Analysis of a putative *crtW* gene of *Myxococcus xanthus*

James Needham

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick

Department of Biological Sciences

September 2009

Contents

List of Figures

List of Tables

Acknowledgements

As with most large pieces of work this project would not have been possible without the help and guidance of a vast host of people. Firstly special thanks have to be extended to the two Daves; my supervisor Professor Dave Hodgson without whom who knows what stage this investigation would still be at, and Dave Whitworth, for an excessively large amount of laboratory and project advice. The two of them really helped to shape the investigation and make the write-up you see before you now.

Everyone in the Micro II lab deserves a mention, but there is just not enough room to mention you all so apologies to the people who aren't here in name. I particularly need to mention Sam Bryan, Sophie Mazzard, Andy Millard and Ed Spence who again offered much advice throughout my three years in the lab, especially at the start when help was needed to get things going. Then of course there were the many distractions – Jackie (top of the list), Nikki, John, Cecile, Ludwig, Yi, Pip (!) – thanks for breaking the days up.

I also need to mention the Hodgson/Whitworth research group who were a crazy bunch of people, but all filled with useful information: Adam '*Streptococcus*' Giles, Colin '*Listeria*' Bateman, Helena '*Myxo*' Wright, Peter 'The Talker' Cock, Antony 'The Modeller' Holmes, and George 'No speech marks could do you justice' Savva. You really made the whole process…interesting. And in George's case downright dangerous.

Once again thanks to all those I haven't mentioned who played a role in three years worth of work like Fran, Martin. Vicky, Prof. Dave Scanlan, Prof. Colin Murrell, the team at Stores, all who provided me with cell samples, the team at the sequencing suite etc. but overall my final thanks has to go to my family – Mum, Dad, Caz and Inky. You all made sure that I was up every day ready to go to Uni, looked half-presentable, was well fed, had all the equipment to do this write-up, enabled me to get to the end of Thistlegreen Road and also occasionally bit my finger before breakfast.

This is for all of you. Thank you.

Declaration

All the results presented in this thesis are my own work, unless otherwise stated. All sources of information are acknowledged by means of a reference. None of the work contained within this thesis has been used for any previous application for a degree.

Signed…….

Abstract

Carotenoids are produced by all photosynthetic organisms and a large number of bacteria and fungi. They are responsible for a lot of pigmentation in nature, as well as often playing an essential role in the provision of light protection to cells and as precursors of vitamin A in higher organisms. Myxobacton is the primary carotenoid ester generated in the photoresistant bacterium *Myxococcus xanthus*. It is created through a complex light-regulated gene expression cascade and acts to protect the bacteria from blue light and the resultant generation of damaging singlet oxygen species in the presence of porphyrins. The final stage in its production is a ketonisation, and the enzyme responsible for this stage was unknown in *M. xanthus*. We propose a possible location for the gene encoding such a ketolase, *crtW*. The gene is found located within a four-gene operon separate from the other known carotenoid biosynthetic genes, and appears to have two alternative promoter regions. The additional genes in the operon were found to encode a putative MutT/Nudix family hydrolase and a periplasmic, molybdopterin and haeme-dependent oxidoreductase, YedYZ. It is also shown that *crtW* transcription is independent of cell exposure to blue light and that the product is an inner membrane, integral membrane protein probable ketolase. The evolutionary origins of *crtW* are considered in conjunction with a number of other carotenoid biosynthetic enzymes, suggesting that the gene was one of the last to be acquired by *M. xanthus*.

Abbreviations

A-system	Adventurous system
a.a.	Amino acid
ABC	ATP-binding cassette
amp	Ampicillin
APS	Ammonium persulphate
AT	Adenine/Thymine
BLAST	Blast local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
cam	Chloramphenicol
cDNA	Complementary deoxyribonucleic acid
conc.	concentration
DCY	Double casitone yeast
DCYA	Double casitone yeast agar
DCY SA	Double casitone yeast soft agar
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
DTT	Dithiothreitol
e-value	expected value
ECF	Extra cytoplasmic function
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory

g	G-force
GC	Guanine/Cytosine
His	Histidine
HPLC	High-performance liquid chromatography
inc	Incompatibility
IPP	Isopentyl phosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kan	Kanamycin
kb	kilobase
KO	Knockout
kV	kilo-volt
L/l	Litres/litres
LB	Luria-Bertani
LBA	Luria-Bertani agar
LC	LB-$CaCl_2$
LGC	LB-Glucose-$CaCl_2$
MCS	Multiple cloning site
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
NTC	Negative template control
OD_{600}	Optical density at wavelength of 600nm
ONPG	Ortho-Nitrophenyl-β-galactoside
ORF	Open reading frame
P-value	Probability value
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

pdp	Putative dipeptidyl peptidase
POP7	Processing of precursor 7
PPIX	Protoporphyrin IX
QRT-PCR	Quantitative real time – polymerase chain reaction
RACE	Random amplification of cDNA ends
RNA	Ribonucleic acid
RT	Reverse transcriptase
S-system	Social system
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sp.	Species
SOE	Splicing by overlap extension
STRING	Search tool for the retrieval of interacting genes/proteins
TBE	Tris-borate-EDTA
TCS	Two component system
TdT	Terminal deoxynucleotidyl transferase
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
tet	Tetracycline
TM	Tris-Magnesium sulphate
U	Units

Chapter 1

Introduction

1.1. The myxobacteria

The myxobacteria are a family of aerobic soil-dwelling, gliding bacteria; forming part of the phylum proteobacteria, class *Deltaproteobacterium*. All phylum members are gram negative; so-called as they do not stain a blue or purple colour during the process of Gram staining (Beveridge, 2001). They instead turn pink or red due to the presence of an additional outer membrane which results in the cells taking up a counterstain, safranin. Individuals are most commonly rod shaped, approximately 4-12μM in length (Reichenbach, 2001), and have been located on each of the major continents of the world, including Antarctica (Reichenbach, 1999), in addition to a number of water samples, both fresh and saline (Dawid, 2000). When studied within individual soil samples, *M. xanthus* populations also display much genetic diversity (Wu *et al*., 2005; Vos and Velicer, 2006). They are a highly predatory organism, feeding on a range of other bacterial species, and also absorbing nutrients directly from the surrounding environment. The feeding mechanism is often observed as a rippling mass of individuals moving towards and then over a food source (Berleman *et al.*, 2006). This is likened to the recorded behaviour of a hunting 'wolf pack' and has been shown to be affected by a number of ecological variables (Hillesland *et al*., 2001). One of the primary distinguishing traits of the family is that when placed under extreme environmental pressure e.g. a lack of adequate nutrition or exposure to extreme temperatures; they form durable myxospores (Wireman and Dworkin, 1975).

