases, branching enzymes have been observed to form multi-enzyme complexes (Koetting et al., 2010), and these complexes seem often to be dependent upon the phosphorylation status of the individual enzymes. For example, within the starch biosynthesis pathway of higher plants a protein complex has been shown to form between BEI, BEIIb and Phosphorylase (see later) when they are phosphorylated by Plastidial Protein Kinase (Tetlow, 2006).

1.11.5 Debranching Isoamylase enzyme (Isa)

The direct debranching enzymes of the starch synthesis pathway are thought to catalyse the integral step that turns a glycogen-like polymer into starch. There are two forms of direct debranching enzyme: Isoamylase and Pullulanase, but the Pullulanase is thought to operate chiefly within starch degradation while Isoamylase is crucial to its synthesis. Unlike Pullulanase, Isoamylase (Isa) is unable to digest tightly spaced branches. It is therefore thought to clip off any loosely spaced, misplaced α-1,6 linkages that are randomly formed by the branching enzymes within the precursor of amylopectin at the surface of the growing starch granule which would otherwise prevent crystallisation (Deschamps et al., 2008b). This should lead to ordered amylopectin chain aggregation, since the debranched structure favours the formation of tightly spaced parallel double helices required at the root of clusters for polymer crystallisation (Ball et al., 2011). The surfaces of immature starch granules have been observed to contain many short chains, consistent with this model (Tetlow, 2006).

Within all members of the Archaeplastida, Isoamylase has at least two major isoforms, all of which are of bacterial phylogeny (Ball et al., 2011). These two isoforms, ISAI and ISAII, form a hetero-oligomeric complex in several species, and ISAII in isolation actually lacks the amino acid residues essential to catalytic activity and probably works to modulate the action or stability of ISAI (Tetlow, 2006). The disruption of either isoform is therefore thought to be responsible for the accumulation of phytoglycogen (a hydrosoluble glycogen-like polymer) rather than starch in members of the Archaeplastida (Tetlow, 2006). In several species, from *Arabidopsis* to *Chlamydomonas*, a lack of Isoamylase activity leads to the complete substitution of granular starch with phytoglycogen. *Arabidopsis* mutants lacking either isoform also exhibit identical phenotypes, confirming their co-dependent nature. The debranching of misplaced branches within the hydrosoluble amylopectin precursor therefore seems to be mandatory for the generation of the tight packing of double helical structures (Deschamps et al., 2008b).

However, the loss of granular starch synthesis observed in the absence of *isa* is not simply a consequence of the loss of Isoamylase activity. Instead, the loss of Isoamylase activity leads to a change in the glucan structure, which enables other enzymes not normally involved in starch biosynthesis to metabolise the altered glucans and affect their structure (Santelia & Zeeman, 2011). This has been evident in recombinant *Arabidopsis*: when Isoamylase activity is disrupted they stop forming starch granules, as expected. However, when a chloroplastic α-amylase (AmyIII), which is normally involved in starch degradation, is also disrupted, starch granule biosynthesis is partially restored (Zeeman et al., 2010), suggesting that the

pre-amylopectin structure synthesised by the soluble starch synthases and branching enzymes already has the potential to crystallise if working in isolation, and that Isoamylase is needed to facilitate this crystallisation against a backdrop of enzymes competing for the glucan ends. Evidence from mutants of *Arabidopsis* and other plant species has also implicated isoamylases in the control of the number of starch granules, because such mutants with down-regulated isoamylases contain a relatively higher number of smaller granules in their plastids (Santelia & Zeeman, 2011).

1.12 Starch degradation

Several enzymes have been isolated that are thought to function in the degradation of the starch granule. These can be loosely divided into enzymes involved in phosphorylation, which include Glucan Water Dikinases (GWD), Phosphoglucan Water Dikinase (PWD) and SEX4 phosphatases, and those involved in hydrolysis, which include pullulanases and isoamylases, glucanotransferases, β-amylases, α-amylases, α-glucosidases and phosphorylases.

The dikinases (Glucan Water Dikinase (GWD) and Phosphoglucan Water Dikinase (PWD)) begin degradation. GWD is responsible for C6 phosphorylation of the amylopectin molecules. Once GWD has started this process, PWD is able to phosphorylate the C3 position. By introducing phosphates, these dikinases loosens the tight crystal packaging of glucans locally within the granule and open up the hydrophobic double helical structure of the starch molecule, allowing for the degradation of amylopectin by amylases and debranching enzymes. The phosphate is possibly then released by SEX4 phosphatase. The exact role of the SEX4 enzyme remains controversial, but a mutation of the *gdp*, *wdp* or *sex4 phosphatase* locus in *Arabidopsis* has been shown to cause an excess of starch accumulation due to reduced starch degradation (Santelia & Zeeman, 2011).

The non-reducing end of the amylopectin cluster is degraded by β-amylase and α-glucosidases, primarily releasing maltose units, but also leading to the production of some linear maltodextrins; in particular maltotriose, since β-amylase is incapable of degrading malto-oligosaccharides of fewer than four glycosyl units. These are too small to be metabolised directly by phosphorylases and amylases, and knockout mutation experiments have suggested that they are acted on by glucanotransferases. The exact mechanism for this has yet to be fully unpicked, and furthermore seems to differ between species. In *A. thaliana* it seems likely that two glucanotransferases exist: one plastidial and one cytosolic. The cytosolic form is thought to aid in the metabolism of this maltotriose, possibly by

transferring one maltosyl unit onto another oligosaccharide, thereby producing glucose and also extending that oligosaccharide. The maltose exported into the cytosol, meanwhile, is thought to be metabolised by the plastidial glucanotransferase, which transfers one glucosyl unit to a branched glucan and releases the other as glucose. The branched glucan is then degraded by Phosphorylase, liberating glucose-1-phosphate used to synthesise ADP-glucose by plastidial AGPase. Within *C. reinhardtii*, however, evidence suggests that glucanotransferase has a role during starch synthesis, possibly by reattaching the malto-oligosaccharides trimmed by debranching enzyme back onto the amylopectin, contributing to amylopectin branch length. (Ball et al., 2011; Critchley et al., 2001 & Chia et al., 2004).

So although a role for each of these enzymes in starch degradation has been proposed, many details remain uncertain. Phenotypes of *Arabidopsis* mutants lacking chloroplastic Phosphorylase suggest that its phosphorylases are used not only for starch degradation (Tetlow, 2006), but also in starch synthesis, transfering glucosyl units from glucose 1-phosphate to the non-reducing end of α-1,4-linked glucan chains. Whether they work for degradation or synthesis seems to depend on the relative concentration of maltodextrins in their environment, where the presence of maltodextrins favours degradative reactions. Maltodextrins are also required for amylose synthesis, and their presence stimulates Granule Bound Starch Synthase (Tetlow, 2006). Whether Phosphorylase influences Granule Bound Starch Synthase and amylose synthesis might therefore be an interesting avenue of study.

Pullulanases have also been shown to compensate for the loss of *isaII* to a certain degree in some species (Deschamps et al., 2008c), and *C. reinhardtii* mutants defective for Glucanotransferase demonstrated a severe decrease in starch content, strongly suggesting that in this organism the enzyme has a role in starch synthesis, which is quite different to the function observed in *Arabidopsis* (Tetlow, 2006).

All the components of the starch degradation pathway are either of eukaryotic or unknown phylogeny. Cyanobacteria do not contain equivalents to GWD, PWD, β-Amylase or Glucanotransferase (Ball et al., 2011). The dikinases (GWP and PWD) have in fact only been found in the Archaeplastida, with no homologues in glycogen metabolism, and it seems likely that their evolution is entirely unique to the starch biosynthesis pathway that came about during plastid endosymbiosis (Ball et al., 2011). It seems likely that PWDs are in fact only found within the Chloroplastida (Deschamps et al., 2008c).

1.13 Synthesis of Starch *Ostreococcus tauri*

One of the key aims of this project was to try to achieve the synthesis of a starch-like polymer in a bacterial chassis through the sequential introduction of starch biosynthesis genes. The simplest form of starch biosynthesis pathway can be found within the floridean starch synthesis pathways of the Glaucophyta or Rhodophyceae. Since these pathways seem very straightforward, it might have been possible to simply clone the five or six genes known to be engaged specifically in biosynthesis in these pathways (*UDPase*, *gbss*, *ss*, *be* and probably both *isaI* and *isaII*) and insert them into a bacterial chassis on a plasmid vector to achieve biosynthesis. However, two problems arise with this strategy. Firstly, the genes are not readily available. The full genome of the Rhodophyte *Cyanidioschyzon merolae* has been sequenced, but this alga has proved to be a wildcard, completing starch synthesis and degradation with possibly fewer than 10 genes and, as such, having the most basic pathway analysed so far; more basic even than that of the subgroup V cyanobacteria Clg1. *C. merolae* also thrives in temperatures so high they would melt typical Rhodophyceae starch granules (Deschamps et al., 2008c). Therefore, although this makes the *C. merolae* pathway an intriguing subject in terms of industrial application, it also makes it a poor candidate for helping to unravel the fundamental nature of the starch biosynthesis pathway.

Of the Glaucophyta, genes for starch biosynthesis are currently being identified in the novel organism *Cyanophora paradoxa*. However, at the time this project was started, these were limited to the complete gene sequence for *gbss* and the partial sequence for *ss* (Plancke et al., 2008), so extensive further analysis would be required to obtain the complete pathway. Glaucophyta are also notoriously rare, so the acquisition of any living organisms would also have posed a problem.

Even if these hurdles were overcome, another significant issue with the use of floridean starch synthesis genes within a gram negative bacterial chassis is their use of UDP-glucose as a substrate. The nucleotide of NDP-glucose in the glycogen pathways of all gram-negative glycogen-accumulating bacteria is Adenine, and UDP-glucose would be neither produced nor recognised by the bacteria. Although it may be interesting to discover whether such a pathway would still be functional, it may not be the most appropriate first step in this analysis.

The starch synthesis genes for the Chloroplastida, meanwhile, are relatively well conserved between species and examples of each enzyme have been sequenced. Many have also already been used for genetic recombination, such as in 2008 when the University of

Edinburgh iGEM team (Haseloff, 2007) designed a model starch production pathway for *Escherichia coli* using *Zea mays* genes for *isaI* and *isaII*. Additionally, the Chloroplastidal pathway uses ADP-glucose as its substrate, as does the glycogen synthesis pathway of gram negative bacteria, so the enzymes should have something to act on. Indeed, the glycogen synthesis pathway of gram negative bacteria already contains genes that are analogous to the first three genes of the starch synthesis pathway. Bacterial *glgC*, *glgA* and *glgB* code for an ADP-Glucose Pyrophosphorylase, Glycogen Synthase and Glycogen Branching Enzyme, respectively (Ball & Morell, 2003). Since the Isoamylase enzymes of the starch pathway act on a glycogen-like amylopectin precursor, it was considered that they might equally work to crystallise the bacterial glycogen molecule, thereby forming 'waxy starch' (since 'true starch' could only be formed through the addition of amylose molecules that require a Granule Bound Starch Synthase enzyme for their synthesis). These *isa* genes, along with an upregulated *AGPase* gene, were the only additions made to the genome of *E. coli* by the University of Edinburgh iGEM team. The expression of waxy starch was not confirmed, however, and previous down-regulation experiments of various starch metabolism genes including those for the *isa* genes have suggested that just their addition to a glycogen synthesis pathway might not be enough to form starch granules (see above). Even so, it was considered a worthwhile experiment to augment the *E. coli* already transformed with *isaI* and *isaII* with a *gbss* gene, as the 'true' starch that, potentially, this organism might produce would be easier to identify.

The use of a Chloroplastidal starch metabolism pathway as the toolbox from which to construct a minimal pathway is not without its own problems, chiefly from the relative complexity of the pathway and the subsequent functional overlaps that exist among duplicate copies of genes. As previously mentioned, it seems very likely that complexes form between these copies, and this has so far confused the interpretation of single-mutant phenotypes (Deschamps et al., 2008b). In addition to these physical complexes, it also seems that the enzymes form disparate but functional assemblies that may also be integral to the synthesis of starch granules, since the product of one enzymic reaction often becomes the substrate for a different enzyme elsewhere in the pathway. It also seems likely that the degree to which each enzyme controls flux will change during the growth of the starch granule, which may affect the structure of the polymer as well as the amount of starch made (Tetlow, 2006). The way these different enzyme functions 'roll around' one another in the formation of the growing starch granule will be difficult to analyse, since the connections will not be as straightforward as physical biochemical connections.

A further complication comes from the fact that there are at least four enzymes within the glycogen degradation pathway of bacteria that again have analogues within the starch degradation pathway. Glucanotransferase, which in plants and algae transfers the maltodextrins produced by amylopectin hydrolysis to an acceptor glucan where they can be further metabolised by Phosphorylase, is a direct homologue of the bacterial Glucanotransferase enzyme. Equally, two phosphorylases are present in glycogen metabolism: a glycogen phosphorylase (GlgP) and a maltodextrin phosphorylase (MalP). The phosphorylation of transferred maltodextrins feeds back into ADP-glucose synthesis by AGPase (Ball et al., 2011), so the activity of these bacterial enzymes will likely have an effect on the rate of starch synthesis. Also, since maltodextrins are required for the synthesis of amylose – and as such their concentration directly influences the action of Granule Bound Starch Synthase – it seems likely that the glycogen degrading enzymes will complicate attempts to control this mechanism. A chance also exists that the glycogen phosphorylases could compete with the recombinant genes in the formation of the starch granule (Tetlow, 2006) (see above). The bacterial glycogen pathway also contains a glycogen debranching enzyme (GlgX). However, since this enzyme has a very specific activity, hydrolysing only exterior glucan chain branches of four residues in length that have been digested by glycogen phosphorylase (Ball & Morell, 2003), it seems unlikely that it will interfere significantly with starch synthesis.

Lastly, the genes in most common use are currently taken from *Arabidopsis* or common crop plants such as *Zea mays* (maize), *Triricum aestivum* (wheat), *Oryza sativa* (rice) or *Solanum tuberosum* (potato), and in these higher plants the pathway is complicated by the fact that copies exist of many of the isoforms, and also by the fact that the different enzymes often seem to have different functions within different tissues. It therefore seems advisable to use an alga to provide the basic model. The unicellular green alga *Ostreococcus tauri* (figure 1.17) is an attractive organism for study due to its simplicity and the compact nature of its genome, which is comparable to that of a bacterium at 12.5 Mbp, with roughly 80% coding DNA, containing about 8,000 genes. *O. tauri* diverged from the Chloroplastidal common ancestor very early on, and has since undergone significant genome simplification, yet it does still accumulate its single starch granule using the same pathway as higher plants, with the exception of a lack of *ssIV* (see above). At 0.8 μm in diameter it is the smallest free-living eukaryote identified to date (Courties et al., 1994; Koetting et al., 2010), with the smallest genome of any photosynthesising eukaryote and the simplest starch metabolism pathway so far identified in the Chloroplastida (Derelle et al., 2006). Its full genome was

published in 2006 (Derelle et al., 2006) (strain OTH95, isolated from high-light, surface waters (Cardol et al., 2007)), including the complete sequence of its chloroplast (Robbens et al., 2007) and living organisms are already grown in laboratories within the University of Edinburgh. Currently available on the NCBI (National Centre for Biotechnology Information) database are several sequences of starch synthesis genes (Table 1.3), with work ongoing to identify the remainder. The use of *O. tauri* gene sequences was therefore considered a viable option for the first step in exploring the minimum number of genes needed for the pathway and the synergy that exists between those genes.

From the work carried out to date on the disruption of starch synthesis genes in plants and algae, it seems possible that a functional Chloroplastidal pathway might be constructed from as few as eight genes (*ADPase*, *gbss*, *ssII*, *ssIV*, *beII*, *beIII*, *isaI* and *isaII*) (Table 1.4). This correlates with the set of genes suggested for the Archaplastidal common ancestor by Ball et al. (2010). By taking advantage of the glycogen synthesis enzymes already present within an *E. coli* host, it is conceivable that this could be reduced to three additional genes (*gbss*, *isaI* and *isaII*). This was therefore deemed a sensible first step, and once completed the addition of subsequent starch synthesis and degradation genes could be carried out.

Previous work in this area, carried out in the French lab at the University of Edinburgh, used the standard high copy BioBrick assembly plasmid pSB1A2 with a lac promoter and lacZ'-α minigene (http://parts.igem.org/Part:BBa_J33207) to express *isaI* and *isaII* transgenes taken from *Zea mays*, along with an upregulated variant of the *E. coli glgC* gene designated *glgC16* (Govons et al., 1973) to create the plasmid pCF-isaI-isaII. This existing plasmid was used as a starting point for this work (see chapter 5) although it was soon replaced with new constructs expressing transgenes from *O. tauri*, which in all cases used the newer standard high copy BioBrick assembly plasmid pSB1C3, in line with current standards of the Registry of Standard Biological Parts. The same lac promoter and lacZ'-α minigene Biobrick (BBa_J33207) was used for expression and screening, which contains the entire lac promoter plus the CAP binding site involved in regulation by different carbon sources such as glucose. This choice of promoter was maintained in order to provide more comparable results between the effects of *Z. mays* and *O. tauri* transgenes on the *E. coli* polysaccharide stores, as well as for the simple reasons that it was readily available and easily inducible. The promoter is known to be repressed by the presence of glucose (see chapter 3). However, it was felt that this obstacle was easily overcome by growth in lactose, which was hoped to have the additional benefit of increasing polysaccharide stores (see Discussion).

Figure 1.17: Diagram of Ostreococcus tauri, showing the nucleus (n) a single chloroplast (c) containing a single starch granule (sg), a single mitochondrion (m) and a single golgi body (g).

If a minimal pathway was unravelled, analysis of the enzyme function of homologue isoforms from different species, and the multiple copies of each isoform found therein, could be carried out. Not only would this be fascinating in its own right, hopefully helping to explain the success of the Chloroplastidal starch metabolism pathway which, on the surface, seems so needlessly complicated, but it may also have a practical application in the development of synthetic pathways for different types of starch. Currently, starches are often modified chemically before use, so the development of a transgenic bacterial host that could generate tailor-made starches would eliminate the need for this process, saving money and reducing environmentally hazardous waste material (Tetlow, 2006). Also, although the dense, semi-crystalline nature of starch granules helps in their extraction from plant tissues, it does require a lot of energy to disrupt, usually by cooking in water to around 60°C, before enzymic digestion and fermentation can occur (Smith, 2008). Starch degradation technology has recently made developments in the conversion of starch granules to glucose without this need for initial gelatinisation (Smith, 2008), but production of a starch that required a lower energy input for gelatinisation (such as low amylose and low phosphate starches or smaller granule sizes) may still reduce the economic and carbon costs of bioethanol production (Smith, 2008).

Finally, a successfully created starch synthesis pathway within *E. coli* could later be transferred to an organism such as the yeast strain *Saccharomyces cerevisiae* (which is extremely well characterised and already used in the large-scale production of products such as alcoholic drinks) or *Pichia pastoris* (which, although less well characterised, is still fully sequenced and can reach extremely high growth densities in very simple conditions) since these can metabolise at much higher levels and may also be able to grow larger granules due to their larger cell size.

Ostreococcus tauri	***Chlamydomonas reinhardtii***	***Arabidopsis thaliana***
ADPase (large unit & small unit)	ADPase (large unit & small unit)	ADPaseI, ADPaseII (small) ADPaseI, ADPaseII, ADPaseIII (large)
GBSSI	GBSSI, GBSSII	GBSS
SSI, SSII, SSIII, SSIII-B, SSIII-C	SSI, SSII, SSIII, SSIV, SSV	SSI, SSII, SSIII, SSIV
BEI, BEII	BEI, BEII, BEIII	BEII-A, BEII-B
ISAI, ISAII (probably)	ISAI, ISAII	ISAI, ISAII, ISAIII

Table 1.3: Gene sequences currently available from the NCBI database for three starch-producing model organisms. The *O. tauri isaI* and *isaII* genes are rather confusingly named 'DBE 2 isoamylase' and 'Ot12g00310 DBEII', respectively, on the NCBI database.

Enzyme type	Cyano-bacteria	Glauco-phytes	Rhodo-phyceae	Chloroplas-tida	Possible Minimum
NDP-glucose pyrophosphorylase (NDPase)	1 (ADP-glucose)	1 (UDP-glucose)	1 (UDP-glucose)	2 to 6, of large and small subunits (ADP-glucose)	1 (UDP-glucose); 1 or 2 (ADP-glucose, already present in gram negative bacteria)
Granule Bound Starch Synthase (GBSS)	1	1	1	1 or 2	1 (UDP-glucose); 1 (ADP-glucose: GBSSI)
Soluble Starch Synthase (SS)	2	1	1	3 to 5, with multiple copies of isoforms in some species.	1 (UDP-glucose); 2 (ADP-glucose: SSII and SSIV)
Branching Enzyme (BE)		1	1	2, with copies in some species	2 (BEI and BEII)
Debranching enzyme (Isa)	3	?	2	3 or 4	2 (ISAI and ISAII)
Glucan Water Dikinase (GWD)	2	1	1	1, with 2 to 4 copies	1
Phosphoglucan Water Dikinase (PWD)	0	0	0 or 1?	1, with 1 to 3 copies	1
Sex4-type phosphatases	0	1	1	1	?
Pullulanase	1	?	?	1	1
Glucanotransferase	0	1	0	1 or 2	1
β-amylases	0	1	1	1, with 2 to 9 copies	1
α-amylases	0	?	?	1, with 2 to 6 copies	?
α-glucosidases	0	?	?	1	1
Phosphorylases	1	multiple	1	1, with 2 or 3 copies/species	1 (already present in gram negative bacteria)

Table 1.4: Summary of the number of isoforms found for each class of starch metabolism enzyme in starch-producing organisms, as well as the suggested number for a minimum pathway. Starch degradation is less well understood than synthesis, but probably requires a copy of each enzyme. The function of Sex4-type phosphatases is largely unknown, and the presence of α-amylase may not be strictly necessary if β-amylase is already present, accounting for the '?' for these enzymes in the minimum pathway. If the goal is only to synthesise starch, no degradation enzymes should be necessary for a minimum pathway. Adapted from Ball (2011) and Deschamps et al. (2008b).

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Growth media

Medium	Composition
Lysogeny Broth (LB)	1% Bacto-tryptone, 0.5% Yeast Extract, 1% NaCl; pH adjusted to 7.5 with NaOH
LB Agar	1% Bacto-tryptone, 0.5% Yeast Extract, 1% NaCl, 1.5% Bacto-agar; pH adjusted to 7.5 with NaOH
Terrific Broth	1.2 % Bacto-tryptone, 2.4 % Yeast Extract, 0.4 % (v/v) glycerol, 72 mM K_2HPO_4, 17 mM KH_2PO_4
Kornberg Medium	3 % Yeast Extract, 2% Lactose (changed from glucose found in the original recipe), 63 mM K_2HPO_4, 62 mM KH_2PO_4
M9 Salts (4×)	28 g Na_2HPO_4, 12 g KH_2PO_4, 2 g NaCl, 4 g NH_4Cl; made up to 1 litre with double-distilled sterile water
M9 Medium	1 × M9 salts, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 1% lactose, 0.05% yeast extract
Artificial Seawater	8.83×10^{-4} M $NaNO_3$, 3.63×10^{-5} M NH_4Cl, 1×10^{-5} M β-glycerolphosphate, 1×10^{-8} M H_2SeO_3, 1×10^{-3} M Tris-base (pH 7.2), 1×10^{-4} M $Na_2EDTA \cdot 2H_20$, 1×10^{-5} M $FeCl_3 \cdot 6H_20$, 1×10^{-8} M $Na_2MoO_4 \cdot 2H_20$, 1×10^{-9} M $ZnSO_4 \cdot 7H_20$, 1×10^{-9} M $CoCl_2 \cdot 6H_20$, 1×10^{-8} M $MnCl_2 \cdot 4H_20$, 1×10^{-8} M $CuSO_4 \cdot 5H_20$, 1×10^{-10} M Vitamin B_{12}, 1×10^{-9} M Biotin, 1×10^{-7} M Thiamine·HCl and sea salts dissolved in deionised water to 30 ppt
Phosphate Buffered Saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH adjusted to 7.4 with HCl

Table 2.1: Bacterial growth media

Concentrations indicated as percentages are weight/volume (w/v) unless otherwise stated. All media was autoclaved before use.

2.1.2 Electroporation solution

Transport and Storage Solution (TSS): 70.8% (v/v) LB, 20.8% (v/v) 40% PEG 3350 (w/v), 4.16% (v/v) DMSO and 0.0417 M $MgCl_2$ (make 24 ml by mixing 17 ml LB, 5 ml 40% PEG 3350 (w/v), 1 ml 1 M $MgCl_2$ and 1 ml DMSO).

All ingredients except DMSO were autoclaved prior to mixing. DMSO was added by filter sterilisation. TSS was then stored at 4 °C.

2.1.3 Antibiotics

Antibiotic	Abbreviation	Solvent	Stock Conc^n	Final Conc^n
Chloramphenicol	Cm	100% ethanol	40 mg ml^{-1}	40 µg ml^{-1}
Carbenicillin	Cb	Water	80 mg ml^{-1}	80 µg ml^{-1}
Kanamycin	Km	Water	50 mg ml^{-1}	50 µg ml^{-1}
Tetracycline	Tc	70% ethanol	10 mg ml^{-1}	10 µg ml^{-1}

Table 2.2: Antibiotics

Stock and working concentrations of the antibiotics used in this study. All solutions were stored at -20 °C.

2.1.4 Induction and selection reagents

Reagent	Abbreviation	Solvent	Stock Conc^n	Final Conc^n
Isopropyl β-D-1-thiogalactopyranoside	IPTG	Water	90 mg ml^{-1}	90 µg ml^{-1}
bromochloroindoxyl galactoside	X-gal	100% DMSO	40 mg ml^{-1}	40 µg ml^{-1}

Table 2.3: Induction and selection reagents

Stock and working concentrations of the induction and selection reagents used in this study. All solutions were stored at -20 °C.

2.1.5 Enzymes

All restriction enzymes were purchased from New England Biolabs (NEB) and used according to the manufacturer's instructions.

2.1.6 Buffers for *O. tauri* genomic DNA extraction

Extraction Buffer: 0.25 M NaCl, 0.2 M Tris (pH 7.5), 0.025 M EDTA, 0.05 % (v/v) SDS.

2.1.7 Buffers and solutions for Agarose Gel Electrophoresis

TAE Buffer (20×): 0.8 M Tris base, 0.02 M Sodium-EDTA (pH 8.0), 22.8 ml glacial acetic acid, made up to 1 litre with deionised water. TAE was used at a working concentration of 0.5×.

2.1.8 Buffers and Solutions for SDS-PAGE

SDS Running Buffer (10×): 15.15 g Tris base, 72 g Glycine, 5 g SDS, 500 ml H_2O.

SDS Stacking Gel Buffer: 7.87 g Tris HCl, 8 g Tris base, 100 ml H_2O

SDS Loading Buffer (2×): 0.25 ml Stacking gel buffer, 0.4 ml SDS (10% v/v), .02 ml Glycerol, 0.1 ml Bromophenol Blue (2% v/v), 0.1 ml DTT (2M)

Destain solution: 450 ml H_2O, 450 ml Ethanol, 100 ml Glacial acetic acid

Staining Solution: 0.25 g Coomassie Blue. 250 ml Destain solution

2.1.9 Buffers and solutions for Polymerase Chain Reaction

GoTaq®: 5× GoTaq® Flexi Buffer, 25 mM $MgCl_2$ and GoTaq® DNA polymerase (5 µg/µl stock) were purchased from Promega

Phusion®: 5× Phusion® High-Fidelity Buffer (containing 7.5 mM $MgCl_2$), Phusion® polymerase (2 µg/µl stock) and 100% DMSO were purchased from New England Biolabs (NEB).

dNTPs: 100 mM stocks of each Deoxyribonucleotide triphosphate was purchased from Promega. These were mixed to a single 10 mM working solution.

2.1.10 Assay solutions

5% (w/v) Lugol's Iodine solution: 5 % (w/v) Iodine (I_2) and 10 % (w/v) Pottasium Iodide (KI) in deionised water. Total iodine content: 130mg/ml

Anthrone reagent: 20 g litre^{-1} Anthrone in H_2SO_4

Z-buffer for β-glactosidase assay: 0.06 M Sodium Hydrogen Phosphate (Na_2HPO_4), 0.04 M Sodium Dihydrogen Phosphate (NaH_2PO_4), 0.01 M Potassium Chloride (KCl), 0.001 M Magnesium Chloride Heptahydrate ($MgCl_2(H_2O)_7$).
Before assay add 10 mg ml^{-1} lysozyme and SDS up to 0.0025% (v/v).

ONPG solution for β-glactosidase assay: Per litre: 305 ml 0.2 M Sodium Hydrogen Phosphate (Na_2HPO_4), 195 ml 0.2M Sodium Dihydrogen Phosphate (NaH_2PO_4), 500 ml H_2O, 4 g o-nitrophenyl-β-D-galactopyranoside (ONPG) (4 mg ml^{-1})

2.2 Microbiology Methods

2.2.1 Storage of bacteria

The various strains of *E. coli* used in this study were stored at -80 °C in 20% (v/v) glycerol by pelleting cells from stationary phase liquid culture and re-suspending them in a 1:4 mixture of glycerol and LB (autoclaved prior to use). Bacteria were stored in the short term at 4 °C as colonies on inverted agar plates.

2.2.2 Chemical transformation

A culture of the strain to be transformed was grown overnight in LB at 37 °C and 200 rpm. The culture was then diluted 1:100 and grown to exponential growth phase, indicated by an optical density of 0.2 - 0.5 at λ600 nm ($OD_{\lambda 600nm}$). The culture was then divided into 1 ml aliquots in microcentrifuge tubes and spun down in a tabletop centrifuge at around 6000 ×g for 3 minutes. The supernatant was discarded and each cell pellet resuspended in 100 µl TSS. The aliquots were left on ice for approximately 30 minutes. At this point the cells could be stored at -80 °C or used immediately for transformations.

For transformations, DNA (5 - 10 µl of ligation DNA or 2µl of supercoiled plasmid DNA) was added to thawed cell aliquots and they were left on ice for 30 - 60 minutes. After this time, they were heat-shocked at 42 °C for 90 seconds and then placed back on ice for 90 seconds. To each cell aliquot was then added 900 µl LB and the cells were left to recover at 37 °C (or 30 °C for temperature-sensitive plasmids) for 30 to 90 minutes, depending on the

antibiotic resistance they were expected to express. Finally, 100 µl of the cell suspension was spread onto selective media and incubated overnight at the appropriate temperature. If for some reason there was concern over the success of a transformation, the remaining 900 µl of each cell aliquot was centrifuged at approximately 6000 ×g for 3 minutes, 750 µl of the supernatant was removed and the cells re-suspended in the remaining 150 µl. From this more concentrated cell suspension, 100 µl was spread onto selective media and incubated overnight as well.

2.2.3 Growth to maximise polysaccharide content

Cultures were grown overnight with appropriate selection in LB at 37 °C and 200 rpm. Overnight cultures were then used to inoculate fresh Kornberg Media or Terrific Broth cultures, with appropriate selection, at a 1:100 ratio. The freshly inoculated cultures were then grown at 37 °C and 200 rpm to a 0.5 – 0.6 $OD_{\lambda 600nm}$. Cultures were then supplemented with IPTG to 1 mM and incubated for a further 20 hours at 22 °C and 200 rpm.

The cultures were then spun down at approximately 3000 ×g for 15 minutes, the supernatant was discarded and the pellet washed with M9 ×1 solution. The cell pellets were then resuspended in M9 ×1 solution with 2 % (w/v) lactose, IPTG and appropriate selection and incubated for a further 4 hours at 37 °C and 200 rpm (Sundberg et al., 2013). Cultures could then be spun down at 3000 ×g for 15 minutes, washed twice in PBS and equalised for optical density at λ600nm, prior to analysis.

2.2.4 Growth Curves

Three overnight cultures of LB chloramphenicol were set up for each transformant under investigation, each inoculated with a different colony from a transformation plate. They were incubated overnight at 37 °C and 200 rpm, then $OD_{\lambda 600nm}$ equalised to 3.0 (± 0.09). From each overnight, 100 µl was used to inoculate four fresh 10 ml cultures of different media: LB + lactose (2% w/v) + IPTG + chloramphenicol; Kornberg media + lactose (2% w/v) + IPTG + chloramphenicol; Terrific Broth + lactose (2% w/v) + IPTG + chloramphenicol and Terrific Broth + chloramphenicol. The cultures were shaken to mix before 100 µl aliquots were transferred from each fresh culture into the wells of a ‘Costar 3628’ flat-bottom 96-well plate in the pattern shown below. The remaining wells surrounding the culture samples were then filled with 100 µl of sterile media, as shown. In the first instance, 100 µl was transferred from each culture to a 96 well plate twice, using two separate plates that were run simultaneously. In the second instance, only one aliquot was transferred from each culture, so that a single plate was used (figures 2.1 & 2.2). The 96-well plates were incubated in a

Tecan 'Sunrise Microplate Absorbance Reader 30041769' at 37 °C and 'normal' shaking (4.4 mm, 9.2 Hz). Optical density readings were taken every 15 minutes for 24 hours using the 'accuracy' measurement mode. At the same time, the remaining 9.98 ml cultures were incubated at 37 °C and 200rpm. Optical density readings were taken from these cultures every two hours for the first 14 hours, then at 24 hours. Each optical density reading was obtained by mixing 100 µl of culture with 900 µl of sterile media in a standard cuvette, then measuring the $OD_{\lambda 600nm}$ against 1 ml sterile media on a Modulus 'Single Tube Multimode Reader' (Turner Biosystems BS040271) fitted with Absorbance Module E6076.

2.2.5 Microscopy

Light Microscopy: For each culture, 10 µl was heat-fixed onto a microscope slight. The heat-fixed smear was then soaked in 20 µl of 5% Lugol's iodine, before rinsing with an excess of 100% Ethanol. Slides were viewed with a Nikon eclipse E200 tabletop microscope under phase contrast at 1000× magnification. Light microscopy images were obtained using either a Leitz Wetzlar Metallic II microscope fitted with a Lamamatsu digital camera C4742-95-12NRB (black and white images) or a Canon 1XUS9501S digital camera directly through the microscope eyepiece (colour images). All images are considered typical of the cultures.

Transmission Electron Microscopy: For each strain, two 50 ml cultures were grown overnight with IPTG induction and appropriate selection in LB with the addition of 1% (w/v) lactose at 37 °C and 200 rpm. Entire cultures were then spun down at 6000 ×g for 10 minutes. The supernatant was discarded and the pellet washed twice in 25 ml PBS. The cell suspension was then spun down again at 6000 ×g for 10 minutes, the supernatant discarded and the pellet finally re-suspended in 5 ml PBS. From each culture, 1 ml was then transferred to a microcentrifuge tube. These were centrifuged at 6000 ×g for 10 minutes and fixed in 3% glutaraldehyde in 0.1M Sodium Cacodylate buffer, pH 7.3, for 2 hours, then washed in three 10 minute changes of 0.1M Sodium Cacodylate. Specimens were then post-fixed in 1% Osmium Tetroxide in 0.1M Sodium Cacodylate for 45 minutes, then washed in three 10 minute changes of 0.1M Sodium Cacodylate buffer. These samples were then dehydrated in 50%, 70%, 90% and 100% normal grade acetones for 10 minutes each, then for a further two 10-minute changes in analar acetone. Samples were then embedded in Araldite resin. Sections, 1µm thick were cut on a Reichert OMU4 ultramicrotome, stained with Toluidine Blue, and viewed in a light microscope to select suitable areas for investigation. Ultrathin sections, 60nm thick were cut from selected areas, stained in Uranyl Acetate and Lead Citrate then viewed in a Philips CM120 Transmission electron microscope. Images were taken on a Gatan Orius CCD camera.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg	Kornberg
B	Kornberg	lacZ 1 Kornberg	glgB 1 TB w/o lactose	glgC 1 Kornberg	lacZ 1 TB	glgC 1 LB	glgCB 1 TB	lacZ 2 TB w/o lactose	glgC 2 Kornberg	Kornberg	Kornberg	Kornberg
C	Kornberg	lacZ 1 TB	glgB 1 LB	glgCB 1 TB	lacZ 1 TB w/o lactose	glgC 1 Kornberg	glgCB 1 TB w/o lactose	lacZ 2 LB	glgC 2 TB	Kornberg	Kornberg	Kornberg
D	Kornberg	lacZ 1 TB w/o lactose	glgC 1 Kornberg	glgCB 1 TB w/o lactose	glgB 1 LB	glgC 1 TB	glgB 2 TB w/o lactose	lacZ 2 Kornberg	glgCB 2 TB w/o lactose	Kornberg	Kornberg	Kornberg
E	LB	LacZ 1 LB	glgC 1 TB	glgCB 1 LB	glgB 1 Kornberg	glgC 1 TB w/o lactose	glgB 2 LB	lacZ 2 TB	glgCB 2 LB	LB	LB	LB
F	LB	glgB 1 Kornberg	glgC 1 TB w/o lactose	lacZ 1 LB	glgB 1 TB	glgCB 1 LB	glgB 2 Kornberg	glgC 2 TB w/o lactose	glgCB 2 Kornberg	LB	LB	LB
G	LB	glgB 1 TB	glgC 1 LB	lacZ 1 Korn	glgB 1 TB w/o lactose	glgCB 1 Kornberg	glgB 2 TB	glgC 2 LB	glgCB 2 TB	LB	LB	LB
H	LB	LB	LB	LB	LB	LB	LB	LB	LB	LB	LB	LB

	1	2	3	4	5	6	7	8	9	10	11	12
A	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB
B	TB	lacZ 2 Kornberg	glgB 2 TB w/o lactose	glgC 2 Kornberg	lacZ 3 TB	glgC 3 LB	glgCB 3 TB	glgB 3 TB w/o lactose	glgC 3 Kornberg	TB	TB	TB
C	TB	lacZ 2 TB	glgB 2 LB	glgC 2 TB	lacZ 3 LB	glgC 3 Kornberg	glgCB 3 TB w/o lactose	glgB 3 LB	glgC 3 TB	TB	TB	TB
D	TB	lacZ 2 TB w/o lactose	glgC 2 Kornberg	glgCB 2 TB w/o lactose	glgB 3 LB	glgC 3 TB	lacZ 3 TB w/o lactose	glgB 3 Kornberg	glgCB 3 TB w/o lactose	TB	TB	TB
E	TB w/o lactose	lacZ 2 LB	glgC 2 TB	glgCB 2 LB	glgB 3 Kornberg	glgC 3 TB w/o lactose	lacZ 3 LB	glgB 3 TB	glgCB 3 LB	TB w/o lactose	TB w/o lactose	TB w/o lactose
F	TB w/o lactose	glgB 2 Kornberg	glgC 2 TB w/o lactose	lacZ 3 LB	glgB 3 TB	glgCB 3 LB	lacZ 3 Kornberg	glgC 3 TB w/o lactose	glgCB 3 Kornberg	TB w/o lactose	TB w/o lactose	TB w/o lactose
G	TB w/o lactose	glgB 2 TB	glgC 2 LB	lacZ 3 Kornberg	glgB 3 TB w/o lactose	glgCB 3 Kornberg	lacZ 3 TB	glgC 3 LB	glgCB 3 TB	TB w/o lactose	TB w/o lactose	TB w/o lactose
H	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose	TB w/o lactose

Figure 2.1: 96-Well Plate Layout, First Set

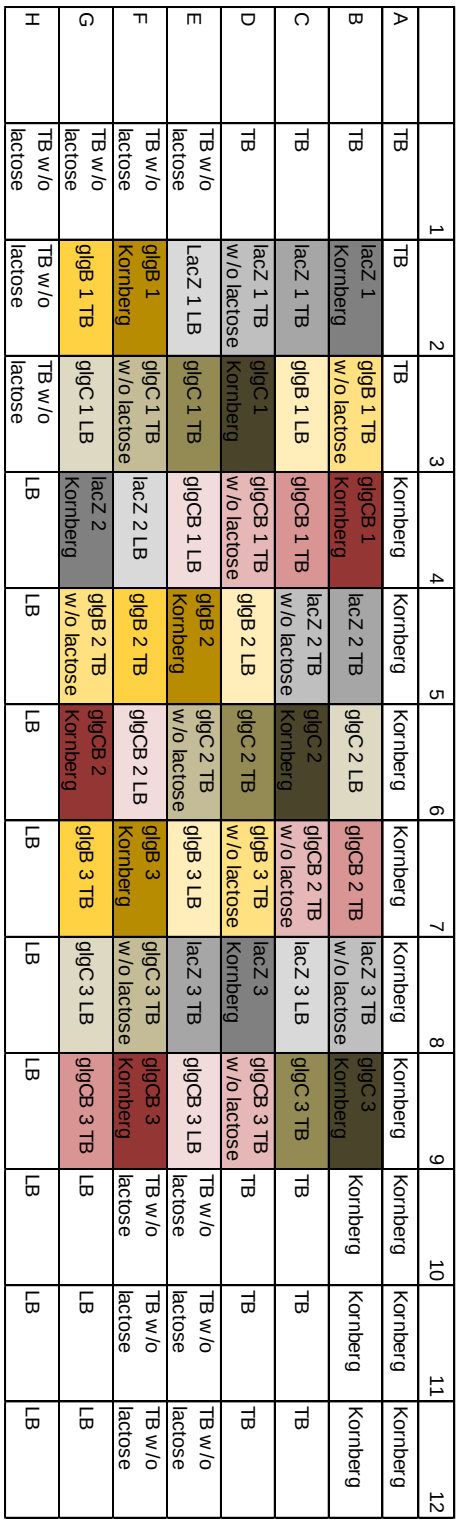


Plate Layout Key:

lacZ	*E. coli* JM109/pJW-lacZ
glgC	*E. coli* JM109/pJW-glgC
glgB	*E. coli* JM109/pJW-glgB
glgCB	*E. coli* JM109/pJW-glgCB
Kornberg	Kornberg Medium + lactose (2% w/v) + IPTG + chloramphenicol
TB	Terrific Broth + lactose (2% w/v) + IPTG + chloramphenicol
TB w/o lactose	Terrific Broth + chloramphenicol
LB	LB + lactose (2% w/v) + IPTG + chloramphenicol

Figure 2.2: 96-Well Plate Layout, Second Set

2.3 Polysaccharide and Enzyme Assays

2.3.1 Iodine

Unless otherwise stated, cultures were grown with IPTG induction and appropriate selection in LB with the addition of 1-2% (w/v) lactose at 37 °C and 200 rpm. For each culture, 1 ml aliquots were spun down in a tabletop centrifuge at approximately 6000 ×g for 3 minutes. The supernatant was discarded and the pellet washed in 1 ml PBS. The cell suspension was spun down again at 6000 ×g for 3 minutes and the supernatant discarded. The remaining cell pellet was then re-suspended in 200 µl of 0.2 % Lugol's iodine solution and the subsequent colour change was observed.

Alternatively, cell pellets were resuspended in 200 µl PBS, to which was added 50 µl $CuSO_4$ (100mM), 50 µl H_2O_2 (6% v/v) and 25 µl Lugol's iodine (0.2%). Suspensions could then be transferred to a 48-well plate for imaging with a flatbed scanner.

Agar plates growing colonies of the appropriate strain were flooded with 0.2% Lugol's iodine solution to test for the presence of starch.

2.3.2 β-Galactosidase

Cultures were grown in the appropriate media at 37°C and 200rpm to an OD_{600nm} of 0.28 – 0.7 (exponential phase) and the OD_{600nm} of each culture was recorder. From each culture, 1 ml was transferred to a microcentrifuge tube, pelleted and resuspended in 800 µl of Z-buffer. The suspensions were incubated at 37°C for 30 minutes. To each suspension was then added 200 µl ONPG solution. The suspensions were incubated at room temperature until a colour change had developed (between a just-visible yellowing and the colour of LB broth). The time taken for this colour change to develop was noted, and the reaction stopped with the addition of 500 µl Sodium Carbonate (Na_2CO_3) (1M). Each suspension was then centrifuged at full speed for 5 – 10 minutes and the top 1 ml of liquid transferred to a cuvette. The $OD_{\lambda 420nm}$ of each sample was then measured against a blank of Z-buffer, ONPG solution and Sodium Carbonate (which had also been spun down to remove any turbidity).

The activity (nmoles o-nitrophenol/min/ $OD_{\lambda 600nm}$) was then calculated for each sample using the following equation (Miller, 1972; Pragai & Harwood, 2002):

$$(SA_{(\lambda 420nm)} \times SV_{(\mu l)}) \div (CV_{(\mu l)} \times CA_{(\lambda 600nm)} \times T_{(min)} \times 0.00486)$$

Where $SA_{(\lambda 420nm)}$ was the final Solution Absorbance

$SV_{(\mu l)}$ was the Solution Volume (1500 µl in all cases)

$CV_{(\mu l)}$ was the Culture Volume (1000 µl in all cases)

$CA_{(\lambda 600nm)}$ was the Culture Absorbance

$T_{(min)}$ was the time taken for a colour change to develop

The molar absorption coefficient of 2-nitrophenyl is 4860 M^{-1} cm^{-1} at pH 10.

2.3.3 Anthrone

Individual Assays: Washed cell pellets were re-suspended in PBS to the volume they had been grown. The optical densities of all cultures were equalised. For each cell suspension, 3 aliquots of 0.33 ml were transferred to microcentrifuge tubes, and to each was added 0.66 ml Anthrone Reagent. Standard glucose concentration solutions (0 µg ml^{-1}, 2 µg ml^{-1}, 10 µg ml^{-1}, 20 µg ml^{-1}, 50 µg ml^{-1} and 100 µg ml^{-1}) were also prepared, and for each standard 3 aliquots of 0.33 ml was mixed with 0.66 ml Anthrone Reagent, in order to provide a standard curve of sugar concentration. The order in which reagent was added to the samples was so arranged that one set of standard glucose solutions was reacted at the start, in the middle, and at the end of the assay process. The order in which reagent was added to each set of culture samples was also alternated. Samples were left on ice for 0.5 – 1 hour. All samples were then transferred to standard cuvettes and measured at $OD_{\lambda 620nm}$ against the 0 µg ml^{-1} standard sample. The glucose equivalent hexose sugar concentration for each sample was estimated against a standard curve of 0 µg ml^{-1}, 10 µg ml^{-1}, 20 µg ml^{-1}, 50 µg ml^{-1} and 100 µg ml^{-1} glucose concentration results. These assays therefore measured the total hexose sugar content of the cells as a glucose equivalent, rather than specifically measuring the glycogen content of the cells.

Starvation Experiment: At the start of the starvation period, cells that had been cultured to maximise their polysaccharide content were recovered by centrifugation, washed and resuspended in M9 chloramphenicol and equalised to an $OD_{\lambda 600nm}$ of 1.5. Each culture was then incubated at 37 °C and 200 rpm, without a carbon source, and sampled at set time points.

At each time point, 3 aliquots of 0.33 ml of each culture were transferred to microcentrifuge tubes. For each glucose sugar standard (0 µg ml^{-1}, 2 µg ml^{-1}, 10 µg ml^{-1}, 20 µg ml^{-1}, 50 µg ml^{-1} and 100 µg ml^{-1}) 3 aliquots of 0.33 ml were also transferred to microcentrifuge tubes. A slight excess of 40 ml Anthrone Reagent was prepared for each assay, in an attempt to mitigate variability, particularly apparent in the final readings of each set. Each sample was mixed with 0.66 ml of reagent and sat on ice for approximately 40 minutes. The order in which reagent was added to the samples was so arranged that one set of standard glucose solutions was reacted at the start, in the middle, and at the end of the assay process. The order in which reagent was added to each set of culture samples was also alternated. All samples were transferred to standard cuvettes and measured at $OD_{\lambda 620nm}$ against the 0 µg ml^{-1} standard sample in a Hitachi 'Digilab U-1800' spectrophotometer. The glucose equivalent hexose sugar concentration for each sample was estimated against a standard curve of 0 µg ml^{-1}, 10 µg ml^{-1}, 20 µg ml^{-1}, 50 µg ml^{-1} and 100 µg ml^{-1} glucose concentration results. The anthrone assays for the starvation experiment therefore measured the total hexose sugar content of the cultures as opposed to the washed cells. As such it was a measure of how much of the hexose sugar found within the original washed cell pellets at the start of the experiment was metabolised over time.

2.4 Molecular Biology Methods

2.4.1 Plasmid purification

A culture of the strain containing the plasmid to be purified was grown overnight with appropriate selection in LB at 37 °C (or 30 °C if the plasmid had a temperature sensitive replication protein) and 200 rpm. Plasmids were extracted from the cells using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Once purified, plasmid DNA was stored in elution buffer (Qiangen) at -20 °C.

2.4.2 Genomic DNA extraction from *E. coli*

A culture of the appropriate strain was grown overnight in LB at 37 °C and 200 rpm. Genomic DNA was extracted from the cells using a Wizard® Genomic DNA purification kit (Promega) according to the manufacturer's instructions. Once purified, genomic DNA was stored in elution buffer (Qiagen) at -20 °C.

2.4.3 Genomic DNA extraction from *O. tauri*

A culture of *Ostreococcus tauri* was obtained with thanks from the Arthur Millar laboratory at the University of Edinburgh, having been grown in filter-sterilised artificial sea water

under constant light in a plant growth incubator fitted with a Moonlight Blue filter for about 7 days, until the culture was dense but still healthy. The culture was maintained at a temperature of 20 °C and a light intensity of approximately 20 µmol m^{-2} s^{-1} and shaken every 2 to 3 days to prevent aggregation.

Once the desired cell density was reached, 10 ml of the culture was centrifuged at maximum speed for 10 minutes and 10 °C. The supernatant was discarded, and the cell pellet immediately re-suspended in 500 µl Extraction Buffer. From this resuspension, 1.5 ml aliquots were transferred to microcentrifuge tubes and heated to 65 °C for 10 minutes and then incubated at 37 °C for 1 hour. After incubation, 500 µl of phenol:chloroform (Sigma) was added to each suspension. They were then vortexed and centrifuged at full speed for 5 minutes.

After centrifugation each suspension had separated into a milky protein-rich layer with an aqueous layer floating on top of it. From the aqueous layer of each suspension, 400 µl was transferred to a fresh tube (being careful not to disturb the protein-rich layer). To this was added 350 µl isopropanol and 50 µl NaAc (3 M), and the mixtures were precipitated at -20 °C for 1 – 3 hours. They were then centrifuged at maximum speed and 4 °C for 30 minutes. The supernatant was removed and each pellet was re-suspended by inversion in 400 µl of 70% (v/v) ethanol and centrifuged again at maximum speed and 4 °C for 2 minutes. The supernatant was removed and the pellets allowed to air-dry, before they were each finally re-suspended in 20 µl elution buffer (Qiagen) and stored at -20 °C.

2.4.4 Polymerase Chain Reaction (PCR)

Standard PCR reactions were used for cloning and analysis. They were carried out using the high-fidelity proof-reading polymerase Phusion® High-Fidelity DNA Polymerase. PCR reactions were carried out in a thermo-cycler according to the polymerase manufacturer's guidelines. The melting temperature of primers was calculated using IDT's free online resource OligAnalyzer 3.1 (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/). Oligonucleotides used in this study are described in table 2.4.