Within the myxobacteria there are a host of different genera which have been studied in greater detail including *Chondromyces*, *Stigmatella* and *Sorangium*. Extensive study of the latter has resulted in the development of a possible treatment for human metastatic breast cancer (Lee *et al.*, 2008). A further genus is *Myxococcus*, which incorporates the species *M. fulvus*, *M. stipitatus* and *M. virescens*. Although all species are similar in their behaviour, the most predominantly studied species of the group is *Myxococcus xanthus*. Its ability to form simple fruiting bodies, complex social interactions and the ease at which it can be cultivated are among the primary reasons why it is currently regarded as the model organism of the myxobacteria (Dworkin, 1996).

1.2. Medical relevance of *Myxococcus* species

As has been mentioned with regard to *So. cellulosum*, there is currently an increased interest in myxobacteria in the field of medicine, principally *M. xanthus*, which is known to produce a number of novel secondary bioactive metabolites. Many of these are involved in aiding *M. xanthus* prey cell lysis. The possible use of the metabolites particularly as antibiotics in order to treat a range of both animal and human conditions is an area of much study. To highlight this potential, a recent study recorded that one of the enzymes encoded by *M. xanthus* could be utilised in the treatment of the human disease celiac sprue (Gass *et al.*, 2005). *M. xanthus* is also capable of producing an antibiotic known as TA, which is bactericidal to other growing bacterial cells. It acts on the bacterial cell wall, preventing efficient wall formation (Zafriri *et al.*, 1981). The action is very similar to that of penicillin, although it functions at a different stage of cell wall production. It is believed that this antibiotic activity allows *M. xanthus* to digest bacterial prey more easily. Other

key antibiotics identified in myxobacteria include the myxalamids (Gerth *et al.*, 1983). These have been shown to inhibit the respiratory electron transport chains in eukaryotes. In total thirteen such antibiotics have now been identified in *M. xanthus* alone (Bode *et al*., 2007).

In addition a family of compounds known as myxovirescins have been isolated from the closely related *Myxococcus virescens* strain Mx v48. These are a family of antibiotics that also prevent bacterial cell wall formation, particularly within other gram negative bacterial species (Gerth *et al.*, 1982). Another specific example of a myxobacterial compound is that of myxochelin A, isolated from *Angiococcus disciformis* An d30. It is an iron chelating compound that was found to play a role in facultative digestion when mixed with a selection of other bacteria (Kunze *et al.*, 1989).

As increased emphasis is placed on the study of *M. xanthus* to aid the advancement of medical research, a greater number of novel metabolites are being identified. In an attempt to establish a measure of the number of metabolites produced by the species a recent study obtained a range of *M. xanthus* strains and compared their metabolite profiles (Krug *et al.*, 2008). It was discovered that despite the samples coming from all around the world there was a high level of convergence between all of the studied strains. 37 novel metabolites were identified, with a number appearing to be similar to the previously identified myxalamids. These are currently being studied in more detail in conjunction with novel secondary metabolite pathways identified in the *M. xanthus* DK1622 genome sequence in an attempt to establish their function. This number also acts to highlight the increasing potential medical benefits *M. xanthus* could offer.

1.3. The novel aspects of *M. xanthus* biology

M. xanthus has been studied intensively for over sixty years (Beebe, 1941), with increasing study following the complete sequencing of the bacterial genome (Goldman *et al.*, 2006). However, the majority of studies continue to focus on three specific areas: development, gliding motility and carotenogenesis.

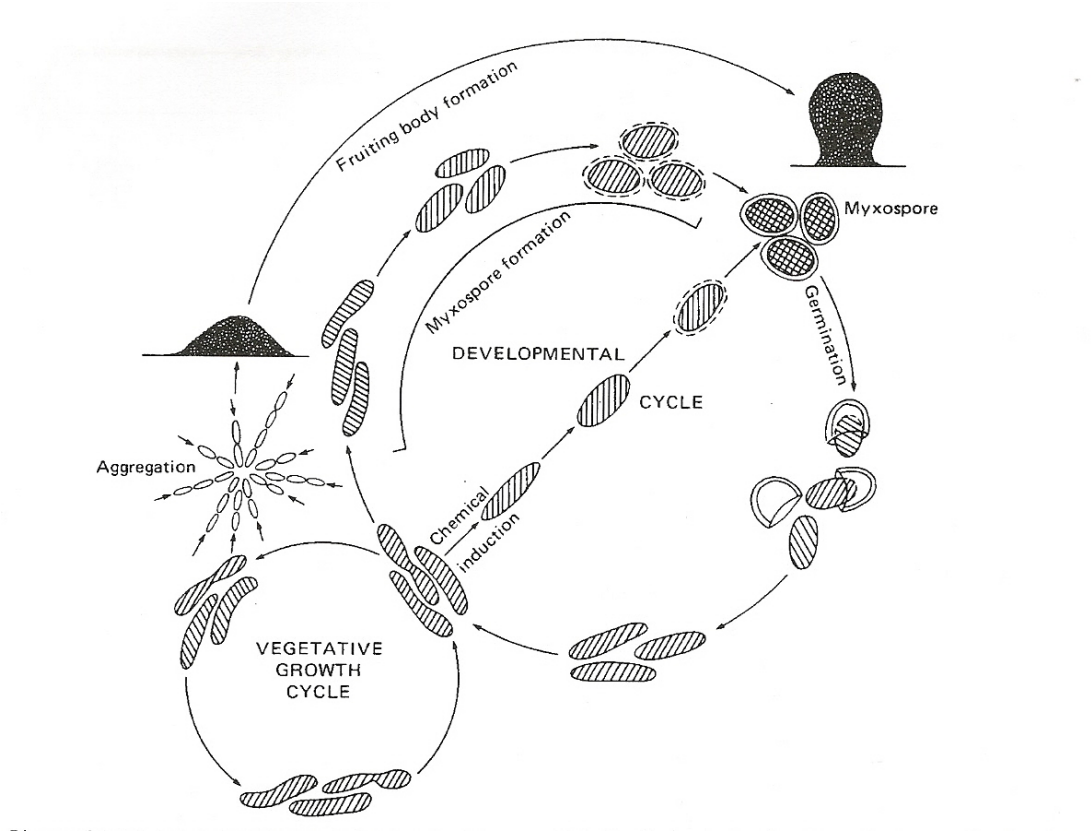


Figure 1.1 The life-cycle of *M. xanthus.* The diagram is not to scale, but the myxospore fruiting body produced is a few hundredths of a millimetre in diameter. This image is reproduced with the kind permission of the original author (Dworkin, 1986).

Development is the system of sporulation, through which relatively large fungal-like fruiting bodies are formed, approximately 100-130μM in size under starvation conditions (Figure 1.1). It is a novel bacterial mechanism, initiated through intercellular signalling (Kuspa *et al.*, 1986) which results in a small number of

individuals clustering together forming a mound of cells. Within the mound individuals then begin to adopt differing roles giving rise to the formation of a durable myxospore (Wireman and Dworkin, 1975). On average each fruiting body consists of a set of approximately 10^6-10^7 *M. xanthus* cells forming the central fruiting body, surrounded by a further protective outer casing layer of slime. Upon forming part of the outer layers, the cells essentially sacrifice themselves in order to ensure the survival of the central cell core. The coccoid myxospores created via this process, can survive in a harsh environment until the growth conditions are more suitable for the bacteria (Sudo and Dworkin, 1968). The specific process of spore creation, along with both how and why the cells adopt different roles, is an area still being studied in much detail (Tzeng & Singer, 2005; Jelsbak *et al.*; 2005, Goldman *et al.*, 2007; Ossa *et al*, 2007).

The mechanisms of *M. xanthus* motility have constantly been a source of great debate and one of increased study at the moment (Starruß *et al*., 2007). The current understanding is that two separate systems of bacterial movement exist – adventurous (A) and social (S). The S-system is adopted during the movement of large groups of individuals and relies upon the activity of type IV pili. These are projected from the leading end of the cells and then retracted; resulting in a pulling action that effectively propels the cell mass forward. By contrast the A-system refers primarily to individual cell movement when they are not found clustered together. In this instance the cells are believed to move through a combination of polysaccharide slime release from the cell extremities and a form of cell adhesion in direct relation to this. The precise motor mechanism of A-motility still remains unknown, although recent proposals suggest that movement is achieved via specific focal cell adhesions and a corkscrew-like movement of the cell body (Mignot *et al*., 2007; Mignot, 2007).

This is in contrast to the previous widely held belief that the cell was propelled along solely through the action of slime release (Wolgemuth *et al.*, 2002).

The final major aspect of *Myxococcus xanthus* behaviour that has been extensively studied is the complex reaction to a light stimulus. This was a study initially prompted by the observed vivid bright orange and yellow *M. xanthus* cell colouration. Some of the very first experiments carried out on the bacterium determined that it was extremely sensitive to blue light, which is approximately 430nM in wavelength (Burchard & Dworkin, 1966a & 1966b). It was established that upon exposure to blue light reactive oxygen species were generated within the cell threatening cell survival (Hodgson & Berry, 1998).

Figure 1.2 A range of the *M. xanthus* carotenoid colours. From top left moving clockwise the strains are: DK101, UWM303, DK718 and DK717.

1.4. Singlet oxygen generation

Singlet oxygen ($^1O_2$) generation is potentially catastrophic to the cell, resulting in protein and membrane damage and if untreated, ultimately cell death. The greatest risk is primarily posed to the cell membrane, as it is here that the damaging species is

initially generated. To protect itself from this highly reactive species, *M. xanthus* has developed a direct response mechanism known as light-induced carotenogenesis (Burchard & Hendricks, 1969). Within the cell a porphyrin is located at the membrane, protoporphyrin IX (PPIX). Porphyrins are molecules which act as photosensitisers and are central elements in a number of common biological compounds, including chlorophyll (Erkoç and Erkoç, 2002) and haemoglobin (Williamson, 1993). Porphyrins are present in the majority of bacteria where they form the base structure of cytochrome haeme (Hodgson and Berry, 1998). They occur naturally as cyclic tetrapyrroles with the ring structure present giving rise to the photosensitising properties (Figure 1.3). These properties are the ones that provide porphyrins with their unique function, being both a benefit when found in chlorophyll in plants and cytochromes in all respiring organisms, but also potentially lethal when occurring in photosensitive bacteria.