2.4.5 Colony PCR

Colony PCR reactions were used for rapidly checking the presence of a gene in one or more transformant. They were carried out in a thermo-cycler using GoTaq® DNA Polymerase according to the manufacturer's guidelines. The appropriate starting concentration of DNA was obtained by suspending a single colony, taken from a plate, in 200 µl sterile water, then

using 2 µl of that suspension in the reaction mixture. Each reaction mixture was held at 95 °C for 5 minutes prior to thermo-cycling, to ensure adequate cell lysis.

2.4.6 Sanger sequencing of DNA using BigDye

Sanger sequencing was used to check individual genes and plasmid constructs. Prior to sequencing, PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), and plasmid constructs were isolated and purified using the QIAprep Spin Miniprep Kit (Qiagen). Sequencing reactions were carried out using the BigDye® Terminator v3.1 Cycle-Sequencing Kit (Applied Biosystems) according to the manufacturer's guidelines. Samples were run on the ABI 3730 DNA analyzer by The GenePool sequencing at the University of Edinburgh (http://genepool.bio.ed.ac.uk/sanger).

2.4.7 Agarose Gel Electrophoresis

DNA from restriction digests, DNA extractions and PCR reactions were visualised using agarose gel electrophoresis. Gel percentage was determined for the specific application and all gels were run in 0.5 × TAE buffer. After electrophoresis, gels were soaked for 20 – 60 minutes in the nucleic acid stain GelGreen™ (Biotium) according to the manufacturer's instructions. They were then viewed in an Epi Chemi II Darkroom (UVP Laboratory Products) using the SYBR®Green setting.

2.4.8 SDS Polyacrylamide Gel Electrophoresis

For each transformant (*E. coli* JM109/pJW-glgC16 or *E. coli* JM109/pJW-lacZ) 3 colonies from a transformation plate were used to inoculate 3 cultures each: LB chloramphenicol; LB chloramphenicol IPTG; LB chloramphenicol IPTG lactose (2% w/v). All cultures were 10 ml, in a 50 ml Falcon tube. They were incubated overnight at 37 °C and 200 rpm. The following day, 1 ml was removed from each culture from which to prepare microscope slides and iodine assays. The remaining 9 ml for each culture was centrifuged, resuspended in 1 ml of PBS + glycerol (15% v/v) and transferred to a 2 ml microcentrifuge tube. The microcentrifuge tubes were placed on ice and sonicated at 15 amplitude microns around 5 times, for 10 seconds each time, with a 20 second pause between each sonication. Sonication was continued beyond 5 bursts if the solution had not yet turned translucent.

Bradford assays were then carried out on all sonicated samples: Cuvettes were each filled with 1 ml Coomassie reagent. Protein standards were created by adding: 0 µl BSA; 2 µl BSA; 4 µl BSA; 6 µl BSA: 8 µl BSA (BSA in all cases is 2 µg ml^{-1} laboratory standard) to Coomassie filled cuvettes, to give standard readings for 0 mg ml^{-1}, 0.2 mg ml^{-1}, 0.4 mg ml^{-1},

0.6 mg ml^{-1} and 0.8 mg ml^{-1} of protein (based on a sample volume of 20 µl). For each standard, 3 samples were prepared. A small amount of each sonicated cell suspension was then added to a fresh cuvette containing 1 ml Coomassie reagent in each case, until the colour change lay somewhere within the range of the standard preparations. Samples were incubated at room temperate for 15 minutes, then measured at $OD_{\lambda 595nm}$ on a Hitachi 'Digilab U-1800' spectrophotometer after zeroing on the 0 mg ml^{-1} protein standard sample. The protein concentration for each sample of sonicated culture was estimated against a standard curve of 0 mg ml^{-1}, 0.2 mg ml^{-1}, 0.4 mg ml^{-1}, 0.6 mg ml^{-1} and 0.8 mg ml^{-1} protein concentration results.

Based on the Bradford assay results, sonicated culture aliquots containing 30 µg of protein in each case were mixed with the same amount of ×2 SDS-PAGE loading buffer and heated to 95 °C for 10 minutes. Proteins were then separated by size on Mini-Protean TGX™ pre-cast SDS polyacrylamide gel (4 – 15%) (BioRad Laboratories Inc) against an 11-245 kDa colour, pre-stained protein standard broad-range ladder (NEB) by electrophoresis at 120 V.

2.4.9 BioBrick™ construction

The majority of the cloning performed in this study was carried out using BioBrick standard assembly parts (as proposed by Tom Knight, MIT, 2003). Each assembly part (promoter, coding region, selection marker etc) is cloned with a BioBrick prefix and suffix, the prefix containing the restriction sites *EcoRI*, *NotI* and *XbaI*, and the suffix containing the restriction sites *SpeI* and *PstI*:

BioBrick prefix: gaattc gcggccgc t tctaga

(For the purposes of this study a *NotI* site was generally not included, while a ribosome binding and a *SacI* site were added, overlapping the *XbaI* site: at gaattcc t tctagagctc a aggagg tactag)

BioBrick suffix: t actagt a gcggccg ctgcag

XbaI and *SpeI* cleavage results in complimentary 'sticky ends' that bind together to leave a Threonine-Arginine 'scar'. Any two BioBrick constructs can therefore be ligated to any other.

The *E. coli* strain JM109 and the standard Synthetic Biology plasmid pSB1C3 were used for all cloning procedures and the propagation of plasmid DNA unless otherwise stated.

2.5 BioBrick Sizes

BioBrick	Reference code	Size (bp)
pSB1A2	pSB1A2	2079
pSB1C3	pSB1C3	2070
P_{LAC}-lacZ	BBa_K523005	618
P_{LAC}-lacZ-*glgC16*	BBa_K118020	1923
E.coli glgC	Gene ID: 947942 (NCBI)	1295
E.coli glgB	Gene ID: 947940 (NCBI)	2186
Z. mays isaI	Gene ID: 542318 (NCBI)	2712
Z. mays isaII	Gene ID: 542679 (NCBI)	2829
O. tauri isaI	Gene ID: 9837697 (NCBI)	2541
O. tauri isaII	Gene ID: 9837007 (NCBI)	2346
O. tauri gbss	Gene ID: 9835561 (NCBI)	1725

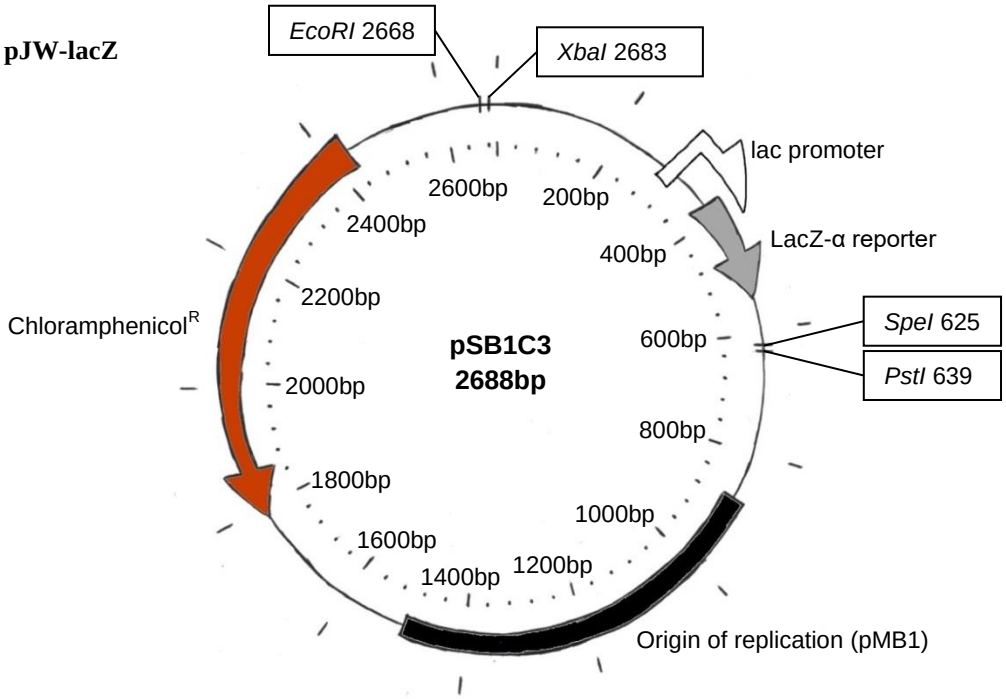
Table 2.4: List of Biobricks used in this study

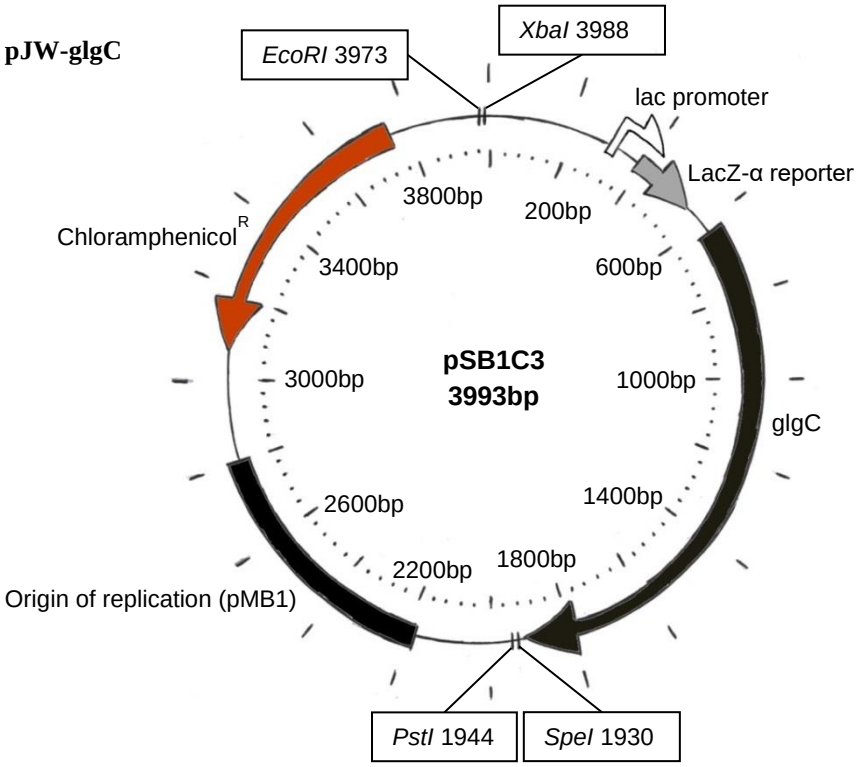
2.6 Plasmids used in this Study

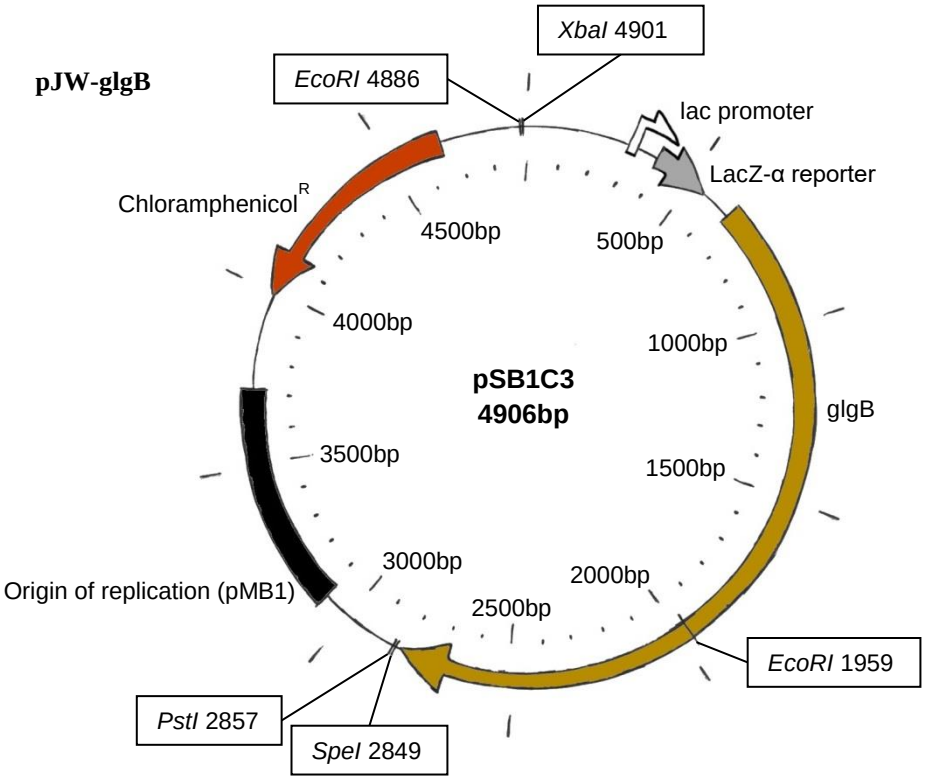
Plasmid	Description	Source
pJW-glgC16a	PSB1A2+ lacZ-α + *glgC16*	iGEM (Edinburgh 2008)
pJW-lacZ	pSB1C3 + lacZ-α	This work
pJW-glgC	pSB1C3+ lacZ-α + *glgC*	This work
pJW-glgC16	pSB1C3+ lacZ-α + *glgC16*	This work
pJW-glgB	pSB1C3+ lacZ-α + *glgB*	This work
pJW-glgCB	pSB1C3+ lacZ-α + *glgC* + *glgB*	This work
pJW-glgC16B	pSB1C3+ lacZ-α + *glgC* + *glgB*	This work
pCF-isaI-isaII	pSB1A2+ lacZ-α + *glgC16* + *isaI* (*Z. mays*)+ *isaII* (*Z. mays*)	iGEM (Edinburgh 2008)
pJW-glgC16-isaI-isaII-gbss	pSB1A2+ lacZ-α + *glgC16* + *isaI* (*Z. mays*)+ *isaII* (*Z. mays*) + *gbss* (*O. tauri*)	This work
pJW-glgC16-isaI-gbss	pSB1C3+ lacZ-α + *glgC16* + *isaI* (*O. tauri*) + *gbss* (*O. tauri*)	This work
pJW-gbss	pSB1C3+ lacZ-α + *gbss* (*O. tauri*)	This work
pJW-isaI	pSB1C3+ lacZ-α + *isaI* (*O. tauri*)	This work

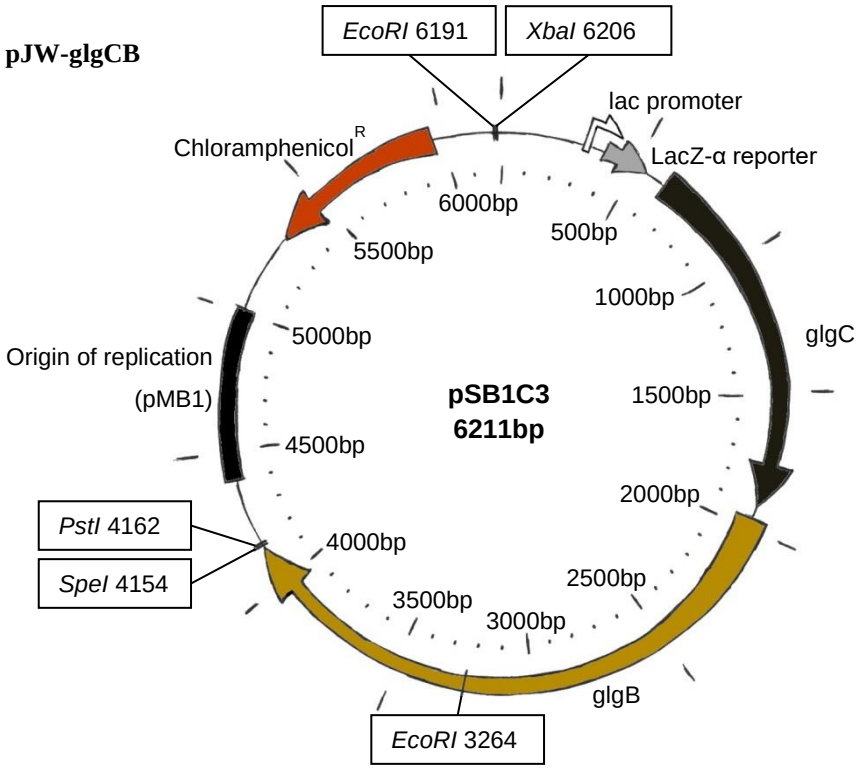
Table 2.5: List of plasmids used in this study

2.7 Plasmid Maps






pJW-glgB
XbaI 4901
EcoRI 4886
lac promoter
LacZ-α reporter
Chloramphenicol R
4500bp
500bp
4000bp
1000bp
pSB1C3
4906bp
glgB
3500bp
1500bp
3000bp
2000bp
Origin of replication (pMB1)
2500bp
EcoRI 1959
PstI 2857
SpeI 2849


pJW-glgCB
EcoRI 6191
XbaI 6206
lac promoter
Chloramphenicol R
LacZ-α reporter
6000bp
500bp
5500bp
1000bp
5000bp
glgC
Origin of replication
(pMB1)
pSB1C3
6211bp
1500bp
4500bp
2000bp
PstI 4162
4000bp
SpeI 4154
2500bp
3500bp
3000bp
glgB
EcoRI 3264

2.8 Sequencing Data

Highlights Key

Shine-Dalgarno sequences

lacZ-α

glgC (Unhighlighted 'G' replaced with 'A' in glgC16)

glgB

Chloramphenicol^R

Restriction sites (ACTAGT:SpeI CTCGAG:PstI GAATTC:EcoRI TCTAGA:XbaI)

pJW-LacZ

```
0     CTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGC
50    TGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCG
100   GTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCAC
150   CCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAT
200   TAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCG
250   CAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
300   ACTTTATGCTTCCGGCTCGTATGTTGTGTGAAATTGTGAGCGGATAACAA
350   TTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCG
400   TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
450   CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
500   CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
550   TTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGA
600   TACTAGAGCTCAAGGAGGTACTAGTAGCGGCCGCTGCAGTCCGGCAAAAA
650   AGGGCAAGGTGTCACCACCCTGCCCTTTTTCTTTAAAACCGAAAAGATTA
700   CTTCGCGTTATGCAGGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
750   TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
800   CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAG
850   CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCACAG
900   GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
950   GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGC
1000  TCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
1050  CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA
1100  GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCAC
1150  GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
1200  TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
1250  GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
1300  AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
1350  CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGAT
1400  CCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
1450  CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
1500  TACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGG
1550  TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
1600  TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG
1650  CTCGAGGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCG
1700  TTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAAC
1750  AGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACTCAT
1800  CGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCA
1850  CAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCC
1900  TTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCA
```

```
1950  TATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCT
2000  GAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTT
2050  TTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGA
2100  AATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCA
2150  TGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACC
2200  GTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAA
2250  GAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTC
2300  TTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTG
2350  AGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATA
2400  TATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTA
2450  GCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTAT
2500  TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCCGATCAACTCGAGTG
2550  CCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAA
2600  TAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAATCCTTAGCTTTCG
2650  CTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAG
```

pJW-glgC

```
0     CTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGC
50    TGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCG
100   GTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCAC
150   CCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAT
200   TAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCG
250   CAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
300   ACTTTATGCTTCCGGCTCGTATGTTGTGTGAAATTGTGAGCGGATAACAA
350   TTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCG
400   TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
450   CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
500   CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
550   TTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGA
600   TACTAGAGCTCAAGGAGGTACTAGATGGTTAGTTTAGAGAAGAACGATCA
650   CTTAATGTTGGCGCGCCAGCTGCCATTGAAATCTGTTGCCCTGATACTGG
700   CGGGAGGACGTGGTACCCGCCTGAAGGATTTAACCAATAAGCGAGCAAAA
750   CCGGCCGTACACTTCGGCGGTAAGTTCCGCATTATCGACTTTGCGCTGTC
800   TAACTGCATCAACTCCGGGATCCGTCGTATGGGCGTGATCACCCAGTACC
850   AGTCCCACACTCTGGTGCAGCACATTCAGCGCGGCTGGTCATTCTTCAAT
900   GAAGAAATGAACGAGTTTGTCGATCTGCTGCCAGCACAGCAGAGAATGAA
950   AGGGGAAAACTGGTATCGCGGCACCGCAGATGCGGTCACCCAAAACCTCG
1000  ACATTATCCGCCGTTATAAAGCGGAATACGTGGTGATCCTGGCGGGCGAC
1050  CATATCTACAAGCAAGACTACTCGCGTATGCTTATCGATCACGTCGAAAA
1100  AGGCGCACGTTGCACCGTTGCTTGTATGCCAGTACCGATTGAAGAAGCCT
1150  CCGCATTTGGCGTTATGGCGGTTGATGAGAACGATAAAATTATCGAATTT
1200  GTTGAAAAACCTGCTAACCCGCCGTCAATGCCGAACGATCCGAGCAAATC
1250  TCTGGCGAGTATGGGTATCTACGTCTTTGACGCCGACTATCTGTATGAAC
1300  TGCTGGAAGAAGACGATCGCGATGAGAACTCCAGCCACGACTTTGGCAAA
1350  GATTTGATTCCCAAGATCACCGAAGCCGGTCTGGCCTATGCGCACCCGTT
1400  CCCGCTCTCTTGCGTACAATCCGACCCGGATGCCGAGCCGTACTGGCGCG
1450  ATGTGGGTACGCTGGAAGCTTACTGGAAAGCGAACCTCGATCTGGCCTCT
1500  GTGGTGCCGGAACTGGATATGTACGATCGCAATTGGCCAATTCGCACCTA
1550  CAATGAATCATTACCGCCAGCGAAATTCGTGCAGGATCGCTCCGGTAGCC
1600  ACGGGATGACCCTTAACTCACTGGTTTCCGGCGGTTGTGTGATCTCCGGT
1650  TCGGTGGTGGTGCAGTCCGTTCTGTTCTCGCGCGTTCGCGTGAACTCATT
1700  CTGCAACATTGATTCCGCCGTATTGTTACCGGAAGTATGGGTAGGTCGCT
1750  CGTGCCGTCTGCGCCGCTGCGTCATCGATCGTGCTTGTGTTATTCCGGAA
1800  GGCATGGTGATTGGTGAAAACGCAGAGGAAGATGCACGTCGTTTCTATCG
1850  TTCAGAAGAAGGCATCGTGCTGGTAACGCGCGAAATGCTACGGAAGTTAG
1900  GGCATAAACAGGAGCGATAATAATACTAGTAGCGGCCGCTGCAGTCCGGC
```

```
1950  AAAAAAGGGCAAGGTGTCACCACCCTGCCCTTTTTCTTTAAAACCGAAAA
2000  GATTACTTCGCGTTATGCAGGCTTCCTCGCTCACTGACTCGCTGCGCTCG
2050  GTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
2100  GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
2150  GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC
2200  CACAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
2250  GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
2300  GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
2350  CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACG
2400  CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
2450  TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTAT
2500  CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
2550  CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
2600  TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGT
2650  ATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
2700  TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
2750  AGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATC
2800  TTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGAT
2850  TTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATT
2900  AAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
2950  GACAGCTCGAGGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACC
3000  GAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTA
3050  TCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCA
3100  CTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGC
3150  CATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTG
3200  TCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTT
3250  GTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGAT
3300  TGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCC
3350  AGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTG
3400  CCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTT
3450  GCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGC
3500  TCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCG
3550  GGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTA
3600  CGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTA
3650  CATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTG
3700  GGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTT
3750  CCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGAT
3800  CTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCCGATCAACTC
3850  GAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTAT
3900  AAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAATCCTTAGC
3950  TTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAG
```

pJW-glgB

```
0     CTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGC
50    TGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCG
100   GTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCAC
150   CCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAT
200   TAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCG
250   CAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
300   ACTTTATGCTTCCGGCTCGTATGTTGTGTGAAATTGTGAGCGGATAACAA
350   TTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCG
400   TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
450   CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
500   CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
550   TTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGA
600   TACTAGAGCTCAAGGAGGTACTAGAGCTCAAGGAGGTAGACAAGCATGTC
```

```
650   CGATCGTATCGATAGAGACGTGATTAACGCGCTAATTGCAGGCCATTTTG
700   CGGATCCTTTTTCCGTACTGGGAATGCATAAAACCACCGCGGGACTGGAA
750   GTCCGTGCCCTTTTACCCGACGCTACCGATGTGTGGGTGATTGAACCGAA
800   AACCGGGCGCAAACTCGCAAAACTGGAGTGTCTCGACTCACGGGGATTCT
850   TTAGCGGCGTCATTCCGCGACGTAAGAATTTTTTCCGCTATCAGTTGGCT
900   GTTGTCTGGCATGGTCAGCAAAACCTGATTGATGATCCTTACCGTTTTGG
950   TCCGCTAATCCAGGAAATGGATGCCTGGCTATTATCTGAAGGTACTCACC
1000  TGCGCCCGTATGAAACCTTAGGCGCGCATGCAGATACTATGGATGGCGTC
1050  ACAGGTACGCGTTTCTCTGTCTGGGCTCCAAACGCCCGTCGGGTCTCGGT
1100  GGTTGGGCAATTCAACTACTGGGACGGTCGCCGTCACCCGATGCGCCTGC
1150  GTAAAGAGAGCGGCATCTGGGAACTGTTTATCCCTGGGGCGCATAACGGT
1200  CAGCTCTATAAATACGAGATGATTGATGCCAATGGCAACTTGCGTCTGAA
1250  GTCCGACCCTTATGCCTTTGAAGCGCAAATGCGCCCGGAAACCGCGTCTC
1300  TTATTTGCGGGCTGCCGGAAAAGGTTGTACAGACTGAAGAGCGCAAAAAA
1350  GCGAATCAGTTTGATGCGCCAATCTCTATTTATGAAGTTCACCTGGGTTC
1400  CTGGCGTCGCCACACCGACAACAATTTCTGGTTGAGCTACCGCGAGCTGG
1450  CCGATCAACTGGTGCCTTATGCTAAATGGATGGGCTTTACCCACCTCGAA
1500  CTACTGCCCATTAACGAGCATCCCTTCGATGGCAGTTGGGGTTATCAGCC
1550  AACCGGCCTGTATGCGCCAACCCGCCGTTTTGGTACTCGCGACGACTTCC
1600  GTTATTTCATTGATGCCGCACACGCAGCTGGTCTGAACGTGATTCTCGAC
1650  TGGGTGCCAGGCCACTTCCCGACTGATGACTTTGCGCTTGCCGAATTTGA
1700  TGGCACGAACTTGTATGAACACAGCGATCCGCGTGAAGGCTATCATCAGG
1750  ACTGGAACACGCTGATCTACAACTATGGTCGCCGTGAAGTCAGTAACTTC
1800  CTCGTCGGTAACGCGCTTTACTGGATTGAACGTTTTGGTATTGATGCGCT
1850  GCGCGTCGATGCGGTGGCGTCAATGATTTATCGCGACTACAGCCGTAAAG
1900  AGGGGGAGTGGATCCCGAACGAATTTGGCGGGCGCGAGAATCTTGAAGCG
1950  ATTGAATTCTTGCGTAATACCAACCGTATTCTTGGTGAGCAGGTTTCCGG
2000  TGCGGTGACAATGGCTGAGGAGTCTACCGATTTCCCTGGCGTTTCTCGTC
2050  CGCAGGATATGGGCGGTCTGGGCTTCTGGTACAAGTGGAACCTCGGCTGG
2100  ATGCATGACACCCTGGACTACATGAAGCTCGACCCGGTTTATCGTCAGTA
2150  TCATCACGATAAACTGACCTTCGGGATTCTCTACAACTACACTGAAAACT
2200  TCGTCCTGCCGTTGTCGCATGATGAAGTGGTCCACGGTAAAAAATCGATT
2250  CTCGACCGCATGCCGGGCGACGCATGGCAGAAATTCGCGAACCTGCGCGC
2300  CTACTATGGCTGGATGTGGGCATTCCCGGGCAAGAAACTACTGTTCATGG
2350  GTAACGAATTTGCCCAGGGCCGCGAGTGGAACCATGACGCCAGCCTCGAC
2400  TGGCATCTGTTGGAAGGCGGCGATAACTGGCACCACGGTGTCCAGCGTCT
2450  GGTGCGCGATCTGAACCTCACCTACCGCCACCATAAAGCAATGCATGAAC
2500  TGGATTTTGACCCGTACGGCTTTGAATGGCTGGTGGTGGATGACAAAGAA
2550  CGCTCGGTGCTGATCTTTGTGCGTCGCGATAAAGAGGGTAACGAAATCAT
2600  CGTTGCCAGTAACTTTACGCCGGTACCGCGTCATGATTATCGCTTCGGCA
2650  TAAACCAGCCGGGCAAATGGCGTGAAATCCTCAATACCGATTCCATGCAC
2700  TATCACGGCAGTAATGCAGGCAATGGCGGCACGGTACACAGCGATGAGAT
2750  TGCCAGCCACGGTCGTCAGCATTCACTAAGCCTGACGCTACCACCGCTGG
2800  CCACTATCTGGCTGGTTCGGGAGGCAGAATGACACAACTCGCTACTAGTA
2850  GCTGCAGTCCGGCAAAAAAGGGCAAGGTGTCACCACCCTGCCCTTTTTCT
2900  TTAAAACCGAAAAGATTACTTCGCGTTATGCAGGCTTCCTCGCTCACTGA
2950  CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
3000  AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
3050  ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
3100  GCTGGCGTTTTTCCACAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
3150  GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
3200  GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
3250  GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
3300  CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
3350  AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
3400  ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
3450  CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
3500  GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
3550  AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
```

```
3600  GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT
3650  TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
3700  AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT
3750  CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
3800  ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGA
3850  GTAAACTTGGTCTGACAGCTCGAGGCTTGGATTCTCACCAATAAAAAACG
3900  CCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTC
3950  ATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACG
4000  CCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTG
4050  CCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGG
4100  CATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGG
4150  GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAA
4200  CTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT
4250  AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA
4300  TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAA
4350  AACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATC
4400  CCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAG
4450  CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC
4500  TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGT
4550  CTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT
4600  TACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTC
4650  TCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATAC
4700  GCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT
4750  GCCCGATCAACTCGAGTGCCACCTGACGTCTAAGAAACCATTATTATCAT
4800  GACATTAACCTATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAA
4850  AAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTT
4900  CTAGAG
```

pJW-glgCB

```
0     CTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGC
50    TGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCG
100   GTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCAC
150   CCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAT
200   TAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCG
250   CAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
300   ACTTTATGCTTCCGGCTCGTATGTTGTGTGAAATTGTGAGCGGATAACAA
350   TTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCG
400   TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
450   CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
500   CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
550   TTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGA
600   TACTAGAGCTCAAGGAGGTACTAGATGGTTAGTTTAGAGAAGAACGATCA
650   CTTAATGTTGGCGCGCCAGCTGCCATTGAAATCTGTTGCCCTGATACTGG
700   CGGGAGGACGTGGTACCCGCCTGAAGGATTTAACCAATAAGCGAGCAAAA
750   CCGGCCGTACACTTCGGCGGTAAGTTCCGCATTATCGACTTTGCGCTGTC
800   TAACTGCATCAACTCCGGGATCCGTCGTATGGGCGTGATCACCCAGTACC
850   AGTCCCACACTCTGGTGCAGCACATTCAGCGCGGCTGGTCATTCTTCAAT
900   GAAGAAATGAACGAGTTTGTCGATCTGCTGCCAGCACAGCAGAGAATGAA
950   AGGGGAAAACTGGTATCGCGGCACCGCAGATGCGGTCACCCAAAACCTCG
1000  ACATTATCCGCCGTTATAAAGCGGAATACGTGGTGATCCTGGCGGGCGAC
1050  CATATCTACAAGCAAGACTACTCGCGTATGCTTATCGATCACGTCGAAAA
1100  AGGCGCACGTTGCACCGTTGCTTGTATGCCAGTACCGATTGAAGAAGCCT
1150  CCGCATTTGGCGTTATGGCGGTTGATGAGAACGATAAAATTATCGAATTT
1200  GTTGAAAAACCTGCTAACCCGCCGTCAATGCCGAACGATCCGAGCAAATC
1250  TCTGGCGAGTATGGGTATCTACGTCTTTGACGCCGACTATCTGTATGAAC
1300  TGCTGGAAGAAGACGATCGCGATGAGAACTCCAGCCACGACTTTGGCAAA
```

```
1350  GATTTGATTCCCAAGATCACCGAAGCCGGTCTGGCCTATGCGCACCCGTT
1400  CCCGCTCTCTTGCGTACAATCCGACCCGGATGCCGAGCCGTACTGGCGCG
1450  ATGTGGGTACGCTGGAAGCTTACTGGAAAGCGAACCTCGATCTGGCCTCT
1500  GTGGTGCCGGAACTGGATATGTACGATCGCAATTGGCCAATTCGCACCTA
1550  CAATGAATCATTACCGCCAGCGAAATTCGTGCAGGATCGCTCCGGTAGCC
1600  ACGGGATGACCCTTAACTCACTGGTTTCCGGCGGTTGTGTGATCTCCGGT
1650  TCGGTGGTGGTGCAGTCCGTTCTGTTCTCGCGCGTTCGCGTGAACTCATT
1700  CTGCAACATTGATTCCGCCGTATTGTTACCGGAAGTATGGGTAGGTCGCT
1750  CGTGCCGTCTGCGCCGCTGCGTCATCGATCGTGCTTGTGTTATTCCGGAA
1800  GGCATGGTGATTGGTGAAAACGCAGAGGAAGATGCACGTCGTTTCTATCG
1850  TTCAGAAGAAGGCATCGTGCTGGTAACGCGCGAAATGCTACGGAAGTTAG
1900  GGCATAAACAGGAGCGATAATAATACTAGAGCTCAAGGAGGTAGACAAGC
1950  ATGTCCGATCGTATCGATAGAGACGTGATTAACGCGCTAATTGCAGGCCA
2000  TTTTGCGGATCCTTTTTCCGTACTGGGAATGCATAAAACCACCGCGGGAC
2050  TGGAAGTCCGTGCCCTTTTACCCGACGCTACCGATGTGTGGGTGATTGAA
2100  CCGAAAACCGGGCGCAAACTCGCAAAACTGGAGTGTCTCGACTCACGGGG
2150  ATTCTTTAGCGGCGTCATTCCGCGACGTAAGAATTTTTTCCGCTATCAGT
2200  TGGCTGTTGTCTGGCATGGTCAGCAAAACCTGATTGATGATCCTTACCGT
2250  TTTGGTCCGCTAATCCAGGAAATGGATGCCTGGCTATTATCTGAAGGTAC
2300  TCACCTGCGCCCGTATGAAACCTTAGGCGCGCATGCAGATACTATGGATG
2350  GCGTCACAGGTACGCGTTTCTCTGTCTGGGCTCCAAACGCCCGTCGGGTC
2400  TCGGTGGTTGGGCAATTCAACTACTGGGACGGTCGCCGTCACCCGATGCG
2450  CCTGCGTAAAGAGAGCGGCATCTGGGAACTGTTTATCCCTGGGGCGCATA
2500  ACGGTCAGCTCTATAAATACGAGATGATTGATGCCAATGGCAACTTGCGT
2550  CTGAAGTCCGACCCTTATGCCTTTGAAGCGCAAATGCGCCCGGAAACCGC
2600  GTCTCTTATTTGCGGGCTGCCGGAAAAGGTTGTACAGACTGAAGAGCGCA
2650  AAAAAGCGAATCAGTTTGATGCGCCAATCTCTATTTATGAAGTTCACCTG
2700  GGTTCCTGGCGTCGCCACACCGACAACAATTTCTGGTTGAGCTACCGCGA
2750  GCTGGCCGATCAACTGGTGCCTTATGCTAAATGGATGGGCTTTACCCACC
2800  TCGAACTACTGCCCATTAACGAGCATCCCTTCGATGGCAGTTGGGGTTAT
2850  CAGCCAACCGGCCTGTATGCGCCAACCCGCCGTTTTGGTACTCGCGACGA
2900  CTTCCGTTATTTCATTGATGCCGCACACGCAGCTGGTCTGAACGTGATTC
2950  TCGACTGGGTGCCAGGCCACTTCCCGACTGATGACTTTGCGCTTGCCGAA
3000  TTTGATGGCACGAACTTGTATGAACACAGCGATCCGCGTGAAGGCTATCA
3050  TCAGGACTGGAACACGCTGATCTACAACTATGGTCGCCGTGAAGTCAGTA
3100  ACTTCCTCGTCGGTAACGCGCTTTACTGGATTGAACGTTTTGGTATTGAT
3150  GCGCTGCGCGTCGATGCGGTGGCGTCAATGATTTATCGCGACTACAGCCG
3200  TAAAGAGGGGGAGTGGATCCCGAACGAATTTGGCGGGCGCGAGAATCTTG
3250  AAGCGATTGAATTCTTGCGTAATACCAACCGTATTCTTGGTGAGCAGGTT
3300  TCCGGTGCGGTGACAATGGCTGAGGAGTCTACCGATTTCCCTGGCGTTTC
3350  TCGTCCGCAGGATATGGGCGGTCTGGGCTTCTGGTACAAGTGGAACCTCG
3400  GCTGGATGCATGACACCCTGGACTACATGAAGCTCGACCCGGTTTATCGT
3450  CAGTATCATCACGATAAACTGACCTTCGGGATTCTCTACAACTACACTGA
3500  AAACTTCGTCCTGCCGTTGTCGCATGATGAAGTGGTCCACGGTAAAAAAT
3550  CGATTCTCGACCGCATGCCGGGCGACGCATGGCAGAAATTCGCGAACCTG
3600  CGCGCCTACTATGGCTGGATGTGGGCATTCCCGGGCAAGAAACTACTGTT
3650  CATGGGTAACGAATTTGCCCAGGGCCGCGAGTGGAACCATGACGCCAGCC
3700  TCGACTGGCATCTGTTGGAAGGCGGCGATAACTGGCACCACGGTGTCCAG
3750  CGTCTGGTGCGCGATCTGAACCTCACCTACCGCCACCATAAAGCAATGCA
3800  TGAACTGGATTTTGACCCGTACGGCTTTGAATGGCTGGTGGTGGATGACA
3850  AAGAACGCTCGGTGCTGATCTTTGTGCGTCGCGATAAAGAGGGTAACGAA
3900  ATCATCGTTGCCAGTAACTTTACGCCGGTACCGCGTCATGATTATCGCTT
3950  CGGCATAAACCAGCCGGGCAAATGGCGTGAAATCCTCAATACCGATTCCA
4000  TGCACTATCACGGCAGTAATGCAGGCAATGGCGGCACGGTACACAGCGAT
4050  GAGATTGCCAGCCACGGTCGTCAGCATTCACTAAGCCTGACGCTACCACC
4100  GCTGGCCACTATCTGGCTGGTTCGGGAGGCAGAATGACACAACTCGCTAC
4150  TAGTAGCTGCAGTCCGGCAAAAAAGGGCAAGGTGTCACCACCCTGCCCTT
4200  TTTCTTTAAAACCGAAAAGATTACTTCGCGTTATGCAGGCTTCCTCGCTC
4250  ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCA
```

```
4300  CTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
4350  AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
4400  GCGTTGCTGGCGTTTTTCCACAGGCTCCGCCCCCCTGACGAGCATCACAA
4450  AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT
4500  ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
4550  CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
4600  GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
4650  GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
4700  GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
4750  ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATG
4800  TAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT
4850  AGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGG
4900  AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
4950  GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCT
5000  CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGA
5500  AAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
5100  CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATA
5150  TATGAGTAAACTTGGTCTGACAGCTCGAGGCTTGGATTCTCACCAATAAA
5200  AAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTG
5250  AGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAA
5300  TTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCA
5350  TTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCC
5400  AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAA
5450  AACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGG
5500  TGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAAC
5500  CCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGA
5600  ATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCG
5650  ATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACA
5700  CTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGG
5750  ATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACT
5800  TGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGA
5850  ACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATG
5900  TTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT
5950  TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAA
6000  AATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCT
6050  TACGTGCCCGATCAACTCGAGTGCCACCTGACGTCTAAGAAACCATTATT
6100  ATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCAGAATTTCAG
6150  ATAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGGAATTCGCGGC
6200  CGCTTCTAGAG
```

2.9 Oligonucleotides for Cloning

Oligonucleotides were designed using IDT's free online resource OligAnalyzer 3.1 (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/) and manufactured by Sigma or IDT. Stock solutions of 100 mM were made by dissolving in sterile water and stored at -20 °C. Working solutions were made up to 10 mM and stored at -20 °C.

Name	Sequence	Summary of use
glgB _F	ATGAATTCCTTCTAGAGCTCAAGGAGGTAGACAAGCATGTCCGATC	BioBrick
glgB _R	ATCCTGCAGCTACTAGTAGCGAGTTGTGTCATTCTG	BioBrick
glgB_MABEL_F	GAGTTCTTGCGTAATACC	
glgB_MABEL_R	AATCGCTTCAAGATTCTC	
OtGBSS _F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGAGTCGAACGGCGTTCGAGG	BioBrick
OtGBSS_tp1_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGCGCGGGCGAAGGAATC	BioBrick
OtGBSS_tp2_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGAAGATTGTTTTTGTTTCCGCGG	BioBrick
OtGBSS _R	ATCCTGCAGCTACTAGTAttattACGTCTGAGCGGCGAGCG	BioBrick
OtISAI_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGCAACCCGTGCAATTG	BioBrick
OtISAI_tp1_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGCACGTGAAGGTACAAGCCTG	BioBrick
OtISAI_tp2_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGGTGATGCAAGCGCACTG	BioBrick
OtISAI_R	ATCCTGCAGCTACTAGTATTATTATGCCAGGCTCACTGCAC	BioBrick
OtISAII_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGTGAGGAGTGGGATG	BioBrick
OtISAII_tp1_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGGAGGGTGGTTGGGACACCGG	BioBrick
OtISAII_R	ATCCTGCAGCTACTAGTAGTTGAAGAACTACGCGCGTCTC	BioBrick
OtISAII_MABEL_F	GAGTTCCGAATCCGCG	
OtISAII_MABEL_R	TCTGGAACTCCATCTCGTC	
OtBEI_tp_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGCGCGTTTTTCACGCGGCATC	BioBrick
OtBEI_R	ATCCTGCAGCTACTAGTACCACATCATCCGCATCGTGC	BioBrick
OtBEII_F	ATGAATTCCTTCTAGAGCTCAAGGAGGGATCGGCACACGATGTC	BioBrick
OtBEII_tp_F	ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATGAGATCGGGGATGGCGACGACG	BioBrick
OtBEII_R	ATCCTGCAGCTACTAGTAGAATCTCATCGGTGCGCGTC	BioBrick
OtBEII_MABEL_F	GAGTTCATCGCCGAAGAC	
OtBEII_MABEL_R	GGGAGCGGTATACGTCC	
OtAdaptin-N-terminal_F	ATGAATTCCTTCTAGAGCTCAAGGAGGGTATCCAATCGACGCATG	BioBrick
OtAdaptin-N-terminal_R	ATCCTGCAGCTACTAGTACTCCAGTCTACGATCGCTC	BioBrick
LacZ_DF	GAAGCGGTGCCGGAAAGCTGGCTGGAGTGATACTAG	Paperclip
LacZ_DR	GGCCTAGTATCACTCCAGCCAGCTTTCCGGCACCGCTTCTGGT	Paperclip
GBSS_UF	GCCAGCTCAAGGAGGTACTAGATGGCGCGGGCGAAGGAATCGAC	Paperclip
GBSS_UR	TTCCTTCGCCCGCGCCATCTAGTACCTCCTTGAGCT	Paperclip
GBSS_DF	GCAGACGCGCTCGCCGCTCAGACGTAATAATACTAG	Paperclip
GBSS_DR	GGCCTAGTATTATTACGTCTGAGCGGCGAGCGCGTCTGCCTTG	Paperclip
ISAI_UF	GCCAGCTCAAGGAGGTACTAGATGGCACGTGAAGGTACAAGCC	Paperclip
ISAI_UR	TGTACCTTCACGTGCCATCTAGTACCTCCTTGAGCT	Paperclip
ISAI_DF	ATTTTTCGTGCAGTGAGCCTGGCATAATAATACTAG	Paperclip
ISAI_DR	GGCCTAGTATTATTATGCCAGGCTCACTGCACGAAAAATAACG	Paperclip
ISAII_UF	GCCGAGCTCAAGGAGGTACTAGATGGAGGGTGGTTGGGACACC	Paperclip
ISAII_UR	TCCCAACCACCCTCCATCTAGTACCTCCTTGAGCTC	Paperclip
ISAII_full_UF	GCCGAGCTCAAGGAGGTACTAGATGGTGAGGAGTGGGATGCGG	Paperclip
ISAII_full_UR	ATCCCACTCCTCACCATCTAGTACCTCCTTGAGCTC	Paperclip
ISAII_DF	CCAAGTGAGAGACGCGCGTAGTTCTTCAACTACTAG	Paperclip
ISAII_DR	GGCCTAGTAGTTGAAGAACTACGCGCGTCTCTCACTTGGACAT	Paperclip
BEI_UF	GCCGAGCTCAAGGAGGTACTAGATGCGCGTTTTTCACGCGGCA	Paperclip
BEI_UR	GCGTGAAAAACGCGCATCTAGTACCTCCTTGAGCTC	Paperclip
BEI_DF	CGAGTACGGCGCACGATGCGGATGATGTGGTACTAG	Paperclip
BEI_DR	GGCCTAGTACCACATCATCCGCATCGTGCGCCGTACTCGGCGC	Paperclip
BEII_UF	GCCGAGCTCAAGGAGGTACTAGATGGATCGGGATGGCGACGAC	Paperclip
BEII_UR	TCGCCATCCCGATCCATCTAGTACCTCCTTGAGCTC	Paperclip
BEII_DF	TCATAGCAATGACGCGCACCGATGAGATTCTACTAG	Paperclip

BEII_DR	GGCCTAGTAGAATCTCATCGGTGCGCGTCATTGCTATGATACA	Paperclip
glgC_up_F	GCCCTGTTAACAACGTTG	Gene Knockout
glgC_up_R	ATCGAATTCGACTAACTCCTTTTTTATCATCTCTG	Gene Knockout
glgC_down_F	ATCCTGCAGTGCAGGTTTTACATGTATGTTC	Gene Knockout
glgC_down_R	CCTTGATAGGCCAGGTTG	Gene Knockout
glgA_up_F	GTTCAGAAGAAGGCATCGTGCTGGTAACGCGCGAAATGCTACGGAAGTTAGGGCATAAACCATAAACAGTAATACAAGGGGTG	Gene Knockout
glgA_down_R	GCTAAGCGTGGGCGATGAATATGTAAACGGAGCATTCATATAGGCGTTTCCTGAAAACTACTTCGGGCTTATTAGAAAAACTC	Gene Knockout
glgB_up_F	TCAAAAAAATGTCACAACCAGAAGTCAAAAATCCAATTGGATGGGGTGACACAATAAAACCATAAACAGTAATACAAGGGGTG	Gene Knockout
glgB_down_R	GCGCCGAGGGGAGCGGGTTTGCCAATGGCGAGTTGTGTCATTCTGCCTCCCGAACCAGCCCTTCGGGCTTATTAGAAAAACTC	Gene Knockout
glgX_up_F	ACGGTCGTCAGCATTCACTAAGCCTGACGCTACCACCGCTGGCCACTATCTGGCTGGTTCCATAAACAGTAATACAAGGGGTG	Gene Knockout
glgX_down_R	GCGCCAACATTAAGTGATCGTTCTTCTCTAAACTAACCATGACTAACTCCTTTTTTATCACTTCGGGCTTATTAGAAAAACTC	Gene Knockout
glgP_down_R	CGAAATGATGGCGGAAAAAAACGGGACCCTTTGGCCCCGTTCTATTTATTGGTGAACTTACTTCGGGCTTATTAGAAAAACTC	Gene Knockout
pSBNX3_ins_F2	AAATAGGCGTATCACGAGGC	Sequencing
LacZ_seq	AGTTGCGCAGCCTGAATG	Sequencing
glgC_seq	GGCGACCATATCTACAAGCAAG	Sequencing
glgC_junctfor	CGTCATCGATCGTGCTTGTG	Sequencing
glgB_seq1	TATCCCTGGGGCGCATAAC	Sequencing
glgB_seq1.2	CGCTTTACTGGATTGAACG	Sequencing
glgB_seq2	GGTGGTGGATGACAAAGAACG	Sequencing
OtGBSS_seq	CGAAGACGGTTACTCCAAGG	Sequencing
OtGBSS_junctfor	GTCTCGTCGATACCGTCAAG	Sequencing
OtISAI_seq1	TGTTTGTGATACACCGCGTC	Sequencing
OtISAI_seq1.1	CAGATGATTGTGGTCGTGAAG	Sequencing
OtISAI_seq2.1	GTTTTACCCTGCGTGATTGTG	Sequencing
OtISAI_seq3	TGATGTTCTGAGCGCAGAAG	Sequencing
OtAdaptin-N-terminal_seq1.1	CTGAATCCGTCCTCGCATC	Sequencing
OtAdaptin-N-terminal_seq2	GGCGTCCTACTTTCAACTCG	Sequencing
OtAdaptin-N-terminal_seq3	GATACCTCGCTTCCACAAACC	Sequencing

Table 2.6: List of oligonucleotides used in this study

Chapter 3: Upregulation of Glycogen Synthesis in *E. coli*

3.1 Introduction

ADP-glucose pyrophosphorylase, coded for by *glgC*, is thought responsible for catalysing the first committed step, which is also the rate-limiting step, of glycogen synthesis in *E. coli*. It achieves this by converting glucose-1-phosphate and ATP into the ADP-glucose used in glycogen's formation. Previously reports (e.g. Govons et al., 1973; Wilson et al., 2010) show that increased GlgC activity leads to higher glycogen levels in *E. coli*. It was therefore hypothesised (perhaps naïvely) that upregulating *glgC* expression would lead to the straightforward accumulation of more glycogen than in wild type *E. coli* cultured the same way. A higher accumulation was indeed observed, but was by no means straightforward.

3.2 Upregulation of the *glgC* gene in *E. coli* leads to an 'inclusion body' phenotype

E. coli were first transformed with pJW-glgC16a (see Methods 2.7); a plasmid made up of the standard synthetic biology backbone pSB1A2 with an additional *lacZ'-α* minigene as a screening marker and the AMP-resistant form of *glgC* (Govons et al., 1973), designated *glgC16*. This feedback-resistant variant is created by a single-base substitution in the 1007th nucleotide, resulting in a G335D substitution. Both the *lacZ'-α* and the *glgC16* genes were expressed under a *lac* promoter. The same construct was later transferred to a pSB1C3 backbone, in line with current standards of the Registry of Standard Biological Parts, to create the plasmid pJW-glgC16 (see Methods 2.7). When transformed with either plasmid and grown in media supplemented with 1-2% lactose, a high proportion of *E. coli* cells were seen to contain large inclusion bodies, easily observed by light microscopy after staining with 5% Lugol's iodine (figures 3.3, 3.8, 3.10, 3.19, 3.24).

The same cells were also found to react when pelleted and resuspended in 0.2% Lugol's iodine, often forming the intense blue colour similar to that which occurs with solutions of long, single helix-containing polysaccharides such as amylose, where untransformed cultures or those transformed with an 'empty' variant of the same plasmid (pJW-lacZ, which lacks the *glgC16* gene. See Methods 2.7) stained yellow-brown (figure 3.1). However, the darkening of the cell suspensions showed a degree of variability and, furthermore, was observed to fade rapidly over time, with a difference being easily observed within minutes.

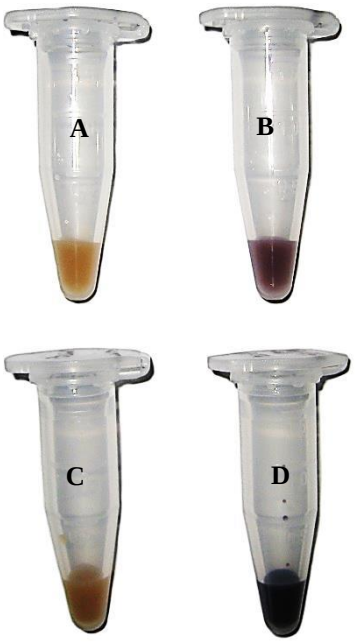


Figure 3.1: Iodine staining of pJW-glgC16a transformed JM109 cells

Cultures were grown in M9 medium supplemented with 1% glucose and 0.05% yeast extract, with or without the addition of IPTG, for 4 hours. Cells from 10ml of culture in each case were recovered by centrifugation and resuspended in 200 µl Lugol's iodine (0.2%)

A) Untransformed JM109, grown without IPTG; B) *E. coli* JM109/pJW-glgC16a, grown without IPTG; C) untransformed JM109, grown with IPTG; D) *E. coli* JM109/JM109, grown with IPTG. [Optical Density (at λ600 nm) reached by each culture: A) 0.111; B) 0.118; C) 0.111; D) 0.116.]

3.3 Reducing variability of iodine assay

The iodine assay was used throughout this work as a quick and simple primary test to look for phenotypic changes within the polysaccharide stores of transformed *E. coli*, which, if found, could be further investigated by a wider arsenal of methods. However, during early experiments this assay was found to show highly variable and unstable results. Efforts were therefore made to reduce this variability.

Initially, cultures grown in LB supplemented with IPTG were simply spun down and resuspended in 0.2% Lugol's iodine. LB is the growth medium for bacterial cultures most commonly used in molecular biology, as it permits fast growth for many species including *E. coli*. However, these iodine assays showed a lack of reproducibility, which was suspected to have been due to variability between different LB batches.

M9 is a defined minimal medium for *E. coli*, providing the essential complement of phosphorous, nitrogen and sulphur. M9 medium can be supplemented with a carbon source and other additives such as yeast extract (containing essential amino acids) to produce higher growth rates. When *E. coli* JM109/pJW-glgC16 were grown in M9 minimal medium, supplemented with 1% glucose, 0.05% yeast extract and IPTG, then resuspended in 0.2% Lugol's iodine, the reaction was again observed to fade over time, but the loss of colour proceeded more slowly than for cells grown in LB.

The switch from LB to M9, however, was not found to significantly increase the reproducibility of these experiments. Indeed, the addition of a simple washing step, by resuspending recovered cells in PBS, prior to resuspension in iodine, was found to have a far greater effect. Cultures grown in M9 did generally seem to show darker staining with iodine (see discussion) so this advantage was balanced against the inconvenience of slower growth times when selecting the growth conditions for each experiment.

One possible reason for the observed colour-fading and variability of this assay was the competition of thiol (SH) groups for the binding of iodine (Manonmani & Kunhi, 1999). To counteract this, it was found that the addition of 100 mM $CuSO_4$ solution and 6% H_2O_2 to the suspensions before adding the iodine solution both increased and prolonged the staining, suggesting it was reducing the competition. Although this addition of protective agents modified the results of the assay by intensifying the iodine stain, this intensification was found by Manonmani & Kunhi (1999) and by this work (figure 3.2) to be linear over different polysaccharide concentrations, so that results were still comparable. Furthermore,

as previously mentioned, the iodine assay was only used as a preliminary indicator in this work, and not used for quantitative results. Even with the addition of these modifiers, suspensions were still seen to fade over time. However, this fading was now found to proceed more slowly, taking hours rather than minutes to show a visible change, and as such seemed to eliminate the variability previously observed.

3.4 The effects of boiling and potassium hydroxide on iodine assays

As part of the development of this assay, cells were often lysed prior to resuspension in iodine in an attempt to acquire a purer suspension of the polysaccharides. This was achieved by harvesting cells from culture by centrifugation and then resuspending them first in 30% potassium hydroxide (KOH), boiling them, then washing with ethanol, before repelleting the lysate and suspending it in PBS. When viewed under the microscope, cell lysate from cultures of *Ecoli* JM109/pJW-glgC16 were observed to still contain intact inclusion bodies staining with iodine (figure 3.3 & 3.4). An issue with this technique, however, was that KOH was found to react dramatically with iodine, bleaching even the iodine's original pale yellow colour without the presence of polysaccharide. When polysaccharides were also present, the addition of KOH led to a bleaching of the solution (equal to that of solutions without polysaccharide) within 15 minutes. However, an extra washing step with PBS before the final resuspension in iodine was found to nullify this effect (figure 3.5). Cells lysed with KOH and washed in ethanol were found to undergo slower colour fading when centrifuged and resuspended in iodine (figure 3.6), possibly again due to a reduction in amino acid groups that would compete with the polysaccharide for iodine binding.

Iodine assays performed on cells lysed with KOH, washed and resuspended in PBS with $CuSO_4$ solution and H_2O_2 prior to the addition of Lugol's iodine showed a reduction in variability as well as a marked increase in the time they took to fade (figure 3.6). After such treatment, *E. coli* Jm109/pJW-glgC16a invariably stained far darker than a pJW-lacZ control and the stain held for easily long enough to obtain qualitative results.

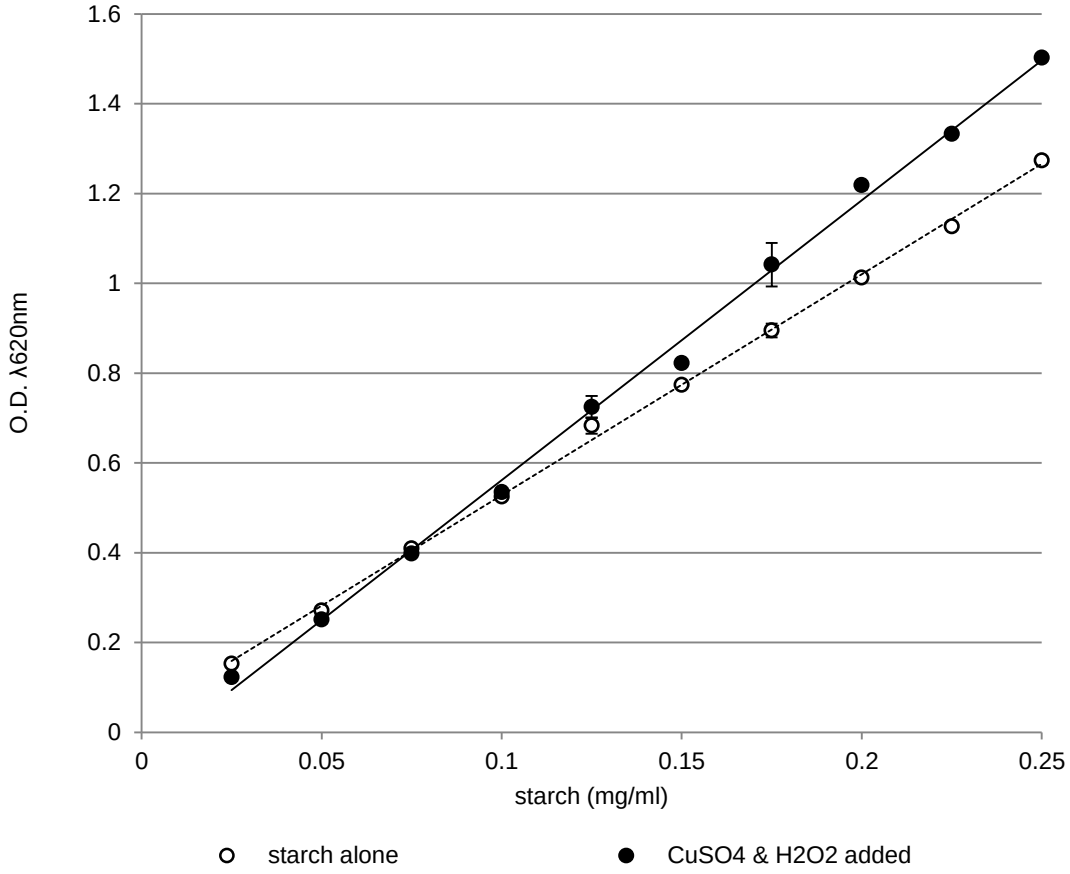


Figure 3.2: The effect of $CuSO_4$ and H_2O_2 as protective agents on the intensity of staining of starch and iodine reactions

1 ml starch solution at different concentrations was mixed with 25 µl Lugol's iodine (0.2%) with or without the prior addition of 50 µl $CuSO_4 \cdot 5H_2O$ (100 mM) and 50 µl H_2O_2 (6%). The optical density of each culture was measured at λ620 nm. Error bars represent the standard error of the mean when n=3. The soluble starch was from 'Scientific Laboratory Supplies' (CHE3620).

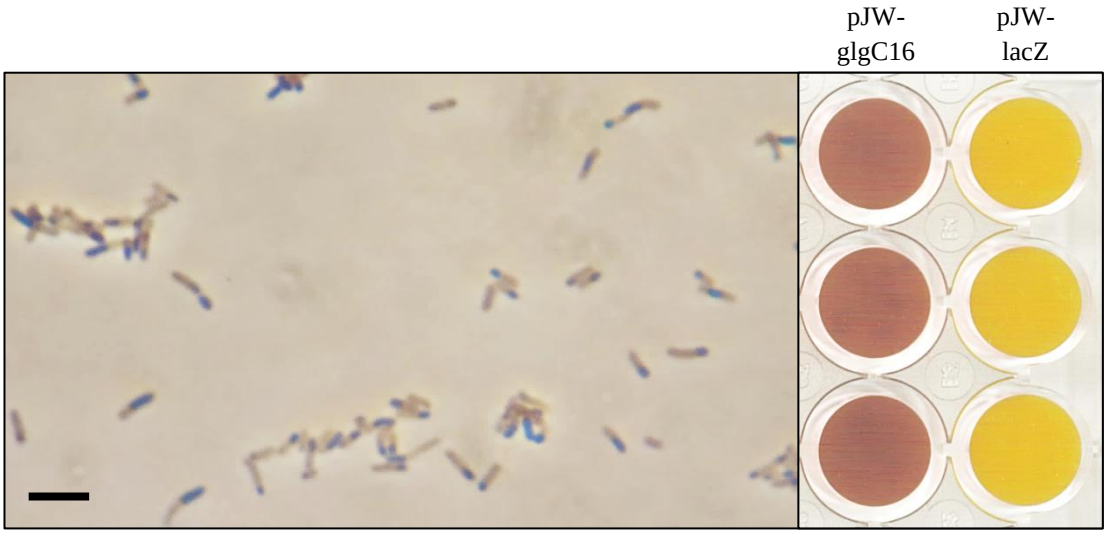


Figure 3.3: Phase contrast light microscopy showing 'inclusion body' phenotype in iodine stained *E. coli* JM109/pJW-glgC16 cells and positive iodine assay results of the same cultures, prior to lysis with KOH

Cultures were grown overnight in LB supplemented with IPTG and 1% lactose, at 37°C and 200 rpm. Cultures imaged under light microscopy. Scale bar represents 5 µm (approx.). Cells from 1 ml of culture in each case were also recovered by centrifugation, washed and resuspended in 200 µl PBS, before adding 50 µl $CuSO_4 \cdot 5H_2O$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%), then transferred to the wells of a 48-well plate and imaged using a flatbed scanner.

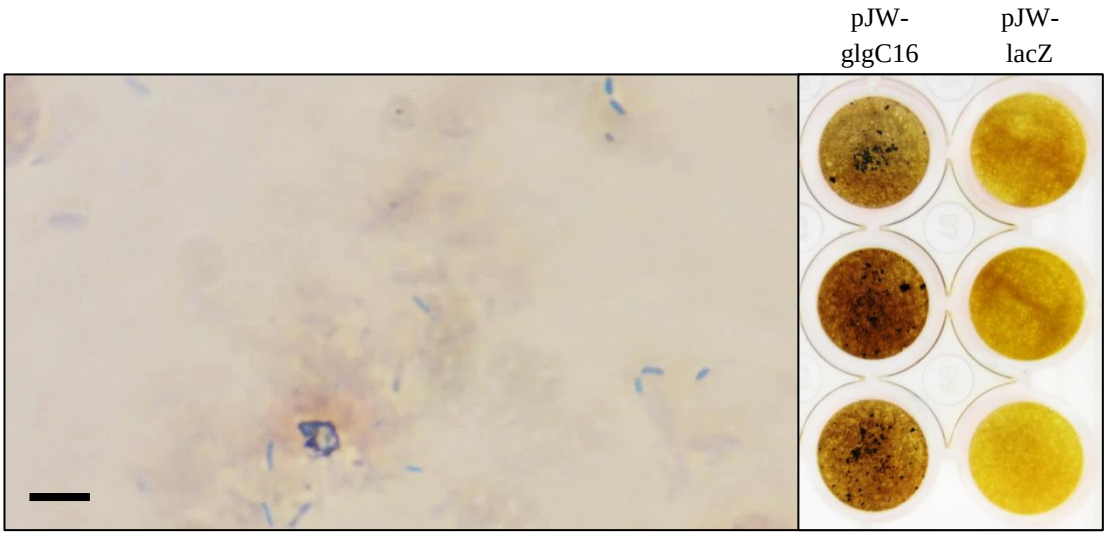


Figure 3.4: Phase contrast light microscopy showing 'inclusion body' phenotype in iodine stained *E. coli* JM109/pJW-glgC16 lysate and positive iodine assay results of the same cultures, post lysis with KOH

The same cultures as above, after lysis with 30% KOH and 5 minutes boiling. Scale bar represents 5 µm (approx.).

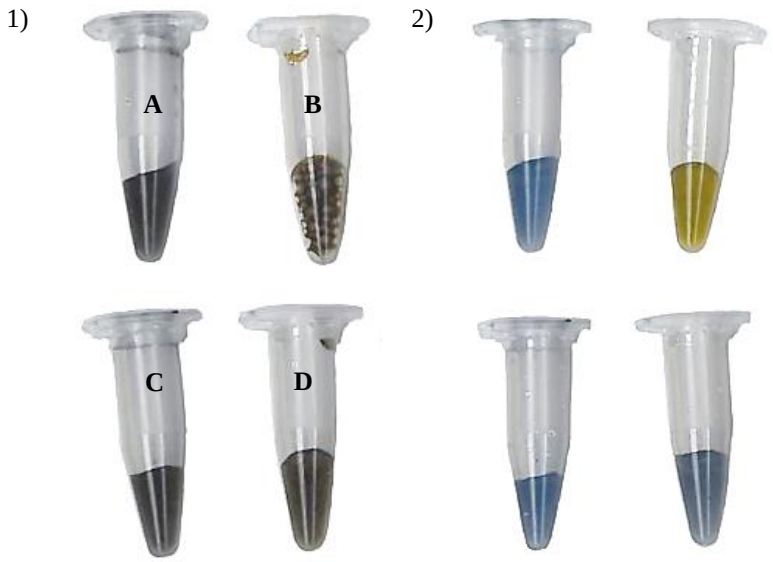


Figure 3.5: Iodine staining of starch suspensions, showing the effects of KOH

Pure starch (50 ng) was subjected to different treatments prior to the addition of 50 µl $CuSO_4 \cdot 5H_2O$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%).

Treatments:

1) A) suspended in 200 µl PBS and boiled (5 min); B) suspended in KOH (30%) and boiled (5 min). Washed with ethanol, centrifuged and resuspended in 200 µl PBS; C) suspended in 200 µl PBS and boiled (5 min). Centrifuged and resuspended in 200 µl PBS; D) suspended in KOH (30%) and boiled (5 min). Washed with ethanol, centrifuged and resuspended in 200 µl PBS. Centrifuged again and resuspended in 200 µl PBS.

2) The same tubes 24 hours later.

Soluble starch obtained from 'Scientific Laboratory Supplies' (CHE3620).

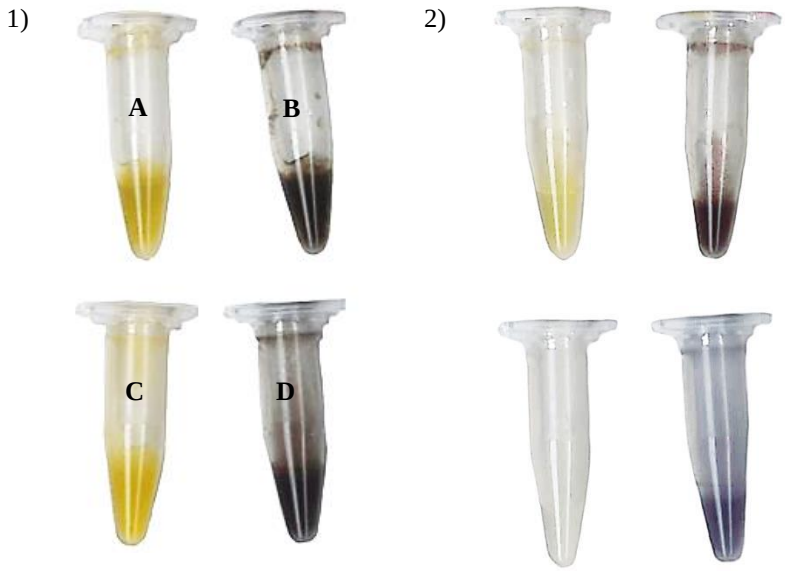


Figure 3.6: Iodine staining of pJW-lacZ and pJW-glgC16 transformed JM109 cells, showing the effects of lysis with KOH and precipitation with ethanol prior to staining
Cultures were grown overnight in M9 medium supplemented with 1% lactose, 0.05% yeast extract and IPTG. Cells from 1.4 ml of culture in each case were recovered by centrifugation, treated as described below and finally resuspended in 200µl PBS, before adding 50 µl $CuSO_4 \cdot 5H_2O$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%).

1) A) *E. coli* JM109/pJW-lacZ cells lysed in 30% KOH, boiled (5 min), washed with ethanol, washed with PBS; B) *E. coli* JM109/pJW-glgC16 cells, lysed in 30% KOH, boiled (5 min), washed with ethanol, washed with PBS; C) *E. coli* JM109/pJW-lacZ cells, suspended in 200 µl PBS and boiled (5 min); D) *E. coli* JM109/pJW-glgC16 cells, suspended in 200 µl PBS and boiled (5 min).
2) The same tubes, 24 hours later.

3.5 Inhibition of the lac promoter by glucose

Because the transgenes under investigation were expressed via a *lac* promoter, it was suspected that the addition of glucose as a carbon source may inhibit their expression, as glucose is a known inhibitor of the *lac* promoter. This was confirmed with a β-galactosidase assay (figure 3.7), where the activity of *E. coli* JM109/pJW-lacZ grown in LB supplemented with IPTG and glucose was significantly lower than that of cells grown in LB supplemented with IPTG alone or with IPTG and lactose ($p<0.05$). However, the activity of the cells grown with glucose and IPTG was still higher than that for cells grown in LB alone ($p<0.05$), and no significant difference was found between the activity of cells grown in the presence of lactose or glucose (without the addition of IPTG) ($p>0.05$).

Microscopy of cells transformed with pJW-glgC16 largely supported these findings, with cultures grown in either LB or M9, supplemented with lactose, showing a high proportion of cells with the 'inclusion body' phenotype, while those grown in media supplemented with glucose appear to show fewer, smaller, inclusion bodies (figure 3.8). This suggests that the transgenes are more strongly expressed in the presence of lactose, but are still not entirely repressed by glucose in the growth medium, despite the fact that the β-galactosidase assay showed no significant difference in the activity levels of *E. coli* JM109/pJW-lacZ grown in LB alone versus those grown in LB supplemented with either lactose or glucose ($p>0.05$). The inclusion bodies in these instances would therefore seem to be more the result of the high sugar content of the medium than of the high expression levels of the transgene (although control cultures of *E. coli* JM109/pJW-lacZ grown in LB supplemented with lactose did not show the inclusion body phenotype, supporting the theory that the glgC16 transgene is still necessary for this phenotype).

Subsequent cultures were still supplemented with lactose instead of glucose, since although glucose did not entirely repress the expression of transgenes under the *lac* promoter it did lead to significantly lower expression than lactose when in conjunction with IPTG. Lactose was also suspected to have the additional benefit of further upregulating storage polysaccharide accumulation (see final discussion).

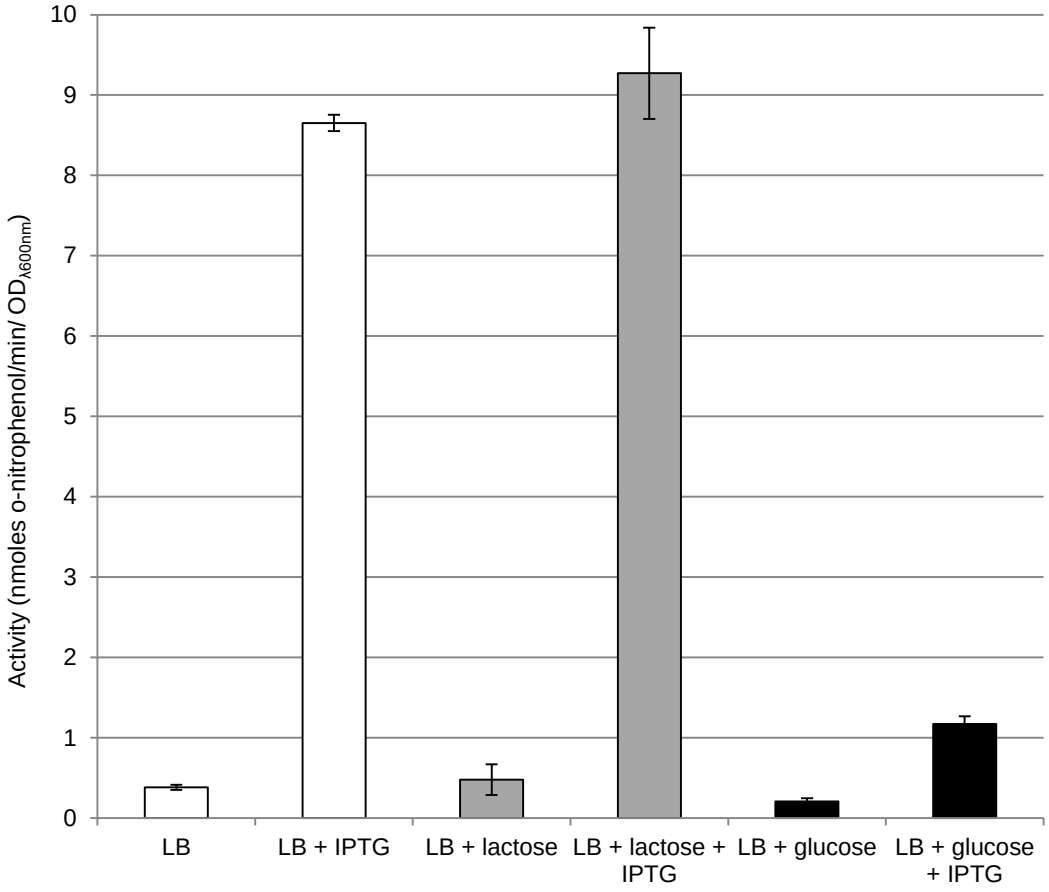


Figure 3.7: β-Galactosidase assay on *E. coli* JM109/pJW-lacZ grown in LB with different supplements

Cultures were grown for 2 hours 10 minutes, at 37°C and 200rpm, to $OD_{\lambda 600nm}$ of ~0.3. Error bars represent the standard error of the mean when n=3.

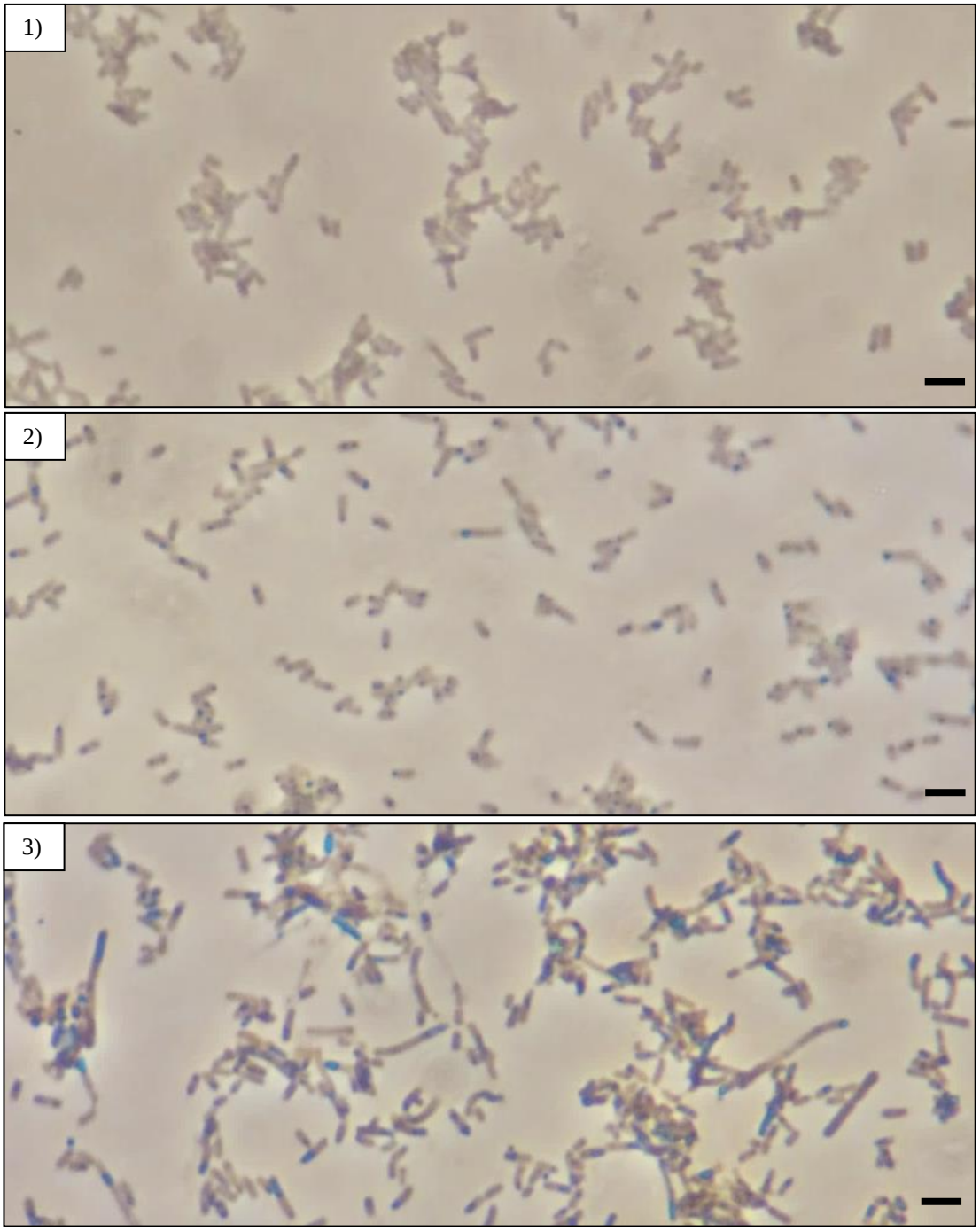


Figure 3.8: Phase contrast light microscopy showing 'inclusion body' phenotype in iodine stained pJW-glgC16 transformed JM109 cells

Cultures transformed with pJW-lacZ or pJW-glgC16 were grown overnight in LB medium supplemented with either glucose or lactose. Scale bars represent 5 µm (approx).

1) *E. coli* JM109/pJW-lacZ grown in lactose; 2) *E. coli* JM109/pJW-glgC16 grown in glucose, showing small inclusion bodies; 3) *E. coli* JM109/pJW-glgC16 grown in lactose, showing inclusion bodies

3.6 SDS-PAGE analysis

To determine whether the inclusion bodies observed under the microscope were composed of proteins, perhaps caused by the overexpression of an inactive enzyme, an SDS-polyacrylamide gel (SDS-PAGE) analysis was performed on cultures of *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgC16, grown in LB with the addition of lactose and IPTG to stimulate inclusion body formation (figure 3.9). Samples from the cultures assayed by SDS-PAGE were also stained with 5% Lugol's iodine and viewed under a light microscope, which confirmed the presence of inclusion bodies in *E. coli* JM109/pJW-glgC16 grown with lactose and IPTG. *E. coli* JM109/pJW-glgC16 grown with IPTG but without additional sugar were also seen to contain inclusion bodies, though to a lesser degree. Those cells also showed visual evidence of flocculation and filamentation, indicating possible cell division defects, although this was not confirmed (figure 3.10). An iodine assay was also performed on 1ml of recovered culture in each case (figure 3.11) and showed a far more intense colour change for *E. coli* JM109/pJW-glgC16 than for *E. coli* JM109/pJW-lacZ, where that colour was at its most intense when cells were grown in the presence of IPTG and lactose, supporting the results of the microscopy. The colour of the iodine complex was found to be dark orange-brown (as it consistently was in the improved assays), suggesting a high concentration of polysaccharide containing relatively short stretches of single-helices.

The *E. coli* GlgC enzyme is a homotetrameric structure, composed of four units of around 50 kDa each (Ballicora et al., 2007). The resulting protein pattern from the SDS-PAGE analysis showed the presence of one or two intense bands at around this size in *E. coli* JM109/pJW-glgC16 compared to the control, *E. coli* JM109/pJW-lacZ, only under growth in IPTG without lactose. However, the inclusion bodies observed were most numerous in cultures grown in IPTG and lactose, in which there is no clear difference between the SDS-PAGE for the two strains.

The SDS-PAGE was therefore considered as evidence against the hypothesis that the inclusion bodies visible under the microscope were composed of a single over-accumulating protein, as no band considered intense enough to account for them were observed within the assays of those cells showing the strongest inclusion body phenotype. The identity of the intense bands seen in the other samples remains unknown.

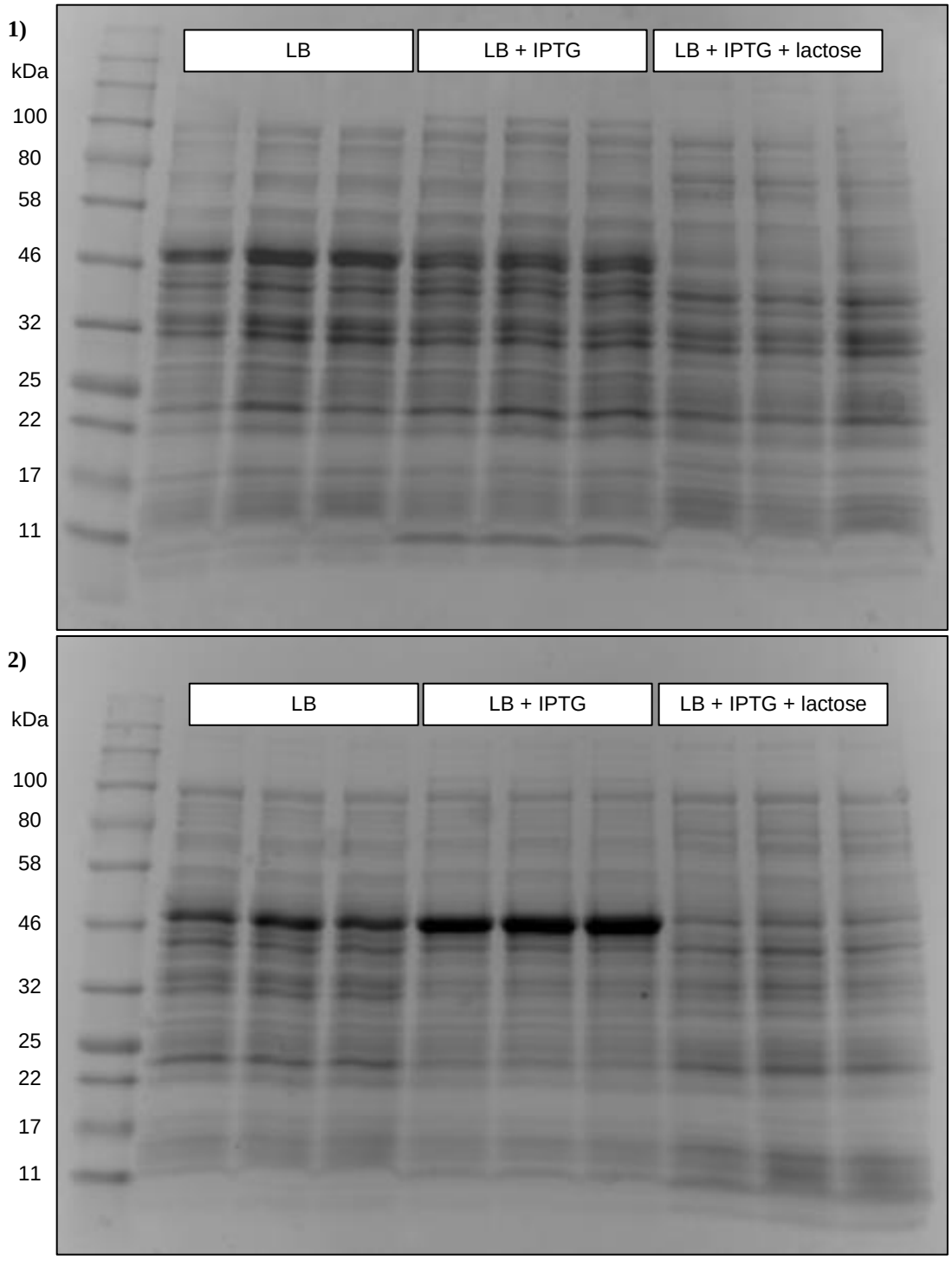


Figure 3.9: SDS-PAGE of transformed JM109 grown with or without IPTG and lactose

1) biological triplicates of SDS-PAGE of *E. coli* JM109/pJW-lacZ grown in three media types

2) biological triplicates of SDS-PAGE of *E. coli* JM109/pJW-glgC16 grown in three media types. Each well was loaded with 30 µg of protein. Data is representative of several independent experiments.

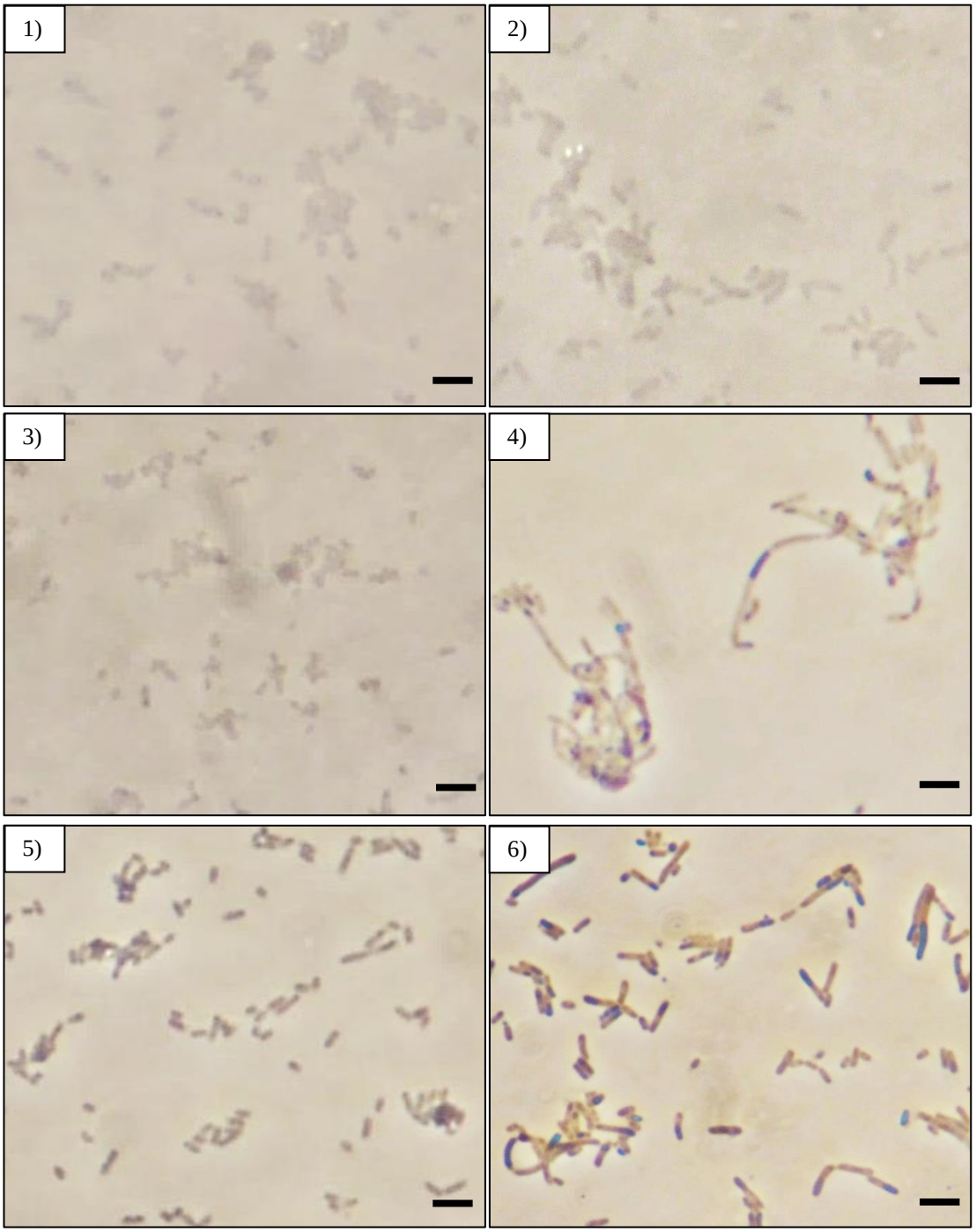


Figure 3.10: Phase contrast light microscopy of iodine stained, transformed JM109 cells for SDS-PAGE

1) *E. coli* JM109/pJW-lacZ grown in LB; 2) *E. coli* JM109/pJW-glgC16 grown in LB; 3) *E. coli* JM109/pJW-lacZ grown in LB with IPTG; 4) *E. coli* JM109/pJW-glgC16 grown in LB with IPTG, showing cells that show evidence of flocculation and filamentation. Some inclusion bodies are also visible; 5) *E. coli* JM109/pJW-lacZ grown in LB with IPTG and lactose; 6) *E. coli* JM109/pJW-glgC16 grown in LB with IPTG and lactose, showing abundant inclusion bodies. Scale bars represent 5 µm (approx.). The iodine stain gives poor resolution for cells not grown in high sugar media.

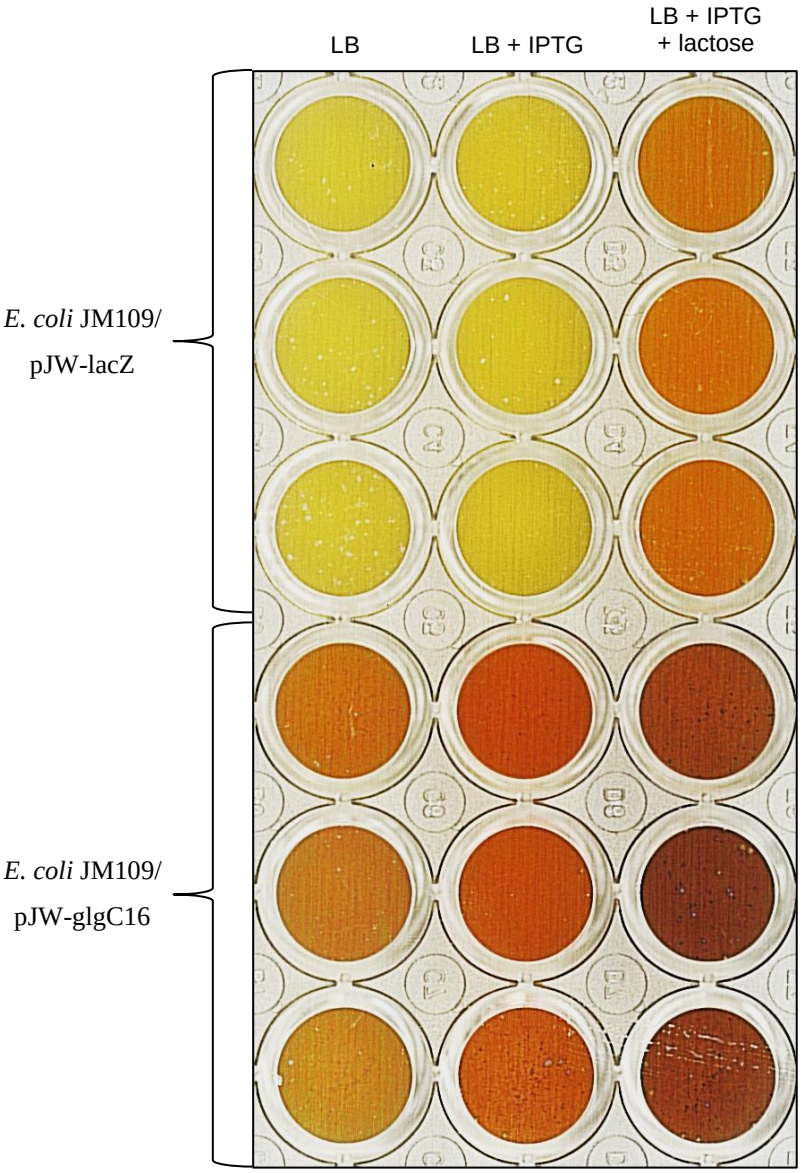


Figure 3.11: Iodine assay of transformed JM109 cells for SDS-PAGE

Cells from 1 ml of culture in each case were recovered by centrifugation, washed and resuspended in 200 µl PBS, before adding 50 µl $CuSO_4 \cdot 5H_2O$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%), then transferred to the wells of a 48-well plate and imaged using a flatbed scanner.

3.7 Analysis of pJW-glgC16 transformed cells by TEM

To visualise more clearly the inclusion bodies formed in JM109 with upregulated *glgC*, they were viewed by transmission electron microscopy. JM109 transformed with either the pJW-lacZ control plasmid or the pJW-glgC16 plasmid were grown in 50 ml cultures of LB supplemented with lactose and IPTG, overnight, before the entire cultures were pelleted and fixed for imaging.

Analysis by TEM suggested that *E. coli* JM109/pJW-glgC16 were phenotypically different from the control cells, in that around a quarter of them seemed to contain some sort of clearly defined inclusion, negatively stained after treatment with uranyl acetate and lead citrate (figure 3.13). Although several of the control cells also appeared to contain negatively-stained inclusive matter, the inclusions of *E. coli* JM109/pJW-glgC16 are considered by this study to be distinct, in that they seem more defined and compact than anything observed within the control cells, also often being localized in the poles of the cell (figure 3.12). This corresponds, at least, with the general peripheral localisation of normal *E. coli* glycogen granules as well as the GlgC and GlgA enzymes, as identified in previous studies (Wilson et al., 2010; Morán-Zorzano et al., 2007).

The relative abundance of the apparent inclusion bodies identified within the TEM images of the *E. coli* JM109/pJW-glgC16 would also suggest that they are the same phenomenon previously identified under light microscopy after staining with iodine. It seems clear, however, that these inclusion bodies do not have a starch-like structure. Previous groups have shown starch granules to have a smooth surface appearance when double-stained with uranyl acetate and lead citrate and viewed under TEM. That smooth appearance has been further revealed by incubation with α-amylase or hydrolysis with HCl to be the result of an underlying tiered structure, as discussed in the introduction (Deschamps et al., 2008a; Smith, 1999; Ral et al., 2004). In contrast to this, the inclusion bodies identified in this study show a far more aggregate appearance, as if roughly assembled (figure 3.12 & 3.14). However, the inclusions also do not seem similar to normal bacterial glycogen, as this tends to positively stain when treated with uranyl acetate and lead citrate (Wilson et al., 2010; Morán-Zorzano et al., 2007).If they are indeed composed of polysaccharide, as hypothesised, these inclusions seem to show a stronger similarity to glycogen α-particles, found predominantly within animal livers, though also observed within insect flight muscles, rat muscle and mouse cardiac tissue (Sullivan et al., 2010 & 2012), although this similarity has not been investigated further.

***E. coli* JM109/pJW-lacZ** ***E. coli* JM109/pJW-glgC16**

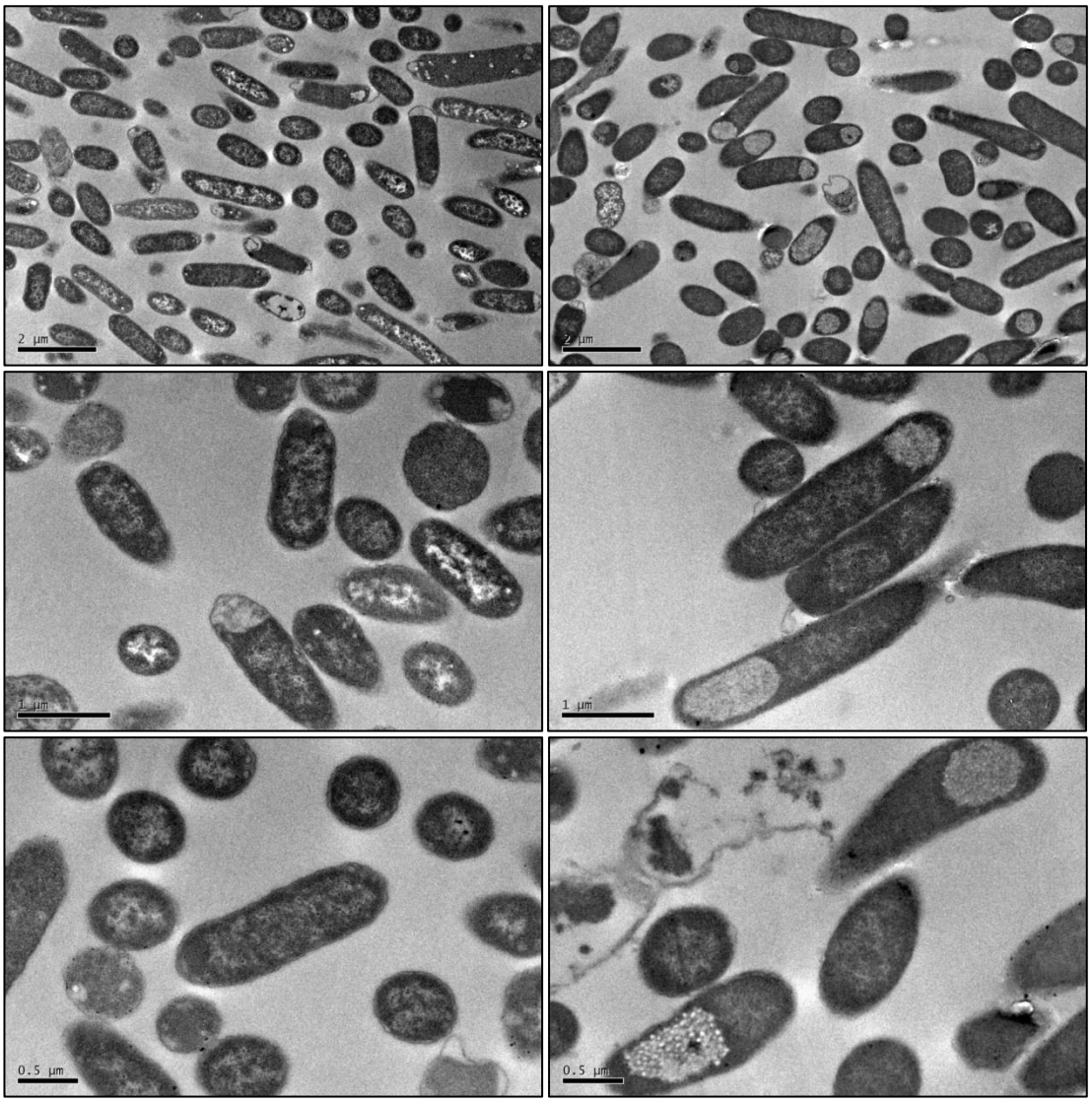


Figure 3.12: Transmission electron micrographs (TEMs) of *E. coli* JM109/pJW-lacZ (left) and *E. coli* JM109/pJW-glgC16 (right)

Cultures were grown overnight in LB supplemented with lactose and IPTG, sectioned from Araldite resin and negatively stained with uranyl acetate and lead citrate.

Many of the *E. coli* JM109/pJW-lacZ cells show negatively-stained inclusive matter. However, it is considered that the defined granular bodies visible within the *E. coli* JM109/pJW-glgC16 are distinct, and correspond to those bodies found to stain with iodine when viewed by light microscopy. Scale bars represent: top images 2 µm; middle images 1 µm; bottom images 0.5 µm Images are selected from 14 *E. coli* JM109/pJW-lacZ and 20 *E. coli* JM109/pJW-glgC16 transmission electron micrographs captured, and are considered to be representative.

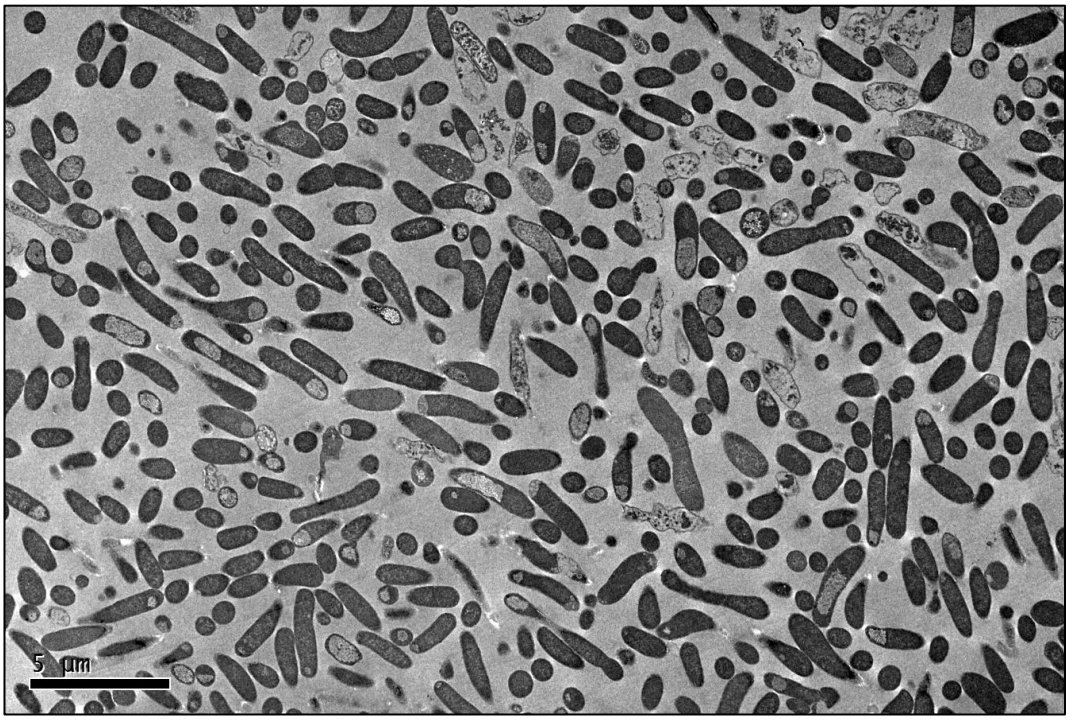


Figure 3.13: TEM of *E. coli* JM109/pJW-glgC16

A wide-field image is provided to give an indication of the propensity of the 'inclusion body' phenotype. A visual analysis of the image suggests it contains 378 valid cells, of which 98 are considered to show the 'inclusion-body'phenotype, giving an approximate ratio of 26%. Scale bar represents 5 µm.