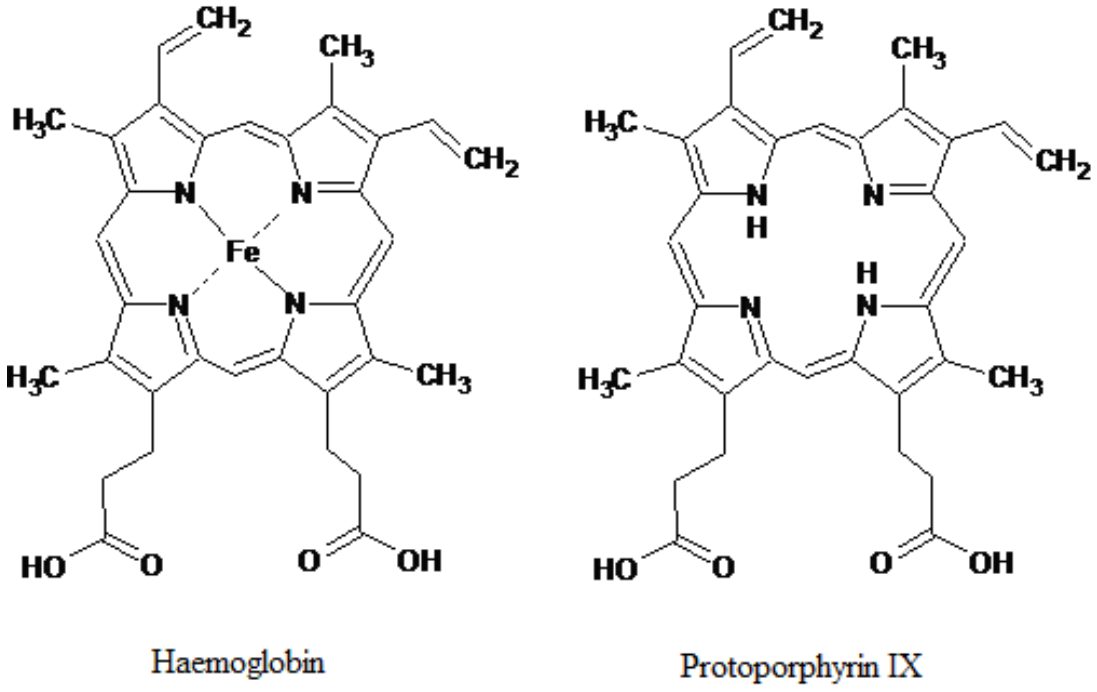


Figure 1.3 The molecular structure of haemoglobin and PPIX. The primary difference is the presence of a central iron element in haemoglobin.

Upon exposure to light, PPIX can absorb a photon of blue wavelength. This results in its conversion to ^{3}PPIX, the triplet state molecule, itself capable of extensive cell damage. More often this in turn reacts with ground state triplet oxygen ($^3O_2$) generating the singlet oxygen species. When created, singlet oxygen can release heat energy, reverting back to its ground state triplet oxygen. However, singlet oxygen can persist in the environment for upwards of 100µS, increasing the potential for it to interact instead directly with protein and lipids in the cell causing oxidation and inactivation of these compounds. A more in depth discussion on the generation and effects of $^1O_2$ can be found in Hodgson & Berry (1998). In order to counteract reactive oxygen generation, *M. xanthus* has evolved light-induced carotenogenesis, ultimately giving rise to the production of photoprotective carotenoids.

1.5. Carotenoids

Carotenoids are compounds produced in all species of plants and a vast array of bacterial species, as well as most fungi (Johnson, 2003). The primary role of carotenoids is the protection against photoactivated oxygen species as discussed above. In plants, algae and cyanobacteria, they are essential, as oxygen-generating photosynthesis brings together light, photosensitisers (chlorophyll, porphyrin and cytochrome haeme) and oxygen. Carotenoids are also responsible for much of the pigmentation observed in nature, ranging from the soft oranges and yellows of autumn leaves, to the more intense pinks of both salmon and flamingos (Armstrong, 1997). The role of carotenoids is not solely as a visual stimulant, although for many of the flowering plant species this is an important function. In a number of birds and fish species it has been reported that carotenoid levels may actually function as an

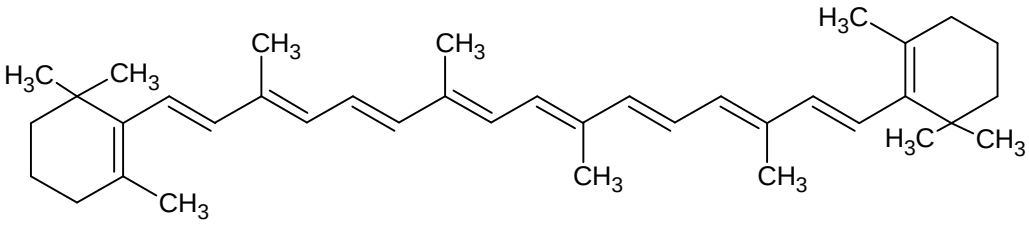
indication of the relative fitness of a given individual (Lozano, 2001; Sandre *et al.*, 2007; Maan *et al.*, 2008). In plants and many photosynthetic bacteria, carotenoids are accessory light harvesting pigments, in conjunction with chlorophyll (Anthony *et al.*, 2005).

Many of the higher organisms, including all mammals, are unable to synthesise carotenoids, so must take them in as an essential dietary requirement. This is primarily due to the fact that they act as precursors of vitamin A (Datta *et al.*, 2007), of which the most commonly known precursor is beta-carotene. They also exhibit very strong antioxidant properties, utilised by a range of species in order to protect both cells and tissue from damage (Valko *et al.*, 2006). It is the latter function that is of particular benefit to the majority of bacterial species, including *M. xanthus*. It is also a factor exploited by the many global drug companies that produce carotenoid based products, advertising them for improving consumer health. Despite the range of products that are currently available on the market, and their nutritional potential (Krinsky and Johnson, 2005), the benefits of increased carotenoid consumption are unconfirmed, and in some instances could actually prove detrimental to human health (Tanvetyanon and Bepler, 2008).

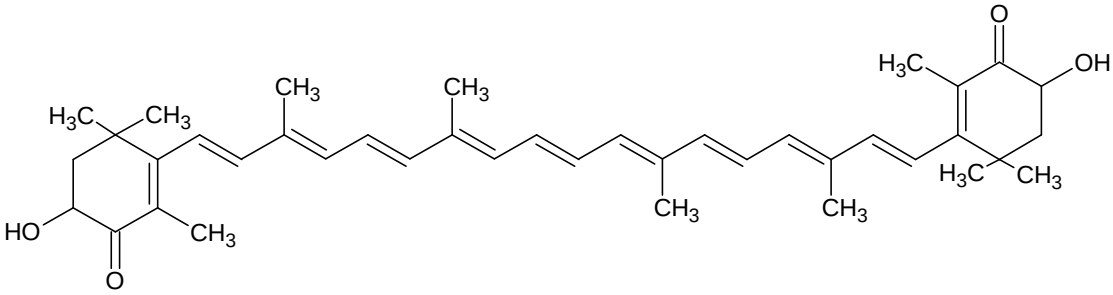
Carotenoids naturally occur in two primary forms, xanthophylls and carotenes. Both are constructed of similar long carbon chains, up to 30-40 carbon atoms in length, joined by a mixture of both single and double covalent bonds. The primary difference between the two is that the carotene chains are made solely of carbon and hydrogen, whereas xanthophylls incorporate oxygen in addition e.g. as hydroxyl groups (Goodwin, 1986), (Figure 1.4) and can also form compounds with more complex elements such as metals.

Most photoresistant bacteria are able to utilise the carotenoids to quench generated damaging reactive oxygen species (Glaeser and Klug, 2005). The presence

of carbon double bonds throughout their structure enable the 'removal' of excitation energy from the primary reactive species; ^{3}PPIX and $^{1}O_2$ in the case of *M. xanthus*. The energy causes a reversible cis or trans isomerisation of the double bonds in the carotenoid, and the energy is then in turn released as heat (Truscott, 1990).



β-carotene



Astaxanthin

Figure 1.4 A carotene (β-carotene) and a xanthophyll (astaxanthin).

The action not only quenches the reactive species, but also ensures that the carotenoids are self-quenching and are able to interact again with reactive species without themselves being damaged or destroyed in the process. Studies have recorded that the effectiveness of carotenoid quenching is directly proportional to the amount of alternating double bonds present in the molecule, e.g. the more double bonds, the more effective the carotenoid (Conn *et al.*, 1991). Besides *M. xanthus* a

host of other bacterial species utilise carotenoids for protection in this way including those of the genera *Bradyrhizobium* (Hannibal *et al.*, 2000), *Erwinia* (Misawa *et al.*, 1990) and *Streptomyces* (Takano *et al.*, 2005). All adopt similar carotenogenesis pathways, although the final carotenoids formed and the levels at which they are produced do vary, as does the control of carotenoid production.

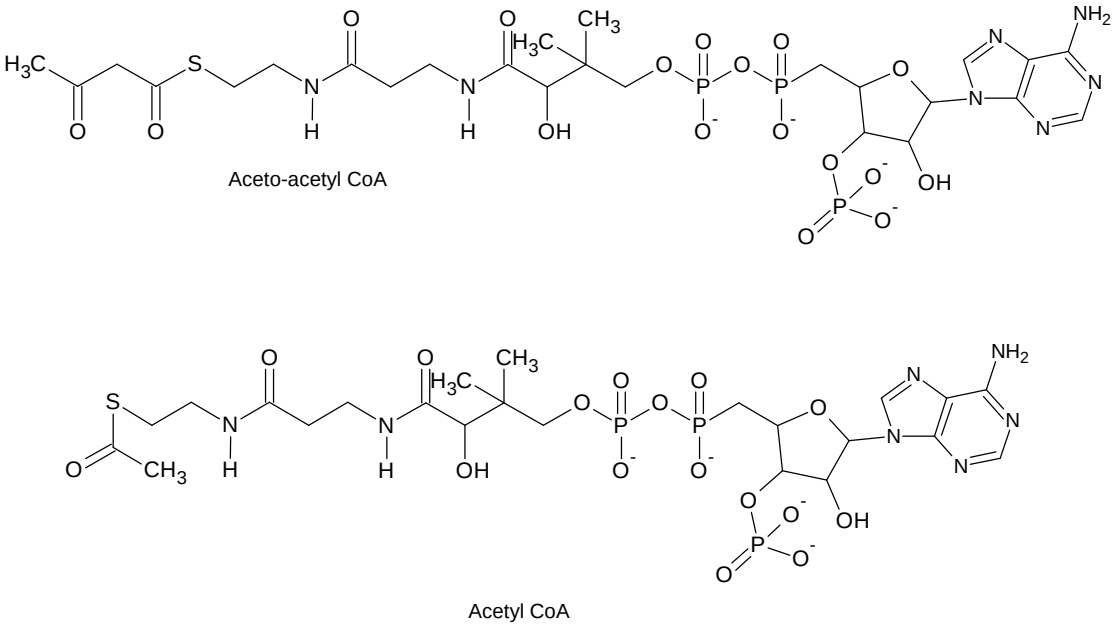


Figure 1.5 The two original molecules used in the generation of IPP (see Figure 1.6).

All bacterial carotenoids are formed by the initial combination of a number of isoprenoid subunits (Umeno *et al.*, 2005). In the myxobacteria these originate as the smaller molecules known as isopentyl phosphates (IPPs), which are generated through the mevalonate pathway (Wilding *et al.*, 2000). The pathway (also known as the acetyl CoA pathway) is integral to the production of cholesterol and lipids within the cell. The process of IPP generation is a six step procedure, originating with the two initial starting molecules of acetyl CoA and acetoacetyl CoA (Figure 1.5), each subsequent step occurring in the cell cytosol. Upon the generation of a reduced molecule of mevalonate, three further ATP dependent steps then occur ultimately

giving rise to the IPP product (Figure 1.6). A series of additional enzyme reactions act to bind a number of the IPP subunits together, creating larger carotenoid intermediates e.g. geranylgeranyl pyrophosphate combines with an identical second molecule through condensation to generate phytoene. The accompanying diagrams (Figures 1.7a and 1.7b) detail the various enzymatic stages which give rise to the final two primary carotenoid esters of *M. xanthus*; myxobacton and myxobactin. These are created in a 3:1 ratio respectively (Reichenbach and Kleinig, 1971) and make up the majority of the carotenoids identified in the cell. A second pair of carotenoids, believed to constitute less than 10% of the total generated carotenoids, are created via a pathway that branches off from the one already detailed. The pair, torulene and 4-keto-torulene, are present at such low levels that their generation does not contribute significantly to *M. xanthus* cell protection.