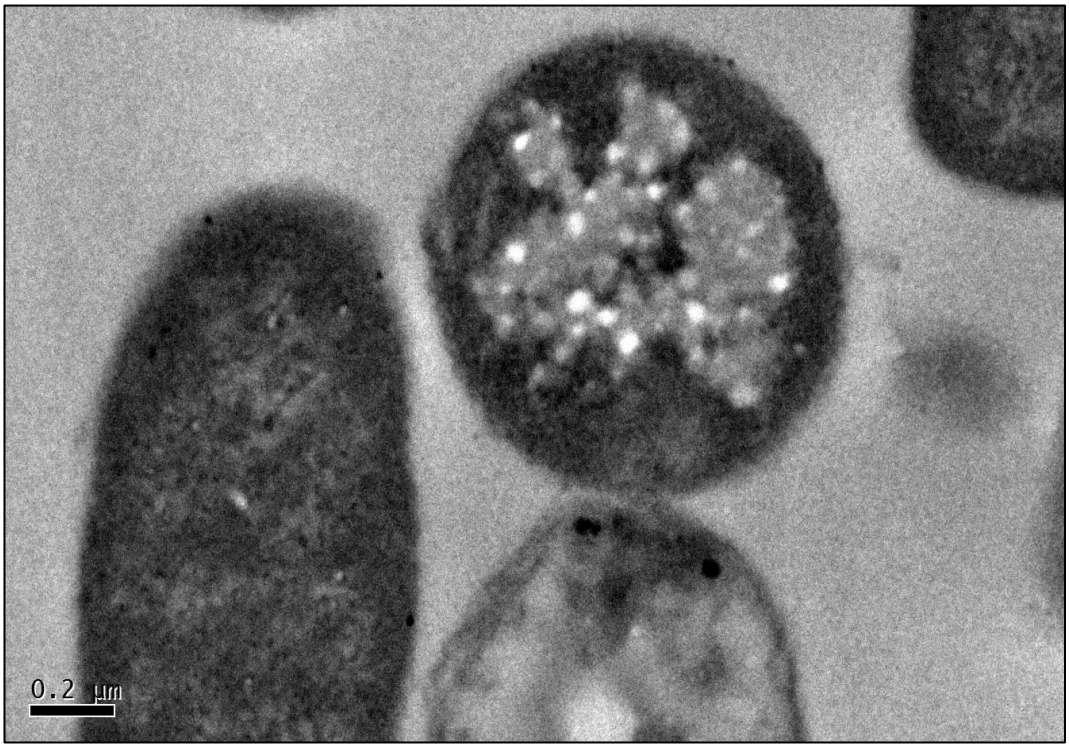


Figure 3.14: Close-up TEM of an 'inclusion body' phenotype *E. coli* JM109/pJW-glgC16, showing the apparent aggregate nature of the inclusion

Scale bar represents 0.2 µm.

3.8 Anthrone assays of *glgC16*-transformed cells

Since analysis by SDS-PAGE suggested that the inclusion bodies described were not protein aggregates, anthrone assays were carried out on cultures of *E. coli* JM109/pJW-glgC16, to discover their total sugar content. If no significant difference was found between the total sugar content of these inclusion-body-containing cells and a control group, it would also suggest that the inclusion bodies were not composed of polysaccharide.

Cultures of JM109 transformed with either pJW-lacZ or pJW-glgC16 were grown overnight in LB supplemented with lactose and IPTG. The 50 ml cultures were then harvested by centrifugation, washed and resuspended in phosphate buffer. The supernatant was checked for polysaccharides using iodine, which showed no colour change (data not shown). Microscopic examination confirmed that inclusion bodies were present in the cultures of *E. coli* JM109/pJW-glgC16 but not in the control cultures. A lysis step was deemed unnecessary in the preparation of the cells, since the anthrone assay itself involved dissolving them in 98% sulphuric acid. The assay therefore gives the total sugar content of the cells, including soluble sugars, rather than focusing specifically on the insoluble sugar of which the inclusion bodies should make up a large proportion.

By the end of the growth period the control cultures had reached a higher optical density, with an average $OD_{\lambda 600nm}$ of 5.21 (+/- 0.1), compared to the *E. coli* JM109/pJW-glgC16 cultures, which had grown to an average $OD_{\lambda 600nm}$ of 4.04 (+/- 0.2) (data not shown), a growth trend that had been noticed throughout this work and is investigated further in the next chapter. The optical densities of all cultures were therefore equalised to an approximate $OD_{\lambda 600nm}$ of 4 (±0.08) prior to the anthrone assay, to achieve a more relative indication of total hexose sugar contents.

The results of the anthrone assay suggest the *E. coli* JM109/pJW-glgC16 to have a significantly higher hexose sugar content than the control cells ($p<0.05$), reaching an average glucose equivalent of 71.81 µg/ml versus 40.19 µg/ml for the controls (figure 3.15).

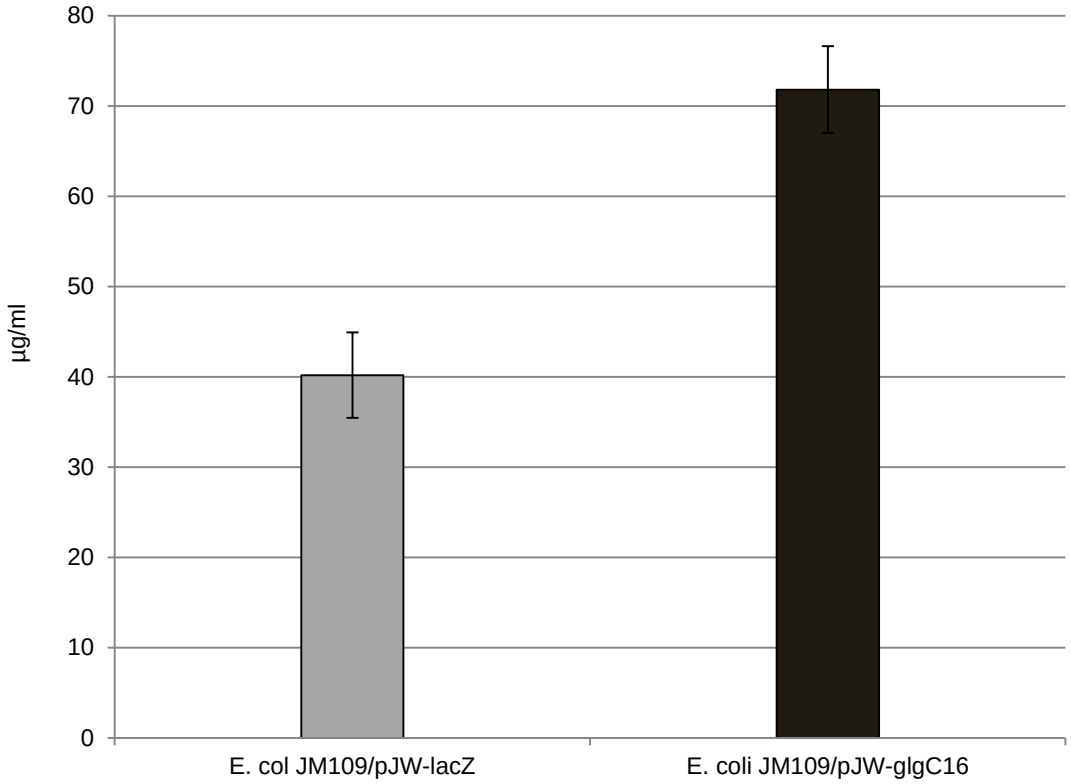


Figure 3.15: Total hexose sugar content of *E. coli* JM109/pJW-lacZ versus *E. coli* JM109/pJW-glgC16

From anthrone assays performed on *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgC16, grown to stationary phase in LB supplemented with lactose and IPTG, then equalized to $OD_{\lambda 600nm}$ 4.0. Error bars represent the standard error of the mean when n=3.

3.9 Effects of upregulating *glgC* and *glgB*

To investigate how much of this phenomenon was due to the AMP-resistance of the altered form of GlgC (GlgC16) used, pJW-glgC was constructed (see Methods 2.7), identical to pJW-glgC16 but bearing wildtype glgC. This was successfully cloned from genomic DNA of the same JM109 strain of *E. coli* that was used in experiments (figure 3.16), and the validity of the constructs was confirmed by sequencing. When grown in medium supplemented with lactose and IPTG, cells containing this construct were again unfailingly found to accumulate inclusion bodies which stained with 5% Lugol's iodine (figure 3.19).

It was therefore hypothesised that the excess building blocks of ADP-glucose were allowing the GlgA enzyme to reach its otherwise unexpressed potential and join glucans together at a far higher rate than it normally does. Extending this hypothesis, it was considered that the branching enzyme might be unable to keep up, so that the α-1,4 linked chains of the glycogen molecule are extending beyond their normal length. If this were the case, they might start to spontaneously form helices, either as extended single helices or double-helices with their neighbours. The colour observed from iodine assays of these transformants tended to show as intense dark brown, suggesting an overall reduction in the lengths of single helices within the structures, as compared to wild type glycogen, but an increase in the proportion of those structures. If the hypothesis was correct, this would suggest that the extended chains are mostly forming double-helices, though this remains hugely speculative.

To try to test the hypothesis, another plasmid was constructed which also contained the *glgB* gene, cloned from genomic DNA of the same JM109 strain of *E. coli* that was used in experiments (figure 3.17 and figure 3.18, also see Methods 2.7). The addition of *glgB* under the same expression as *glgC* was expected to upregulate the branching enzyme to roughly the same degree as the ADP-glucose pyrophosphorylase. If the hypothesis is correct, the following phenotypes would be expected: the cells should show the same sugar-excess phenotype, since the rate-limiting step of glycogen synthesis has still been upregulated. However, by upregulating the branching enzyme to the same degree, branches are hypothesised to occur as frequently as in wild-type glycogen, thereby preventing the formation of novel helices. This should also result in an iodine reaction that is similar to that of wild-type glycogen. Lastly, it was hypothesised that this additional branching would bring back the steric interference that – at least theoretically – stops the growth of glycogen granules once they reach a certain size, thereby also abolishing the 'inclusion body' phenotype.

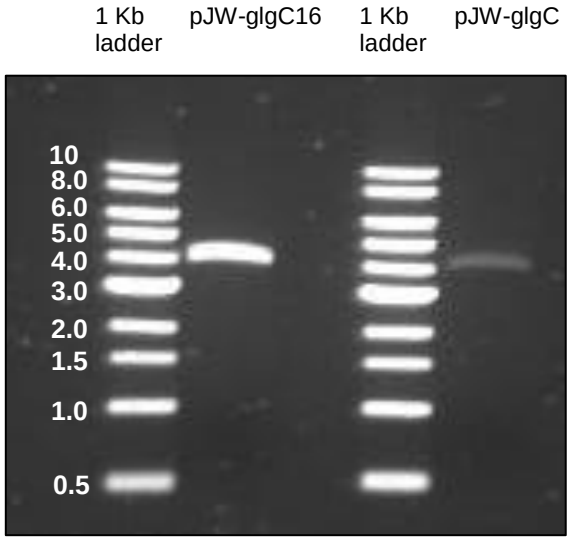


Figure 3.16: Gel electrophoresis of linear pSB1C3 constructs containing *glgC* and *glgC16*
Plasmid DNA was prepared from *E. coli* JM109/pJW-glgC and *E. coli* JM109/pJW-glgC16 and digested with EcoRI prior to gel electrophoresis. Successful cloning was further confirmed by sequencing.
Expected sizes: pJW-glgC16 – 3993 bp; pJW-glgC – 3993 bp.

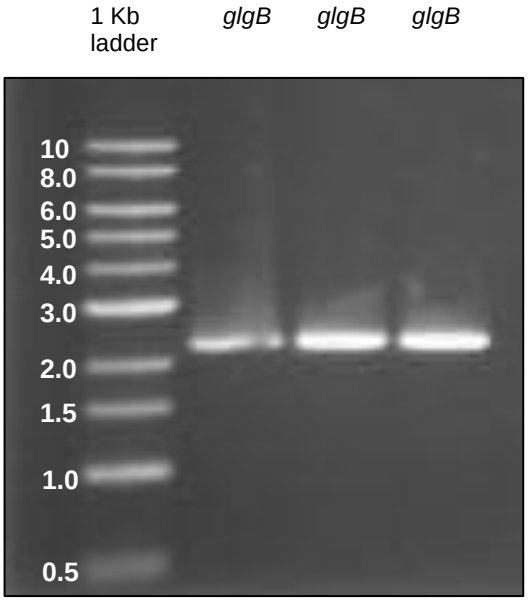


Figure 3.17: Gel electrophoresis of the PCR product of the *glgB* gene
Successful cloning was further confirmed by sequencing.
Expected sizes: glgB – 2186 bp

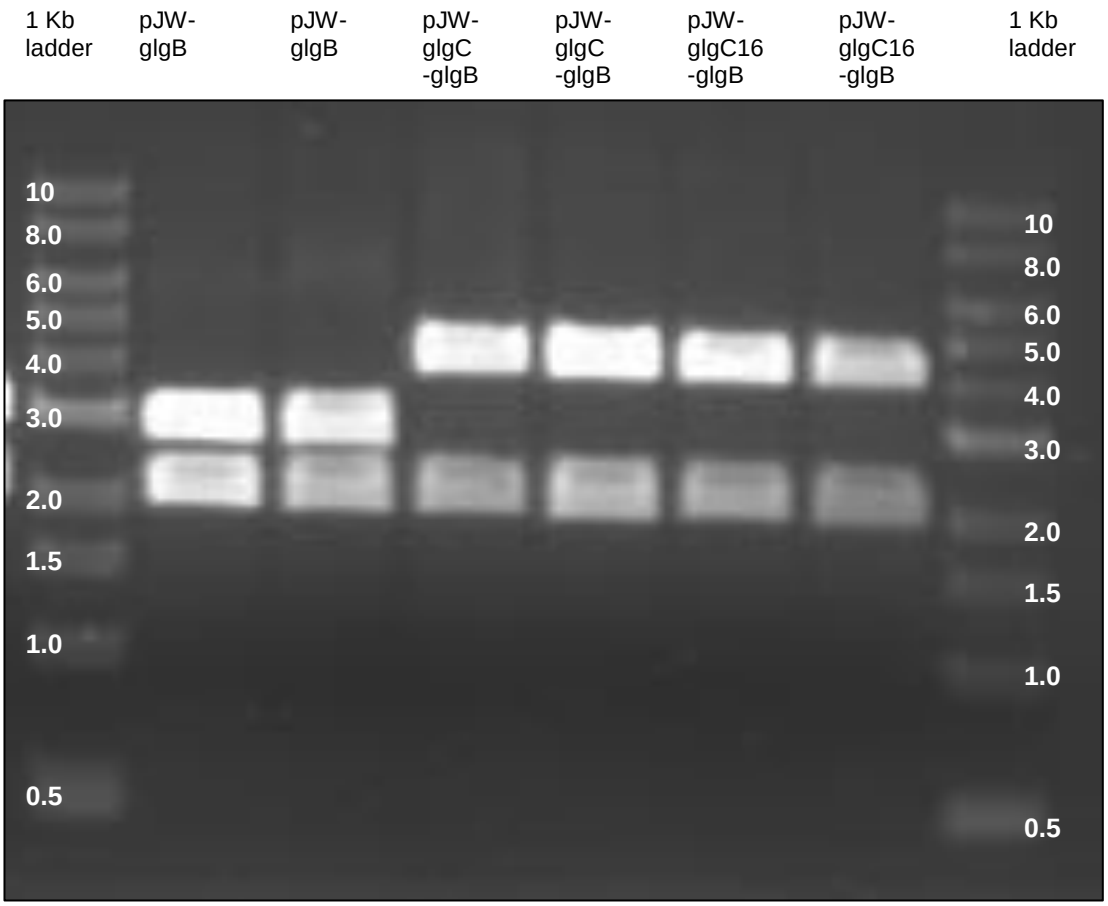


Figure 3.18: Gel electrophoresis of pSB1C3 constructs containing *glgB*, alone and in tandem with *glgC* or *glgC16*

Plasmid DNA was prepared from transformed cells and digested with XbaI and PstI prior to gel electrophoresis. All constructs were confirmed by sequencing.

Expected sizes: pSB1C3 – 2070 bp; lacZ-glgB – 2857 bp; lacZ-glgC-glgB – 4162 bp; lacZ-glgC16-glgB – 4162 bp;

3.10 Phenotypes of pJW-*glgCB* transformed cells

As predicted, *E. coli* JM109/pJW-glgCB grown in LB supplemented with lactose (1%) and yeast extract (0.05%) did not show the presence of inclusion bodies when stained with Lugol's iodine and viewed under the microscope, while *E. coli* JM109/pJW-glgC grown in the same conditions did. Cells transformed with the pJW-glgB plasmid appeared the same as the control under the microscope (figure 3.19).

Iodine assays of the constructs also met with predictions. The *E. coli* JM109/pJW-glgCB in fact stained consistently darker than the control cells, however this was with the rust-red colour more suggestive of high levels of glycogen than the very dark brown colour of the *E. coli* JM109/pJW-glgC. *E. coli* JM109/pJW-glgB, meanwhile, tended to stain a paler yellow than the control cells, suggestive of even fewer binding sites for the iodine, perhaps as a result of slightly increased branching and a subsequent shorter average chain length within its polysaccharides (figure 3.20).

Although both predictions were therefore met, it was additionally found that while *E. coli* JM109/pJW-glgC and *E. coli* JM109/pJW-glgCB showed slower growth than the control, the *E. coli* JM109/pJW-glgCB unfailingly grew to a higher optical density (figure 3.22. Also see chapter 4). Furthermore, the total hexose sugar content of the *E. coli* JM109/pJW-glgCB, rather than being as high as was found in the *glgC* upregulated mutants, was over twice as high, after the culture densities had been equalized (figure 3.21 & 3.23). The results of the initial anthrone assay (figure 3.21), from cultures grown to an $OD_{\lambda 600nm}$ of 1.0 in LB supplemented with lactose (1%) and IPTG, showed a significantly higher hexose sugar content for *E. coli* JM109/pJW-glgC over *E. coli* JM109/pJW-lacZ (though by a smaller degree than seen before, which is thought to be due to the relatively short incubation time for these cultures) and also for *E. coli* JM109/pJW-glgCB over both *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgC ($p<0.05$ in all cases). No significant difference was found between *E. coli* JM109/pJW-glgB over *E. coli* JM109/pJW-lacZ ($p>0.05$) which remains consistent with their similar appearance and reaction to iodine.

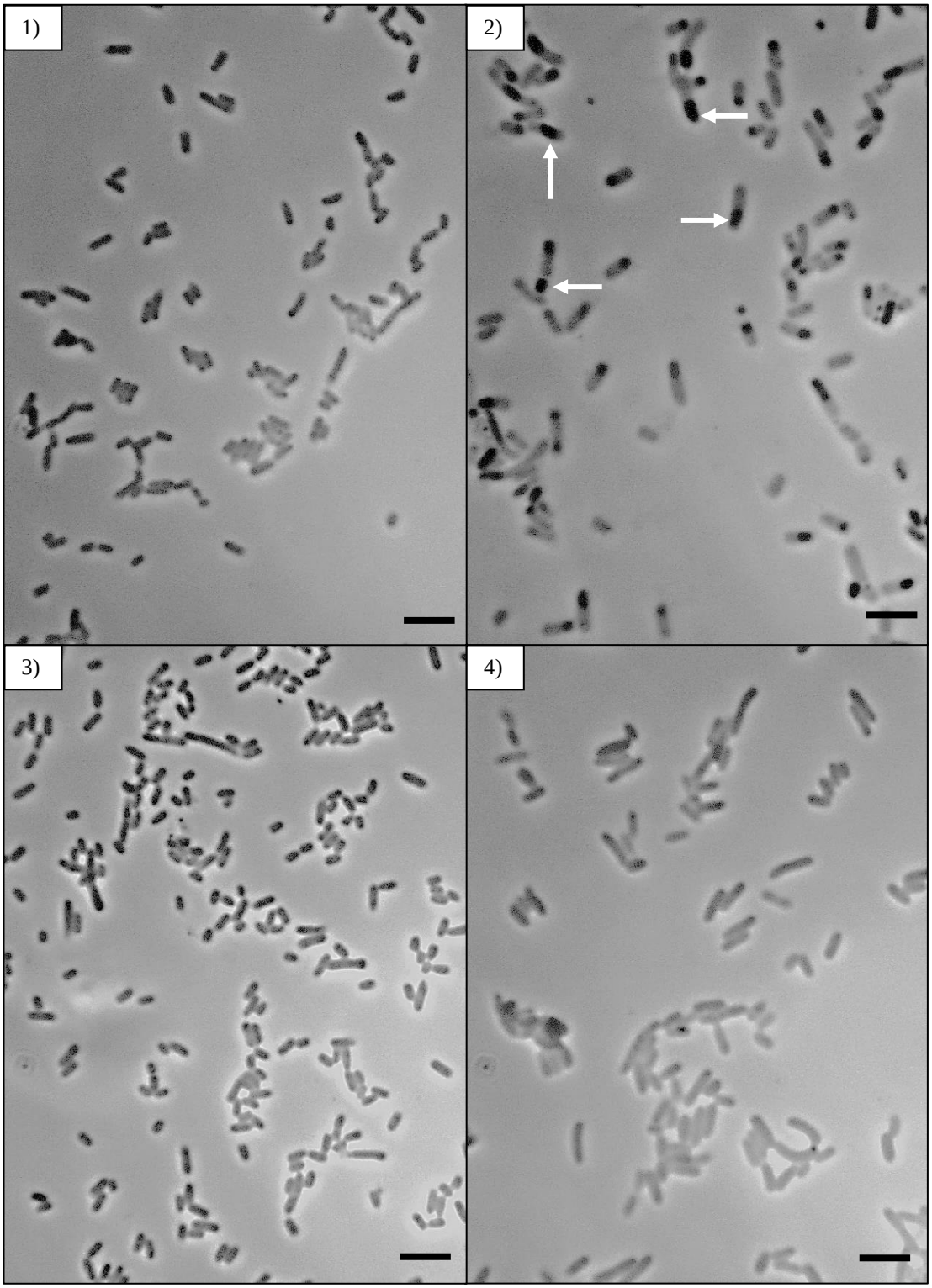


Figure 3.19: Phase contrast light microscopy of iodine-stained, transformed JM109 cells

Cultures grown to stationary phase in LB supplemented with lactose & IPTG

1) *E. coli* JM109/pJW-lacZ; 2) *E. coli* JM109/pJW-glgC; 3) *E. coli* JM109/pJW-glgB;

4) *E. coli* JM109/pJW-glgCB.

Arrows point to inclusion bodies within *E. coli* JM109/pJW-glgC cells. None were visible in the cells of any other transformants. Scale bars represent 5 µm (approx)

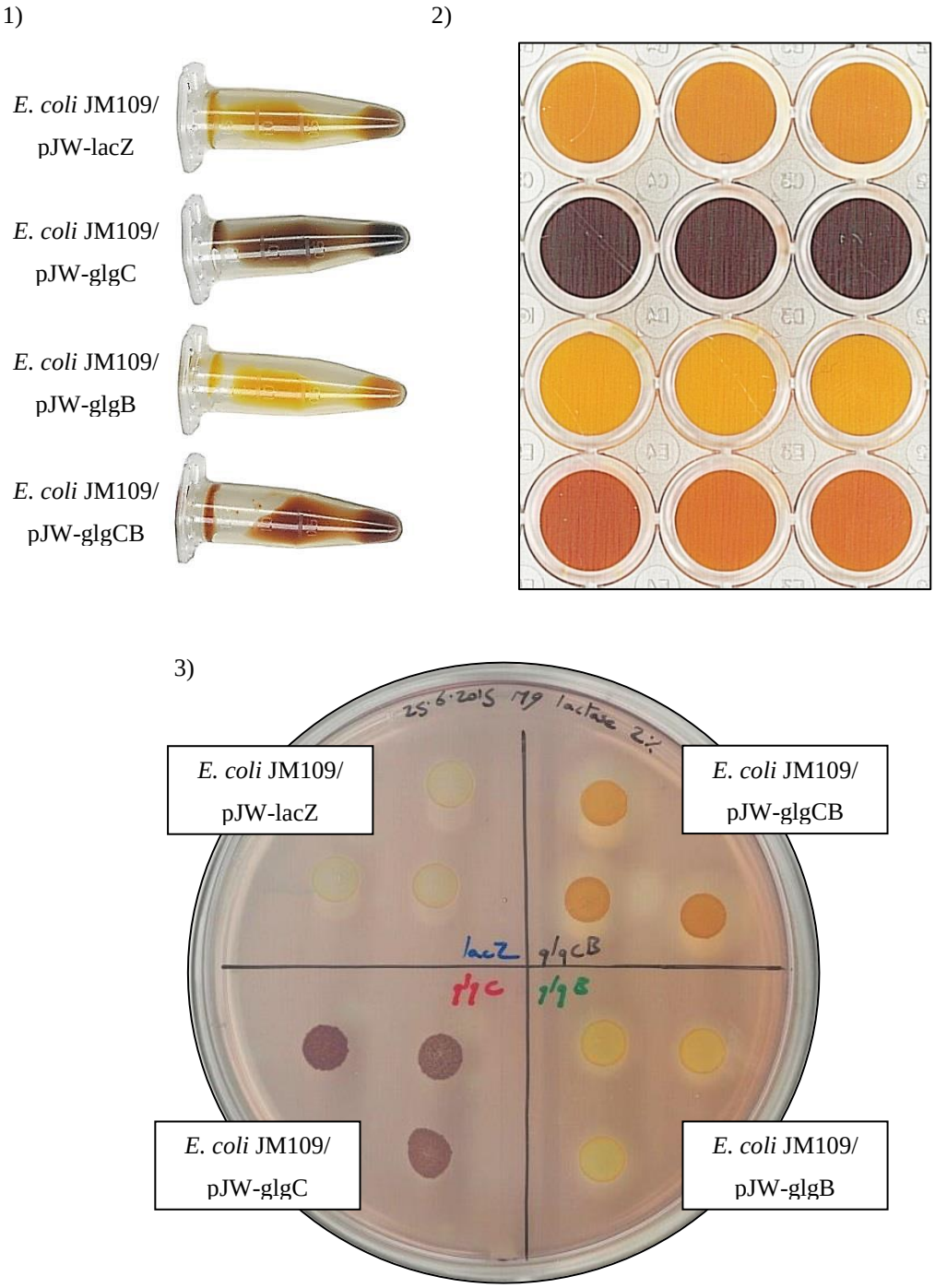
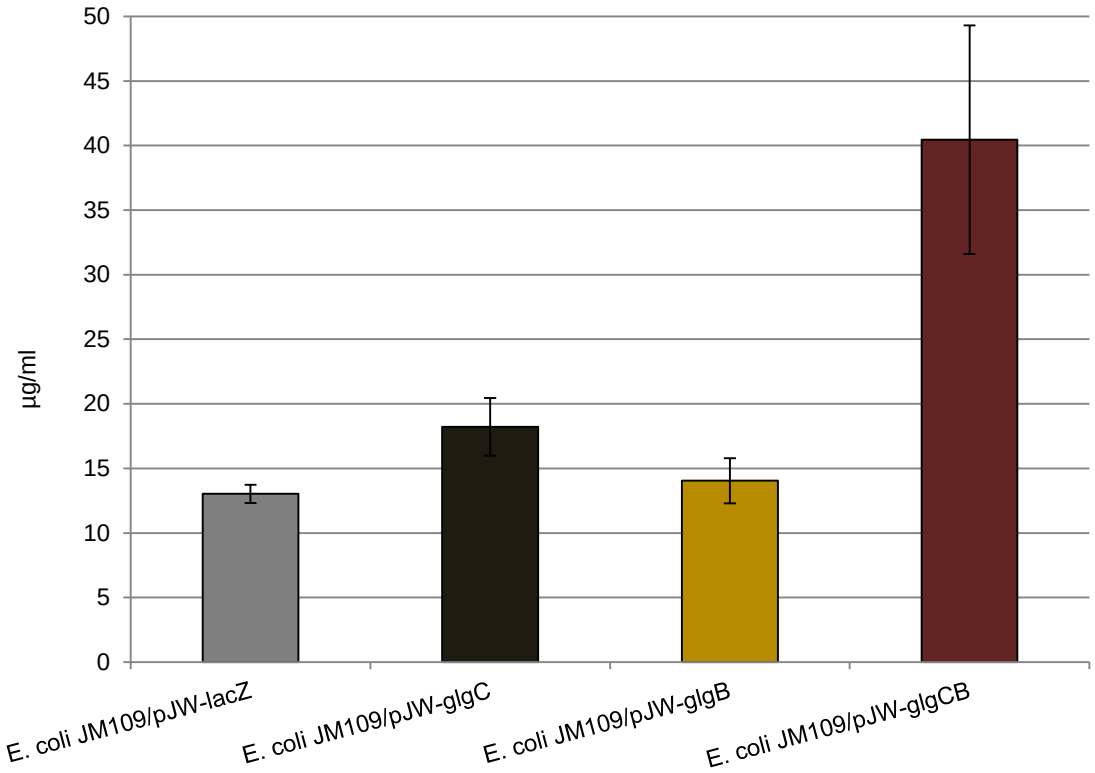


Figure 3.20: Iodine assays of transformed JM109

1) Cell pellets from 1.4 ml of the same cultures as figure 3.19, resuspended in 100 µl Lugol's iodine (0.2%); 2) Cell pellets from fresh cultures of the same strains (3 biological replicates of each) grown under the same conditions, pelleted and resuspended in 200 µl PBS, plus 50 µl $CuSO_4$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%), transferred to the wells of a 48-well plate and imaged using a flatbed scanner; 3) patches from the same cultures used in image 2, spotted onto an M9 plate, grown overnight at 37°C and flooded with Lugol's iodine (0.2%).



Figures 3.21: Total hexose sugar content of the same cultures

From anthrone assays performed on the same cultures, after growth to $OD_{\lambda 600nm}$ 1.0. Error bars represent the standard error of the mean when n=3.

Of the three *E. coli* JM109/pJW-glgCB cultures assayed for figure 3.21, one was observed to grow significantly faster, reaching an $OD_{\lambda 600nm}$ of 1.0 in around 11 hours, compared to around 16 hours for the other two biological replicates (data not shown). This faster growing culture also showed a significantly higher total hexose sugar content than the two that grew more slowly ($p<0.05$), with an average glucose equivalent of 60.06 µg/ml, nearly double that of the other two cultures with their average glucose equivalent of 30.66 µg/ml (figure 3.21 shows only the combined data for all three replicates).

The plasmids of the two different phenotype expressing versions of the pJW-glgCB transformed cells were sequenced and no difference was found between them. The reason for the difference in growing times and sugar content of these cultures therefore remains unknown, and this level of variability was not observed again.

A second anthrone assay was carried out, this time cultured in a manner expected to maximise polysaccharide content (Sundberg et al., 2013, also see Methods). The cells were then washed twice, resuspended in PBS and equalised to an $OD_{\lambda 600nm}$ of 3.5. It was found that, at the end of this growth time, all three of the *E. coli* JM109/pJW-glgCB cultures had grown to a significantly higher density than those of the other transformants, reaching an average $OD_{\lambda 600nm}$ of 6.53 while the other cultures only grew to an $OD_{\lambda 600nm}$ of around 4 (figure 3.22). Unlike the last assay, each transformant in this instance showed a single phenotypic trend.

This second anthrone assay show that once again, a significantly higher hexose sugar content was reached in *E. coli* JM109/pJW-glgC than in *E. coli* JM109/pJW-lacZ. However, under these growth conditions the difference was more extreme, with the final glucose equivalent being nearly three times as high as that of the control cells, at 89 µg/ml compared to 31 µg/ml. Meanwhile, the *E. coli* JM109/pJW-glgCB showed a significantly higher hexose sugar content than both the *E. coli* JM109/pJW-lacZ and the *E. coli* JM109/pJW-glgC ($p<0.05$ in all cases), with a final glucose equivalent that was 2.5 times greater than the *E. coli* JM109/PJW-glgC and over seven times higher than the control cells, at 233 µg/ml. (figure 3.23). Since these cultures were also diluted to compensate for the higher cell densities reached after the 28 hours of growth (around 1.5 times that of the other cultures) the sugar accumulation of the culture prior to this equalisation can be safely assumed to be even greater still.

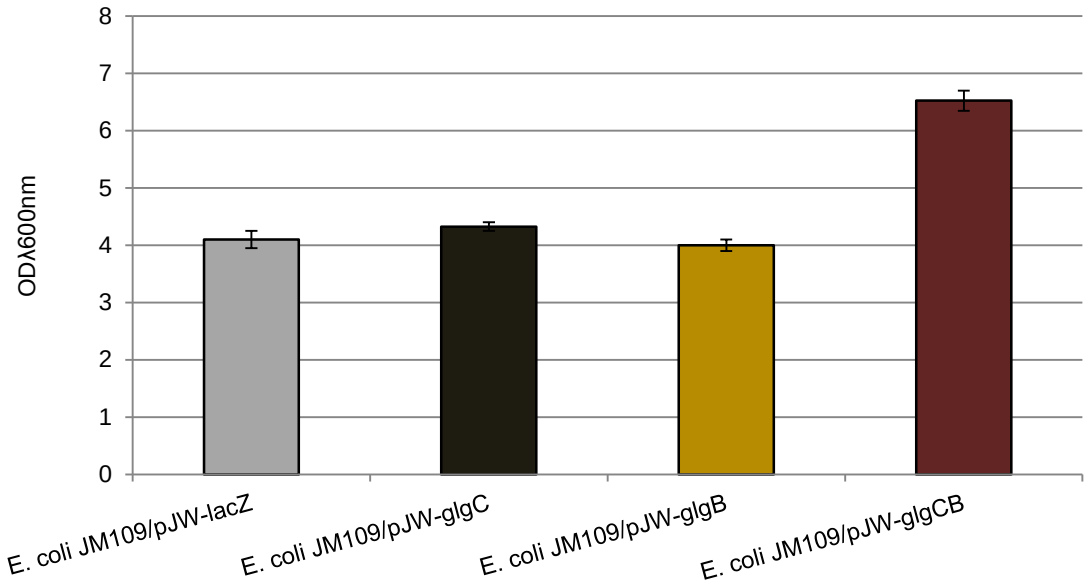


Figure 3.22: Optical densities ($OD_{\lambda 600nm}$) reached by transformed JM109 cells

Readings taken after around 30 hours growth in Kornberg medium followed by M9, in a manner expected to maximise their polysaccharide content. Error bars represent the standard error of the mean when n=3.

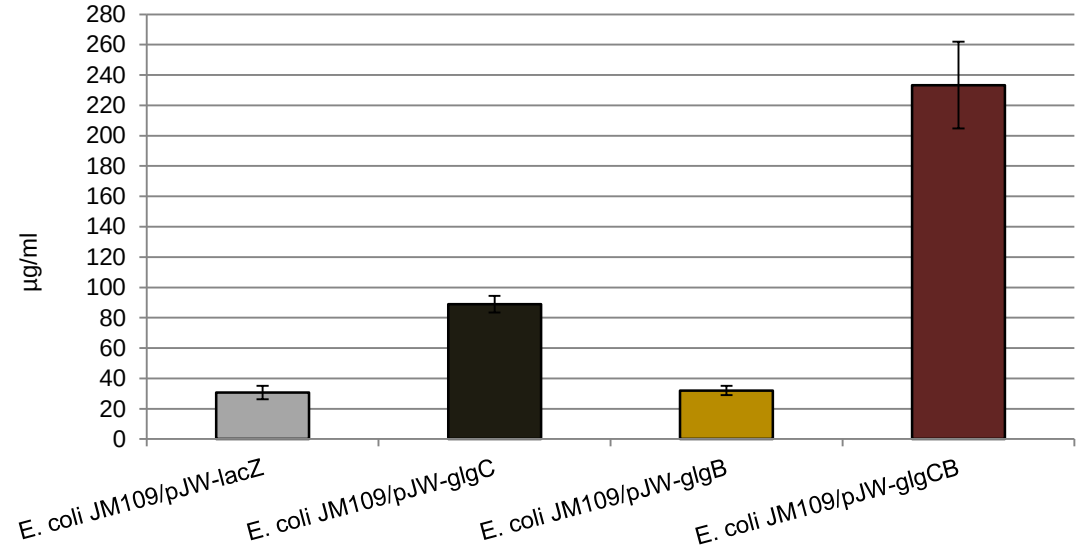


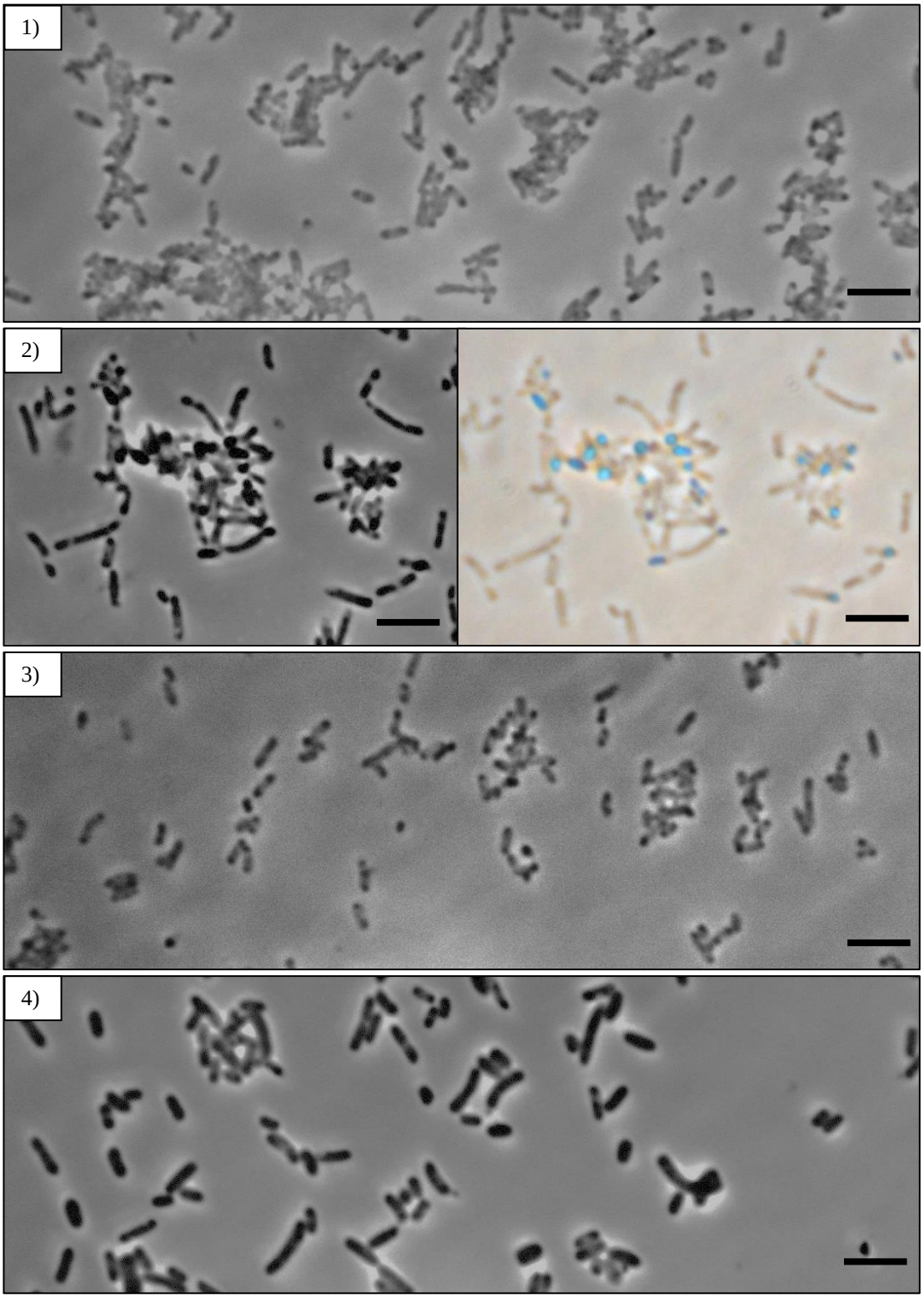
Figure 3.23: Total hexose sugar content of the same cultures

From anthrone assays performed on the same cultures, equalised to $OD_{\lambda 600nm}$ 3.5. Error bars represent the standard error of the mean when n=3.

Microscopy performed on the cultures correlated with previous findings: *E. coli* JM109/pJW-glgC showed inclusion bodies that stained with iodine and were clearly visible under a light microscope, while neither those transformed with pJW-glgB or pJW-glgCB showed any visible inclusion bodies (figure 3.24). Cells patched onto a Kornberg medium agar plate, grown at 37°C overnight and soaked with Lugol's iodine (0.2%) also showed a dark staining of the *E. coli* JM109/pJW-glgC colonies, a slight staining of the *E. coli* JM109/pJW-glgCB colonies and no visible staining of either the *E. coli* JM109/pJW-glgB colonies or the control (figure 3.25). An Iodine assay performed on cell pellets from the cultures after they had been grown for around 30 hours, washed and equalised, again showed a similar trend to before: the *E. coli* JM109/pJW-glgC stained a very dark brown colour and the *E. coli* JM109/pJW-glgCB stained deep rust red. Under these growing conditions there was no visible discrepancy between the staining of the *E. coli* JM109/pJW-glgB and the control cells, which both turned a dark orange colour (figure 3.25).

3.11 Discussion

This work showed that the transformation of *E. coli* with additional copies of its native *glgC* gene, expressed on a high copy number plasmid under a lac promoter, led to the formation of inclusion bodies within the cells when they were grown in culture supplemented with lactose and IPTG. Inclusion bodies were also observed in the same transformants when grown in media supplemented with glucose in place of lactose, although to a lesser extent, which is thought to be due to the inhibitory effect of glucose on the lac promoter. The same 'inclusion body' phenotype was not observed in cells transformed with the same plasmid but lacking the *glgC* gene, under any of the growth conditions tested. When present, inclusion bodies were observed to stain a deep brown in reaction with iodine. A brown stain with iodine is generally more typical of amylopectin than glycogen, which tends to stain red, though the likely reason for this is only that the highly structured nature of the amylopectin crystal inhibits the formation of long single helices. So while it is tempting to assume that this novel iodine staining is suggestive of a more amylose-like structure, all that can reasonably be assumed is that the single helices within the mutant polysaccharide structures are shorter than they are in wild-type glycogen. The intensity of the stain also suggests a high proportion of these short chains, which seems to match the high total hexose sugar content observed via anthrone assay. However, the anthrone assay did not distinguish between the structures that any hexose sugars might be present in, and therefore any correlation between those results and hypotheses for polysaccharide structure can only be speculative at this point.



Figures 3.24: Phase contras light microscopy of iodine-stained, transformed JM109 cells
From cell cultures used in anthrone assay (figure 3.21). Scale bars represent 5 µm (approx)
1) *E. coli* JM109/pJW-lacZ; 2) *E. coli* JM109/pJW-glgC photographed with different cameras;
3) *E. coli* JM109/pJW-glgB; 4) *E. coli* JM109/pJW-glgCB. Only *E. coli* JM109/pJW-glgC showed inclusion bodies, appearing blue when stained with iodine and viewed by phase contrast as shown.

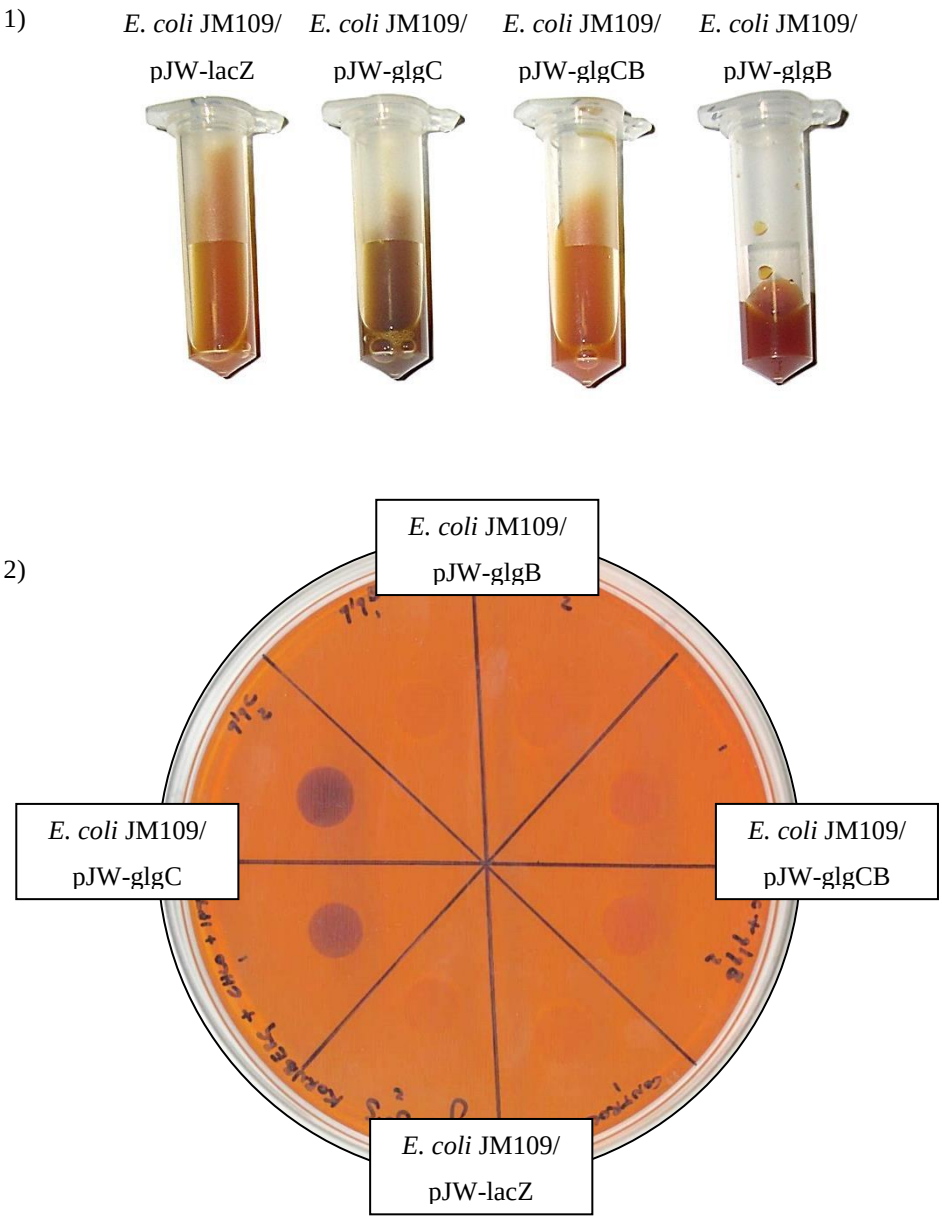


Figure 3.25: Iodine assay of the same transformant cultures

1) Cell pellets from 1.4 ml of the cultures used in the anthrone assay (figure 3.22), resuspended in 500 µl Lugol's iodine (0.2%); 2) The same cultures spotted onto a Kornberg Medium plate, grown overnight at 37°C and flooded with Lugol's iodine (0.2%)

Cells transformed with pJW-glgC16 or pJW-glgC and grown in M9 as opposed to LB tended to stain darker with iodine, perhaps as a result of forming more polysaccharides in the face of nitrogen starvation, since LB is rich in nitrogen while M9 is nitrogen poor. Additionally, the reaction of polysaccharide helices with iodine is known to be interfered with by components of LB broth, in particular by thiol (SH) groups present in the medium which compete for the binding of iodine (Manonmani & Kunhi, 1999), so this may have contributed to the weaker staining and the faster fading of the stain for LB-grown cells. Metal acts as a catalyst in the auto-oxidation of thiol groups, resulting in the formation of -S-S- bonds, thereby relieving the competition (Manonmani & Kunhi, 1999). The addition of $CuSO_4$ prior to iodine was therefore found to work as a protective agent, maintaining the iodine stain of cells grown in either type of media.

An SDS-PAGE analysis of the *E. coli* JM109/pJW-glgC16 cells suggested that the inclusions observed were not composed of a single over-expressed protein, since those cultures grown in media supplemented with lactose and IPTG, confirmed by microscopy to contain a relatively high proportion of the inclusions, did not show a comparatively large single band on the gel. A large band was observed in the analysis of cells grown in media supplemented with IPTG but no additional sugar. This band was around the 50 KDa size expected of the GlgC subunits. However, a band of similar molecular weight was observed for the control cultures grown under the same conditions (in media supplemented with IPTG, but not in media supplemented with both IPTG and lactose). The identity of this band has not been confirmed. It may be that the extremely large band seen in the case of *E. coli* JM109/pJW-glgC16 grown in the presence of IPTG includes the GlgC16 protein, perhaps in conjunction with another native protein of a similar size, explaining its appearance, to a lesser extent, in the analysis of the control cells. However, this would not explain the lack of a similar band in the cultures expressing the strongest transgenic phenotype. The SDS-PAGE cultures were grown in LB, which is only weakly buffered, so the absence of the large protein band observed in cultures grown in the presence of IPTG and lactose is perhaps a result of the acidification of the culture, though the possible mechanism for this is unknown. Further work is therefore needed to interpret these results.

It seems reasonable to assume that, since a novel phenotype was observed in *E. coli* JM109/glgC but not those *E. coli* transformed with an empty plasmid and grown under the same conditions, the transgene was responsible for the phenotype. However, the lack of any unusually intense band in the SDS-PAGE of cells showing the phenotype, while supporting

the hypothesis that the inclusion bodies observed are not composed of a single protein, also seem to suggest that the transgene is not being heavily transcribed. A comparison of gene expression – through Northern blotting or qRT-PCR – with enzyme concentration – through Western blotting or his-tagging – with enzyme activity – through measurement of substrate and product concentrations – should therefore be extremely revealing.

A Western blot could be used to identify the presence of the GlgC16 protein on the SDS-PAGE gel (although it would be unable to distinguish it from the bacteria's native GlgC). However, to the best of the author's knowledge, there are no commercially available antibodies specific to *E. coli* GlgC. Work by other groups have included Western blots for this protein, but only after raising their own antibodies against GlgC in rabbits, which was beyond the scope of this project (Eydallin et al., 2007; Sweetlove et al., 1996a). An alternative method might be to his-tag the protein so that it could be purified through a nickel or cobalt column, before running a subsequent SDS-PAGE on the purified protein to give an impression of its relative abundance in the different cultures.