Many of the enzymes believed to be responsible for the creation of the carotenoid intermediates within the system still remain unstudied. In the case of *M. xanthus* this is in stark contrast to the greater detail already obtained for other similar systems (Fraser *et al.*, 1996; Takano *et al.*, 2005) and also despite the fact that the majority of the products determined in the pathway have been identified for a substantial period of time. Many of the enzymes and their associated genes have not yet been located in *M. xanthus*, existing only hypothetically, and identified only through a combination of comparing other known synthesis pathways and online genome searches (Figure 1.8).

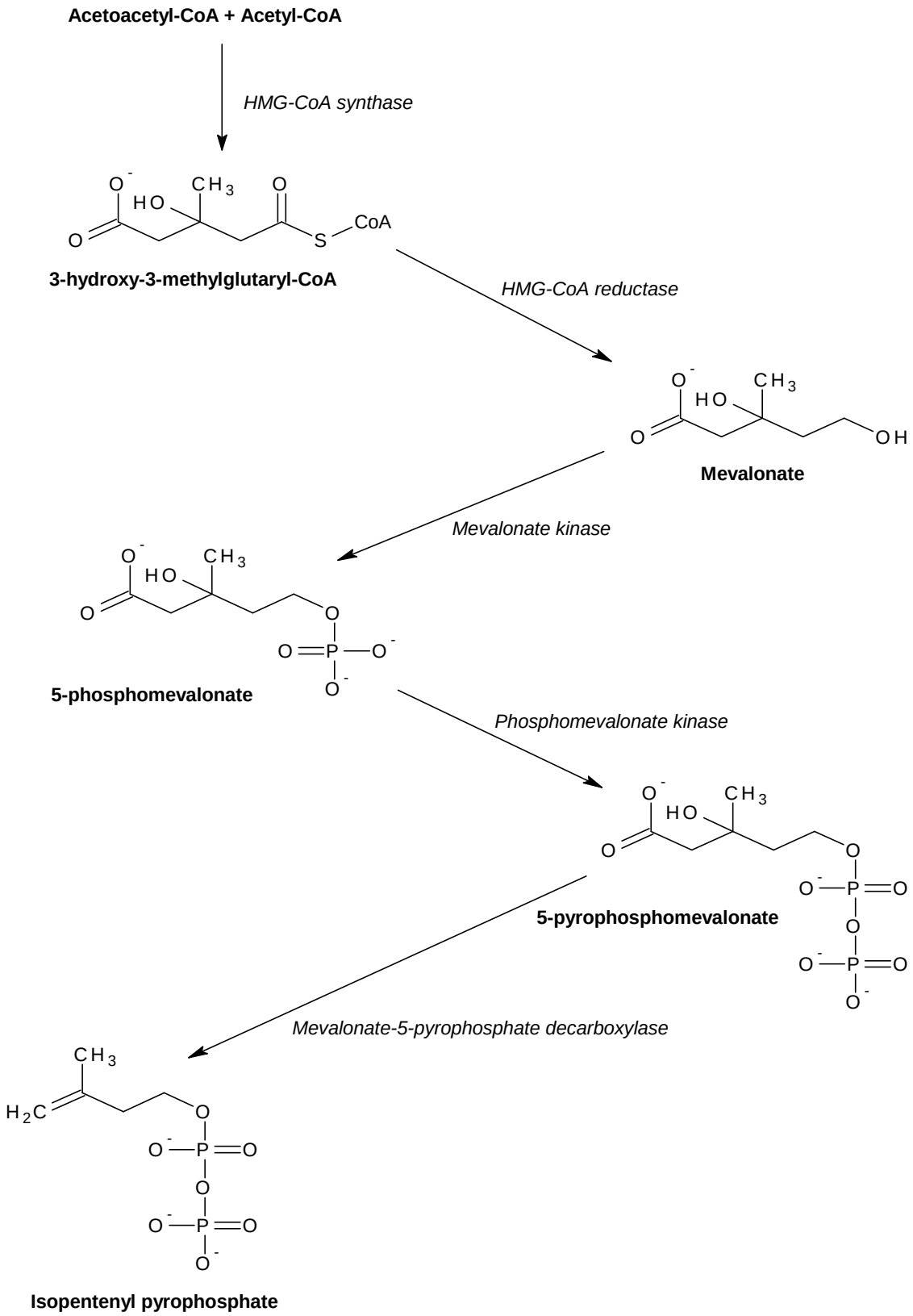


Figure 1.6 The mevalonate pathway within *M. xanthus* which generates IPP, with each enzyme activity labelled.

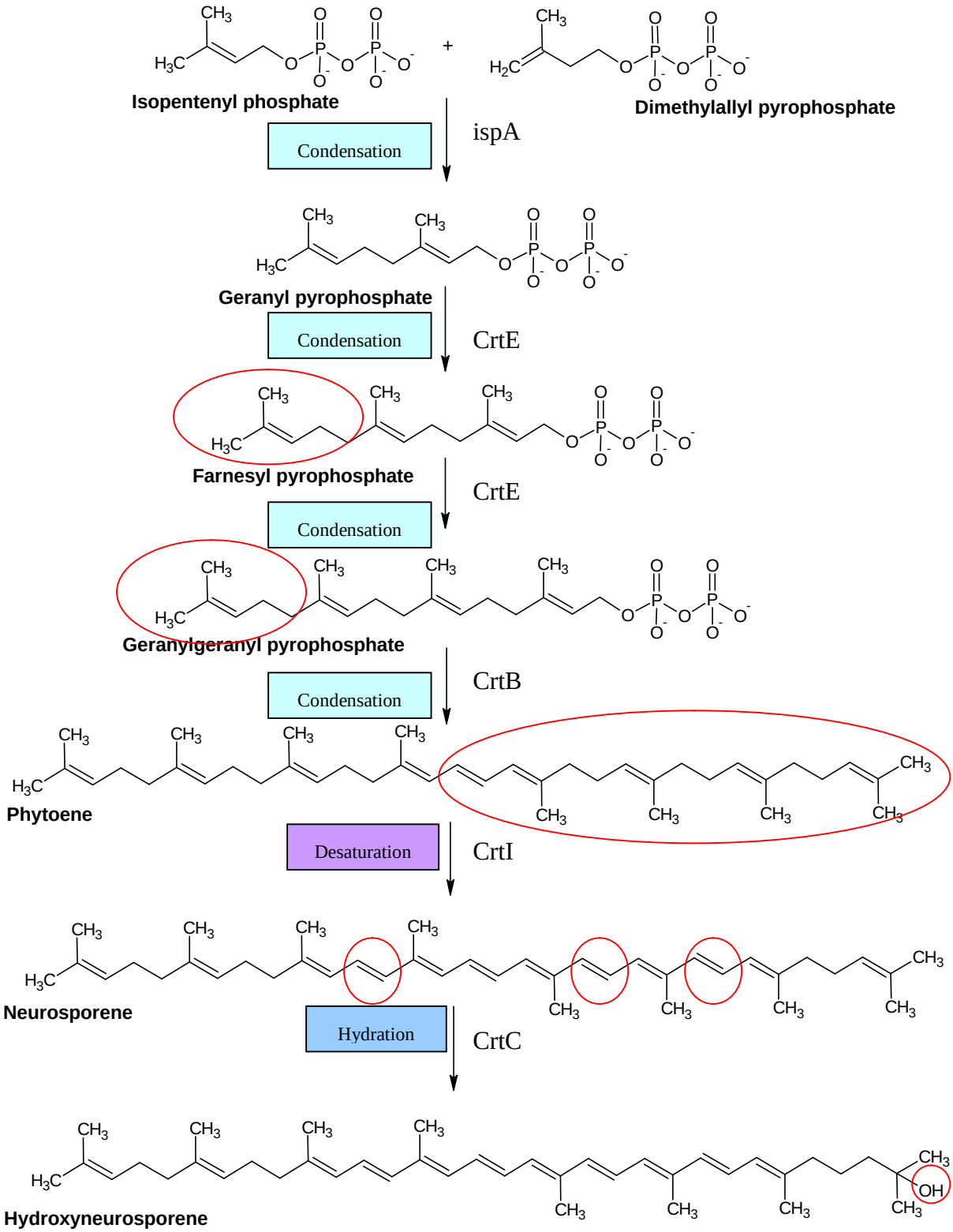


Figure 1.7a The first part of the carotenoid pathway of *M. xanthus*, from the original IPP subunits up to the generation of hydroxyneurosporene. Each arrow is labelled with the enzyme that is believed to be responsible for that stage of the process, and the basic enzymatic process that occurs. The structural change at each stage is circled.