Such an analysis might be particularly revealing in conjunction with an activity assay, to give an indication of whether GlgC activity increases relative to expression, for example by measuring the formation of diphosphates (a biproduct of GlgC activity) versus the decrease in NADH concentration (Boehlein et al., 2015). The kinetic constants of the enzyme could then also be analysed, by repeating the activity assay under different substrate concentrations.

An activity assay for GlgC could also be compared to the abundance of the inclusion body phenotype, to get an indication of whether GlgC activity is directly responsible for this phenomenon. The strength of this transgenic phenotype could be measured by purifying the inclusion bodies, for example through lysis followed by separation of the insoluble inclusion bodies from the soluble fraction of the cell by centrifugation. However, the *E. coli* cells contain other insoluble polysaccharides which may also be present in greater concentrations in cells expressing pJW-glgC16 and pJW-glgC. The inclusion bodies might therefore be purified further, perhaps by separation in a Percol gradient, before performing a sugar assay. In this study, total hexose sugar assays were performed on aliquots of washed whole cells, using anthrone reagent followed by a spectrophotometric analysis. The higher sugar content observed in the *glgC16* and *glgC* transformed cells therefore cannot be confidently ascribed to the presence of the inclusion bodies.

Furthermore, the anthrone assay itself is perhaps not an optimal method for measuring cellular sugars. It was noticed during this work that readings taken using this method consistently dipped towards the end of assays. Attempts were made to mitigate this phenomenon through repetition, the preparation of excess reagent and the assaying of the standard solutions used to calculate total sugar concentrations of the washed cells at the beginning, middle and end of each assay (see Methods) and as such the standard error calculated for the average of each set of readings was felt to be acceptable. However, alternative total sugar assays do exist that may produce more accurate measurements, such as the enzymatic analysis of cellular sugars digested to glucose, as outlined by Smith & Zeeman (2006). This would have the added benefit of only giving the concentration of carbohydrates composed of glucose, which should be the case for any storage polysaccharides present. Although the anthrone assays performed in this work were calibrated using known glucose concentrations, the reagent does react with other monosaccharides to different degrees. The assay in this case was designed to indicate the concentration of hexose sugars. However, the presence of pentose sugars within the samples is known to affect the results (Johanson, 1954), meaning that the presence of such sugars within the bacteria transformants may have skewed the outcome of this assay. With this in mind, it would also be interesting to perform chromatography on the cell extract, to discover whether indeed any significant amounts of other types of sugar are present.

Despite these issues, anthrone assays performed on transformants in this work have been consistent in showing a higher hexose sugar content for *glgC16* and *glgC* transformed cells than for a control. An experiment was therefore devised to test the hypothesis underlying this observation: That GlgC activity normally performs the rate-limiting step of glycogen synthesis and, for this reason, upregulating its expression in JM109 *E. coli* grown in sugar-rich media leads to the appearance of inclusion bodies in many of the cells, as well as a higher average intracellular concentration of sugar than in a control group grown under the same conditions, because GlgC increases the substrate availability for GlgA, the Glycogen Synthase. This increased substrate is thought to allow the Glycogen Synthase to meet a previously unrealised potential and synthesise the linear glucan chains within the polysaccharide granule at a faster rate than the GlgB enzyme can add branches to them, so that they grow long unbranched chains able to spontaneously form double-helices, and that this interweaving of adjacent chains is causing the aggregation of granules into the large inclusion bodies observed.

If this hypothesis were correct, it was thought that the addition of *glgB* to the same plasmid as *glgC* would allow the branching of the polysaccharide to keep speed with the synthesis of linear chains. This was expected to stop the formation of inclusion bodies, since unbranched regions of glucan chain would no longer be able to grow long enough to wind together into helices, and furthermore the steric interference caused by the dense packing of glucan branches, which theoretically limits the growth of native glycogen granules, would be reintroduced. However, the cells were still expected to show the high sugar content seen in those transformed with *glgC*, since the Glycogen Synthase would still be able to synthesise chains at the increased rate. All these phenotypes were observed, with the added observation that, rather than synthesising the same high levels of polysaccharide as the *glgC* transformed cells, those transformed with both *glgC* and *glgB* showed more than double their total sugar content. The reason for this remains unknown. The hypothesis might be tested more conclusively with comparative assays of GlgC and GlgA, since for it to be correct would require the increased expression of GlgC to result in a proportional increase in GlgA activity.

It would also help confirm or disprove the theories formed about these polysaccharides if the chain length distribution for each could be discovered. This might be achieved through the digestion of each polysaccharide type with Isoamylase and Pullulanase, followed by their analysis by High Performance Liquid Chromatography or High Performance Anion-Exchange Chromatography (Sundberg et al., 2013; Corradini et al., 2012).

Chapter 4: Potential Applications of Storage Polysaccharides in *E. coli*

4.1 Introduction

There has recently been much interest in the potential of cyanobacteria as an alternative to land crops as the carbon source for the next generation of biofuels (Angermayr et al., 2009; Ducat, 2011). These organisms convert solar energy to biomass with more than twice the efficiently of energy crops such as switchgrass, at 0.5% compared to 0.2% (Melis, 2009), and have the added advantage that they grow in aquatic environments, allowing for year-round cultivation without competing for arable land that is also badly needed for food production. The glycogen they use as a carbon store can also be more readily converted into bioethanol than the cellulose of grasses, using existing biorefinery techniques. However, a drawback with the use of bacteria as an industrial-scale carbon source is that, in general, they accumulate glycogen when starved of nutrients other than carbon. Since these conditions also limit biomass accumulation, it is important to identify a suitable strain which combines high biomass productivity with a satisfactory glycogen content (Aikawa et al., 2014).

In light of this search, the phenotypes observed in *E. coli* cells transformed with copies of their own *glgC* and *glgB* genes seem particularly interesting. The combination of high cell density growth that *E.coli* JM109/pJW-glgCB seem capable of, coupled with their ability to synthesise over seven times the amount of polysaccharide as control cells, suggest that such genetic manipulation could be useful in the production of high carbon-yield organisms for industries such as biofuel production. Meanwhile, the high polysaccharide accumulation in the form of inclusion bodies produced by cells transformed with *glgC* also suggest potential for industrial applications, since if this single mutation can lead to the creation of polysaccharide granules that the bacteria are unable to digest, higher yields might be made more attainable and easier to harvest. An added attraction of these results is that there are currently no EU laws prohibiting the growth in the field of organisms transformed with genes from their own genome. If the same mutations were therefore found to produce similar effects in an organism such as a cyanobacterium, there is theoretically no reason why they couldn’t be grown en masse.

To investigate the viability of these transformations for further application, experiments were run to measure aspects of growth in different media as well as viability and polysaccharide content under starvation conditions.

4.2 Growth of *glgC*-upregulated and *glgC-glgB*-upregulated *E. coli*

It had been previously observed that JM109 transformed with *glgC* and/or *glgB* showed different growth rates and reached stationary phase at different optical densities from either each other or a control strain. To investigate this further, three cultures of each transformant (*E.coli* JM109/pJW-lacZ (control); *E.coli* JM109/pJW-glgC; *E.coli* JM109/pJW-glgB; and *E.coli* JM109/pJW-glgCB) were grown overnight, equalised to the same density and used to inoculate 3 cultures each of four different types of media: LB supplemented with lactose (2%) and IPTG; Kornberg medium supplemented with lactose (2%) and IPTG; Terrific Broth (TB) supplemented with lactose (2%) and IPTG; and TB without supplements. TB, like Kornberg medium, is extremely rich and should allow dense growth of all cultures. Since the transgenes under investigation are expressed via a lac promoter, a lack of supplementary lactose or IPTG should mean that they are not activated, so that growth rates of all cultures remains equal. Two aliquots from each inoculated culture were transferred to 96-well plates and their optical densities were measured every 15 minutes in a plate reader. The remaining cultures were also incubated at 37°C and 200rpm, and $OD_{\lambda 600nm}$ readings were taken from them by hand every other hour for 14 hours, then again at 24 hours. The experiment was run a second time, taking readings from just the plate reader, after taking just one aliquot from each of 48 fresh cultures (figures 4.1 to 4.4, also see Methods).

Results show that, without IPTG or lactose to activate expression of the transgenes, the different cultures grew at the same rate (figure 4.4). In all media supplemented with lactose and IPTG, *E. coli* JM109/pJW-glgCB grew at a slightly slower rate than the control, but as observed previously, reached stationary phase at higher cell densities, most noticeably when grown in supplemented TB. *E. coli* JM109/pJW-glgB showed growth rates extremely similar to the control in all circumstances. *E. coli* JM109/pJW-glgC, which were confirmed by microscopy to show the inclusion body phenotype (data not shown), also showed a slower growth rate than the control. However, when grown in supplemented TB they too reached a higher cell density than the control cultures, although this trend was not seen when they were grown in supplemented LB. When grown in supplemented Kornberg medium, plate reader results showed *E. coli* JM109/pJW-glgC reach stationary phase at a lower optical density than the control cultures, while the shake flask readings showed it reach a higher final optical density. Apart from this discrepancy, the pattern of growth for all cultures tended to be similar across the plate reader and the shake flask readings, despite the fact that in all cases the shake flask readings showed far higher densities.

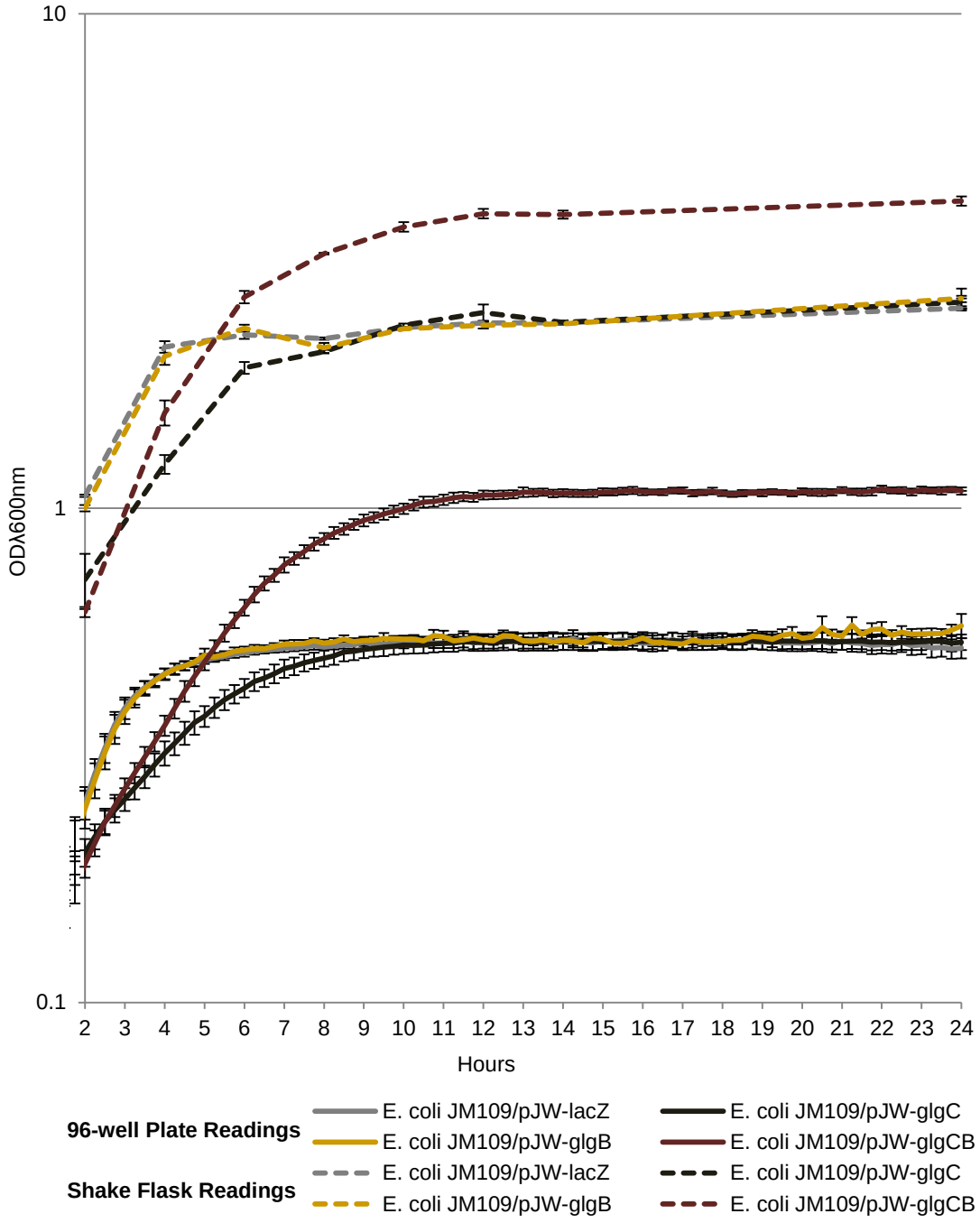


Figure 4.1: Growth curves for transformants, in LB + lactose (2%) + IPTG

Solid lines show 96-well plate readings from 3 plates. Error bars represent the standard error of the mean when n=6.

Dashed lines show 50 ml shake flask readings. Error bars represent the standard error of the mean when n =3.

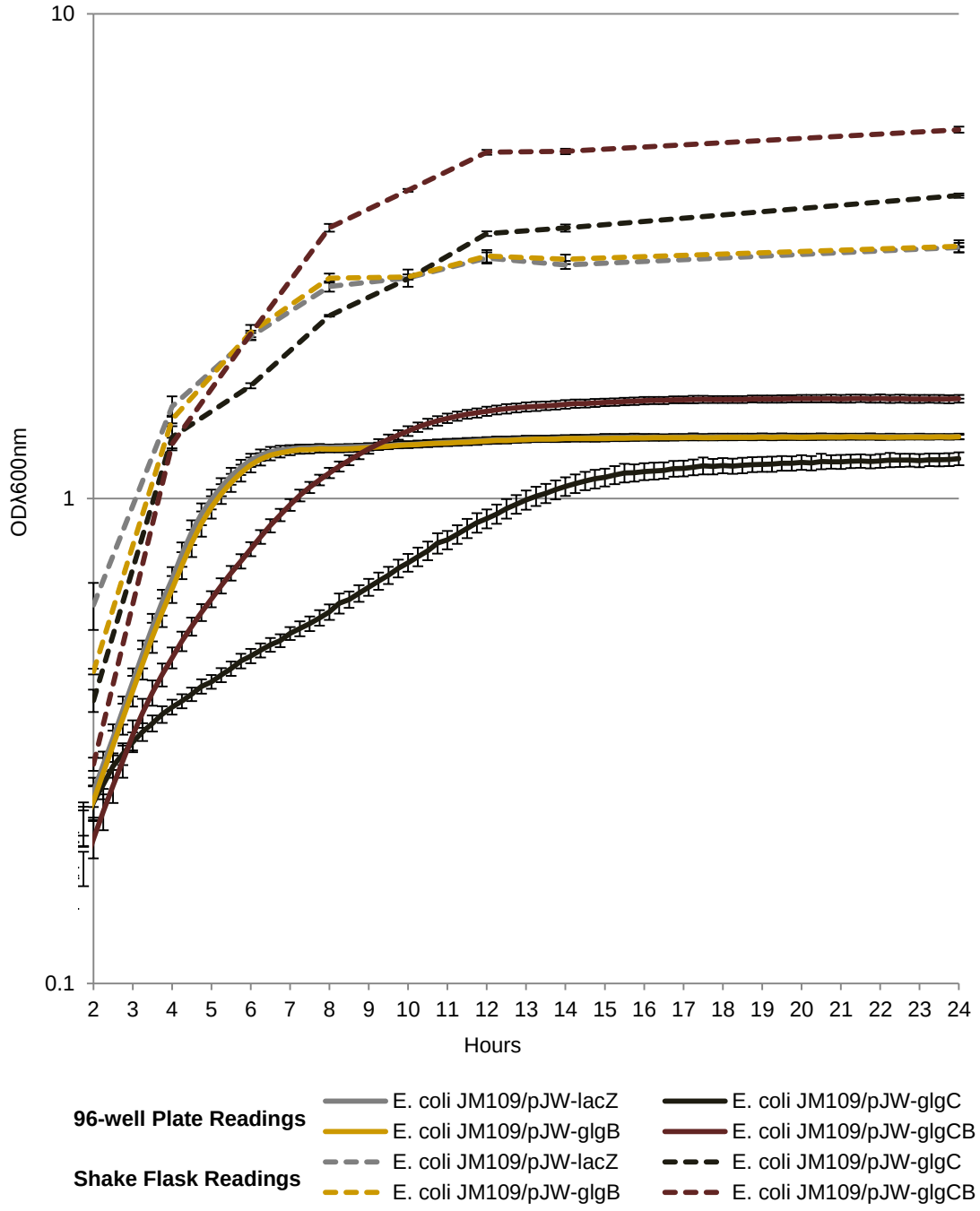


Figure 4.2: Growth curves for transformants, in Kornberg Medium + lactose (2%) + IPTG

Solid lines show 96-well plate readings from 3 plates. Error bars represent the standard error of the mean when n=6.

Dashed lines show 50 ml shake flask readings. Error bars represent the standard error of the mean when n=3.

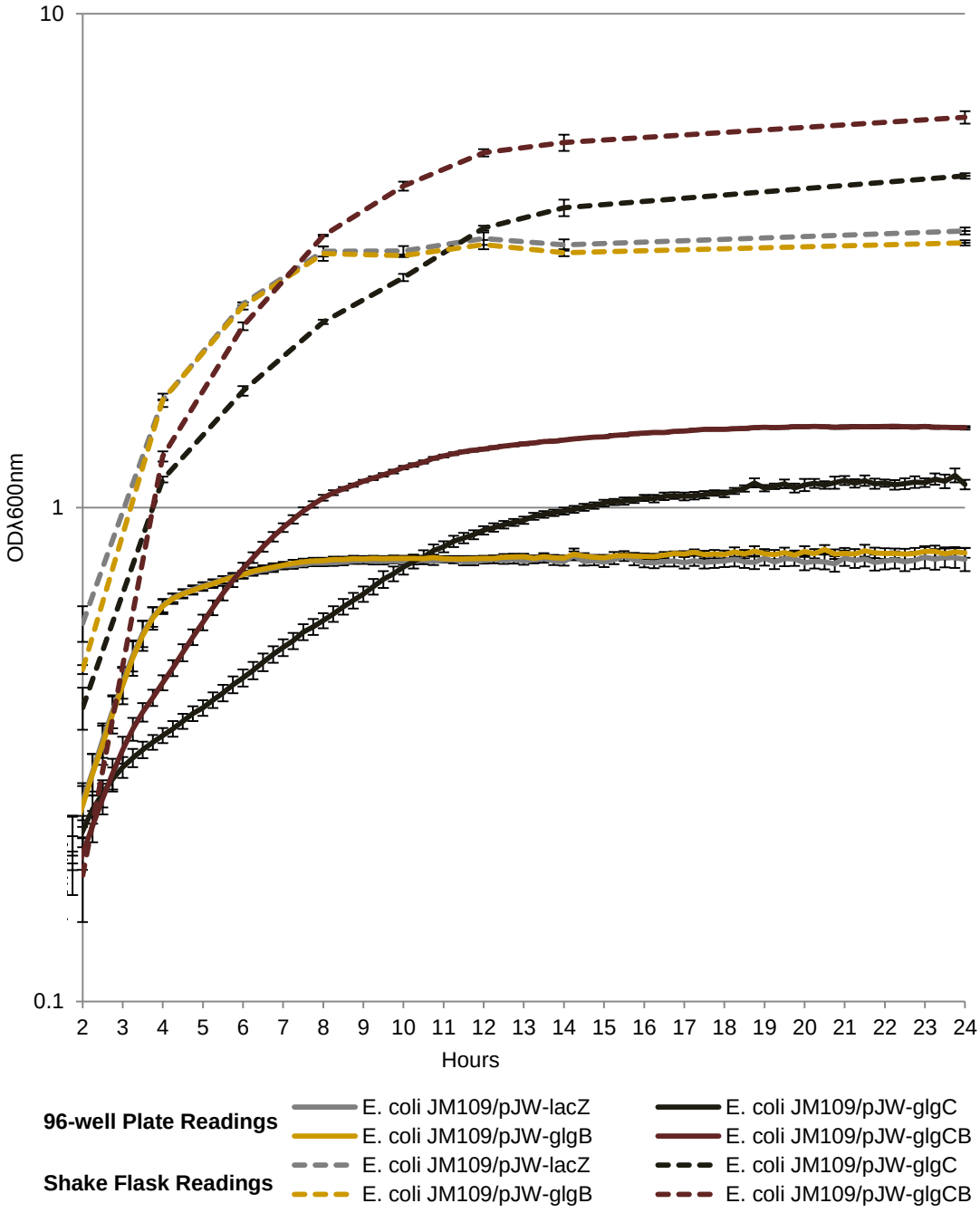


Figure 4.3: Growth curves for transformants, in Terrific Broth + lactose (2%) + IPTG

Solid lines show 96-well plate readings from 3 plates. Error bars represent the standard error of the mean when n=6.

Dashed lines show 50 ml shake flask readings. Error bars represent the standard error of the mean when n=3.

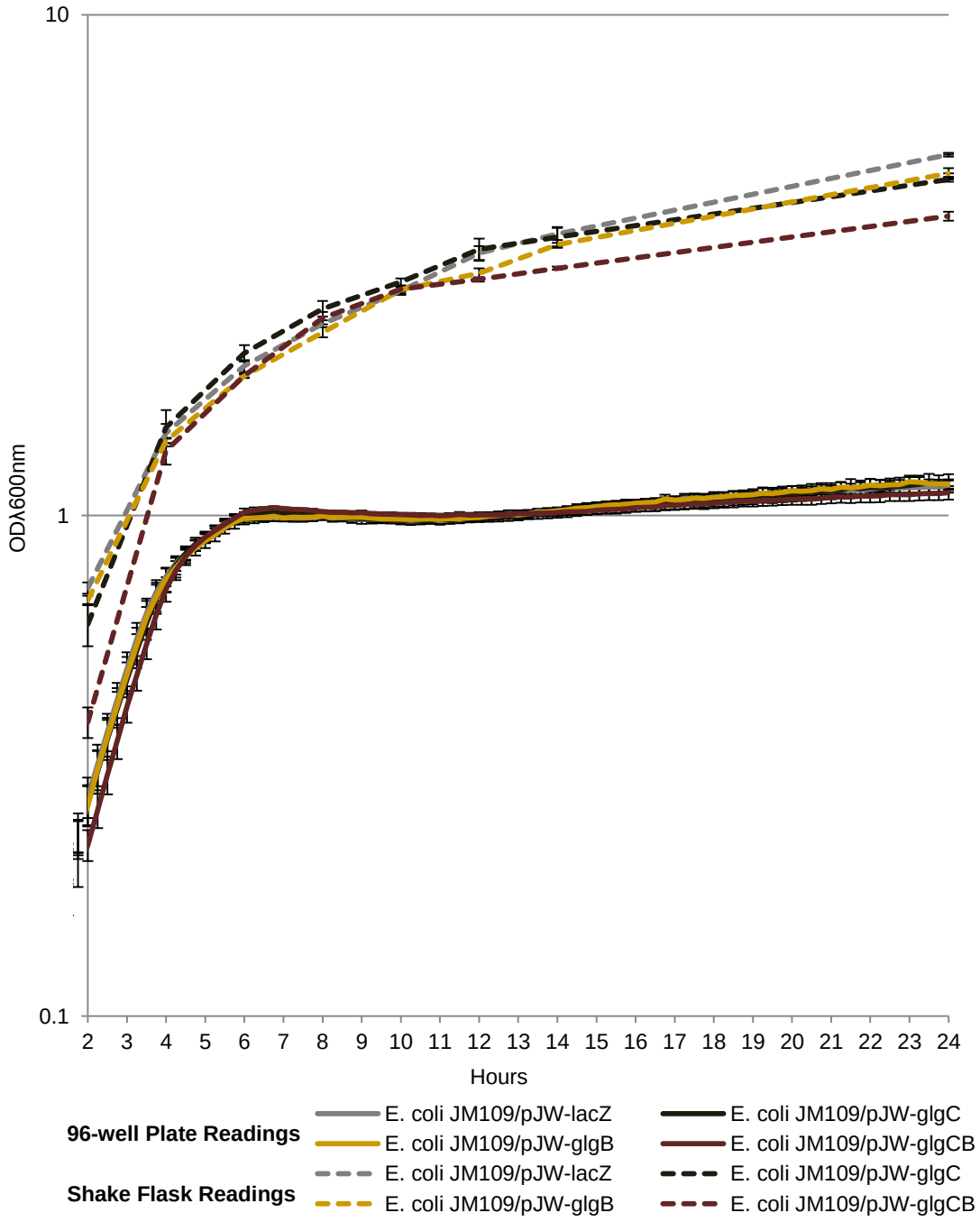


Figure 4.4: Growth curves for transformants, in Terrific Broth without supplements

Solid lines show 96-well plate readings from 3 plates. Error bars represent the standard error of the mean when n=6.

Dashed lines show 50 ml shake flask readings. Error bars represent the standard error of the mean when n=3.

4.3 Viability of glycogen upregulated cells

Glycogen accumulation is thought to be an adaptation to improve bacterial survival under stressful conditions, such as when enteric species are shed into the outside environment. As such, it was predicted that the changes to the polysaccharide content of cells brought about by transformation with highly expressed copies of their own *glgC* and *glgB* glycogen synthesis genes would have a marked effect on survival rates under such conditions. In particular, it was suspected from previous observations that *E. coli* JM109/pJW-glgC would be unable to digest the polysaccharide inclusion bodies that they accumulate under growth in high sugar media. The working hypothesis for these inclusion bodies is that upregulated GlgC activity is permitting higher-than-normal activity of the Glycogen Synthase enzyme, and that this is leading to the creation of long unbranched regions of polysaccharide chains that are spontaneously intertwining, as they are thought to do within the crystalline lamella of amylopectin granules. Starch accumulating organisms are known to express dikaneses (GWD and PWD) specifically to help unzip these double-helices during enzymatic starch degradation (Santelia & Zeeman, 2011). Since glycogen accumulating bacteria lack genes for these enzymes it seems conceivable that, if the iodine staining seen in *E. coli* JM109/pJW-glgC is indeed caused by an accumulation of helices, those bacteria would be prohibited from metabolising the interwoven chains. It was therefore predicted that, although these mutants accumulate higher concentrations of polysaccharide during growth in high sugar media (see previous chapter), much of that polysaccharide would be unavailable during times of stress. Furthermore, growth experiments have suggested that the accumulation of these inclusion bodies might be a burden on the cells (figures 4.1 to 4.4), so that under stressful conditions bacteria that have accumulated them may show a higher mortality rate than a control strain.

Meanwhile, *E. coli* JM109/pJW-glgCB have been found to accumulate relatively large stores of polysaccharide when grown in a high sugar medium. Furthermore, they do not express the inclusion body phenotype seen in *E. coli* JM109/pJW-glgC, and they stain a deep red with iodine in a manner suggestive of high levels of phenotypically normal glycogen. It was therefore hypothesised that these cells would be able to metabolize the polysaccharide stores they acquire during growth in high sugar medium when they are transferred to a stressful, low nutrient environment. Furthermore, they were observed to reach higher cell densities than control strains when grown in rich medium. It is unknown how this might relate to growth under stressful conditions. However, it seems conceivable that their increased

polysaccharide content would provide a carbon supply during this period, allowing them maintain greater cell viability.

Cells transformed with a highly expressing copy of *glgB* have so far shown little phenotypic difference from control strains. Iodine assays have suggested that these cells form even shorter polysaccharide chains than the control, since they tend to stain a paler yellow. This would be consistent with the enzyme creating a higher degree of branching. Previous work by other groups has shown that bacteria accumulating polysaccharides with shorter average chain lengths tend to have higher survival rates under stressful conditions, possibly as a side-effect of their glycogen being more difficult to metabolise, causing it to be more sparingly rationed (Wang & Wise, 2011). These mutants may therefore show greater viability under stressful conditions, though this prediction is tenuous at best, being based only on a few unquantified observations.

To test these hypotheses, three overnight cultures for each transformant (*E. coli* JM109/ pJW-lacZ; *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/ pJW-glgCB) were $OD_{\lambda 600nm}$ equalised and used to inoculate cultures of Kornberg medium supplemented with lactose (2%). These were grown at 37°C to an $OD_{\lambda 600nm}$ of around 0.6, then supplemented with IPTG and grown at 30°C for a further 20 hours. Cells were then washed, resuspended in M9 supplemented with lactose (2%) and grown at 37°C for a further 4 hours. This culture method is expected to maximise the accumulation of polysaccharides (Sundberg et al., 2013). Cells were then washed, resuspended in M9 and equalised to an $OD_{\lambda 600nm}$ of 1.5. Each culture was then incubated at 37°C and 200rpm, without a carbon source, for nine days, and $OD_{\lambda 600nm}$ readings were taken at intervals during this time to track the changing cell densities of each culture (figure 4.5).

Contrary to expectations, *E. coli* JM109/pJW-glgCB showed a rapid decline in culture density during this period, dropping from the starting $OD_{\lambda 600nm}$ of 1.5 to 0.4 by the end of the nine days; over three times the decline in cell density seen in the control (figure 4.5). *E. coli* JM109/pJW-glgC showed the second steepest decline, dropping from the starting $OD_{\lambda 600nm}$ of 1.5 to 0.7 by the end of the nine days (figure 4.5), supporting the hypothesis that the accumulation of inclusion bodies does not aid in cell survival. The accumulation of inclusion bodies within the cells of these cultures was confirmed by microscopy at the start of the starvation period (figure 4.10). *E. coli* JM109/pJW-glgB were observed to maintain a slightly higher culture density than the control (figure 4.5), which perhaps supports the hypothesis that survival under starvation conditions is improved by a shorter average chain length of

glycogen storage granules (Wang & Wise 2011), although the difference in cell survival between these transformants and the control strain is not significant at any point ($p>0.05$).

Protein assays and colony counts were carried out alongside optical density readings at 132 and 216 hours (the final two time points of the experiment). The protein assays showed a relative difference between cultures that correlated well with the culture optical density results (figures 4.6 and 4.7). However, the colony forming unit count told a slightly different story, showing a greater difference between the transformants, and also a significant decline in colony forming units over the 84 hour period, for all strains ($p<0.05$) (figure 4.8), which was not observed for any strain in either the OD or protein assay results ($p>0.05$ in all cases). Unfortunately none of the methods provides a perfect means to estimate viable cell numbers. Both the culture density measurements and the protein assay have the disadvantage of not being able to distinguish between live and dead cells. Since there was certainly a significant mortality rate for all strains over the course of this experiment, it could therefore easily be that these methods underestimated the decline in viable cell numbers. However, the *cfu* count was also vulnerable to bias: readings were taken after 24 hours growth on LB agar plates, but further growth was observed beyond this period which was unquantifiable due to overlapping of the colonies on the plate after this time. It therefore seems possible that after the extended period of starvation, viable cells were present in the cultures that were unable to grow fast enough to be counted.

Despite the differences between results obtained via culture density readings, protein assays and colony counts, overall a similar trend was observed by all three methods: cultures of *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgB showed no significant difference from each other at either time period by any method of detection ($p>0.05$), while both showing significantly higher culture densities than either *E. coli* JM109/pJW-glgC or *E. coli* JM109/pJW-glgCB at both time times, by all methods of detection ($p<0.05$).

The larges discrepancy between the different methods of analysis appear in the comparisons of *E. coli* JM109/pJW-glgC and *E. coli* JM109/pJW-glgCB. Despite the similarity of the graphs, an ANOVA test on the optical density readings showed no significant difference among the readings for either transformant at either time point ($p>0.05$), while a significant difference was found between *E. coli* JM109/pJW-glgC and *E. coli* JM109/pJW-glgCB at both time points in the protein assay data ($p<0.05$). A significant difference was also found between these two strains in the *cfu* readings at 216 hours ($p<0.05$) though not for the readings taken at 132 hours.

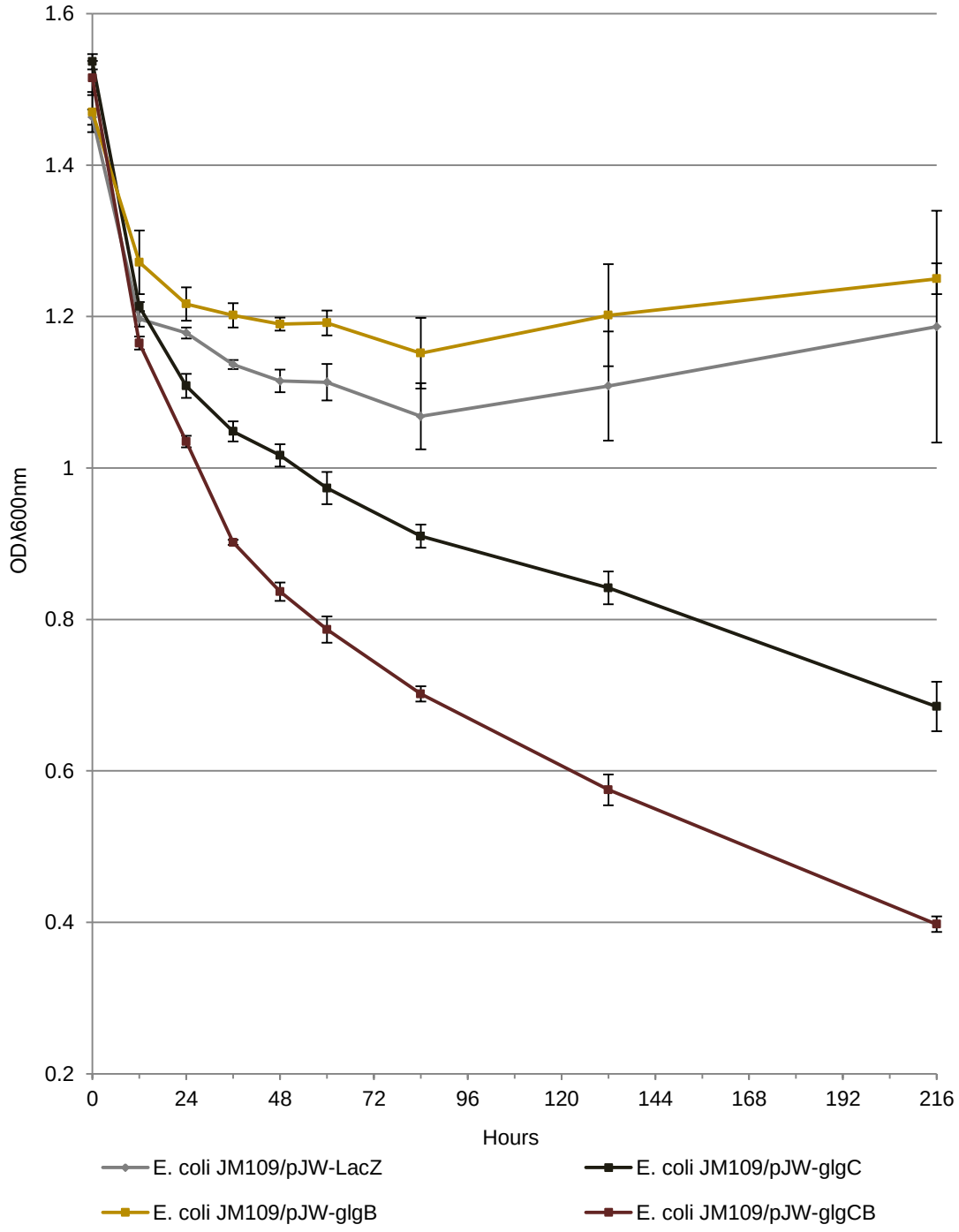


Figure 4.5: Growth curves for transformed JM109 over 216 hours of starvation conditions
For *E. coli* JM109/pJW-lacZ (control); *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/pJW-glgCB under starvation conditions in M9 without a carbon source, starting from cell densities of $OD_{\lambda 600nm}$ of 1.5. Error bars represent the standard error of the mean when n=3.

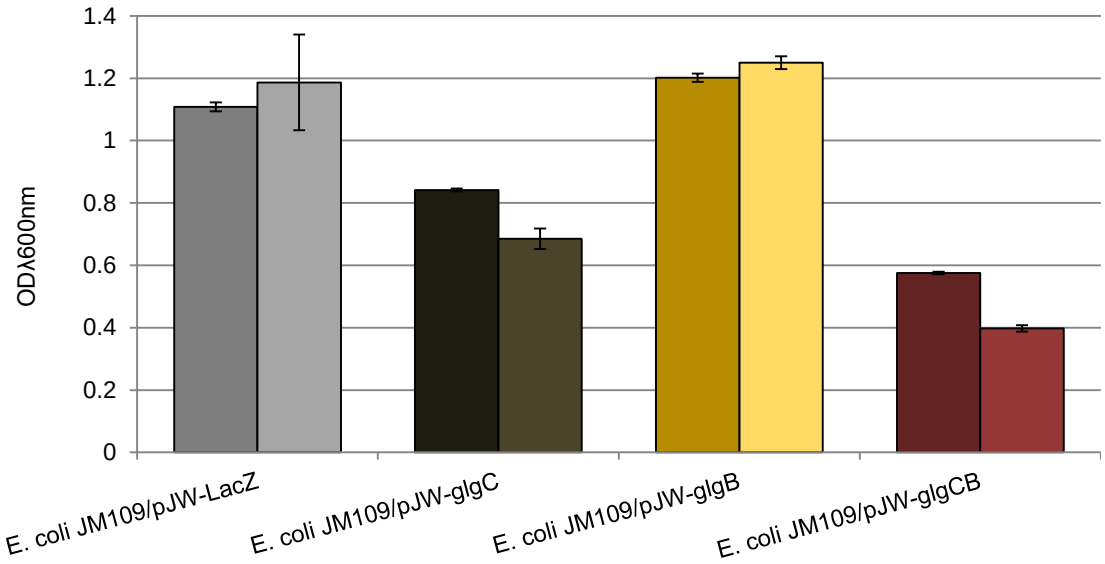


Figure 4.6: Optical density readings for 132 hours and 216 hours of the starvation experiment
For *E. coli* JM109/pJW-lacZ (control); *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/pJW-glgCB under starvation conditions in M9 without a carbon source. Error bars represent the standard error of the mean when n=3.

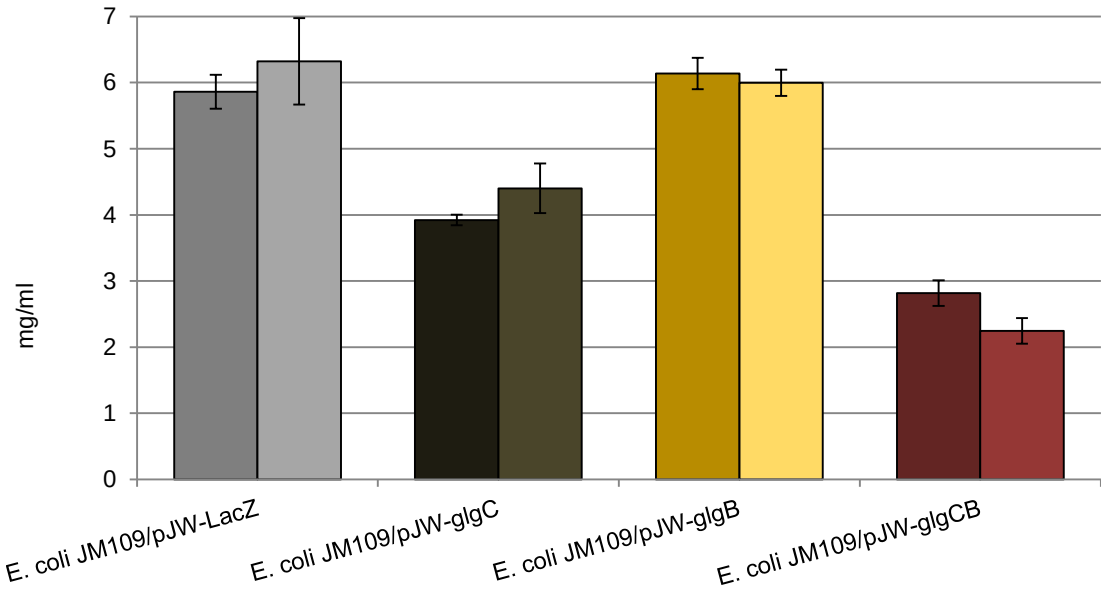


Figure 4.7: Protein assay readings for 132 hours and 216 hours of the starvation experiment
For *E. coli* JM109/pJW-lacZ (control); *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/pJW-glgCB under starvation conditions in M9 without a carbon source. Error bars represent the standard error of the mean when n=3. Readings correlate well with culture $OD_{\lambda 600nm}$ readings.

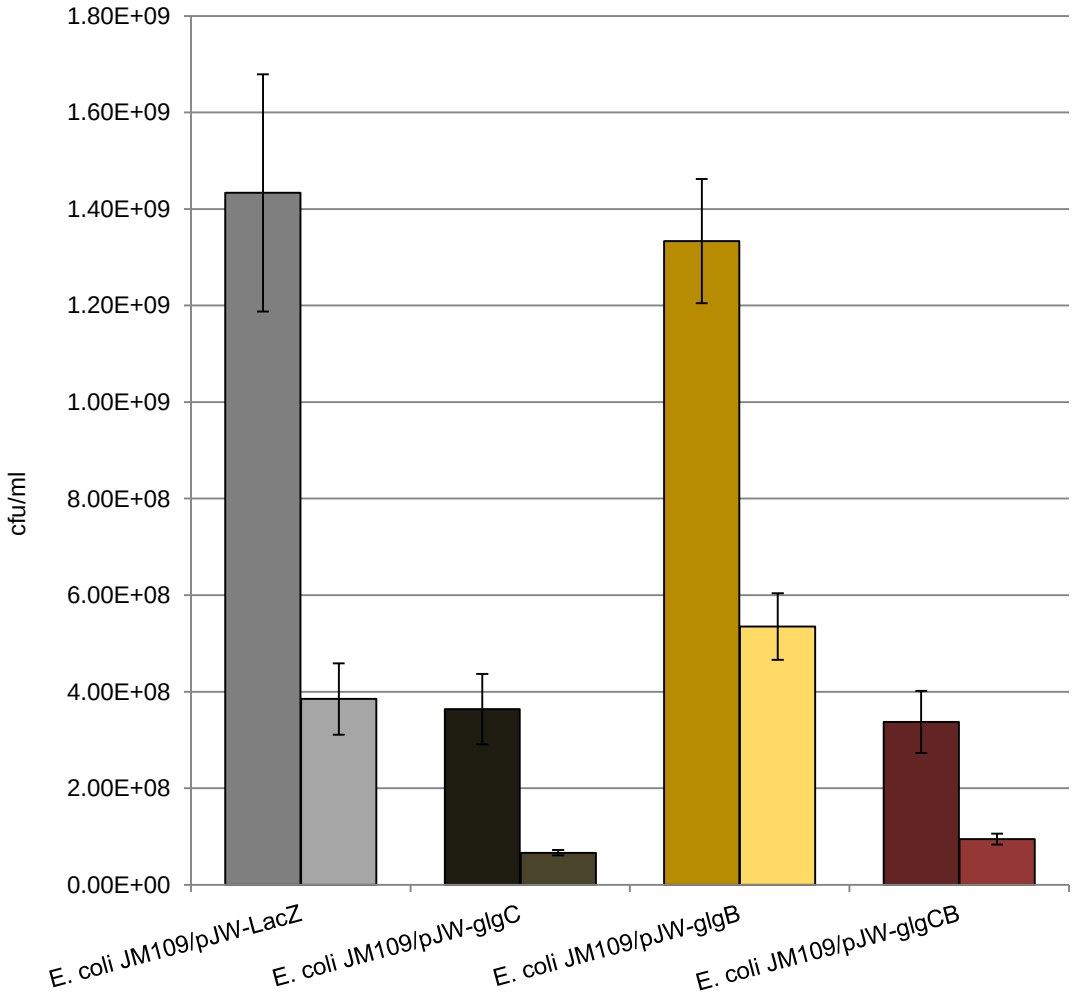


Figure 4.8: Colony counts for 132 hours and 216 hours of the starvation experiment
For *E. coli* JM109/pJW-lacZ (control); *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/pJW-glgCB under starvation conditions in M9 without a carbon source. Error bars represent the standard error of the mean when n=3.

4.4 Can *glgC*-upregulated *E. coli* be starved of their granules?

The relatively high mortality rate of *E. coli* JM109/pJW-glgC when cultured under starvation conditions seems to fit with the hypothesis that these cells are unable to metabolize the inclusion bodies they form during prior growth in a high sugar medium. Meanwhile, the even higher mortality rate of *E. coli* JM109/pJW-glgCB under the same conditions was surprising, as it was considered that the high concentrations of polysaccharide accumulated by these cells would provide an accessible carbon store to aid survival during times of nutrient stress. To investigate the depletion (or lack of it) of storage polysaccharides within these transformants, anthrone assays were carried out during the starvation experiment described above, at the same time points as the culture density measurements, to reveal the total sugar concentrations of the different transformants over this period (figure 4.9, 1). The experiment was later repeated, but with measurements taken at different time points and an extended running time, in an attempt to clear up the uncertainty seen at the tail end of the first experiment (figure 4.9, 2).

As in previous experiments (see chapter 3), the *E. coli* JM109/pJW-glgC cells contained over double the total sugar concentration of the control cells at the start of the starvation experiment (after all cultures had been grown for 28 hours in Kornberg medium and M9 supplemented with lactose, respectively). Meanwhile, *E. coli* JM109/pJW-glgCB cells contained over seven times the total sugar concentration of the control, again in keeping with previous findings. As before, *E. coli* JM109/pJW-glgB showed sugar concentrations similar to those of the control.

Over the course of the nine days (for the first experiment) or 11 days (for the second experiment) under starvation conditions, both *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgB showed a slight decline in total sugar content (in the first experiment: significant for *E. coli* JM109/pJW-glgB ($p<0.05$), but not for *E. coli* JM109/pJW-lacZ ($p>0.05$); in the second experiment: significant for both transformants ($p<0.05$)), which appeared to level off and remain constant after around 24 hours, at around 20 µg/ml in the first run of the experiment and around 10 µg/ml in the second run. This seems to correlate with the culture densities observed over the course of the first experiment (figure 4.5). Both strains showed a similar total sugar concentration at every time point (for the first experiment there was a significant difference between *E. coli* JM109/pJW-lacZ and *E. coli* JM109/pJW-glgB for the first two data points ($p<0.05$), but not for any others. For the

second experiment there was no significant difference between the two strains at any data point ($p>0.05$)), which would seem to refute the hypothesis that *E. coli* JM109/pJW-glgB were metabolizing their glycogen stores more slowly.

Meanwhile, the total sugar content of *E. coli* JM109/pJW-glgC, which was over twice as high as that of the control at the beginning of both experiments, was observed to decline much more rapidly during the first 24 hours of starvation conditions. After this period the sugar content seemed to stabilize, at close to 50 µg/ml in the first experiment and 30 µg/ml in the second experiment, and remained around that level for the remainder of each experiment. Cells from each culture were stained with iodine and observed under the microscope at the start and end of these experiments and, in the case of *E. coli* JM109/ pJW-glgC, the inclusion bodies they contained were clearly visible at both time points (figures 4.10 and 4.11). It therefore still seems conceivable that these inclusion bodies are composed of polysaccharide that is inaccessible to the bacteria, and as such they account for the decline in total sugar content of these cells stopping at a higher concentration to that of the control, and remaining higher over the successive eight days under starvation conditions (significant for every time point in both experiments ($p<0.05$)).

In support of this theory, in the first experiment *E. coli* JM109/pJW-glgCB started the starvation period with over seven times the total hexose sugar content of the control cells, but also saw a rapid decline in these sugar levels to the point where, at around six days of starvation, the level dropped below that of *E. coli* JM109/pJW-glgC. By day nine the total sugar content of these cells was observed to be similar to that of the control, at 27 µg/ml compared to 28 µg/ml. This suggests that these cells were able to rapidly metabolize all of the additional polysaccharide they had accumulated during prior growth in high sugar media.

Unfortunately the final anthrone assay readings of the first experiment showed a high degree of variability, so that even though the difference between results for *E. coli* JM109/ pJW-glgC and *E. coli* JM109/pJW-glgCB were found by t-test to be significant ($p<0.05$), it was felt that they remained untrustworthy. The experiment was therefore repeated and extended, in order to better test the conclusions drawn. In the second experiment, the total hexose sugar levels of *E. coli* JM109/pJW-glgCB never dropped below those of *E. coli* JM109/pJW-glgC, despite the experiment being extended for another 48 hours. At the end of that experiment, there was no significant difference between the sugar levels of these transformants ($p>0.05$), while both remained significantly higher than the control ($p<0.05$).

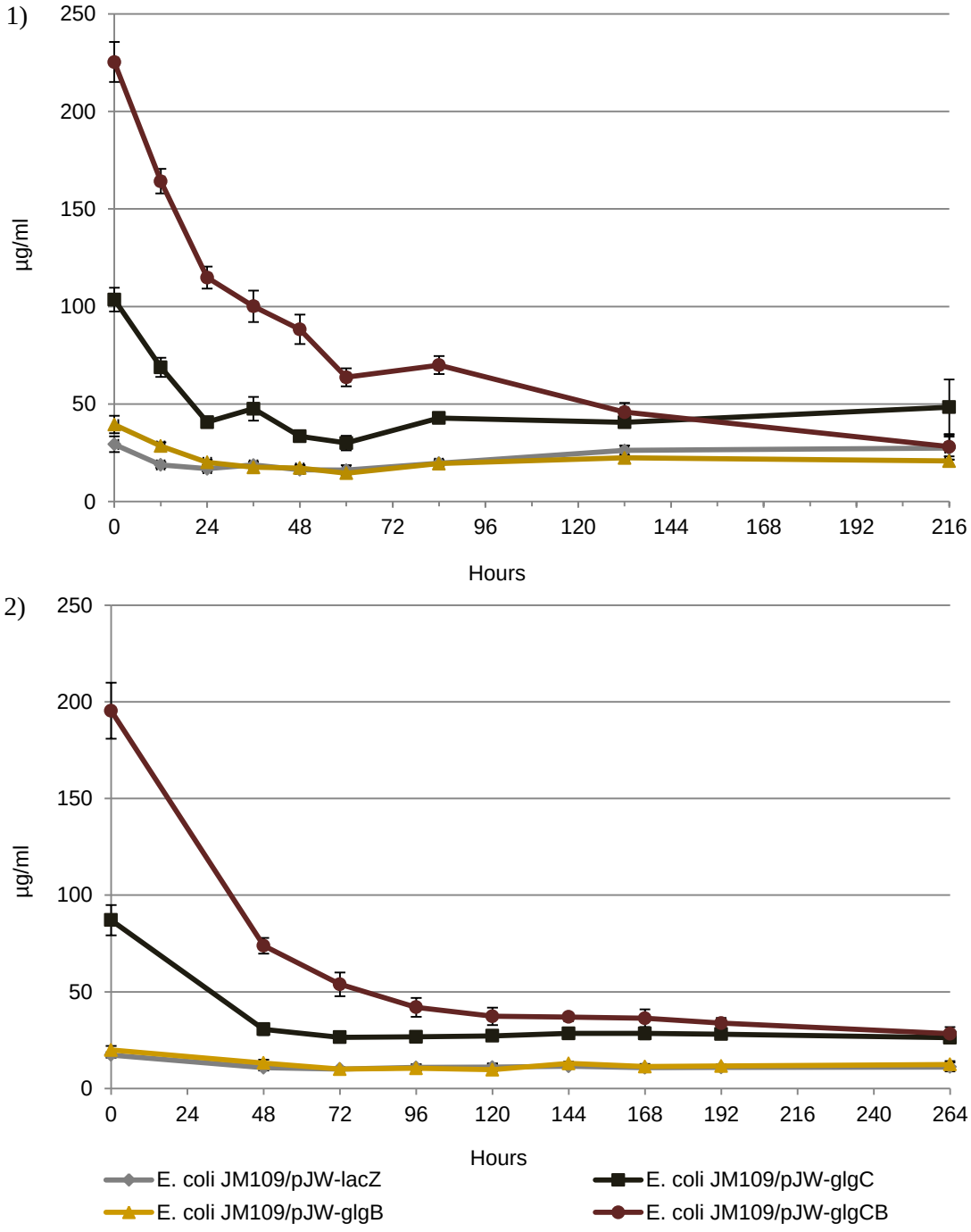


Figure 4.9: Total hexose sugar content of cells over the course of the starvation experiment for *E. coli* JM109/pJW-lacZ (control); *E. coli* JM109/pJW-glgC; *E. coli* JM109/pJW-glgB; and *E. coli* JM109/pJW-glgCB under starvation conditions in M9 without a carbon source. Error bars represent the standard error of the mean when n=3. Graphs 1) and 2) show results of two separate experiments, run under the same conditions.

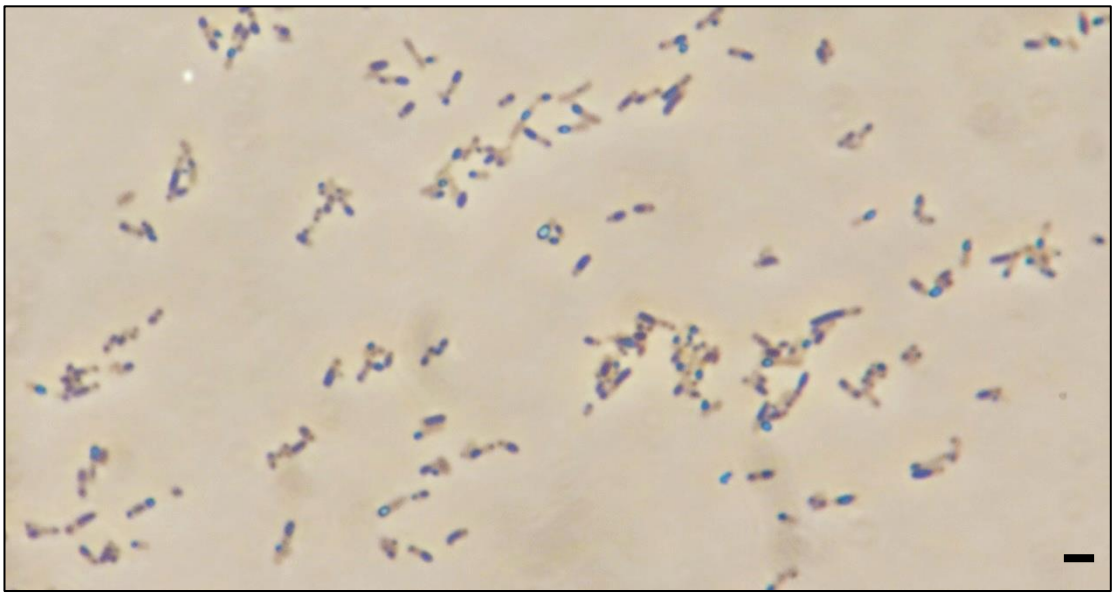


Figure 4.10: Light microscopy of iodine stained, *E. coli* JM109/pJW-glgC before starvation
From cultures of *E. coli* JM109/pJW-glgC used in the starvation experiment, imaged on day 1, showing clear inclusion bodies. Scale bar represents 5 µm (approx.)

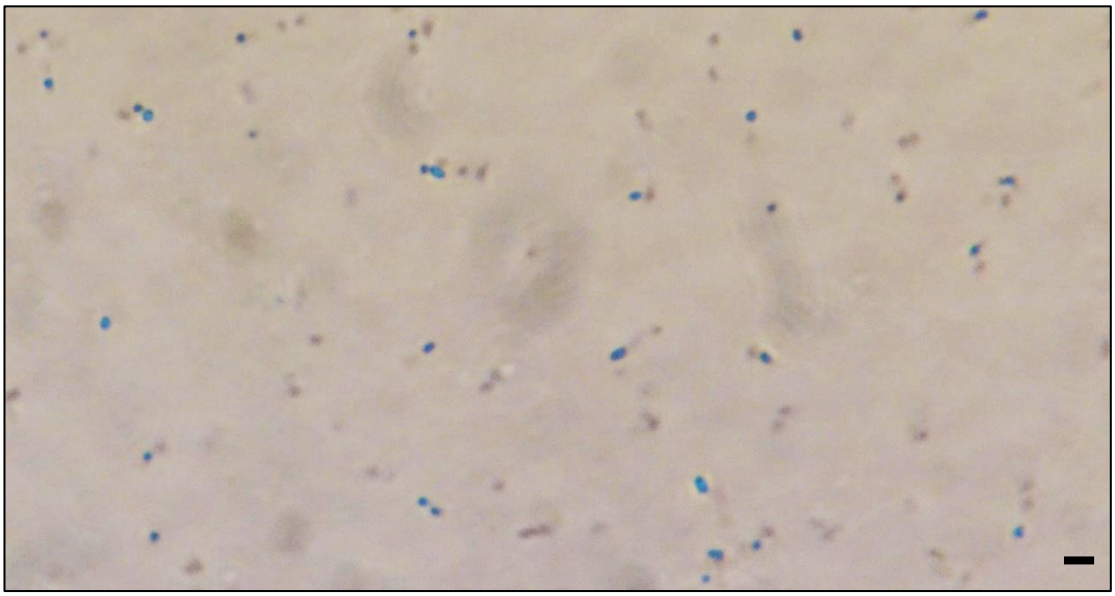


Figure 4.11: : Light microscopy of iodine stained, *E. coli* JM109/pJW-glgC after starvation
From cultures of *E. coli* JM109/pJW-glgC used in the starvation experiment, imaged on day 8, still showing clear inclusion bodies. Scale bar represents 5 µm (approx.)

4.5 Discussion

This work has shown that *E. coli* JM109/pJW-glgCB do indeed grow to a significantly higher culture density than *E. coli* JM109/pJW-lacZ grown under the same conditions. These transformants also show a consistently higher total sugar content than *E. coli* JM109/ pJW-lacZ when grown in carbon-rich media. However, contrary to predictions, they have also proved to be far more vulnerable when exposed to starvation conditions, with culture densities dropping to below three times less than a control after nine days. The excess storage sugar is therefore perhaps an added stress to these cells under such conditions.

E. coli JM109/pJW-glgC were also more vulnerable to starvation conditions than the *E. coli* JM109/pJW-lacZ control, and also showed the slowest growth rate of any of the transformants. If indeed the inclusion bodies formed during prior growth in carbon-rich media are formed of polysaccharides, it seems as though these transformants were incapable of digesting them. Again, further testing is needed to confirm this, but if confirmed it could have implications for the synthesis of polysaccharides for industry, since large inclusion bodies that the bacteria are unable to digest should be easier to harvest than native glycogen granules.

The second starvation experiments performed (which was felt to be more trustworthy and conclusive than the first, with an extended experimental period and lower standard errors for most time points, in particular the final time point) additionally seemed to suggest that *E. coli* JM109/pJW-glgCB contained hexose sugars it was unable to digest, since anthrone assays for this transformant showed its total hexose sugar content level off at around the same concentration as that of the *E. coli* JM109/pJW-glgC. No inclusion bodies are observed in these cells under the microscope, so a hypothesis for this phenomenon has not as yet been formed. It may simply be a result of the very low cell viability by the end of the starvation period. Further work is therefore needed to clarify these results.

E. coli lack the dikinases (Glucan Water Dikinase and Phosphoglucan Water Dikinase) that starch metabolising organisms use to prize the α-1,4 linked glucan chains out of their double-helices in order to make them accessible to water soluble enzymes such as phosphorylase and debranching enzyme (GlgP and GlgX, respectively, in *E. coli*). Therefore the addition of *gwd* and *pwd* to the battery of transgenes expressed in these cells may allow for the digestion of the inclusion bodies. Such cells could therefore be tested alongside *E. coli* JM109/pJW-glgC, to see if indeed the additional enzymes cause a greater reduction of total sugar content during a period of starvation. If this turned out to be the case, it would add

much weight to the theory that the additional *glgC* is leading to the growth of longer unbranched regions of glucan chains, which then spontaneously twist into double-helices with adjacent chains, causing an aggregation of polysaccharide.

It was thought that *E. coli* JM109/pJW-glgB might show better survival rates under starvation conditions than the *E. coli* JM109/pJW-lacZ control, since work by other groups has indicated that bacteria containing more branched forms of glycogen survive for longer under stressful conditions, possibly because their glycogen granules are harder to digest and therefore rationed more stringently. However, no clear result to this effect was obtained; rather, the *E. coli* JM109/pJW-glgB behaved as the control in all conditions.

An analysis of the chain length distribution of the polysaccharides formed by all three transformants would be extremely enlightening. To this end, the polysaccharides could be digested to completion with isoamylase and pullulananse, and the mix of linear chains produced analysed by High Performance Liquid Chromatography or High Performance Anion-Exchange Chromatography (Sundberg et al., 2013; Corradini et al., 2012). This would also help to confirm the 'long chain' theory of the *E. coli* JM109/pJW-glgC, and help to unveil what is happening to the storage sugars of the *E. coli* JM109/pJW-glgCB cells.

Chapter 5: Starch Synthesis in a Bacterial Cell

5.1 Introduction

The interest in the genetic and biochemical pathways used in the creation of starch has been growing steadily over recent years, driven by practical concerns regarding food security and the need for sustainable energy with a low environmental impact. The world population is predicted to exceed nine thousand million by 2050, with a possible plateau at around 10 thousand million occurring within this century. As such, the challenge of providing enough food is becoming paramount, particularly if we also want to improve the lives of the 868 million people currently suffering from chronic hunger (Johnston et al., 2014).

Improvements are being seen in terms of a worldwide trend of increasing affluence, bringing with it higher food consumption. An analysis by UNICEF showed that the worldwide proportion of under-five year olds who are underweight declined between 1990 and 2011 from 28% to 17%, and this positive trend looks set to continue (Johnston et al., 2014). However, this increase in affluence presents a challenge for future food production, not only because people are eating more, but also because increased income has also shown a trend towards diets richer in meat and dairy, which require greater amounts of land and resources to produce. Furthermore, as the worldwide trend towards economic prosperity continues, the demand for energy also rises. Although the increasing availability of oil reserves continues to outstrip consumption, the rate of increasing production is slower than the rate of increasing demand, so that a deficit looks likely to soon become reality (USGS World Petroleum Assessment 2000). At the same time, concerns over anthropogenic global warming brought about by fossil-fuel emissions, coupled with the growing energy demands from our ever-increasing population, have triggered a search for alternative sources of energy. One such alternative is biofuel, which has been successfully implemented in many countries. However, the large-scale use of current biofuels, made from existing plant energy stores such as corn starch, creates a demand for even more arable land. Such vast land requirements are unworkable: food security becomes too heavily compromised, even allowing for the further destruction of natural habitats. The existing biofuel industry is comparatively extremely small, but still the rise in food prices caused by this competing land use has probably been felt.

A population growing in both size and affluence is therefore on trend towards a perfect storm of resource deficiency. There is a demand for more food of a type that also requires more

land and resources to produce. The current global food production has been estimated to contribute to 30% of the global greenhouse gas emissions (Johnston et al., 2014), and this will continue rise. Meanwhile, the same land is also being competed for to grow crops for biofuel, to feed growing energy demands, while the use of fossil fuels also continues to rise. One of the many negative effects of the increase in greenhouse gas emissions, and the anthropogenic global warming that accompanies it, is a significant reduction of the land able to produce either food or biofuel crops. Even allowing for the widescale destruction of natural ecosystems, demands will soon become unsustainable.

A huge amount of research is being carried out to try to address these issues, among which is ongoing work to unpick the complexities of the starch biosynthesis pathway. Starch currently accounts for around 70% of the global diet, while also being used as the primary stock for most biofuel production, so any improvements in yield, or novel and sustainable forms of production, would provide a significant benefit. Most groups working in this field are focusing on *Arabidopsis thaliana* or common crop plants such as *Zea mays* (maize), *Triricum aestivum* (wheat), *Oryza sativa* (rice) or *Solanum tuberosum* (potato). The understanding of the functions of the various enzymes involved in starch biosynthesis has, as a result of this work, been hugely improved, so that we now have a model of the enzymes and reactions necessary for biosynthesis. However, the precise means by which the tiered structure of starch is produced have not yet been deduced. Furthermore, the results garnered from the study of one organism often seem to conflict with the results obtained by another group working on a different organism. This is perhaps unsurprising, as the starch pathway shows a great deal of complexity and variability. The number of enzymes involved can range from as few as 10 to around 40, and numbers close to this top end are often observed within the higher plants most subjected to study. This increase in the number of genes involved seems likely to permit fine-tuning of the pathway, so that many of the enzymes perform subfunctional roles or have synergistic effects on their counterpart enzymes. This complexity and plasticity serves to obfuscate attempts to unravel the pathway, as gene deletions in one organism will not necessarily produce the same results in a different organism.

For this reason, a different approach has been taken in this project. Rather than study gene deletions within the starch synthesis pathway of a higher plant, a simple model organism – *E. coli* – has been chosen as the object of study. *E. coli* does not naturally produce starch, but instead synthesises its far simpler precursor, glycogen, using a core pathway of just five enzymes. Through the sequential addition of genes scavenged from the starch synthesis

pathway of the Chloroplastida, it was hoped that a point would be reached where the transformed *E. coli* started to produce polysaccharides that were more starch-like than glycogen-like, thereby elucidating a minimum pathway required for starch synthesis through this 'bottom-up' approach.

Initially, the chloroplastidial gene used to transform *E. coli* were the *isoamylase I* and *II* genes from *Zea mays*. There is much evidence to suggest that the isoamylases play a key role in starch synthesis, and the genes from *Zea mays* are both well characterised and readily available. However, in an attempt to reduce complexity in the system, a different species from within the Choroplastida was subsequently chosen as the donor of starch genes. *Ostreocuccus tauri* is the smallest free-living eukaryote, comparable to bacteria in both its cells size and its genome size. It is not as well characterised as species such as *Zea mays* or *Arabidopsis thaliana*. However, its overall reduced complexity is also reflected in its starch synthesis pathway, which seems to be composed of only around 10 genes, despite the organism synthesising true starch within its chloroplast. It is assumed that selective pressures have led to an overall reduction in the *O. tauri* genome, such that many of its metabolic pathways, including starch synthesis, are already near to the minimum required for function. It was therefore hoped that the introduction of *O. tauri* genes into the *E. coli* chassis would provide a more straightforward model for starch synthesis, which could act as a platform for further gene introductions, including genes from more species with complex pathways, in order to try to understand the effects those genes have on the fine-tuning of starch synthesis.

For this project, then, it was hypothesised that the addition of *isoamylase* genes, initially taken from *Zea mays*, might alter the structure of the transformed *E. coli*'s native glycogen, perhaps even by reducing the steric interference of the glucan branches of the outer tier and allowing the granules to grow beyond their normal size. It was further considered that, subsequently, harvesting all the starch synthesis genes from *Ostreococcus tauri* would lead to an arsenal of enzymes that worked harmoniously with each other, and might also provide results that were easier to interpret. The *isoamylase I* and *isoamylase II* genes were therefore cloned from the *O. tauri* genome, along with its *granule bound starch synthase* gene, which was thought to have had a role in the evolutionary transition from glycogen to starch, and is possibly still used in the synthesis of the long, tier-spanning amylopectin chains in some algae. After the addition of these transgenes, the confirmation of their expression and activity, and the analysis of any phenotypic changes observed, the sequential addition of further starch synthesis genes could be considered.

5.2 Transformation of *E. coli* with *Zea mays* isoamylase genes and the *Ostreococcus tauri* granule bound starch synthase gene

It was hypothesised that the major difference between starch and glycogen synthesis, which caused the formation of amorphous and crystalline tiers within the starch granule and allowed it to grow to such a large size, might be down to the action of the Isoamylase enzymes found within all starch-accumulating organisms but lacking from those that accumulate glycogen. Furthermore, the loss of *isoamylase* gene function has been shown in a number of organisms to lead to the synthesis of a glycogen-like polymer (termed phytoglycogen) rather than starch (Ball & Morell, 2003). These enzymes are thought to periodically trim the outer chains of a growing starch granule, relieving it from the steric interference that is known to limit the growth of glycogen granules (see introduction). Previous work in the French lab had led to the creation of a pSB1A2 plasmid containing a *lacZ*'-*α* minigene as well as the *isoamylase I* and *isoamylase II* genes of *Zea mays*, all under a lac promoter, termed pCF-isaI-isaII (see Methods 2.7). The plasmid also contained the *glgC16* gene, under the assumption that this would simply upregulate glycogen synthesis, giving the isoamylase transgenes more substrate to work with and therefore making the job of polysaccharide analysis an easier one (see chapters 3 and 4 for why this is not the whole story). *E. coli* were transformed with this plasmid, in the hope that the additional isoamylase enzymes would perform the same function on the bacteria's glycogen granules as they did on the maize's starch granules, intermittently trimming their outer tiers to permit extended growth. Although some change in polysaccharide content was thought to have occurred in these transformed cells, the results were inconclusive and did not result in the formation of a starch-like polymer.

Another enzyme possessed by starch accumulating organisms but not by glycogen accumulators is Granule Bound Starch Synthase. The function of this enzyme in higher plants is thought to be the synthesis of amylose chains within the amorphous regions of the starch granule. The existence of this amylose is considered to improve the storage efficiency of the granule, so that mutations lacking this gene still produce tiered starch granules, but of a less dense, 'waxy' variety that lack amylose (Zeeman et al., 2010). However, work by Ral et al. (2006) has shown that this enzyme is integral to the synthesis of the amylopectin structure of the starch granule in *Chlamydomonas reinhardtii*. They therefore suggest that this enzyme may have had a more vital role in early starch-accumulating organisms, which has been slowly usurped by the other starch synthesis enzymes over evolutionary time.

With this in mind, it was considered that the addition of a *granule bound starch synthase* gene to the arsenal of transgenes might aid the formation of a starch-like polymer in *E. coli* cells. Unfortunately, *Z. mays*, like many higher plants, possesses more than one version of this gene, which are thought to be responsible for the synthesis of starch in different tissues of the mature plant. Since this seemed to add an unnecessary complication to the experiment, and also considering that the change in the function of Granule Bound Starch Synthase from an enzyme integral to amylopectin synthesis in at least some algae to an enzyme with only a subsidiary function in higher plants such as maize might show corresponding differences when used to transform *E. coli*, it was decided that an algal variant of *granule bound starch synthase* gene would be used instead.

Chlamydomonas reinhardtii also possesses two *granule bound starch synthase* genes, for reasons that are still unconfirmed (Deschamps et al., 2008). However, the pikoeukaryotic algae *Ostreococcus tauri* was being investigated by several labs within the university, and as described in the introduction, seemed a strong candidate for harvesting genes, thanks largely to its reduced genome. A genomic prep was therefore made from cultures of this alga and the full *gbss* gene cloned from it (figure 5.2), the validity of which was later confirmed by sequencing.

The full caravan of genes: *lacZ*, *glgC16*, *isaI* (*Z. mays*), *isaII* (*Z. mays*) and *gbss* (*O. tauri*), was then constructed on a pSB1C3 plasmid backbone, rather than pSB1A2, in line with current standards of the Registry of Standard Biological Parts, to create the plasmid pJW-glgC16-isaI-isaII-gbss (see Methods 2.7). JM109 *E. coli* were transformed with this construct, and early iodine assays seemed extremely promising (figure 5.1).

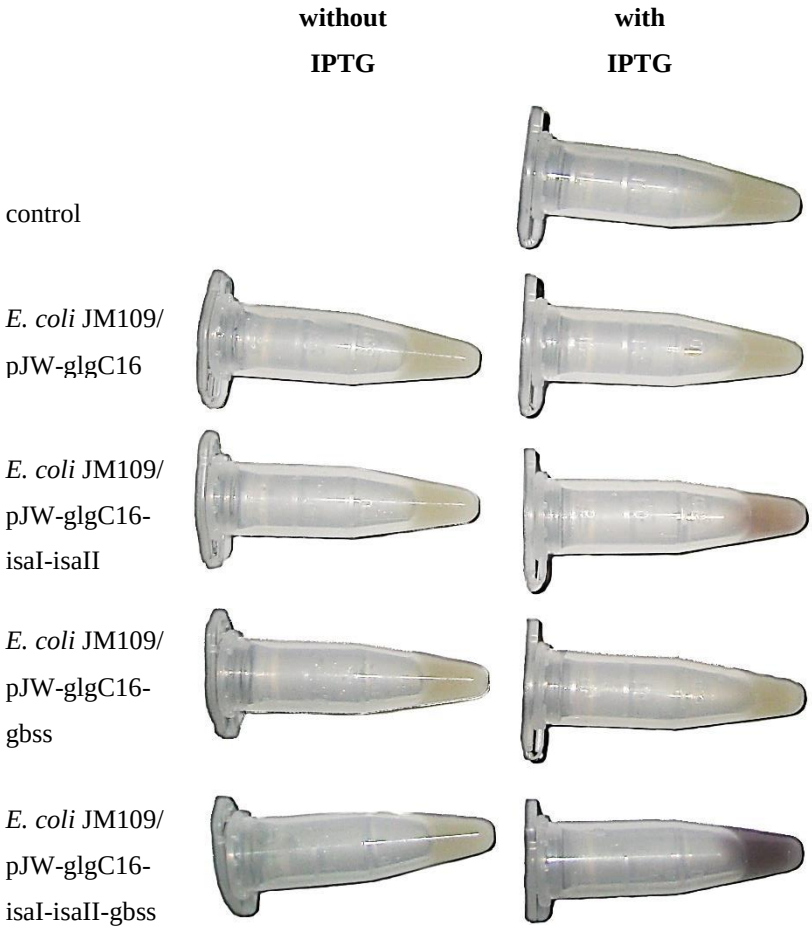


Figure 5.1: Iodine staining of JM109 *E. coli* transformed with various combinations of polysaccharide synthesis genes

Cultures were grown in LB, with or without IPTG, for 3 hours until reaching an $OD_{\lambda 600nm}$ of approximately 0.5. Cells from 2.4 ml of culture in each case were recovered by centrifugation and resuspended in 50 µl Lugol's iodine (0.2%). Cells transformed with the full pJW-glgC16-isaI-isaII-gbss construct were observed to hold the dark stain of the iodine for longer than those transformed with other constructs.

However, these results were extremely variable, and the darker staining of *E. coli* JM109/pJW-glgC16-isaI-isaII-gbss seen in figure 5.1 was not an immediate result. Rather, all the cell pellets except the control stained initially when resuspended in iodine, and the cells containing the full construct held the stain for longest. Subsequent repeating of the assay gave different, often utterly contradictory results.

To try to reduce this variability, cell culture and assay techniques were improved, as already described in chapter 3, most notably through the switch from glucose to lactose as the supplementary carbon source within growth media, the washing of pelleted cells prior to resuspension in Lugol's iodine and the addition of $CuSO_4$ solution and H_2O_2 to these suspensions in order to maintain the colour change observed in iodine-resuspended cells for long enough to obtain comparative results. However, a further problem was discovered.

Cells containing the full pJW-glgC16-isaI-isaII-gbss construct showed a marked polymorphic colony phenotype when grown on agar plates. Analysis of the plasmids within different sized colonies from these plates revealed that the construct was extremely unstable (figure 5.3). A fresh approach was therefore decided on, whereby all starch synthesis transgenes would be harvested from *O. tauri*, in the hope that this would improve reproducibility and lead to the expression of enzymes that had coevolved to function alongside one another, and thereby hopefully show a stronger and more reliable phenotypic change in the *E. coli* transformant.

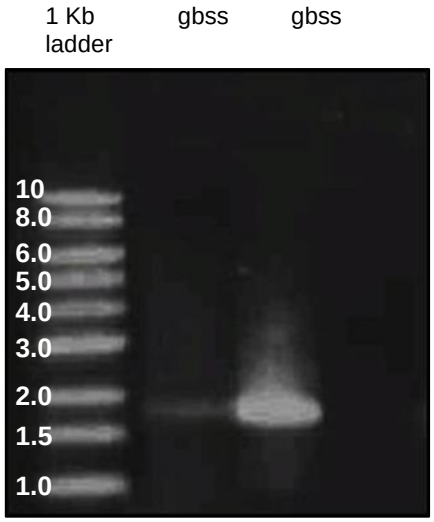


Figure 5.2: Gel electrophoresis of the PCR product of the *gbss* gene

Successful cloning from an *O. tauri* genomic prep was further confirmed by sequencing.

Expected sizes: glgB – 1725 bp

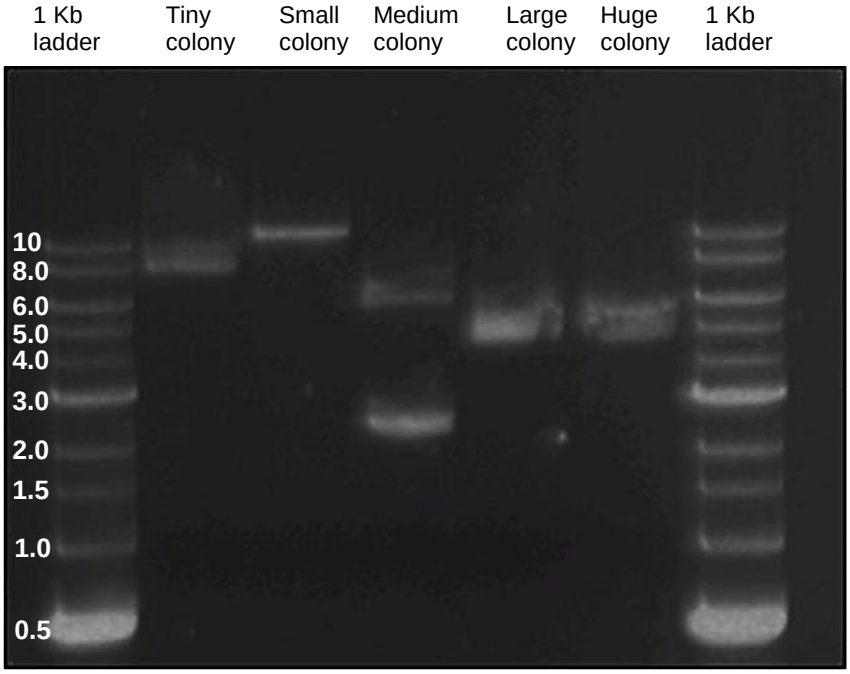


Figure 5.3: Gel electrophoresis of linearised plasmids from polymorphic colonies

Plasmid DNA was prepared from cultured colonies and linearised by digestion with *EcoRI*. All colonies were expected to contain pJW-glgC16-iasI-isaII-gbs, with expected size of ~11250bp (pSB1C3 – 2070; Plac-lacZ – 618; glgC16 – 1295; isaI – 2712; isaII – 2829; gbss – 1725)

5.3 Transformation of *E. coli* with *Ostreococcus tauri* isoamylase genes

The full *O. tauri isoamylase I* gene was ordered from GeneART, and used in the construction of a new pSB1C3 plasmid, in tandem with the *lacZ*'-*α* minigene, *glgC16* and *gbss*, to create pJW-glgC16-isaI-gbss (see Methods 2.7). This plasmid was found to be far more stable when used for transformations, so the polymorphic phenotype of colonies was no longer observed. Analyses of minipreps taken from transformants also showed a consistent plasmid size. JM109 cells transformed with this plasmid were observed to darken when resuspended in Lugol's iodine. However, the same reaction was seen to occur in cells transformed with the same plasmid but lacking either of the starch synthesis genes (figure 5.4). Furthermore, by this point the inclusion body phenotype had been observed in cells transformed with the *glgC16* gene alone (see chapters 3 and 4), and microscopy performed on cells transformed with *glgC16* and/or the *O. tauri* starch synthesis genes revealed that this phenotype was due to the presence of *glgC16* (figure 5.5). The starch synthesis genes therefore seemed to be having little, if any, effect on the polysaccharide synthesis of transformed *E. coli*. However, some phenotypic differences were observed. Cells transformed with *gbss* seemed to show a slower growth rate than control cells, and when streaked on M9 agar supplemented with lactose (1% w/v), cells transformed with *isaI* appeared to show a mucoid colony phenotype (figure 5.6). However, despite multiple repetitions, this phenotype was not observed again.

The multiple Isoamylase enzymes found in many plants and algae have been suggested to operate as a dimer. Isoamylase I is thought to be responsible for the catalytic activity of this dimer, but the presence of Isoamylase II has also been shown to be vital for starch synthesis in most higher plants investigated to date (Tetlow et al., 2006). An exception to this is Maize, from where the original *isaI* and *isaII* genes used in this study originated. Here, the absence of Isoamylase II alters the phenotype of the starch granule, but does not abolish accumulation (Kubo et al., 2010). The same applies for the Isoamylase II of the algae *C. reinhardtii* (Ball & Morell, 2003). Although the function of these enzymes has not been fully investigated in *O. tauri*, it was expected that Isoamylase I alone would affect polysaccharide formation. However, since there is also a distinct possibility that the two forms of Isoamylase identified in this organism form a dimer (see later), the *isoamylase II* gene was also cloned and expressed in *E. coli*. However, a problem encountered with the genes harvested from *O. tauri* first needed to be addressed: the identification and removal of any transit peptides, which may well affect enzyme function within bacterial cells.

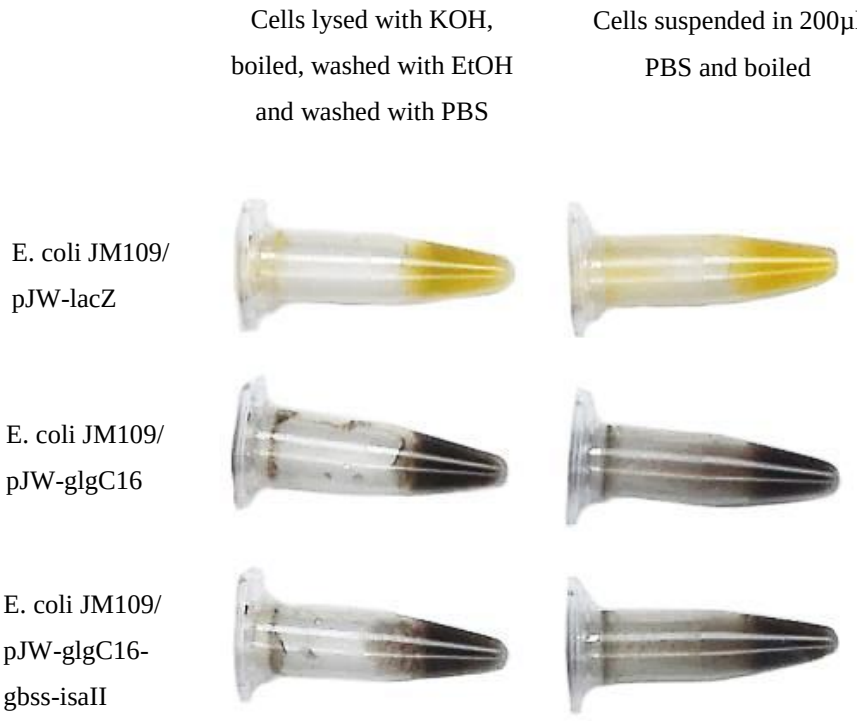


Figure 5.4: Iodine staining of JM109 *E. coli* transformed with *O. tauri* starch synthesis genes or two different control plasmids

Cultures were grown overnight in M9 + lactose (1% w/v), yeast extract (0.05% w/v) and IPTG, at 27°C and 200 rpm. Cells from 2.8 ml of culture in each case were recovered by centrifugation and treated as specified, before being resuspended in 200 µl PBS, plus 50 µl $CuSO_4$ (100 mM), 50 µl H_2O_2 (6%) and 25 µl Lugol's iodine (0.2%)

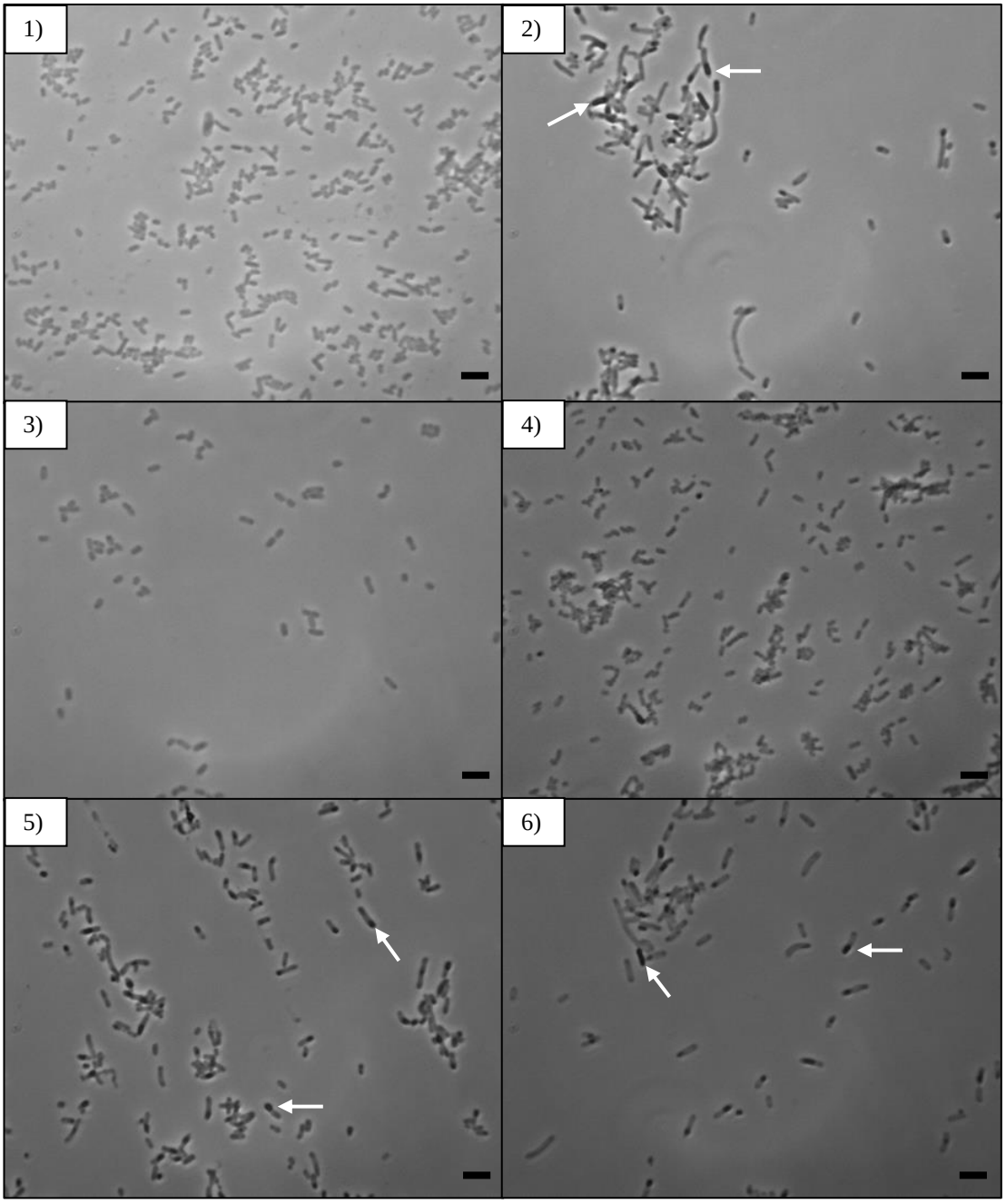


Figure 5.5: Light microscopy of iodine-stained, transformed JM109 cells

Cultures grown overnight in LB + lactose (1% w/v) and IPTG, at 37°C and 200 rpm.

1) *E. coli* JM109/pJW-lacZ; 2) *E. coli* JM109/pJW-glgC16; 3) *E. coli* JM109/pJW-gbss;

4) *E. coli* JM109/pJW-isaI; 5) *E. coli* JM109/pJW-glgC16-gbss; 6) pJW-glgC16-isaI.

Arrows point to inclusion bodies. Scale bars represent 5 µm (approx.)

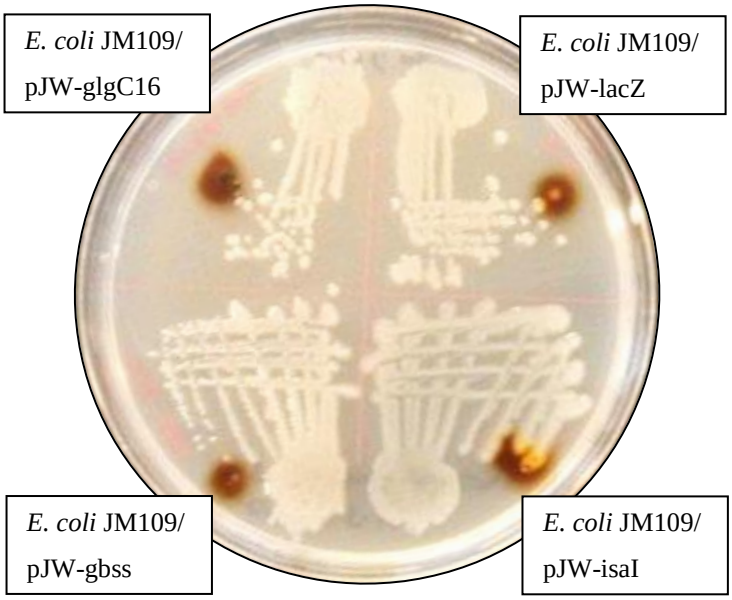


Figure 5.6: Iodine assays of transformed JM109 showing different colony phenotypes
Streaked JM109 transformants, on M9 agar supplemented with lactose (1% w/v) after overnight incubation at 37°C. Colonies transformed with pJW-glgC16 seem to stain with Lugol's iodine. Colonies transformed with pJW-isaI seems to show a mucoid phenotype, though this was never reproduced.

5.4 Transit Peptides

The theory that chloroplasts evolved from a free-living cyanobacterium that was engulfed by a eukaryotic cell is supported by a large body of evidence and now widely accepted. However, the transition from free-living cell to organelle does not seem to be an easy one, as suggested by the infrequency with which it appears to have occurred over the three and a half thousand million years (give or take) of life on Earth. For one thing, the internalised cell has to avoid being digested. In the case of the ancestor of the chloroplast, it is still unclear whether this was achieved by its escape from the phagosome, or by some other means. It might be that the cyanobacterium did not actually enter the host via a phagosome, but that instead a process inhibiting phagocyte fusion to an endoplasmic reticulum-derived membrane enabled this particular endosymbiosis (Nassoury & Morse, 2004). Whatever the means by which the cyanobacterium was engulfed, a second problem soon presented itself. The process of endosymbiosis involves the transfer of almost all genetic control from the symbiont to the host nucleus; it is really this that creates the distinction between an endosymbiont and an organelle. However, once it has occurred, the proteins needed for the endosymbiont to persist and function must be redirected back from the host to the endosymbiont (Patron & Waller, 2007). Typical free-living cyanobacteria contain around 3,200 genes. Chloroplasts, meanwhile, contain a genome encoding around 100 proteins that are synthesised semi-autonomously, but the function of these organelles still relies on around 3000 proteins. So more than 95% of the genes they need are located within the host nucleus, and the proteins which they encode are synthesised in its cytosol (Shi & Theg, 2012).

All primary plastids are surrounded by two membranes, which correspond to the membranes of the cyanobacterial endosymbiont and the eukaryotic host. Cyanobacteria are gram negative and as such have an outer periplasmic membrane as well as a plasma membrane. The process of engulfment would therefore almost certainly have led to the formation of three membranes around the organelle, corresponding to the two bacterial membranes and the additional membrane of the host vacuole. However, while the inner membrane still shows many similarities to the plasma membrane of cyanobacteria, the outer membrane looks most likely to be a chimera of the bacterial periplasmic membrane and the host's vacuole membrane (Nassoury & Morse, 2004). Regardless of their origin, they present a significant challenge. Firstly, proteins need to be targeted specifically to the correct organelle, as any mistargeting is likely to be cytotoxic (so proteins destined for the chloroplast must not end up, for example, in a mitochondrion). Even assuming this first step

is achieved, the proteins must then breach the membranes in order to function within the chloroplast.

The most common means by which proteins enter the chloroplast is via the Toc and Tic membrane channels (acronyms for Translocon of the Outer membrane of Chloroplasts and Translocon of the Inner membrane of Chloroplasts). These translocons are each composed of a number of proteins which recognise and actively transport the appropriate proteins through the membranes, at the cost of ATP. It seems entirely possible that the first steps in the assemblage of these transport systems was through the recruitment of existing secretory proteins normally found within the cyanobacterial membrane. Toc75, for example, is one of the most abundant proteins in the outer membrane of chloroplasts, and is homologous to a protein (SynToc75) found in the membranes of all gram negative bacteria, including cyanobacteria. Its role in prokaryotes is thought to be involved in the secretion of proteins, particularly virulence factors. However, a difference between these protein homologues is that the Toc75 of chloroplasts seems to be inserted into the chloroplastic membrane in the opposite orientation to its counterpart in free-living bacteria. Within plants and algae it has become a nuclear-encoded gene, so it seems plausible that, since it is now synthesised in the cytoplasm of the host cell, outside of the organelle, it is inserted into the organelle membrane in the opposite orientation to that if it had been synthesised inside the organelle. As such, its role has changed from being an excretory system to being an importer of proteins into the plastid (Bruce, 2000; Nassoury & Morse, 2004).

Another Toc protein, Toc34, also has some homology to a family of proteins in prokaryotes. These are the only small G proteins in the bacteria, and their region of homology corresponds to the GTP-binding domain of Toc34. One of the possible roles of Toc34 may be for recognition of preprotein sequences, and GTPase activity has been shown to indirectly stimulate association with the translocon complex (Nassoury & Morse, 2004).

However, the two major types of protein traslocation systems in prokaryotes, responsible for exporting proteins to the periplasmic space, are the SecY complex and the Tat pathway. The heterotrimeric SecY complex is the most highly conserved, and also has a direct homologue in eukaryotes, called the Sec61 complex (Rapoport, 2007). However, despite its widespread use, this protein translocation complex does not seem to have been recruited for protein import into organelles. Within non-green algae the genes that encode its proteins are still found within the chloroplast genome and synthesised within the chloroplast cytosol. Unlike Toc75, then, this complex will not embed itself in the chloroplast membrane in the opposite

orientation to that of free-living bacteria, but would instead remain a protein export system. For that reason, homologues of the complex found in chloroplasts are restricted to translocating proteins across the thylakoid membrane. Although it is synthesised outside the chloroplast in green algae and land plants, its function is still restricted to the thylakoid. Whether this is because, by the time this transition to the host occurred, the traslocation system into the chloroplast had already been established and there was no need for the role of this complex to be altered, or because a mechanism for protein translocation into the chloroplast was required to be separate from that used for secreting proteins from the cell, remains unclear. Likewise, the Tat pathway (acronymous for Twin Arginine Translocation, because the signal identifying proteins translocated by this pathway contain an obligatory pair of arginine residues immediately upstream of the hydrophobic region), which is found in prokaryotes including *E. coli*, has a homologue within chloroplasts that is also limited to protein transport across the thylakoid membrane, possibly for the same reasons, although its structure and mechanism of translocation remain unknown (Nassoury & Morse, 2004).

One theory suggests, then, that bacterial precursor of Toc75 moved from the endosymbiont to the host genome early in the evolution of the chloroplast. Due to its new orientation in the plastid membrane, it then started to import proteins rather than excrete them. Additional components were then added to this import machinery to increase efficiency, which would become more and more necessary as subsequent genes migrated from the plastid to the host. Most of these additional components seem to have been derived from host proteins.

Importantly, all the translocation systems of prokaryotes, including the SecY complex and the Tat pathway, require an identifying signal on the protein to be translocated, which, like chloroplast transit peptides, involve a tripartite design with a short, positively charged amino-terminal segment, a central hydrophobic segment of around 7 to 12 amino acids, and a more polar C-terminal segment that is recognised by the signal peptidase enzyme (Emanuelsson & von Heijne, 2001; Rapoport, 2007). The signal responsible for targeting proteins to the bacterial ancestor of Toc75 may therefore have been adapted to form some or all of what are now the transit peptides found in plants and algae. This is supported by the fact that the stromal peptidase responsible for cleaving the leader peptide of a plastid precursor protein shows sequence similarity to bacterial peptidases. Transit peptides have been shown to be organised as functional domains, suggesting that they were created through the shuffling of exons derived from cyanobacterial genes. Over time, selective pressure could therefore have led to the existence of transit peptides targeting nucleus-encoded

proteins back into the chloroplast. Early exons are thought to have encoded protein molecules of 15 to 30 amino acids in length, so a modern transit peptide could have evolved through the linking of three separate exons. Indeed, the transit peptides of several chloroplast precursors are still encoded by three exons. This shuffling may have also led to the creation of transit peptides with alternative domain organisations (e.g. coil-helix versus helix-coil). (Bruce, 2000). So targeting signals could have been attached to the N-terminus of the appropriate proteins through exon shuffling. Alternatively, over the course of evolutionary time, each gene from the endosymbiont may have transferred to the host nucleus many times, never resulting in a protein that could return to its site of effect until the gene found itself adjacent to a sequence able to function as a targeting signal (Nassoury & Morse, 2004).

Due to the similarity in the broad structure of plastid transit peptides from the Chloroplastida, Rhodophyta and Glaucophyta, it can be assumed that they evolved early on in the transition from endosymbiont to plastid (Nassoury & Morse, 2004). In addition to the dangers inherent in mistargeted proteins, this would seem to make it likely that all transit peptides would share well-defined primary or secondary structural motifs. However, modern transit peptides instead seem remarkably heterogeneous. Their lengths can vary from 13 to more than 100 residues, and no extended blocks of sequence conservation have been found within them. However, all of them share properties of an abundance of hydroxylated residues and few acidic residues, which gives them an overall positive charge (Shi & Theg, 2012; Jarvis & Robinson, 2004). Generally, stromal targeting transit peptides all share a structure of three distinct regions: an uncharged N-terminal domain of around 10 residues, beginning with methionine-alanine (MA) and terminating with glycine (G) or proline (P); a central domain enriched in threonine (T) and even more so in serine (S), and lacking negatively charged residues (aspartic acid (D) & glutamic acid (E)); and a C-terminal domain enriched in arginine (R). Plant transit peptides also contain a generally high level of alanine (around 18%) The N-terminal region, from residues 6 to 14, is thought to be crucial for the initial interaction with the translocation complex. The second region may be primarily responsible for interacting with the components of the inner membrane (Emanuelsson & von Heijne, 2001; Patron & Waller, 2007; Bruce, 2000). A large proportion of the transit peptides studied to date also show predicted binding motifs for molecular chaperones (Patron & Waller, 2007). The N-terminal signal peptides targeting the secretory systems of eukaryotes and prokaryotes, meanwhile, also show a tripartite design. The have a short, positively-charged N-terminal domain, a central hydrophobic domain and a C-terminal domain that is recognised by the signal peptidase enzyme.

In aqueous environments, transit peptides tend to be unstructured. However, portions of transit peptides have been shown to adopt α-helical structures when associated with lipid environments such as plastid membranes. These membrane-induced conformations may be an important feature for recognition, allowing seemingly disparate sequences to form common conformations at the chloroplast surface, enabling a receptor in the membrane to bind and facilitate the transport of many different precursors (Bruce, 2000).

Most of these findings have been made studying the transit peptides of higher plants, while far less research has been carried out to date on the transit peptides of algae. Studies that have been made have shown that algal transit peptides are more difficult to predict than those of higher plants. They still show a net positive charge and an abundance of alanine, but the hydroxylated content is lower than that of plant transit peptides, at 18.3% (rather than 26.6%), though serine is still typically the most common of these residues. Algal transit peptides also tend to be shorter (around 32 residues) and their secondary structures only composed of single helical elements (Patron & Waller, 2007).

5.5 Identification of transit peptides in *O. tauri* genes

The nature of transit peptides has caused unforeseen problems in the use of starch synthesis genes taken from *O. tauri*. The products of these genes are expected to naturally occur within the chloroplast, as this is where the single starch granule of the *O. tauri* cell is located. They are therefore predicted to contain N-terminal transit peptides. Due to the net positive charge of these transit peptides, and their similarity to signal peptides targeting bacterial secretory systems, it is thought that the synthesis of these proteins in their entirety within *E. coli* will lead to their possible secretion, and certainly result in reduced activity, since the enzyme is not present in its normal mature form. This may explain some part of the variability observed in results up to this point. Efforts were therefore made to identify and remove them, but as noted above, algal transit peptides are notoriously difficult to predict, and this was not a straightforward endeavour.

To add to the challenge, the N-terminal region of α-glucosidases, such as isoamylase, seems to be highly variable even after removal of the transit peptide, since it is this region that determines the size of the glucan chain cleaved during branching or debranching. The extended N-terminal regions of many of the enzymes investigated in this work are therefore likely to be a combination of transit peptides and specialized chain-length recognition sites, meaning that any overzealous cleaving could easily affect the enzymes' substrate specificity and lead to erroneous conclusions.

Several programmes exist that are used to identify N-terminal signals, including transit peptides. However, the use of these has failed to predict chloroplast transit peptides in some cases, and made unlikely predictions in others. In fact, the transit peptides of these enzymes often seem to show more similarity to mitochondrial presequences than higher plant chloroplast transit peptides. Such mitochondrial presequences are known to also be enriched in positively charged residues, in particular arginine (R), to lack negatively charged residues and to have the ability to form α-helices. The differences between these and chloroplast transit peptides therefore seem ostensibly very subtle, where the only major discrepancy is a rarity of arginine in the first 20 residues of chloroplast transit peptides, along with the common feature of an alanine at position two (Emanuelsson & von Heijne, 2001; Patron & Waller, 2007). A number of *O. tauri* starch synthesis proteins also possess arginine (R) and lysine (K) rich N-terminal regions.

As can be seen in figures 5.7 to 5.10, transit peptide prediction programmes seem to have a very limited use regarding these protein sequences. Only two of the programmes (TargetP and ChloroP, based on the same system) predict that any of the three proteins investigated (Isoamylase I, Isoamylase II and Granule Bound Starch Synthase) contain chloroplast-targeting transit peptides. Both programmes only make this prediction for one of the enzymes (Isoamylase I) and do so with low probability and with a cleavage site that would result in a suspiciously short signal sequence (16 residues long, where algal transit peptides are expected to be around 32 residues). The results from these analyses seem to suggest that two of the sequences (Isoamylase I and Granule Bound Starch Synthase) are more likely to contain mitochondrial presequences; an error of signal sequence predictors that has been noted previously for the signal sequences of *Chlamydomonas reinhardtii* chloroplast targeted proteins (Franze et al., 1990). The third sequence, for Isoamylase II, shows a generally very low probability for having any sort of target sequence at all.

targetp v1.1 prediction results
Cleavage site predictions included.
Using PLANT networks.

Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
O.tauri_ISAI	851	0.864	0.753	0.001	0.007	C	5	16
O.tauri_ISAII	571	0.025	0.782	0.026	0.622	M	5	105
O.tauri_GBSSI	574	0.276	0.613	0.014	0.096	M	4	74

Figure 5.7: TargetP signal sequence predictions for *O. tauri* IsaI, IsaII & GBSS
Results from TargetP (http://www.cbs.dtu.dk/services/TargetP/) on the sequences of three *O.tauri* starch synthesis enzymes: Isoamylase I, Isoamylase II and Granule Bound Starch Synthase. Isoamylase I is predicted to contain an N-terminal sequence slightly more likely to target the chloroplast than the mitochondria. However, the length of this predicted sequence is around half that which would be expected. Meanwhile, Isoamylase II and Granule Bound Starch Synthase are both predicted to have an N-terminal sequence targeting the mitochondria. Isoamylase II in particular has a very low score for the presence of a chloroplast transit peptide. All three predictions show a very low reliability. Emanuelsson & von Heijne (2001) have shown that TargetP predicts signal peptides with high sensitivity but performs less well on mitochondrial and chloroplast targeting peptides

cTP: the sequence contains a chloroplast transit peptide; **mTP:** the sequence contains a mitochondrial targeting peptide; **SP:** the sequence contains a signal peptide.

RC: Reliability class, from 1 to 5, where 1 indicates the strongest prediction. RC is a measure of the size of the difference ('diff') between the highest (winning) and the second highest output scores.

Predotar prediction results

Sequence	Mitochondrial	Plastid	ER	Elsewhere	Prediction
O.tauri_ISAI	0,67	0,03	0,00	0,32	Mitochondrial
O.tauri_ISAII	0,07	0,00	0,00	0,92	None
O.tauri_GBSSI	0,27	0,01	0,00	0,72	Possibly mitochondrial

Figure 5.8: Predotar signal sequence predictions for *O. tauri* IsaI, IsaII & GBSS
Results from Predotar (https://urgi.versailles.inra.fr/Tools/Predotar) on the sequences of three *O.tauri* starch synthesis enzymes: Isoamylase I, Isoamylase II and Granule Bound Starch Synthase. For all three, the programme shows a very low score for the presence of chloroplast transit peptides. Isoamylase I is predicted to contain an N-terminal mitochondrial target sequence and Granule Bound Starch Synthase is predicted to have an N-terminal sequence targeting it to a non-organelle location.

chlorop v1.1 prediction results

Name	Length	Score	cTP	CS-score	cTP-length
O.tauri_ISAI	851	0.539	Y	5.409	16
O.tauri_ISAII	571	0.432	-	4.997	3
O.tauri_GBSSI	574	0.483	-	10.360	21

Figure 5.9: ChlorP signal sequence predictions for *O. tauri* IsaI, IsaII & GBSS
Results from ChloroP (http://www.cbs.dtu.dk/services/ChloroP/), designed specifically to identify chloroplast transit peptides, on the sequences of three *O.tauri* starch synthesis enzymes: Isoamylase I, Isoamylase II and Granule Bound Starch Synthase. Isoamylase I is predicted to contain an N-terminal sequence that targets the chloroplast. However, the length of this predicted sequence is around half that which would be expected, as with the prediction from TargetP.

cTP: tells whether or not this is predicted as a chloroplast transit peptide containing sequence. This prediction is based solely on the 'score'.
CS-score: is the MEME scoring matrix score for the suggested cleavage site.
cTP-length: is the predicted length of the presequence.

MitoProt II - v1.101 prediction results for ISAI

Net charge of query sequence	: -23
Analysed region	: 49
Number of basic residues in targeting sequence	: 14
Number of acidic residues in targeting sequence	: 1
Cleavage site	**: 43**
PROBABILITY of export to mitochondria	**: 0.9939**

Cleaved sequence :

MATRAIASARSRAPARAREGTSLARARRWTRSPHRGVSGRRS

MitoProt II - v1.101 prediction results for ISAII

Net charge of query sequence	: -14
Analysed region	: 14
Number of basic residues in targeting sequence	: 3
Number of acidic residues in targeting sequence	: 0
Cleavage site	**: 13**
PROBABILITY of export to mitochondria	**: 0.0979**
Cleaved sequence	: MVRSGMRYGYRI

MitoProt II - v1.101 prediction results for GBSS

Net charge of query sequence	: -9
Analysed region	: 25
Number of basic residues in targeting sequence	: 8
Number of acidic residues in targeting sequence	: 1
Cleavage site	**: 23**
PROBABILITY of export to mitochondria	**: 0.9755**
Cleaved sequence	: MSRTAFEAKTNARRAGRAGVRA

Figures 5.10: MitoProt II signal sequence predictions for *O. tauri* IsaI, IsaII & GBSS
Results from MitoProt II (http://ihg.gsf.de/ihg/mitoprot.html), designed specifically to identify mitochondria-targeted signals, on the sequences of three *O.tauri* starch synthesis enzymes: Isoamylase I, Isoamylase II and Granule Bound Starch Synthase. Isoamylase I and Granule Bound Starch Synthase are predicted with high probability to contain N-terminal sequences that target the mitochondria. Isoamylase II shows a very low probability of possessing a mitochondrial target sequence.

Given the limited nature of these results, attempts were made to analyse the sequences by eye. First, a protein-protein BLAST search was carried out for each of the three enzymes, through the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the databank of non-redundant protein sequences. This method of estimation is obviously extremely rough, but does nonetheless seem to offer some insight. One of the more startling aspects of the BLAST results obtained was that for two of the proteins, Isoamylase I and Granule Bound Starch Synthase, there appeared to be an N-terminal 'overhang' of non-homologous residues before the amino acids start correlating with those of the corresponding proteins from different organisms (figures 5.11 & 5.13). The length of the Granule Bound Starch Synthase 'overhang' is 40 residues, which corresponds relatively well to the 32 residue region that is considered the average for an algal transit peptide. The 'overhang' for Isoamylase I, meanwhile, is much larger, at 115 residues. However, this may well be a consequence of variability at the N-terminus of α-amylase enzymes, as a result of their different site specificities.

This 'overhang' phenomenon was not seen in the case of Isoamylase II (figure 5.12). However, this is interesting in itself, since it matches the lack of signal sequence predictions for this enzyme from the various prediction programmes tried earlier (figures 5.7 To 5.10). It has been shown that Isoamylase II may form a dimer with Isoamylase I in several species (Tetlow, 2006), so it may be that the protein sequence for this enzyme in fact lacks any transit peptide sequence, and instead it dimerizes with Isoamylase I outside of the chloroplast and then piggybacks through the traslocation system, with one transit peptide working for the pair. However, this theory is problematic, as the proteins would be unlikely to dimerize until they had folded, which is only expected to occur once they are within the chloroplast. Further work is therefore needed to clarify this.

Due to the dimerization, this enzyme may also not need the extended N-terminal region necessary for determining the size of the glucan chain cleaved during debranching, since this function can be fulfilled by Isoamylase I, which would go some way to explain the relatively small size of the Isoamylase II sequence. Another possibility, however, is simply that the sequence for *O. tauri* Isoamylase II provided on the NCBI database is incomplete.

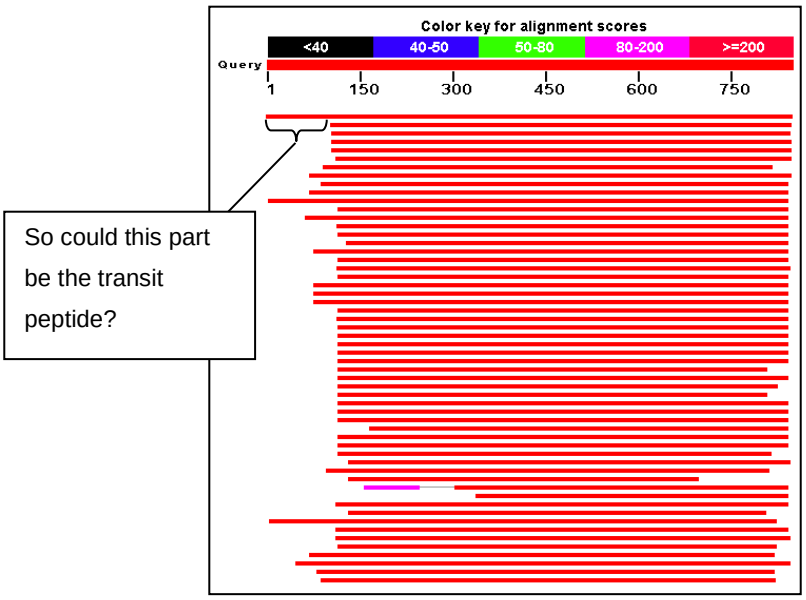


Figure 5.11: *Ostreococcus tauri* ISAI (Gene ID 9837697) protein-protein BLAST result
Schematic against a databank of non-redundant protein sequences. Most of the homologous sequences start at exactly residue 115 of the *O. tauri* sequence.

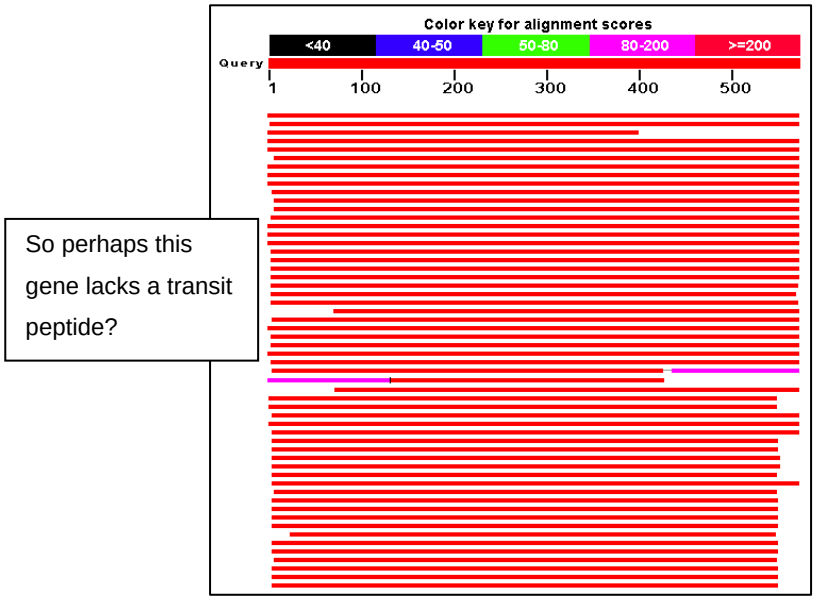


Figure 5.12: *Ostreococcus tauri* ISAII (Gene ID 9835977) protein-protein BLAST result
Schematic against a databank of non-redundant protein sequences.

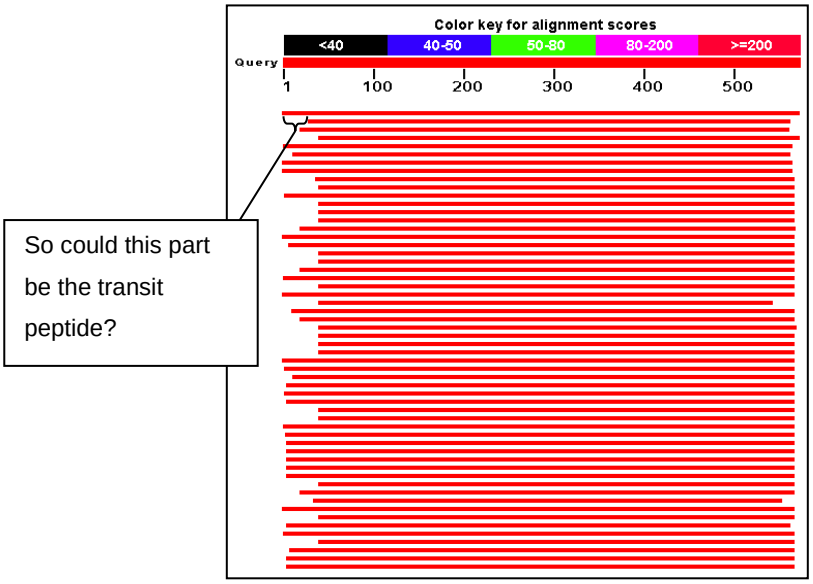


Figure 5.13: *Ostreococcus tauri* GBSS (Gene ID 9835561) protein-protein BLAST result
Schematic against a databank of non-redundant protein sequences. About half of the sequences start at exactly residue 40 of the *O. tauri* sequence.

A second 'by-eye' analysis was carried out, this time just on the α-amylase members of the *O. tauri* starch synthesis enzymes. This analysis was made possible by the fact that all α-amylase proteins contain conserved residues, responsible for the catalytic activity of these enzymes, which have been identified in all members of the family analysed to date. It was considered that the relative positioning of these conserved residues compared to those of α-amylase sequences with identified signal sequences may provide a further clue as to the location of the cleavage sites within the *O. tauri* starch synthesis enzymes. Figures 5.14 and 5.15 show the amino acid sequences of two α-amylase enzymes, *Arabidopsis thaliana* Isoamylase I and the *E. coli* debranching enzyme (which, in the case of *Arabidopsis thaliana*, is shown without its signal sequence, and in the case of *E. coli* does not have one). It can be seen from figures 5.16 and 5.17 that the conserved residues were identified in both of the *O. tauri* isoamylase sequences. As previously noted, the *O. tauri* Isoamylase II sequence is shorter than that of the other isoamylases, and the first conserved residue is found at amino acid 203, compared to position 335 in the *A. thaliana* sequence. Again, this may be due to the lack of an extended N-terminal region in the *O. tauri* protein, as it may not need to recognise specific polysaccharide branch lengths. The *O. tauri* Isoamylase I sequence, meanwhile, has its first conserved residue at position 389, so perhaps the 54 residue difference between this and the position of the conserved residue in the *A. thaliana* sequence corresponds to the cleavage site of the transit peptide. However, 54 residues is still long for an algal transit peptide, and as previously noted, the N-terminal region of these enzymes shows high variability even when ignoring the transit peptides. Furthermore, a general divergence in the sequences over evolutionary time will also obviously create variability in the sequence lengths before and after residues, as can be seen by the different sequence lengths of the different proteins in between their conserved residues. This means of analysis is therefore not a trustworthy one, but does still serve to contradict the previous suggested transit peptide cleavage points arrived at by the other methods of analysis.

```
0.    MDAIKCSSSF LHHTKLNTLF SNHTFPKISA PNFKPLFRPI SISAKDRRSN
50.   EAENIAVVEK PLKSDRFFIS DGLPSPFGPT VRDDGVNFSV YSTNSVSATI
100.  CLISLSDLRQ NKVTEEIQLD PSRNRTGHVW HVFLRGDFKD MLYGYRFDGK
150.  FSPEEGHYYD SSNILLDPYA KAIISRDEFG VLGPDDNCWP QMACMVPTRE
200.  EEFDWEGDMH LKLPQKDLVI YEMHVRGFTR HESSKIEFPG TYQGVAEKLD
250.  HLKELGINCI ELMPCHEFNE LEYYSYNTIL GDHRVNFWGY STIGFFSPMI
300.  RYASASSNNF AGRAINEFKI LVKEAHKRGI EVIMDVVLNH TAEGNEKGPI
350.  FSFRGVDNSV YYMLAPKGEF YNYSGCGNTF NCNHPVVRQF ILDCLRYWVT
400.  EMHVDGFRFD LGSIMSRSSS LWDAANVYGA DVEGDLLTTG TPISCPPVID
450.  MISNDPILRG VKLIAEAWDA GGLYQVGMFP HWGIWSEWNG KFRDVVRQFI
500.  KGTDGFSGAF AECLCGSPNL YQGGRKPWHS INFICAHDGF TLADLVTYNN
550.  KNNLANGEEN NDGENHNYSW NCGEEGDFAS ISVKRLRKRQ MRNFFVSLMV
600.  SQGVPMIYMG DEYGHTKGGN NNTYCHDNYM NYFRWDKKEE AHSDFFRFCR
650.  ILIKFRDECE SLGLNDFPTA KRLQWHGLAP EIPNWSETSR FVAFSLVDSV
700.  KKEIYVAFNT SHLATLVSLP NRPGYRWEPF VDTSKPSPYD CITPDLPERE
750.  TAMKQYRHFL DANVYPMLSY SSIILLLSPI KDP
```

Figure 5.14: *Arabidopsis thaliana* Isoamylase I amino acid sequence

The red highlighted letters are conserved residues, which are found in all α-isoamylases.

```
0.    MTQLAIGKPA PLGAHYDGQG VNFTLFSAHA ERVELCVFDA NGQEHRYDLP
50.   GHSGDIWHGY LPDARPGLRY GYRVHGPWQP AEGHRFNPAK LLIDPCARQI
100.  DGEFKDNPLL HAGHNEPDYR DNAAIAPKCV VVVDHYDWED DAPPRTPWGS
150.  TIIYEAHVKG LTYLHPEIPV EIRGTYKALG HPVMINYLKQ LGITALELLP
200.  VAQFASEPRL QRMGLSNYWG YNPVAMFALH PAYACSPETA LDEFRDAIKA
250.  LHKAGIEVIL DIVLNHSAEL DLDGPLFSLR GIDNRSYYWI REDGDYHNWT
300.  GCGNTLNLSH PAVVDYASAC LRYWVETCHV DGFRFDLAAV MGRTPEFRQD
350.  APLFTAIQNC PVLSQVKLIA EPWDIAPGGY QVGNFPPLFA EWNDHFRDAA
400.  RRFWLHYDLP LGAFAGRFAA SSDVFKRNGR LPSAAINLVT AHDGFTLRDC
450.  VCFNHKHNEA NGEENRDGTN NNYSNNHGKE GLGGSLDLVE RRRDSIHALL
500.  TTLLLSQGTP MLLAGDEHGH SQHGNNNAYC QDNQLTWLDW SQASSGLTAF
550.  TAALIHLRKR IPALVENRWW EEGDGNVRWL NRYAQPLSTD EWQNGPKQLQ
600.  ILLSDRFLIA INATLEVTEI VLPAGEWHAI PPFAGEDNPV ITAVWQGPAH
650.  GLCVFQR
```

Figure 5.15: *Escherichia coli* MG1655 GlgX (debranching enzyme) amino acid sequence

The red highlighted letters are conserved residues, which are found in all α-isoamylases. The yellow highlighted letters show regions of general homology to the *A. thaliana* sequence around the conserved residues, picked out by eye.

```
0.    ATRAIASARS RAPARAREGT SLARARRWTR SPHRGVSGRR SRATRWTLDA
50.   VSTEDAEGWR RETSKAKETR PTRTTVADAG RGVDDVHART RVAREAKRRT
100.  APRGVPGVCD TPRRGDASAL GATRVVGCPD DTVNFAVYTS AATAVSLVLW
150.  TPEGLARGEI AGEIELDETT NKTGSVWHVA LPRCAEDVLY GYRVDGPYEP
200.  EAGHRFDKSK ILLDPYAKFT VSRPEYGVAS KKEDGTEDCW PQYAGGVPKK
250.  LRSDGKEDFD WEGVTSPKRP MRDLVVYEAH ARGLTADLET KAKPGTYAAI
300.  EEALPYLKQL GVNAIELMPC HEFNEMEYHS LNHVTGEFRR NFWGYSTVNF
350.  FSPMTRYAEA GADDCGREAA REFKRMIREC HRAGIEVIMD VVFNHTAEGN
400.  EQGLTLSFRG LDNRVYYMVA PEGQFYNYSG CGNTMNCNHP VVREFILECL
450.  RYWVLEYHID GFRFDLASIL TRASSMWDRA NIFGEPTAET PMLEEVVIGT
500.  PLQDPPLIDA ISNDPVLAGT KLIAEAWDAG GLYQVGSFPH YGVWSEWNGK
550.  FRDDVRNFIK GVDGYAGLFA ERLCGSPNLY ADRSPSASIN FVTAHDGFTL
600.  RDCVSYNEKQ NHANGEENRD GEEHNASWNC GLSCDDDGEC WDPEIVALRD
650.  RQMRNFVVAL FVAQGVPMMY MGDEYGHTKC GNNNTYCHDN ALNWIDWSEA
700.  SSPLAGDGLA RFTKQVIALR KKHSAFRLDS FPSADNIQWH GHLPDTPMWD
750.  EESRFVAFTL QDKPETDKFY IAFNSHHEPA MLKLPSPPER CKWKLILDTS
800.  LESPFDVLSA EDIAEADSYT AEAMFLPGLR KNTYLCADRS AVIFRAVSLA
```

Figure 5.16: *Ostreococcus tauri* Isoamylase I amino acid sequence

Listed on the NCBI database as DBE 2 'isoamylase (IC)', Gene ID 9837697. The underlined region indicates the 'overhang' of the N-terminal region that shows no similarity to the sequences of the corresponding protein in different organisms. The grey highlighted region indicates a possible transit peptide, based on lining up the conserved residues with those of *A. thaliana*. The red highlighted letters are the conserved residues. The yellow highlighted letters show regions of general homology to the *A. thaliana* sequence around the conserved residues, picked out by eye.

```
0.    VRSGMRYGYR INGEGGWDTG QRFDATKVLM DPYAPLVEAR RKVFGEWSKH
50.   KVSGDTNDPD MLSGYDFESA PFDWRGVESP QIDEKDMIVY EMTVRAFTAD
100.  ASSGLDADAR GSYAGVAAKV EHLKSLGVNV VELLPVFEYD EMEFQRIPNP
150.  RDHMVNTWGY STMSFFAPMT RFGKKGCSAR EASREFKEMV RALHAAGIQV
200.  ILDVVYNHTG EMNDELPNTC SMRGIDNKTY YMTDTNQYVQ MLNFTGCGNT
250.  LNANNPYVSQ FILDSLKHWV KEYHVDGFRF DLASALCRDE QGHPMNSPPL
300.  IRAIAKDPEL AHVKLIAEPW DCGGLYQVGS FPNWDRWSEW NGAYRDVLRR
350.  FIKGDEGMKS DFARRISGSS DMYHHNNRKP YHSVNFITAH DGFTLRDLVS
400.  YNTKHNMANG EFNNDGANDN YSWNCXXXXX XXRLNNFQWN ELDSQKEHYF
450.  RFASEMVKFR RLHPLLGRET FLTDADVTWH EDRWDDPQSK FLAFTLHDSA
500.  GYGCGDLYIA FNAHEFYVDA ALPSPPNGKR WARIVDTNLP SPEDFIVKGK
550.  FGVESRYNVA PRGCVILMSK
```

Figure 5.17: *Ostreococcus tauri* Isoamylase II amino acid sequence

Listed on the NCBI database as 'DBEII (ISS)', Gene ID 9835977. The red highlighted letters are conserved residues. The yellow highlighted letters show regions of general homology to the *A. thaliana* sequence around the conserved residues, picked out by eye. The sequence is notable for being considerably shorter than that of other α-isoamylase proteins, and may lack both a transit peptide and an extended N-terminal domain used to dictate substrate specificity.

```
0.    MSVTHTSLSS INESLPRLAR ARAHVDAGRR RDGLIIRKSF GRSRTPIARA
50.   RAVSDVDVPD ATILRVRYHR TDGSSYAGWG AHAWGDVRAP TTWDAPLAAT
100.  LDDYSTWATF DVELVSGASR VSVLIHRGED QDCRAEDLDV SAGTREVWLV
150.  SGYSCAFEQE PDLKALPKGD VDKHRAIWVS ASVIAVPPDF ALSAAARESV
200.  KYSLVSSESA ELRVTGEGVE GGDGEPQTLE LIPTGLSKGA REKFPHIAAA
250.  DYRALELPPS APVRDLLRRQ LAVAAVDEDG APVDATGVQL QGALDELYAY
300.  DGPLGAEFGS DNLVTLRVWA PTAQNVTLSV FDEPRGDATR IEKSMTRDSR
350.  TGVWSVSGEF DGKFYNYNVT VFNPSTGQVS TNVSSDPYAR SLAADGRRAH
400.  VCDISKDDLK PAGWESFEKP TFTHPVDCAI YELHIRDFSA MDDTVSSAAR
450.  GKYVAFTESS STCVSHLRKL AQAGLTHVHL LPSYDFGSVP ELPENQKSVD
500.  FEQLAALPSN SSKQQELIEE CNWCDAFNWG YDPVHYGVPE GSYSTDADGS
550.  RRVLEYREMV KSLAENGLRV ICDVVYNHTL SSGPSDVNSV LDKIVPGYYH
600.  RRNFDGFIEA STCCNNTASE HYMMDRLIVD DLVHWARSYK VDGFRFDLMG
650.  HLMLSTMLKA KKALNSLTLE RDGVDGKSLY LYGEGWDYAE VAQGRVGKNA
700.  SQLNLAHTGI GSFNDRVREG CIGGSPFADP RLQGFLTGLY YSPNGVVQQG
750.  DEESQRYRMM EDGERILAAL AGNVRDFAYV DRHGVEVMAS SAVWPNSNVA
800.  YAGEPEETVN YVSAHDNETL FDSITLRVAD SISLAERCRI NHVATAIVAL
850.  AQGVPFFHAG DEILRSKSLD RDSYSSGDWF NRLDYSGETH NFGVGLPPAR
900.  KNGDRYAFIT EMLADLSMRP TKENIHAATK NLCELLSIRR STPLLRLRSS
950.  HDIQRRMKFY NRGPAQTPGL IIASINDGDA STPGLQALDS NYKRIVLAIN
1000.ATPRRISHEE AGLKVDFAGV ELKLHPLISD AVAAESSVVD GAYTIPPYTW
1050.AVFVQCR
```

Figure 5.18: *Ostreococcus tauri* BEII sequence

Listed on the NCBI database as '1,4-alpha-glucan branching enzyme/starch branching enzyme II (ISS)', Gene ID 9834842. The red highlighted letters are conserved residues, which are found in all α-isoamylases. The yellow highlighted letters show regions of general homology to the *A. thaliana* sequence around the conserved residues, picked out by eye.

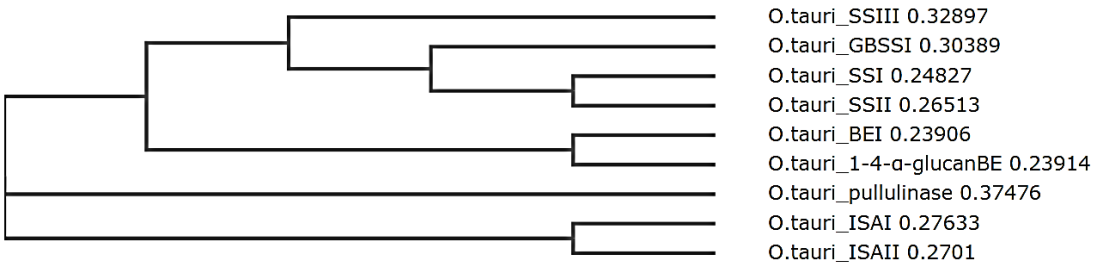


Figure 5.19: Phylogenetic relationship of *O. tauri* starch genes using Clustal

'*O.tauri*_pullulanase' is listed as '1,4-alpha-glucan branching enzyme/starch branching enzyme II' in the NCBI database (Gene ID 9834842).

The same method of comparing conserved residues was also applied to the *O. tauri* branching enzymes, since they too belong to the α-amylase family (figure 5.18). Although this yielded no further clues as to the location of the transit peptide cleavage sites, it did throw up the interesting finding that a BLAST analysis of the protein listed on the NCBI database as '1,4-alpha-glucan branching enzyme/starch branching enzyme II' (Gene ID 9834842) shows it has a far stronger homology to the pullulanases of other organisms than it does to branching enzymes. A further analysis of the *O. tauri* starch synthesis gene using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) provided a phylogenetic tree (figure 5.19) that placed this protein on a different branch to the other listed branching enzymes, in between them and the isoamylases, further supporting the idea that this protein is in fact a de-branching pullulanase. Although this was not investigated further during the course of this project, the existence of a third form of debranching enzyme within the *O. tauri* arsenal would make an interesting area for further study.

Finally, a third by-eye analysis of the protein sequences was carried out on the three *O. tauri* protein sequences (Isoamylase I, Isoamylase II and Granule Bound Starch Synthase), looking at the properties of the first 250 residues of each sequence to try to identify the features cited in the literature (an abundance of alanine as well as a high arginine/lysine content and serine/threonine content, with an overall hydroxylated residue content of around 18%, and a paucity of the negatively charged residues aspartic acid and glutamic acid). Figures 5.20 to 5.22 Show these sequences colour-coded to highlight the different types of residue. One of the first notable observations from this analysis is the general paucity of signal-sequence specific residues (alanine, positively charged residues and hydroxylated residues) within the first 50 amino acids of the Isoamylase II sequence, when compared to the other two enzymes (tables 5.1 to 5.3). Once again, this seems to support the hypothesis that, at least judging by the sequence available, this enzyme seems to lack its own transit peptide. Judgement of the other two enzyme sequences is less straightforward. Certainly trends are apparent, such as a low number of negatively charged residues at the start of each sequence. The first 50 or so residues of the Isoamylase I sequence certainly look to be fulfilling the vaguely-known requirements of an algal transit peptide, as do the first 30 or so residues of the Granule Bound Starch Synthase sequence. However, the data is very noisy, and no obvious cleavage point was considered to be apparent.

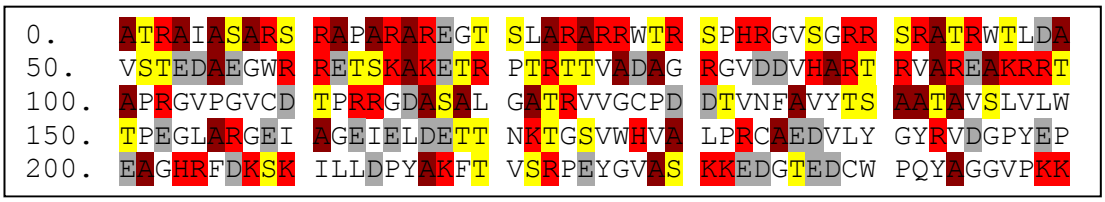


Figure 5.20: First 250 amino acids of the *Ostreococcus tauri* Isoamylase I

Listed on the NCBI database as DBE 2 'isoamylase (IC)', Gene ID 9837697. Alanine (A) residues are highlighted in dark red, the positively charged residues arginine (R), histidine (H) and lysine (K) are highlighted in bright red, and the hydroxylated residues serine (S) and threonine (T) are highlighted in yellow. Negatively charged residues (aspartic acid (D) and glutamic acid (E)) are highlighted in grey.

	A	R/H/K	S/T	D/E	Charge
1-10	4	2	3	0	2
11-20	3	3	1	1	2
21-30	2	4	2	0	4
31-40	0	4	2	0	4
41-50	2	2	3	1	1
51-60	1	1	2	3	-2
61-70	1	3	3	2	1
71-80	2	1	3	1	0
81-90	1	3	1	2	1
91-100	2	5	1	1	4
101-110	1	1	0	1	0
111-120	2	2	2	1	1
121-130	1	1	1	1	0
131-140	1	0	3	1	-1
141-150	3	2	0	0	2
151-160	1	1	1	2	-1
161-170	1	0	2	4	-4
171-180	1	2	2	0	2
181-190	1	1	0	2	-1
191-200	0	1	0	2	-1
201-210	1	4	1	2	2
211-220	1	1	1	1	0
221-230	1	1	2	1	0
231-240	0	2	1	4	-2
241-250	1	2	0	0	2

Table 5.1 Table of specific residues in the first 250 amino acids of *O. tauri* Isoamylase I

A: alanine; R: arginine; H: histidine; K: lysine; S: serine; T: threonine; D: aspartic acid; E: glutamic acid. The last column ('Charge') shows the number of positively charged residues minus the number of negatively charged residues.

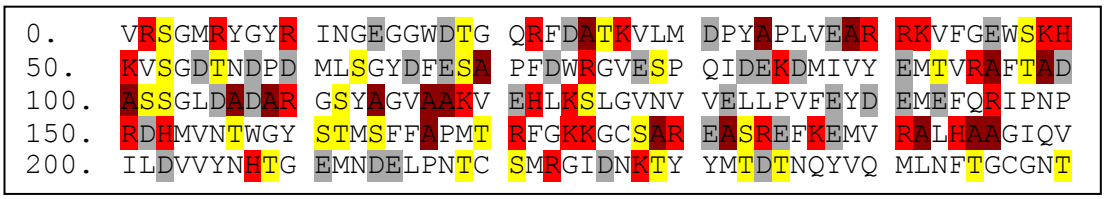


Figure 5.21: First 250 amino acids *of the Ostreococcus tauri* Isoamylase II

Listed on the NCBI database as 'DBEII (ISS)', Gene ID 9835977. Alanine (A) residues are highlighted in dark red, the positively charged residues arginine (R), histidine (H) and lysine (K) are highlighted in bright red, and the hydroxylated residues serine (S) and threonine (T) are highlighted in yellow. Negatively charged residues (aspartic acid (D) and glutamic acid (E)) are highlighted in grey.

	A	R/H/K	S/T	D/E	Charge
1-10	0	3	1	0	3
11-20	0	0	1	2	-2
21-30	1	2	1	1	1
31-40	2	1	0	2	-1
41-50	0	4	1	1	3
51-60	0	1	2	3	-2
61-70	1	0	2	2	-2
71-80	0	1	1	2	-1
81-90	0	1	0	3	-2
91-100	2	1	2	2	-1
101-110	3	1	2	2	-1
111-120	3	1	1	0	1
121-130	0	2	1	1	1
131-140	0	1	0	3	-2
141-150	2	1	2	2	-1
151-160	0	2	1	1	1
161-170	1	0	4	0	0
171-180	1	4	1	0	4
181-190	1	2	1	3	-1
191-200	3	2	0	0	2
201-210	0	1	1	1	0
211-220	0	0	1	3	-3
221-230	0	2	2	1	1
231-240	0	0	2	1	-1
241-250	0	0	2	0	0

Table 5.2 Table of specific residues in the first 250 amino acids of *O. tauri* Isoamylase II

A: alanine; R: arginine; H: histidine; K: lysine; S: serine; T: threonine; D: aspartic acid; E: glutamic acid. The last column ('Charge') shows the number of positively charged residues minus the number of negatively charged residues.

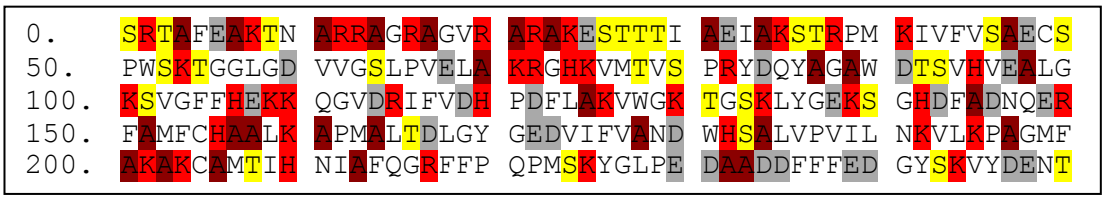


Figure 5.22: First 250 amino acids of the *O. tauri* Granule Bound Starch Synthase

Listed on the NCBI database as 'GBSS (ISS)', Gene ID 9835561. Alanine (A) residues are highlighted in dark red, the positively charged residues arginine (R), histidine (H) and lysine (K) are highlighted in bright red, and the hydroxylated residues serine (S) and threonine (T) are highlighted in yellow. Negatively charged residues (aspartic acid (D) and glutamic acid (E)) are highlighted in grey.

	A	R/H/K	S/T	D/E	Charge
1-10	2	2	3	1	1
11-20	3	4	0	0	4
21-30	2	2	4	1	1
31-40	2	2	2	1	1
41-50	1	1	2	1	0
51-60	0	1	2	1	0
61-70	1	0	1	1	-1
71-80	0	4	2	0	4
81-90	2	1	0	1	0
91-100	1	1	2	2	-1
101-110	0	4	1	1	3
111-120	0	2	0	2	0
121-130	1	2	0	1	1
131-140	0	2	2	1	1
141-150	1	2	0	3	-1
151-160	3	2	0	0	2
161-170	2	0	1	1	-1
171-180	1	0	0	3	-3
181-190	1	1	1	0	1
191-200	1	2	0	0	2
201-210	3	3	1	0	3
211-220	1	1	0	0	1
221-230	0	1	1	1	0
231-240	2	0	5	0	0
241-250	0	1	2	2	-1

Table 5.3 Table of specific residues in the first 250 amino acids of *O. tauri* Granule Bound Starch Synthase

A: alanine; R: arginine; H: histidine; K: lysine; S: serine; T: threonine; D: aspartic acid; E: glutamic acid. The last column ('Charge') shows the number of positively charged residues minus the number of negatively charged residues.

It was therefore decided that a definitive answer on the location of the cleavage point for the transit peptides of these proteins was beyond the scope of this study, and a more relaxed approach should be used. Since disruption of the transit peptide (in particular the beginning of it, which the literature concludes is most important for recognition by protein excretion systems) may well be enough to prevent the *E. coli* from excreting these enzymes, primer sets were designed to clone both *isoamylase I* and *granule bound starch synthase* from *O. tauri* genomic DNA, missing the first few residues (16 in the case of *isoamylase I* and 21 in the case of *granule bound starch synthase*), to create transgenes that would code for genes almost certainly still in possession of a small portion of their transit peptides, but hopefully not enough to affect activity. A second set of primers was also designed for each gene, this time knocking off 115 residues in the case of *isoamylase I* and 40 residues in the case of *granule bound starch synthase*, thereby knocking off the N-terminal regions that showed a lack of homology to similar genes from different organisms from the BLAST analysis (figures 5.23 and 5.25).

Figures 5.24 and 5.26 show that all sets of primers cloned DNA products of the expected size. However, upon insertion into the pSB1C3 backbone for transformation, the *granule bound starch synthase* lacking its first 21 residues was found to be extremely unstable (figure 5.27), and a culture was never grown that contained a single, appropriately sized version of this plasmid. Furthermore, transformation of *E. coli* with these plasmids was never found to elicit an obvious phenotype, and did not lead to colonies that stained with iodine any more than a control culture lacking *O. tauri* transgenes. However, *E. coli* have not so far been transformed with composite plasmids containing more than one *O. tauri* gene lacking its suspected transit peptide. In particular, *isoamylase II* has been cloned (in its entirety) from *O. tauri* genomic DNA (figure 5.28), but *E. coli* has not yet been transformed with a composite of this gene along with either of the *isoamylase I* variants lacking N-terminal (potential transit peptide) regions. The genes have also not yet been tested in union with starch branching enzymes.

```
0.     ATG GCA ACC CGT GCA ATT GCC AGC GCA CGT AGC CGT
36.    GCA CCG GCA CGC GCA CGT GAA GGT ACA AGC CTG GCA
72.    CGT GCA CGC CGT TGG ACC CGT AGT CCG CAT CGT GGT
108.   GTT AGC GGT CGT CGT AGT CGT GCA ACC CGC TGG ACC
144.   CTG GAT GCA GTT AGC ACC GAA GAT GCC GAA GGT TGG
180.   CGT CGT GAA ACC AGC AAA GCA AAA GAA ACC CGT CCG
216.   ACC CGT ACC ACC GTT GCA GAT GCA GGT CGT GGT GTG
252.   GAT GAT GTT CAT GCA CGT ACC CGT GTT GCC CGT GAA
288.   GCA AAA CGT CGT ACC GCA CCG CGT GGT GTT CCG GGT
324.   GTT TGT GAT ACA CCG CGT CGT GGT GAT GCA AGC GCA
360.   CTG GGT GCC ACC CGT GTG GTT GGT TGT CCG GAT GAT
396.   ACC GTT AAT TTT GCA GTT ……................................
2493.  …… AAA AAT ACC TAT CTG TGT GCC GAT CGT TCA GCC
2526.  GTT ATT TTT CGT GCA GTG AGC CTG GCA TAA TAA

F-1:   ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATG GCA CGT GAA
       GGT ACA AGC CTG
F-2:   ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATG GGT GAT GCA
       AGC GCA CTG
R:     ATCCTGCAGCTACTAGTA TTA TTA TGC CAG GCT CAC TGC AC
```

Figure 5.23: *Ostreococcus tauri isaI* primer design

The red letters are suspected to code for a transit peptide. Underlined and highlighted letters indicate primer homology sequences.

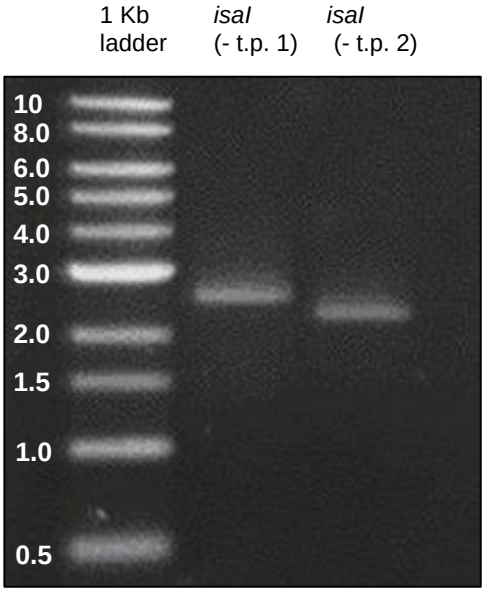


Figure 5.24: Gel electrophoresis of two PCR products of the *Ostreococcus tauri isaI* gene, each missing a suspected transit peptide region

Cloned from genomic DNA using two different sets of primers, designed to remove the suspected transit peptides. Expected sizes: isaI (-t.p. 1) – 2511 bp; isaI (-t.p. 2) – 2367 bp

```
0.      ATG AGT CGA ACG GCG TTC GAG GCG AAG ACG AAT GCG
36.     CGT CGG GCG GGA CGC GCG GGC GTT CGC GCG CGG GCG
72.     AAG GAA TCG ACG ACG ACG ATC GCG GAG ATC GCG AAG
108.    AGT ACG CGA CCG ATG AAG ATT GTT TTT GTT TCC GCG
216.    GAG TGC TCG CCG TGG AGT AAG ACC GGG GGG CTC GGG
252.    GAC GTC GTC GGG TCG CTT CCG GTG ..................
1656.   ..... GAG ACG ACC AAG AAG GCT GAG GTT GTC GTT CCC
1689.   ATC GCC AAG GCA GAC GCG CTC GCC GCT CAG ACG TGA

F:      ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAG ATG AGT CGA ACG
        GCG TTC GAG G
F-1:    ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAGATG GCG CGG GCG
        AAG GAA TC
F-2:    ATGAATTCCTTCTAGAGCTCAAGGAGGTACTAG ATG AAG ATT GTT
        TTT GTT TCC GCG G
R:      ATCCTGCAGCTACTAGTA TTA TTA CGT CTG AGC GGC GAG CG
```

Figure 5.25: *Ostreococcus tauri gbss* primer design

The red letters are suspected to code for a transit peptide. Underlined and highlighted letters indicate primer homology sequences.

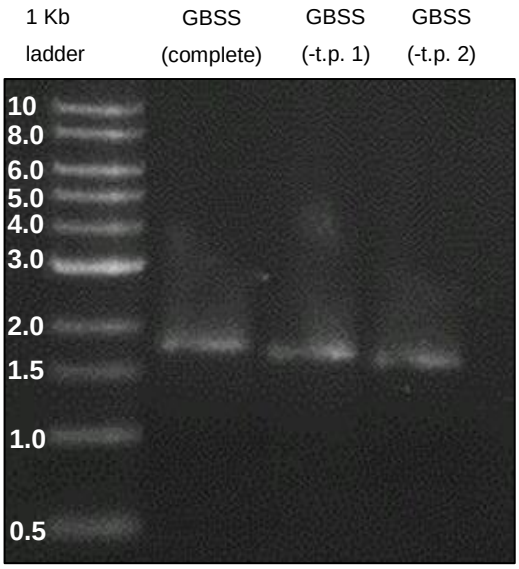


Figure 5.26: Gel electrophoresis of three PCR products of the *Ostreococcus tauri gbss* gene, missing two suspected transit peptide regions

Gel electrophoresis of *Ostreococcus tauri gbss* cloned from genomic DNA using three different sets of primers, designed to remove the suspected transit peptide.

Expected sizes: GBSS – 1725 bp; GBSS (-t.p. 1) – 1662 bp; GBSS (-t.p. 2) – 1605 bp

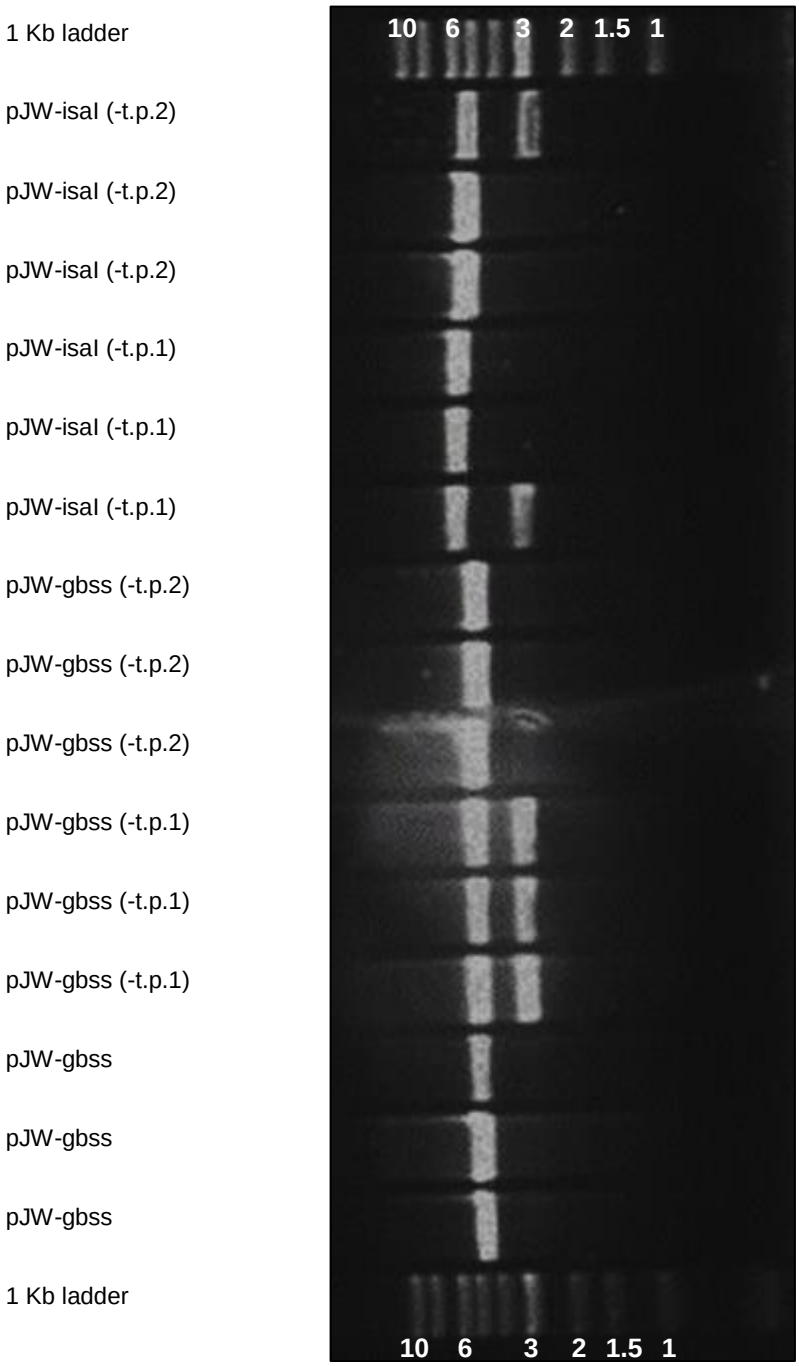


Figure 5.27: Gel electrophoresis of pSB1C3 constructs containing *isaI* & *gbss* cloned to remove some or all of their suspected transit peptides

Plasmid DNA was prepared from transformed cell cultures and linearised by EcoRI digestion prior to gel electrophoresis.

Expected sizes: pJW-gbss – 4422 bp; pJW-gbss (-t.p. 1) – 4359 bp;pJW-gbss (-t.p. 2) – 4302 bp; pJW-isaI (-t.p. 1) – 5205 bp; pJW-isaI (-t.p. 2) – 4908 bp.

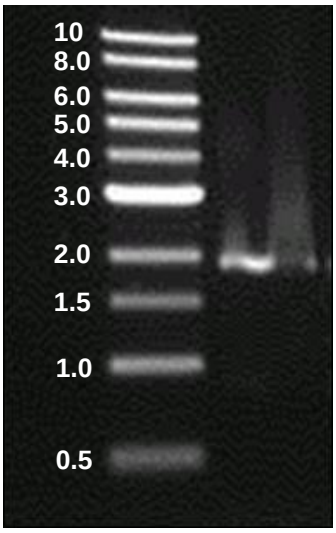


Figure 5.28: Gel electrophoresis of PCR product of *Ostreococcus tauri* Isoamylase II
Listed on the NCBI database as 'DBEII (ISS)', Gene ID 9835977. Cloned from genomic DNA
Expected size: 1736 bp

5.6 Discussion

Work to transform *E. coli* with starch synthesis genes from the pikoeukaryotic algae *Ostreococcus tauri*, in the hope of catalysing the syntheses a starch-like polymer, has so far yielded extremely limited and inconclusive results. However, the work to date has elucidated many issues that were unknown at the start of the project which suggest multiple avenues for further work.

Initial results seemed to strongly indicate an effect on polysaccharide synthesis in *E. coli* by the introduced Isoamylase I and Isamylase II enzymes from *Z. mays*. However, these results were later concluded to be the result of variability with culture and assay techniques, confusing the observations of phenotypic changes brought about by the tandemly expressed *glgC16* gene, which is not a starch synthesis gene, but rather a feedback resistant form of one of the *E. coli*'s own glycogen synthesis genes, which was added in the – retrospectively misguided – belief that it would simply lead to the production of more glycogen, which could then be used as a substrate by the Isoamylases. The inconclusive nature of these results does not prove the Isoamylases are not affecting polysaccharide synthesis, however, and these *isoamylase* transformants (without the additional *glgC16*) might still show a more obvious phenotype if cultured in media designed to maximise polysaccharide synthesis, and subjected to further assays, in particular a Native PAGE assay to confirm the activity of these enzymes.

A continuation of the work on the expression of *O. tauri* starch synthesis genes within *E. coli* is also still expected to yield results, which will hopefully give a clearer picture of the functions of each enzyme than would similar transformations using genes from higher plants. The next step in this work is the creation of a plasmid containing the gene for the version of Isoamylase I that lacks the first 16 residues of its suspected transit peptide in tandem with the *isoamylase II* gene, and its subsequent expression in *E. coli*. A Western blot, or an SDS PAGE assay after his-tagging the enzymes, could then be carried out on washed cell pellets of these transformants to confirm or refute the presence of the enzymes within the bacterial cells, prior to further assays to investigate polysaccharide structure. These protein assays could also be compared to gene expression assays, such as qRT-PCR or Northern blotting, to get an impression of how much of the protein is being maintained in the cell. For a more precise analysis of the effect of the suspected transit peptide region on the location of the enzymes expressed in bacterial cells, Western blots or SDS-PAGE assays could be carried out on both cytoplasmic and periplasmic extracts of cells containing the truncated protein

and the original transit peptide-containing protein, to see if there is a difference in the final location of the two forms of the enzyme. Once the location of the truncated enzymes had been confirmed, an activity assay would be required to confirm that they were still functional. In the case of Isoamylase, this would be best achieved with a Native PAGE assay, as mentioned previously.

Failing these methods, or simply to support the results obtained from work in transformed *E. coli*, the surest means of elucidating the exact location of the transit peptide in these protein sequences would be to go back to the original algae from which they were harvested. If the mature proteins were isolated from the chloroplasts of *O. tauri* they could be analysed by N-terminal sequencing, to determine which part of the N-terminal present in the database is missing from the mature protein. Once a putative transit peptide sequence has been identified – either by this method or testing one of the suspected transit peptide sequences already suggested in this chapter – the same sequence could be attached to the N-terminal of a reporter protein. This could then be localised within the algal cell, for example by fluorescence microscopy, to confirm if it had indeed been transported to the chloroplast.

The cytoplasm of an *E. coli* cell is still a very different environment from that of an algal chloroplast, so even if certainty could be achieved in trimming the transit peptide from these proteins while leaving the enzyme fully functional, there is still no guarantee that this alien protein would operate within its new host in the same way as in its native one. However, work by other groups (e.g. Sundberg et al., 2013) has shown that starch synthesis transgenes have had an effect on bacterial glycogen structure, even if it wasn't the expected effect. It may also be possible to extract the enzymes in question and test their function in vitro, which should be made considerably easier by the proteins being expressed in their mature form by *E. coli* cells. The systematic transformation of *E. coli* with starch synthesis genes harvested from *O. tauri* is therefore still considered to be a worthwhile pursuit, which is hoped to yield more viable results in the future.

Chapter 6: Conclusions and Future Work

6.1 Results presented in this work

Although the initial aim of this project – to elucidate the effects of starch synthesis genes by using them to systematically transform glycogen-synthesising *E. coli* – has not yielded significant results and is still ongoing, the work has led to the discovery of several distinct phenotypes in cells transformed with copies of their own glycogen synthesis genes.

E. coli JM109 that were transformed with an additional copy of their native *glgC* (*ADP-glucose pyrophosphorylase*) gene, expressed under a lac promoter on the high copy BioBrick assembly plasmid pSB1C3, were found to contain inclusion bodies, clearly visible under light microscopy, when grown in sugar-rich media (normal bacterial growth media supplemented with 1-2% (w/v) lactose or glucose) in the presence of IPTG. The presence of these inclusion bodies as physical entities within the mutant cells was felt to be confirmed by Transmission Electron Microscopy. The same phenotype was not observed in a control strain under any of the same conditions. It was therefore assumed that the presence of the *glgC* transgene was leading to the presence of the inclusion bodies.

That the upregulation of GlgC leads to an increase in glycogen accumulation in *E. coli* is a widely cited phenomenon (Govons et al., 1973; Ballicora et al., 2003), and has led to the belief that GlgC performs the rate-limiting step of glycogen synthesis. However, to the author's knowledge the presence of the inclusion bodies observed in this work is not cited in the literature. The working hypothesis for these inclusions is they are composed of polysaccharides, present in high quantity due to the upregulation of GlgC and therefore glycogen synthesis: GlgC upregulation is providing more substrate for GlgA (Glycogen Synthase), allowing it to meet a hitherto unrealised potential and extend the α-1,4 linked chains of the bacteria's glycogen storage polymers at a faster rate than normal. An extension of this hypothesis is that the glycogen Branching Enzyme (GlgB), which is under slightly different operonic control (the entire suite of *glgBXCAP* glycogen genes probably exist on a single operon, but with an additional suboperonic promoter within the *glgC* gene providing additional control of *glgA* and *glgP* – see Introduction) is not able to match the new rate of synthesis. This is leading to the formation of unbranched regions long enough to start to intertwine as double helices, which are thought to form spontaneously, for example within the crystalline lamella of amylopectin. This intertwining is perhaps leading to the aggregation of the mutant glucan polymers, visible as the inclusion bodies observed under

microscopy. Alternatively, the reduced branching rate is allowing individual granules to extend far beyond their normal size. The hypothetical maximum diameter of glycogen molecules is around 42 nm, at which point the steric interference between branches is thought to inhibit further growth (see Introduction). If this is the case, then increasing the rate of α-1,4 linked chain growth without increasing the rate of branching may be removing this steric interference, so that long external chains are able to grow uninhibited and crystallise among themselves.

In support of this hypothesis, the inclusion bodies are found to stain dark brown in the presence of iodine solution while the surrounding cell remains unstained, suggesting that the inclusions are composed of polysaccharide. Furthermore, an SDS-PAGE was performed on *E. coli* JM109/pJW-glgC16 grown in conditions to promote inclusion body formation (confirmed by microscopy), which showed no intense band in those cells compared to controls, suggesting that the inclusions are not composed of a single over-expressing protein. Anthrone assays were also performed on the *E. coli* JM109/pJW-glgC, showing them to have a significantly higher total hexose sugar content than *E. coli* JM109/pJW-lacZ grown under the same conditions.

However, results of the SDS-PAGE were not felt to be entirely conclusive, and the hypothesis put forward in this work would be better tested by more specific protein assays, such as a Western Blot, or an SDS-PAGE of the GlgC protein alone, made possible by his-tagging. In particular, although the SDS-PAGE performed in this work seemed to show that the inclusion bodies under investigation are not composed of a single protein, it was surprising that no particularly intense band was seen on the gel for the inclusion-containing cells, as this would seem to suggest that the transgene is not being expressed at as high a rate as expected. To clarify these issues, a comparison of gene expression – through Northern blotting or qRT-PCR – with enzyme concentration – through Western blotting or his-tagging – with enzyme activity – through measurement of substrate and product concentrations – is considered to be a valuable next step in this work.

Additionally, the Anthrone assays performed in this work were carried out on recovered, washed, whole cell pellets. They were therefore not a measure of the hexose sugars present within the inclusion bodies, but instead measured all hexose sugars within the cells. It therefore cannot be concluded that the higher sugar content found in *E. coli* JM109/pJW-glgC by this test is accounted for by the inclusion bodies, or even by an increase in glycogen content. Assays performed on the insoluble fraction of the cells might provide

more conclusive results. However, the cells contain other insoluble polysaccharides (for example, peptidoglycan) that may still be thought to confuse the results. More conclusive tests might be possible if a method was developed to purify the inclusion bodies, for example through a Percoll gradient. The inclusion could then be tested for sugar or protein content in isolation. Further work to investigate this phenomenon would be particularly enlightening, since the inclusion bodies formed by these transformants bear a striking resemblance to the glycogen α-particles found in the liver and muscle of animals. The means by which these α-particles are held together is also unknown. However, a link has been suggested between the size of glycogen α-particles in mammalian livers and the presence of diabetes, so any work that would help elucidate their formation and structure could provide medical benefits (Sullivan et al., 2012).

Within this work, it was felt that the hypothesis proposed might be tested by the introduction of an additional copy of *glgB* onto the same high copy BioBrick assembly plasmid as the *glgC* transgene, under the same promoter. It was thought that, if the proposed hypothesis was correct, then upregulating the Branching Enzyme to roughly the same degree might allow it to keep speed with the new high rate of linear glucan chain synthesis. It would therefore abolish the inclusion body phenotype, while leaving the mutant cells with the same increased total hexose sugar content as the *E. coli* JM109/pJW-glgC, since *glgC* is still upregulated to the same extent in both, but the increased branching within the glucan polymer should prevent the growth of linear chains to the point where they spontaneously start to form helices.

The abolition of the inclusion body phenotype was indeed observed in *E. coli* JM109/ pJW-glgCB cells. Furthermore, when assaying with iodine solution, *E. coli* JM109/ pJW-glgCB cells consistently stained a deep red colour, typical of wild-type glycogen, unlike the dark brown of the *E. coli* JM109/pJW-glgC cells. Although this is not considered in any way to be quantifiable data, it is felt that it supports the hypothesis that the branching ratio in *E. coli* JM109/pJW-glgCB cells is similar to that of wild type glycogen, while being dissimilar in *E. coli* JM109/pJW-glgC cells. However, the colour difference only indicates that the *E. coli* JM109/pJW-glgCB cells contain polysaccharides with slightly longer regions of single-helices than the *E. coli* JM109/pJW-glgC cells, so any conclusions regarding the overall structure of the polymers are considered extremely tenuous.

Furthermore, the prediction that *E. coli* JM109/pJW-glgCB cells would show the same increased total hexose sugar content as the *E. coli* JM109/pJW-glgC cells was not met.

Instead, these cells showed a total hexose sugar content that was consistently significantly higher than that of both the *E. coli* JM109/pJW-lacZ and the *E. coli* JM109/pJW-glgC cells. The reason for this is not clear.

It was also considered that if the inclusion bodies observed in *E. coli* JM109/pJW-glgC cells grown in high sugar media are indeed formed from polysaccharide that has spontaneously crystallised, the cells might well lack the machinery needed to unravel them, which would restrict digestion by GlgP. To test this, *E. coli* JM109/pJW-glgC, *E. coli* JM109/pJW-glgCB and *E. coli* JM109/pJW-lacZ transformants were exposed to starvation conditions for several days, with regular total hexose sugar assays performed during that time. If the total hexose sugar content of the *E. coli* JM109/pJW-glgC cells was found to plateau at a higher level than that of the *E. coli* JM109/pJW-lacZ cells, this was expected to add weight to the theory that they are unable to digest the inclusion bodies. This was indeed observed, and furthermore the inclusion body phenotype was still clear under light microscopy within cultures of *E. coli* JM109/pJW-glgC at the end of the starvation period. However, the total hexose sugar content of *E. coli* JM109/pJW-glgCB cells – which did not show the presence of inclusion bodies at any time – was also found to plateau at the same higher level as that of *E. coli* JM109/pJW-glgC. The reason for this remains unknown. However, both strains were found to be extremely vulnerable to starvation compared to a control, despite their hypothetically higher polysaccharide reserves. The plateau seen in the total hexose sugar content of the cultures may therefore have been an artefact caused by the reduction in cell viability; in other words, it may have been that there were simply not enough live cells left in the culture to make a significant difference to the sugar content by the end of the experiment. Further work is needed to clarify these findings.

In the work presented in Chapter 5, on the introduction of starch synthesis genes into an *E coli* host, it was hypothesised that the transformation of *E. coli* with genes synthesising Isoamylase enzymes might lead to the modification of the bacteria's native glycogen granules into a more starch-like polymer, since these enzymes are integral to starch formation. It is thought that this might occur through the trimming of loosely packed chains at the edge of the expanding polysaccharide granules, thereby removing the stress of steric interference that normally caps the growth of glycogen, allowing it to continue growing indefinitely with the tiered structure that is characteristic of starch. Unfortunately these experiments were met with a number of setbacks. *O. tauri* was selected as the donor species for the enzyme genes, due mostly to the relatively small number of genes in its starch

synthesis pathway, which was expected to provide the best chance of yielding uncomplicated results. However, the presence of transit peptides at the N-terminus of at least one of the Isoamylase enzymes being used was suspected to be interfering with results. It was thought that this could be confused within the bacteria for a signal sequence of the excretion system, leading to the possible excretion of the enzyme. The transit peptides of algae such as *O. tauri* are notoriously difficult to predict, and indeed the identification of a cleavage site was beyond the scope of this project. However, work has been carried out to remove at least part of these transit peptides, in the expectation that this will be enough to stop the confusion with proteins designated for excretion. These shortened enzyme structures have still failed to elicit a reproducible phenotype, but work is ongoing to test the effects of transformation with multiple starch genes, including the two forms of *isoamylase*, *granule bound starch synthase* and the two forms of the *branching enzyme*, and hope remains that progress will be made.

Granule Bound Starch Synthase is considered to be a possible key-player in starch formation. Although most previous investigations have found its role in starch formation to be confined to the synthesis of amylose chains within the amylopectin granule, it has been suggested that its role in synthesis has been usurped over time and that during the initial evolution of starch formation it may have played a crucial role. Therefore, the transformation of *E. coli* with a *granule bound starch synthase* gene taken from a primitive alga such as *O. tauri* is still hoped to contribute to the formation of a starch-like polymer.

Meanwhile, models developed by Milana Filatenkova (working at the University of Edinburgh, unpublished data) suggest that it may in fact be the presence of two different branching enzymes – one working on short chains and one working on long chains – that are the minimum requirement of a tiered polysaccharide granule structure. *O. tauri* branching enzymes have therefore been identified and are in the process of being cloned.

Lastly, *E. coli* were initially transformed with *isoamylase* genes taken from *Zea mays*. Although this transformation was never found to produce a reliable phenotype, the culturing and assay methods had not been optimised at that time. Since work by Sundberg et al. (2013) have shown that a similar transformation of *E. coli* with *A. thaliana isoamylase* genes does lead to a novel phenotype (although not the synthesis of a starch-like polymer), it is expected that a reproducible phenotype can also be obtained with the *Z. mays isoamylase* transformed genes, if cells are grown under conditions that maximise the synthesis of polysaccharides.

6.2 Choice of sugar supplement in the growth medium

In the experiments described in this report, *E. coli* was cultured in medium containing lactose supplemented with IPTG. These conditions have been shown to result in a 1.5-fold increase in LacZ production compared to cultures containing just lactose (Deutscher et al, 2012). However, several reports have indicated that IPTG has a toxic effect on *E. coli*, and that *lacA* expression leads to the synthesis of thiogalactoside transacetylase, used by the cell to acetylate IPTG, whereupon it no longer functions as an inducer of the *lac* operon. It seems likely that the IPTG concentrations used in this research were high enough to maintain expression throughout the culturing periods of each experiment. However, this has not been confirmed analytically, and furthermore the toxic effect of IPTG on the bacteria has not been taken into account, and would be an enlightening subject for further study (Roderick, 2005; Marbach & Battenbrock 2012).

Lactose was chosen as a carbon source predominantly because the transgenes being investigated were expressed under a lac promoter, so while their translation would be repressed by the presence of glucose (which was used initially as the carbon source) it would be activated by the presence of lactose (although this function of activation was already being served by IPTG). However, it was also expected that growth in lactose would increase the formation of polysaccharide stores, due to its effect on the phosphate transport system and the carbon storage response.

A lack of glucose in the growth medium leads to the enzymes of the phosphotransferase system being predominantly phosphorylated. This lifts the inhibition of Enzyme IIA^{Glc} from the LacY permease – which only occurs when Enzyme IIA^{Glc} is dephosphorylated – so that lactose can be actively transported into the cell. At the same time, the phosphorylated form of Enzyme IIA^{Glc} acts to stimulate the expression of the *cya* gene. This leads to increased production of cAMP, which binds to CRP and upregulates the expression of many genes, among them the *lac* operon, which is at no point repressed by LacI, thanks to the presence of IPTG. The *lac* genes are therefore fully expressed, allowing the maximum entry and metabolism of environmental lactose. LacZ cleaves lactose into galactose and glucose. The galactose is converted into glucose-1-phosphate by the enzymes of the *gal* operon, which is also upregulated by the presence of cAMP/CRP (Parks et al., 1971). The glucose and glucose-1-phosphate can then be converted by glucokinase and phosphoglucomutase, respectively, to form glucose-6-phophate, which can be fed into glycolysis to supply the

immediate needs of the cell while also maintaining the supply of PEP which ensures the phosphorylated states of the phosphotransferase system. Alternatively, excess glucose-6-phosphate can be converted into ADP-glucose by GlgC and stored as glycogen. Meanwhile *mlc* is also expressed, through upregulation from the cAMP/CRP complex. Since Enzyme IIB^{Glc} is predominantly phosphorylated, Mlc is unable to bind to it and instead binds to the promoter regions of the genes coding for the phosphotransferase enzymes, repressing their expression. However, it also represses its own expression, thereby ensuring that its repression of the phosphotransferase enzyme genes never becomes absolute.

The phosphorylated state of the phosphotransferase system concurrently results in the inhibition of GlgP, due to its bound state with the phosphorylated form of HPr, resulting in a reduction of glycogen degradation. The cAMP/CRP complex also stimulates expression of *glgS*/*scoR*, which represses expression of surface composition genes, leading to a reduction of AMP, which lifts repression of GlgC, as well as a significant increase in the cell's pool of ATP and glucose, which allows for an increase of glycogen accumulation as a means to store those surplus metabolites. The cAMP/CRP complex also stimulates the expression of RpoS and small, regulatory RNAs including Hfq, which have a similar knock-on effect of stimulating *glgS*/*scoR* and subsequently increasing glycogen accumulation.

In theory, then, this choice of growth medium should increase glycogen accumulation even before the introduction of additional genes. However, no comparison of the glycogen accumulation of cells grown in lactose and IPTG has been made to that of cells grown in other media, so this conclusion remains hypothetical. Furthermore, it is complicated by observations that, contrary to expectations, the Enzyme IIA^{Glc} of cells grown in lactose actually remains approximately 50% (according to Hogema et al., 1999) or even 70% (according to Deutscher et al., 2006) dephosphorylated. They surmise that this surprisingly low level of phosphorylation of the phosphotransferase system is due to the relatively low levels of PEP compared to pyruvate in cells grown on lactose. A low PEP-to-pyruvate ratio is thought to lead to poor phosphorylation of the phosphotransferase proteins, because the first phosphoryl transfer steps are reversible (Meadow & Roseman., 1996). Because growth on lactose lowers the PEP-to-pyruvate ratio while also relying on the expression of genes induced by cAMP-CRP and on gene-products that are sensitive to inhibition by dephosphorylated Enzyme IIA^{Glc}, its metabolism slows its own uptake (Deutscher et al., 2006).

However, the continuation of this cycle shows that the reduction in lactose taken in will soon result in an increase in the PEP-to-pyruvate ratio, leading once again to phosphorylation of the phosphotransferase proteins. This causes an increase in cAMP/CRP expression as well in the liberation of LacY from Enzyme IIA^{Glc}, resulting in an increase of lactose intake. The system is therefore suggested to be self-regulating, alternating between conditions of increasing and decreasing lactose due to its indirect inhibition of its own transport protein, thereby maintaining a safe influx of lactose (Hogema et al., 1999).

Hogema et al. (1999) suggest that the reduction in lactose intake during growth on lactose has evolved as a means to protect the cells against the effects of excessive lactose influx, allowing the cells to adjust the uptake rate to their metabolic capacity. In evidence of this, they showed that cells with a mutated form of LacY that is insensitive to repression by the dephosphorylated form of Enzyme IIA^{Glc} were found to excrete large quantities of glucose; roughly one molecule for every molecule of lactose they consumed. When grown in the presence of supplemented cAMP, the lactose uptake rate of the mutant strain was higher still. However, the strain also showed a reduction in growth rate, suggesting that the excess lactose was taken up too quickly for optimal growth, causing waste of the carbon source through excretion of metabolites and an apparent stressing of the bacteria.

It is possible, therefore, that cells which are able to accumulate greater quantities of glycogen, such as those described in this study, will to a degree circumvent the negative feedback of lactose intake, since a large proportion of the lactose they consume will be converted into glycogen and not fed into glycolysis. It may therefore not create such a marked reduction in the ratio of PEP-to-pyruvate, so resulting in a higher overall rate of phosphorylation of the phosphotransferase system while reducing the deleterious effects of metabolite excess. Again, this has not been investigated, but it would be a fascinating avenue for further research.

6.3 Future work

A number of avenues for future work have been discussed throughout this report. Chief among the further work needed on this project are assays to investigate the activity of the enzymes under investigation, and to better describe the polysaccharides being formed within *E. coli* JM109/pJW-glgC and *E. coli* JM109/pJW-glgCB. As stated previously, a comparison of gene expression with enzyme concentration and enzyme activity is now considered a vital next step if any of the hypotheses presented within this work are to be taken forward. Furthermore, a chain length analysis of the polysaccharides found within the different strains

discussed in Chapters 3 and 4, by High Performance Liquid Chromatography or High Performance Anion-Exchange Chromatography, is considered the most conclusive way to test the 'extended glucan chain' hypothesis put forward to explain the presence of the inclusion bodies.

The starvation experiment described in chapter 4 could also be investigated further. It was felt that the hypothesis under investigation – that the inclusion bodies observed within *E. coli* JM109/pJW-glgC grown in high carbon media are composed of polysaccharide indigestible to the host cell – was neither corroborated nor refuted by this experiment. However, before any further experiments are carried out to this effect, it must be better confirmed whether or not the inclusion bodies observed really are composed primarily of polysaccharide, as hypothesised.

If this was confirmation was made, perhaps through the testing of inclusion bodies purified in solution, then the starvation experiment might be extended by further transforming mutant cells with the genes encoding the starch degradation enzymes Glukan Water Dikinase and Phosphoglucan Water Dikinase (*gwd* and *pwd*). If the hypothesis is correct that *E. coli* JM109/pJW-glgC are producing polysaccharides with longer unbranched regions of glucan chains than native *E. coli* glycogen, and that these chains are winding into double-helices which then prohibit degradation, then the addition of these dikinase genes might allow the inclusion bodies to be digested, since they are used by starch accumulating organisms specifically for the purpose of unzipping polysaccharide double-helices for subsequence digestion. However, these additional enzymes alone may not be enough to allow GlgP access to the non-reducing ends of the glucan chains, so that a negative result obtained from this experiment could not be considered enough to prove or refute the hypothesis.

As stated previously, a comparison of growth rate and total sugar content of untransformed *E. coli* grown in glucose with and without IPTG versus lactose with and without IPTG would indicate whether growth in lactose leads to significant increases in glycogen, even without the inclusion of extra *glg* genes, and whether the IPTG used to supplement cultures throughout this work has any significant toxic effect. A test of IPTG-induced expression over time, to see if modification of the IPTG by the product of *lacA* leads to a reduction of transgene expression, would also be important for clarifying the results presented here. A comparison of the total sugar content of *E. coli* transformed with *glgC* versus those transformed with *glgC16* has also not been carried out to date, and may yield significant results.

Expressing the *glgC* gene on the plasmid pSB1C3, under a P_{lac} promoter, gives enough upregulation to lead to granule formation. However, initial experiments were carried out using *glgC16*; a variant of the gene that encodes a feedback-resistant version of the protein, which was expected to result in a greater phenotypic change, and no comparative studies were made of the phenotypes of cells transformed with these two gene variants. Furthermore, attempts were made to replace the *glgC* gene with *glgC16* in the genome of JM109 *E. coli* by homologous recombination (not reported). However, problems with the counter-selection method used in this homologous recombination have so far prevented the completion of the work. Time permitting, it would be nice to try again with an alternative strategy, to investigate whether the substitution of *glgC* with *glgC16* in the genome could still lead to the formation of inclusion bodies.

In the investigation of starch synthesis genes, an important first step would be to re-examine the phenotype of cells transformed with the *Zea mays* form of *isoamylase I* and *isoamylase II*, after growth in conditions that maximise polysaccharide synthesis, to clarify whether or not these transgenes have any effect of polysaccharide synthesis. Work is also ongoing to clone versions of each *O. tauri* starch synthesis gene that are missing at least the N-terminal region of any potential transit peptide. Various methods have been suggested to confirm whether these attempts have been successful, and if they are then the systematic transformation of *E. coli* with these genes is still expected to have an effect of polysaccharide synthesis, and should hopefully elucidate aspects of the starch synthesis pathway that have so far been obfuscated by the complexities of the pathway in higher plants.

Lastly, it has been suggested in previous work that Soluble Starch Synthase IV may be the primer for starch synthesis (D'Hulst & Merida, 2010). In support of this theory, *O. tauri* lacks an *ssIV* gene, but also only ever contains one starch granule, which is cleaved during cell division so that a part is inherited by each daughter cell, thereby providing the substrate for subsequent starch growth. It is therefore hypothesised that the transformation of *O. tauri* with *ssIV* from a different alga such as *C. reinhardtii* could lead to the synthesis of multiple starch granules within the *O. tauri* chloroplast. If this was observed to be the case, it would lend much support to the role of SSIV as the starch primer.

6.4 Concluding remarks

There is a pressing need to either improve the efficiency of production or find alternative means of supply for the polysaccharides required as both food and an energy source for the growing global population. Work in this project has shown that the transformation of *E. coli* with copies of its own glycogen synthesis genes can drastically alter polysaccharide synthesis and lead to a highly significant increase in the total sugar content of cells, while also affecting the cell densities reached by cultures. In the case of cells transformed with copies of both the *glgC* and *glgB* gene, the combination of high cell density and high total sugar content of the cells leads to an overall increase in the polysaccharide content of a culture to be more than ten times that of a control. This work may well be applicable to other organisms, providing the potential to improve polysaccharide yields, perhaps for use as a carbon source in biofuel production.

References

Aikawa, Shimpei, et al. "Glycogen production for biofuels by the euryhaline cyanobacteria *Synechococcus* sp. strain PCC 7002 from an oceanic environment." *Biotechnology for biofuels* 7.1 (2014): 88.

Alonso-Casajús, Nora, et al. "Glycogen phosphorylase, the product of the *glgP* gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*." *Journal of bacteriology* 188.14 (2006): 5266-5272.

Angermayr, S. Andreas, et al. "Energy biotechnology with cyanobacteria." *Current opinion in biotechnology* 20.3 (2009): 257-263.

Baker, Carol S., et al. "CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli*." *Molecular microbiology* 44.6 (2002): 1599-1610.

Ball, Steven G., and Matthew K. Morell. "From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule." *Annual review of plant biology* 54.1 (2003): 207-233.

Ball, Steven, et al. "The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis." *Journal of experimental botany* 62.6 (2011): 1775-1801.

Ballicora, Miguel A., Alberto A. Iglesias, and Jack Preiss. "ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis." *Microbiology and Molecular Biology Reviews* 67.2 (2003): 213-225.

Barker, Melanie M., et al. "Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation *in vivo* and *in vitro*." *Journal of molecular biology* 305.4 (2001): 673-688.

Binderup, Kim, René Mikkelsen, and Jack Preiss. "Limited Proteolysis of Branching Enzyme from *Escherichia coli*." *Archives of biochemistry and biophysics* 377.2 (2000): 366-371.

Binderup, Kim, René Mikkelsen, and Jack Preiss. "Truncation of the amino terminus of branching enzyme changes its chain transfer pattern." *Archives of biochemistry and biophysics* 397.2 (2002): 279-285.

Boehlein, Susan K., et al. "Enhancing the heat stability and kinetic parameters of the maize endosperm ADP-glucose pyrophosphorylase using iterative saturation mutagenesis." *Archives of biochemistry and biophysics* 568 (2015): 28-37.

Bruce, Barry D. "Chloroplast transit peptides: structure, function and evolution." *Trends in cell biology* 10.10 (2000): 440-447.

Cardol, Pierre, et al. "An original adaptation of photosynthesis in the marine green alga Ostreococcus." *Proceedings of the National Academy of Sciences*105.22 (2008): 7881-7886.

Cenci, Ugo, et al. "Convergent Evolution of Polysaccharide Debranching Defines a Common Mechanism for Starch Accumulation in Cyanobacteria and Plants." *The Plant Cell Online* 25.10 (2013): 3961-3975.

Cenci, Ugo, et al. "Transition from glycogen to starch metabolism in Archaeplastida." *Trends in plant science* 19.1 (2014): 18-28.

Chia, Tansy, et al. "A cytosolic glucosyltransferase is required for conversion of starch to sucrose in Arabidopsis leaves at night." *The Plant Journal* 37.6 (2004): 853-863.

Chiba, Seiya. "Molecular mechanism in alpha-glucosidase and glucoamylase." *Bioscience, biotechnology, and biochemistry* 61.8 (1997): 1233-1239.

Cho, Kye Man, et al. "Comparative analysis of the *glg* operons of *Pectobacterium chrysanthemi* PY35 and other prokaryotes." *Journal of molecular evolution* 67.1 (2008): 1-12.

Corradini, Claudio, Antonella Cavazza, and Chiara Bignardi. "High-performance anion-exchange chromatography coupled with pulsed electrochemical detection as a powerful tool to evaluate carbohydrates of food interest: principles and applications." *International Journal of Carbohydrate Chemistry* 2012 (2012).

Courties, Claude, et al. "Smallest eukaryotic organism." (1994): 255-255.

Critchley, Joanna H., et al. "A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in Arabidopsis." *The Plant Journal* 26.1 (2001): 89-100.

Dauvillée, David, et al. "Two loci control phytoglycogen production in the monocellular green alga *Chlamydomonas reinhardtii*." *Plant physiology* 125.4 (2001): 1710-1722.

Dauvillée, David, et al. "Role of the Escherichia coli *glgX* gene in glycogen metabolism." *Journal of bacteriology* 187.4 (2005): 1465-1473.

Dauvillée, David, et al. "Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate *Crypthecodinium cohnii*." *Proceedings of the National Academy of Sciences* 106.50 (2009): 21126-21130.

Delrue, B., et al. "Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin." *Journal of bacteriology* 174.11 (1992): 3612-3620.

Derelle, Evelyne, et al. "Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features." *Proceedings of the National Academy of Sciences* 103.31 (2006): 11647-11652.

Deschamps, Philippe, et al. "Metabolic symbiosis and the birth of the plant kingdom." *Molecular biology and evolution* 25.3 (2008a): 536-548.

Deschamps, Philippe, et al. "Early gene duplication within chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts." *Genetics*178.4 (2008b): 2373-2387.

Deschamps, Philippe, et al. "The relocation of starch metabolism to chloroplasts: when, why and how." *Trends in plant science* 13.11 (2008c): 574-582.

Deutscher, Josef, Christof Francke, and Pieter W. Postma. "How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in acteria." *Microbiology and Molecular Biology Reviews* 70.4 (2006): 939-1031.

Devillers, Claire H., et al. "Characterization of the branching patterns of glycogen branching enzyme truncated on the N-terminus." *Archives of biochemistry and biophysics* 418.1 (2003): 34-38.

D'Hulst, Christophe, and Ángel Mérida. "The priming of storage glucan synthesis from bacteria to plants: current knowledge and new developments." *New phytologist* 188.1 (2010): 13-21.

Dippel, Renate, and Winfried Boos. "The maltodextrin system of *Escherichia coli*: metabolism and transport." *Journal of bacteriology* 187.24 (2005): 8322-8331.

Dippel, Renate, et al. "The maltodextrin system of *Escherichia coli*: glycogen-derived endogenous induction and osmoregulation." *Journal of bacteriology*187.24 (2005): 8332-8339.

Ducat, Daniel C., Jeffrey C. Way, and Pamela A. Silver. "Engineering cyanobacteria to generate high-value products." *Trends in biotechnology* 29.2 (2011): 95-103.

Emanuelsson, Olof, and Gunnar von Heijne. "Prediction of organellar targeting signals." *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1541.1 (2001): 114-119.

Escalante, Adelfo, et al. "Current knowledge of the *Escherichia coli* phosphoenolpyruvate–carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation." *Applied microbiology and biotechnology* 94.6 (2012): 1483-1494.

Eydallin, Gustavo, et al. "An *Escherichia coli* mutant producing a truncated inactive form of GlgC synthesizes glycogen: Further evidences for the occurrence of various important sources of ADPglucose in enterobacteria."*FEBS letters* 581.23 (2007): 4417-4422.

Facchinelli, Fabio, et al. "Chlamydia, cyanobiont, or host: who was on top in the ménage à trois?" *Trends in plant science* 18.12 (2013): 673-679.

Falcon, Luisa I., et al. "Ultrastructure of unicellular N_2 fixing cyanobacteria from the tropical North Atlantic and subtropical North Pacific oceans." *Journal of phycology* 40.6 (2004): 1074-1078.

Figueroa, Carlos M., et al. "Understanding the allosteric trigger for the fructose-1, 6-bisphosphate regulation of the ADP-glucose pyrophosphorylase from *Escherichia coli*." *Biochimie* 93.10 (2011): 1816-1823.

French, Christopher E. "Synthetic biology and biomass conversion: a match made in heaven?" *Journal of the royal society interface* (2009): rsif20080527.

Fujii, Kazutoshi, et al. "Bioengineering and application of novel glucose polymers." *Biocatalysis and biotransformation* 21.4-5 (2010): 167-172.

Gosset, Guillermo, et al. "Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*." *Journal of bacteriology* 186.11 (2004): 3516-3524.

Govons, Sydney, et al. "Biosynthesis of Bacterial Glycogen XI. Kinetic characterization of an altered adenosine diphosphate-glucose synthase from a 'glycogen-excess' mutant of *Escherichia coli* B." *Journal of biological chemistry* 248.5 (1973): 1731-1740.

Groisman, Eduardo A., et al. "Bacterial Mg^{2+} homeostasis, transport, and virulence." *Annual review of genetics* 47 (2013): 625-646.

Haseloff, J. "Editorial: IET Synthetic Biology." *Synthetic biology, IET* 1.1.2 (2007): 1-2.

Henrissat, Bernard, Emeline Deleury, and Pedro M. Coutinho. "Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria?" *Trends in genetics* 18.9 (2002): 437-440.

Hollands, Kerry, Stephen JW Busby, and Georgina S. Lloyd. "New targets for the cyclic AMP receptor protein in the *Escherichia coli* K-12 genome." *FEMS microbiology letters* 274.1 (2007): 89-94.

Horn, Matthias. "Chlamydiae as symbionts in eukaryotes." *Annual review of microbiology*. 62 (2008): 113-131.

Hou, Zhenglin, et al. "Effectors of the stringent response target the active site of *Escherichia coli* Adenylosuccinate Synthetase." *Journal of biological chemistry* 274.25 (1999): 17505-17510.

Huang, Jinling, and Johann P. Gogarten. "Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids?" *Genome biology* 8.6 (2007): R99.

Izumo, Asako, et al. "Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in *Chlamydomonas*." *Plant science* 180.2 (2011): 238-245.

Jarvis, Paul, and Colin Robinson. "Mechanisms of protein import and routing in chloroplasts." *Current biology* 14.24 (2004): R1064-R1077.

Jobling, Stephen A., et al. "Production of a freeze–thaw-stable potato starch by antisense inhibition of three starch synthase genes." *Nature biotechnology* 20.3 (2002): 295-299.

Johanson, Ronould. "Anthrone in estimation of hexose sugars with special reference to pentose interference." *Analytical chemistry* 26.8 (1954): 1331-1333.

John, Michael, Jürgen Schmidt, and Helmut Kneifel. "Iodine-maltosaccharide complexes: relation between chain-length and colour." *Carbohydrate research*119 (1983): 254-257.

Johnston, Jessica L., Jessica C. Fanzo, and Bruce Cogill. "Understanding sustainable diets: a descriptive analysis of the determinants and processes that influence diets and their impact on health, food security, and environmental sustainability." *Advances in nutrition: an international review journal* 5.4 (2014): 418-429.

Kötting, Oliver, et al. "Regulation of starch metabolism: the age of enlightenment?" *Current opinion in plant biology* 13.3 (2010): 320-328.

Kolb, A., et al. "Transcriptional regulation by cAMP and its receptor protein." *Annual review of biochemistry* 62.1 (1993): 749-797.

Koo, Byoung-Mo, and Yeong-Jae Seok. "Regulation of glycogen concentration by the histidine-containing phosphocarrier protein HPr in *Escherichia coli*." *The journal of microbiology* 39.1 (2001): 24-30.

Kozlov, Guennadi, et al. "Structure of GlgS from *Escherichia coli* suggests a role in protein–protein interactions." *BMC biology* 2.1 (2004): 10.

Kubo, Akiko, et al. "Functions of heteromeric and homomeric isoamylase-type starch-debranching enzymes in developing maize endosperm." *Plant physiology* 153.3 (2010): 956-969.

Kuo, Jong-Tar, Yu-Jen Chang, and Ching-Ping Tseng. "Growth rate regulation of *lac* operon expression in *Escherichia coli* is cyclic AMP dependent." *FEBS letters* 553.3 (2003): 397-402.

Landini, Paolo, et al. "sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: role, regulation and mechanisms of promoter recognition." *Environmental microbiology reports* 6.1 (2014): 1-13.

Lane, Nick, John F. Allen, and William Martin. "How did LUCA make a living? Chemiosmosis in the origin of life." *BioEssays* 32.4 (2010): 271-280.

Lane, Nick. “Life ascending: The ten great inventions of evolution.” *Profile Books Ltd* (2009).

Lange, Roland, Daniela Fischer, and Regine Hengge-Aronis. "Identification of transcriptional start sites and the role of ppGpp in the expression of rpoS, the structural gene for the sigma S subunit of RNA polymerase in *Escherichia coli*." *Journal of bacteriology* 177.16 (1995): 4676-4680.

Lepek, Viviana C., et al. "Analysis of Mesorhizobium loti glycogen operon: effect of phosphoglucomutase (pgm) and glycogen synthase (glgA) null mutants on nodulation of *Lotus tenuis*." *Molecular plant-microbe interactions* 15.4 (2002): 368-375.

Lomako, Joseph, Wieslawa M. Lomako, and William J. Whelan. "Proglycogen: a low-molecular-weight form of muscle glycogen." *FEBS letters* 279.2 (1991): 223-228.

Lucchetti-Miganeh, Céline, et al. "The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts." *Microbiology* 154.1 (2008): 16-29.

Mabee, Warren E., et al. "Canadian biomass reserves for biorefining." *Applied biochemistry and biotechnology* 129.1-3 (2006): 22-40.

Madsen, N. B., and Carl F. Cori. "The binding of glycogen and phosphorylase."*Journal of biological chemistry* 233.6 (1958): 1251-1256.

Magnusson, Lisa U., Anne Farewell, and Thomas Nyström. "ppGpp: a global regulator in *Escherichia coli*." *Trends in microbiology* 13.5 (2005): 236-242.

Manonmani, H. K., and A. A. M. Kunhi. "Interference of thiol-compounds with dextrinizing activity assay of α-amylase by starch-iodine colour reaction: Modification of the method to eliminate this interference." *World journal of microbiology and biotechnology* 15.4 (1999): 485-487.

Marbach, Anja, and Katja Bettenbrock. "*lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA." *Journal of biotechnology* 157.1 (2012): 82-88.

Marroquí, Silvia, et al. "Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants." *Journal of bacteriology* 183.3 (2001): 854-864.

Martin, William. "Evolutionary origins of metabolic compartmentalization in eukaryotes." *Philosophical transactions of the royal society B: biological sciences* 365.1541 (2010): 847-855.

McGrance, Scott J., Hugh J. Cornell, and Colin J. Rix. "A simple and rapid colorimetric method for the determination of amylose in starch products."*Starch-Stärke* 50.4 (1998): 158-163.

Meadow, Norman D., and Saul Roseman. "Rate and equilibrium constants for phosphoryltransfer between active site histidines of *Escherichia coli* HPr and the signal transducing protein IIIGlc." *Journal of biological chemistry* 271.52 (1996): 33440-33445.

Mehdi, Rahimpour, et al. "GlgS, described previously as a glycogen synthesis control protein, negatively regulates motility and biofilm formation in *Escherichia coli*." *Biochemical journal* 452.3 (2013): 559-573.

Meléndez, Ruth, et al. "Physical constraints in the synthesis of glycogen that influence its structural homogeneity: a two-dimensional approach." *Biophysical journal* 75.1 (1998): 106-114.

Melendez-Hevia, Enrique, T. G. Waddell, and E. D. Shelton. "Optimization of molecular design in the evolution of metabolism: the glycogen molecule." *Biochemical journal* 295 (1993): 477-483.

Melis, Anastasios. "Solar energy conversion efficiencies in photosynthesis: minimizing the chlorophyll antennae to maximize efficiency." *Plant science* 177.4 (2009): 272-280.

Miller, J. H. (1972). *Experiments in molecular genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Montero, Manuel, et al. "*Escherichia coli* glycogen metabolism is controlled by the PhoP-PhoQ regulatory system at submillimolar environmental Mg^{2+} concentrations, and is highly interconnected with a wide variety of cellular processes." *Biochemical journal* 424 (2009): 129-141.

Montero, Manuel, et al. "*Escherichia coli* glycogen genes are organized in a single *glgBXCAP* transcriptional unit possessing an alternative suboperonic promoter within *glgC* that directs *glgAP* expression." *Biochemical journal* 433 (2011): 107-117.

Morán-Zorzano, María T., et al. "Occurrence of more than one important source of ADP glucose linked to glycogen biosynthesis in *Escherichia coli* and *Salmonella*." *FEBS letters* 581.23 (2007): 4423-4429.

Nakamura, Yasunori, et al. "Some cyanobacteria synthesize semi-amylopectin type α-polyglucans instead of glycogen." *Plant and cell physiology* 46.3 (2005): 539-545.

Nassoury, Nasha, and David Morse. "Protein targeting to the chloroplasts of photosynthetic eukaryotes: getting there is half the fun." *Biochimica et biophysica Acta (BBA)-molecular cell research* 1743.1 (2005): 5-19.

Pacala, Stephen, and Robert Socolow. "Stabilization wedges: solving the climate problem for the next 50 years with current technologies." *science* 305.5686 (2004): 968-972.

Park, Jong-Tae, et al. "Role of maltose enzymes in glycogen synthesis by *Escherichia coli*." *Journal of bacteriology* 193.10 (2011): 2517-2526.

Parks, John S., et al. "Regulation of galactokinase synthesis by cyclic adenosine 3', 5'-monophosphate in cell-free extracts of *Escherichia coli*." *Journal of Biological Chemistry* 246.8 (1971): 2419-2424.

Patron, Nicola J., and Ross F. Waller. "Transit peptide diversity and divergence: a global analysis of plastid targeting signals." *Bioessays* 29.10 (2007): 1048-1058.

Paul, Brian J., et al. "rRNA transcription in *Escherichia coli*." *Annual review of genetics* 38 (2004): 749-770.

Plancke, Charlotte, et al. "Pathway of cytosolic starch synthesis in the model glaucophyte *Cyanophora paradoxa*." *Eukaryotic cell* 7.2 (2008): 247-257.

Plumbridge, Jacqueline. "Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc." *Current opinion in microbiology* 5.2 (2002): 187-193.

Prágai, Zoltán, and Colin R. Harwood. "Regulatory interactions between the Pho and σB-dependent general stress regulons of *Bacillus subtilis*." *Microbiology*148.5 (2002): 1593-1602.

Rapoport, Tom A. "Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes." *Nature* 450.7170 (2007): 663-669.

Reinhart, Richard A. "Magnesium metabolism." *Archives of international medicine* 148.2415 (1988): 20.

Robbens, Steven, et al. "The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction." *Molecular biology and evolution* 24.4 (2007): 956-968.

Roderick, Steven L. "The *lac* operon galactoside acetyltransferase." *Comptes rendus biologies* 328.6 (2005): 568-575.

Rodríguez-Ezpeleta, Naiara, et al. "Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes." *Current biology* 15.14 (2005): 1325-1330.

Romeo, Tony, and Jack Preiss. "Genetic regulation of glycogen biosynthesis in *Escherichia coli*: in vitro effects of cyclic AMP and guanosine 5'-diphosphate 3'-diphosphate and analysis of in vivo transcripts." *Journal of bacteriology* 171.5 (1989): 2773-2782.

Romeo, Tony, Jill Black, and Jack Preiss. "Genetic regulation of glycogen biosynthesis in *Escherichia coli*: In vivo effects of the catabolite repression and stringent response systems in *glg* gene expression." *Current microbiology* 21.2 (1990): 131-137.

Saibene, Debora, et al. "Iodine-binding in granular starch: Different effects of moisture content for corn and potato starch." *Starch-stärke* 60.3-4 (2008): 165-173.

Sakulsingharoj, Chotipa, et al. "Engineering starch biosynthesis for increasing rice seed weight: the role of the cytoplasmic ADP-glucose pyrophosphorylase."*Plant science* 167.6 (2004): 1323-1333.

Santelia, Diana, and Samuel C. Zeeman. "Progress in Arabidopsis starch research and potential biotechnological applications." *Current opinion in biotechnology* 22.2 (2011): 271-280.

Seok, Yeong-Jae, et al. "Regulation of *E. coli* glycogen phosphorylase activity by HPr." *Journal of molecular microbiology and biotechnology* 3.3 (2001): 385-394.

Shearer, Jane, and Terry E. Graham. "New perspectives on the storage and organization of muscle glycogen." *Canadian journal of applied physiology* 27.2 (2002): 179-203.

Sheng, Fang, et al. "Oligosaccharide binding in *Escherichia coli* glycogen synthase." *Biochemistry* 48.42 (2009): 10089-10097.

Shi, Lan-Xin, and Steven M. Theg. "The chloroplast protein import system: from algae to trees." *Biochimica et biophysica acta (BBA)-molecular cell research* 1833.2 (2013): 314-331.

Sims, Ralph EH, et al. "An overview of second generation biofuel technologies."*Bioresource technology* 101.6 (2010): 1570-1580.

Smith, Alison M. "Making starch." *Current opinion in plant biology* 2.3 (1999): 223-229.

Smith, Alison M. "The biosynthesis of starch granules." *Biomacromolecules* 2.2 (2001): 335-341.

Smith, Alison M. "Prospects for increasing starch and sucrose yields for bioethanol production." *The plant journal* 54.4 (2008): 546-558.

Smith, Alison M., and Samuel C. Zeeman. "Quantification of starch in plant tissues." *Nature protocols* 1.3 (2006): 1342-1345.

Song, Hyung-Nam, et al. "Structural rationale for the short branched substrate specificity of the glycogen debranching enzyme GlgX." *Proteins: structure, function, and bioinformatics* 78.8 (2010): 1847-1855.

Srivatsan, Anjana, and Jue D. Wang. "Control of bacterial transcription, translation and replication by (p) ppGpp." *Current opinion in microbiology* 11.2 (2008): 100-105.

Stryer, Lubert. “Biochemistry, fourth edition.” *W. H. Freeman and company, New York* (1995).

Sullivan, Mitchell A., et al. "Nature of α and β particles in glycogen using molecular size distributions." *Biomacromolecules* 11.4 (2010): 1094-1100.

Sullivan, Mitchell A., et al. "Molecular insights into glycogen α-particle formation." *Biomacromolecules* 13.11 (2012): 3805-3813.

Sundberg, Maria, et al. "The heteromultimeric debranching enzyme involved in starch synthesis in Arabidopsis requires both Isoamylase1 and Isoamylase2 subunits for complex stability and activity." *PloS one* 8.9 (2013): e75223.

Suzuki, Eiji, et al. "Role of the GlgX protein in glycogen metabolism of the cyanobacterium, *Synechococcus elongatus* PCC 7942." *Biochimica et biophysica acta (BBA)-general subjects* 1770.5 (2007): 763-773.

Sweetlove, L., M. Burrell, and T. Ap Rees. "Characterization of transgenic potato (*Solanum tuberosum*) tubers with increased ADP glucose pyrophosphorylase." *Biochemical journal* 320 (1996): 487-492.

Takata, Hiroki, et al. "Purification and characterization of α-glucan phosphorylase from *Bacillus stearothermophilus*." *Journal of fermentation and bioengineering* 85.2 (1998): 156-161.

Tetlow, Ian J. "Understanding storage starch biosynthesis in plants: a means to quality improvement." *Botany* 84.8 (2006): 1167-1185.

Thiemann, Volker, et al. "Heterologous expression and characterization of a novel branching enzyme from the thermoalkaliphilic anaerobic bacterium *Anaerobranca gottschalkii*." *Applied microbiology and biotechnology* 72.1 (2006): 60-71.

Tian, Zhongyuan, et al. "Identification of key regulators in glycogen utilization in *E. coli* based on the simulations from a hybrid functional Petri net model." *BMC systems biology* 7.6 (2013): 1-10.

Traxler, Matthew F., et al. "The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*." *Molecular microbiology* 68.5 (2008): 1128-1148.

Tutar, Yusuf. "Syn, anti, and finally both conformations of cyclic AMP are involved in the CRP-dependent transcription initiation mechanism in *E. coli* lac operon." *Cell biochemistry and function* 26.4 (2008): 399-405.

Ugalde, Juan E., Armando J. Parodi, and Rodolfo A. Ugalde. "De novo synthesis of bacterial glycogen: *Agrobacterium tumefaciens* glycogen synthase is involved in glucan initiation and elongation." *Proceedings of the national academy of sciences* 100.19 (2003): 10659-10663.

Véscovi, Eleonora García, Fernando C. Soncini, and Eduardo A. Groisman. "Mg^{2+} as an extracellular signal: Environmental regulation of Salmonella virulence." *Cell* 84.1 (1996): 165-174.

Vrinten, Patricia L., and Toshiki Nakamura. "Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues." *Plant physiology* 122.1 (2000): 255-264.

Wang, Liang, and Michael J. Wise. "Glycogen with short average chain length enhances bacterial durability." *Naturwissenschaften* 98.9 (2011): 719-729.

Wendrich, Thomas M., et al. "Dissection of the mechanism for the stringent factor RelA." *Molecular cell* 10.4 (2002): 779-788.

Wilson, C. J., et al. "The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding." *Cellular and molecular life sciences* 64.1 (2007): 3-16.

Wilson, Wayne A., et al. "Regulation of glycogen metabolism in yeast and bacteria." *FEMS microbiology reviews* 34.6 (2010): 952-985.

Yamamotoya, Tomoaki, et al. "Glycogen is the primary source of glucose during the lag phase of *E. coli* proliferation." *Biochimica et biophysica acta (BBA)-proteins and proteomics* 1824.12 (2012): 1442-1448.

Yang, Xiaoming, and Edward E. Ishiguro. "Involvement of the N terminus of ribosomal protein L11 in regulation of the RelA protein of *Escherichia coli*."*Journal of bacteriology* 183.22 (2001): 6532-6537.

Yang, Honghui, Mu Ya Liu, and Tony Romeo. "Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the CsrA gene product." *Journal of bacteriology* 178.4 (1996): 1012-1017.

Yep, Alejandra, et al. "Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli*." *Journal of biological chemistry* 279.9 (2004): 8359-8367.

Yu, Xiaochun, Carl Houtman, and Rajai H. Atalla. "The complex of amylose and iodine." *Carbohydrate research* 292 (1996): 129-141.

Zeeman, Samuel C., Jens Kossmann, and Alison M. Smith. "Starch: its metabolism, evolution, and biotechnological modification in plants." *Annual review of plant biology* 61 (2010): 209-234.

Zmasek, Christian M., and Adam Godzik. "Phylogenomic analysis of glycogen branching and debranching enzymatic duo." *BMC evolutionary biology* 14.1 (2014): 183.