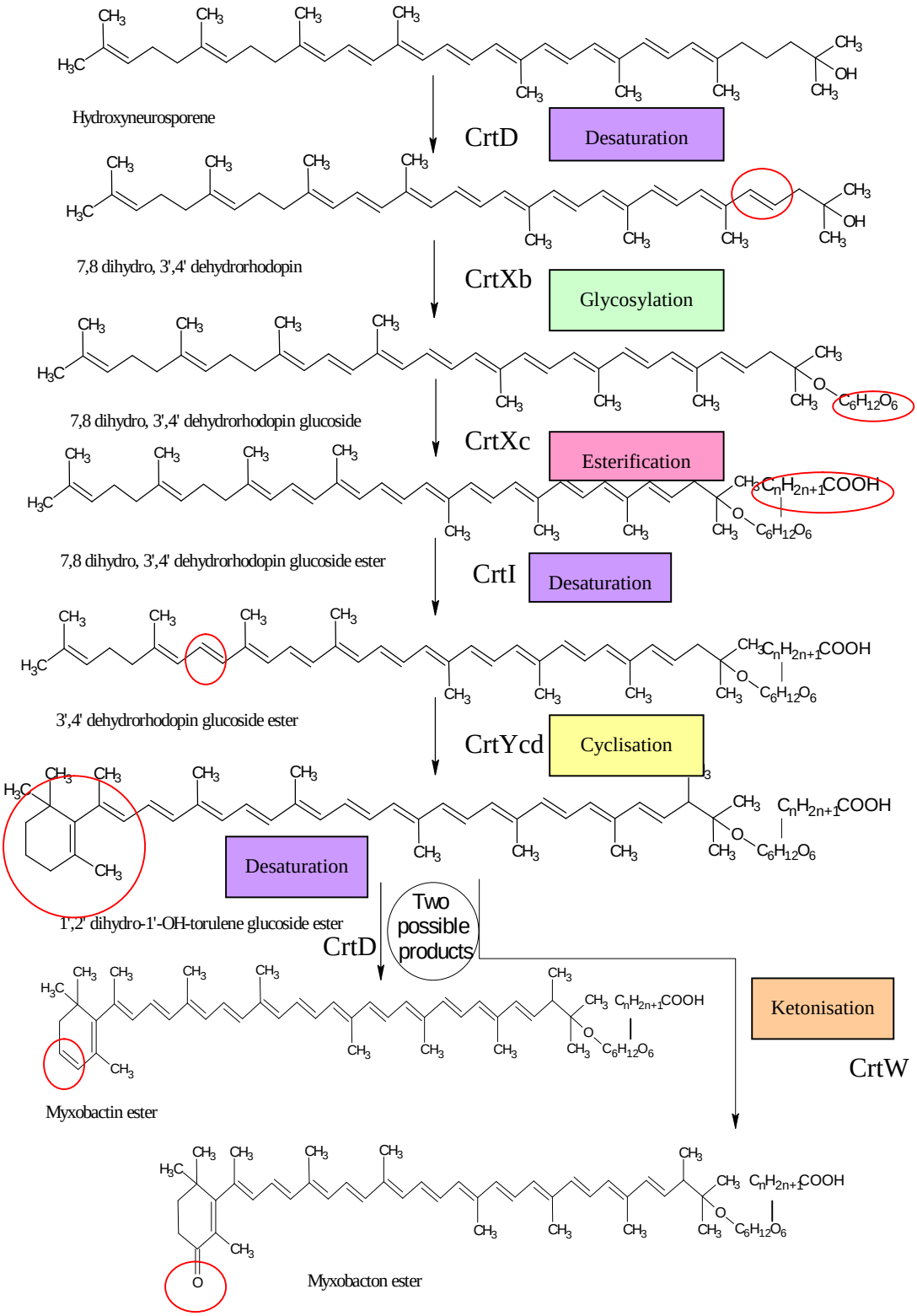
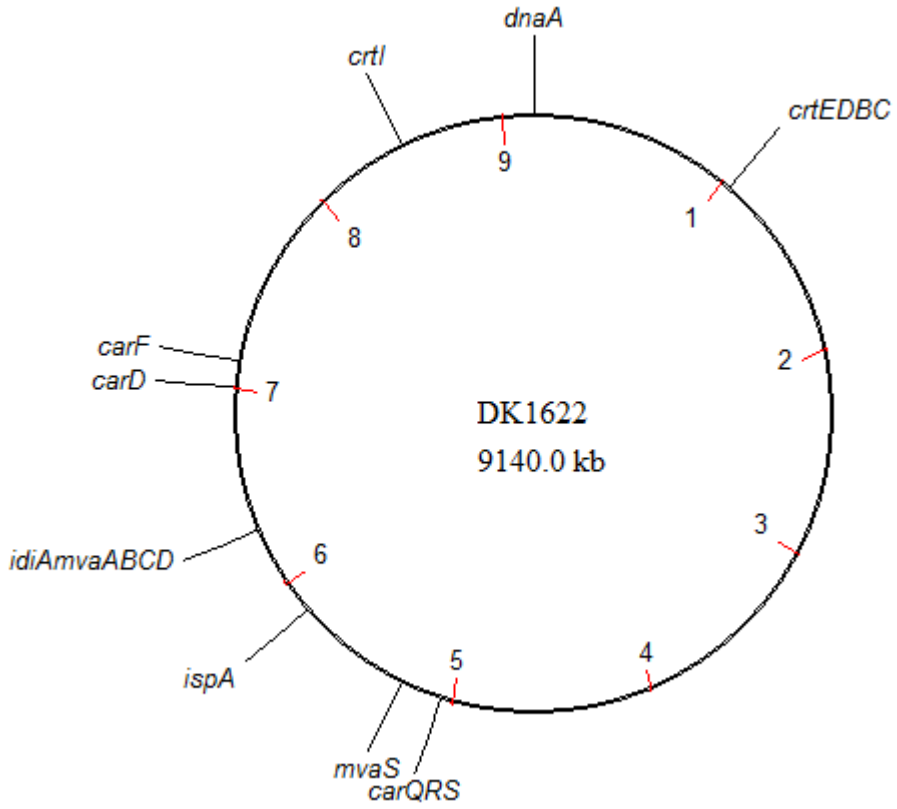


Figure 1.7b The second part of the carotenoid pathway of *M. xanthus*, from the original hydroxyneurosporene up to the generation of the final carotenoid products. Each arrow is labelled with the enzyme that is believed to be responsible for that stage of the process, and the basic enzymatic process that occurs. The structural change at each stage is circled.



dnaA:	1bp	*crtEDBC*:	1019276-1029673bp
carQRS:	5034991-5036509bp	*mvaS*:	5230137-5231393bp
ispA:	5827659-5828552bp	*idiAmvaABCD*:	6271711-6272142bp
carD:	6986642-6987592bp	*carF*:	7118705-7119550bp
crtI:	8493471-8495060bp		

Figure 1.8 The *Myxococcus xanthus* DK1622 genome with known carotenoid and mevalonate pathway genes highlighted. Each red numbered spoke indicates 1000kbp. *dnaA* encodes a chromosomal replication initiation factor; *mvaS* encodes HMG-CoA synthase; *ispA* encode geranyl pyrophosphate synthetase; and *idiAmvaABCD* encodes the four enzymes responsible for converting 3-hydroxy-3-methylglutaryl-CoA to isopentenyl pyrophosphate (Figure 1.6). All other *car* and *crt* gene roles are detailed in the text.

1.6. Synthesis pathway

In order to produce carotenoids, *M. xanthus* has evolved a biosynthetic pathway, under light control, ultimately resulting in the formation of the carotenoid esters (Figure 1.9). The majority of the proteins are located on the inner membrane surface of the cell with two, CarR and CarF, being integral membrane proteins. The principle carotenoid biosynthetic enzymes are encoded by the *crtEBDC* operon (Botella *et al.*, 1995; Ruis-Vazquez *et al.*, 1993), but the whole system is almost entirely controlled through the direct activity of the *carQRS* operon. This is the initial gene region transcribed following cell exposure to a damaging blue light source.

1.6.1. The CarR complex

When a *M. xanthus* cell is grown in the dark, in a resting state, the protein CarR is attached to the inner cytoplasmic membrane (Hodgson, 1993). Here it sequesters CarQ (Gorham *et al.*, 1996), a known sigma factor and a member of the ECF family (extra cytoplasmic functions family), produced from the same operon, *carQRS*. As a result of this interaction CarR can thus be defined as an anti-sigma factor (McGowan *et al.*, 1993). CarQ is the primary protein responsible for controlling gene transcription during the carotenogenic process, including its own transcription, by promoting transcription of *carQRS* (Whitworth *et al.*, 2004). In order to promote transcription, CarQ first has to be released from the membrane-bound complex it forms with CarR. The precise mechanism by which this occurs is not currently known, although it is certain that exposure of the cells to blue light does result in the disruption of CarR (Kraft *et al.*, 2003) and the subsequent release of CarQ (Browning *et al.*, 2003).

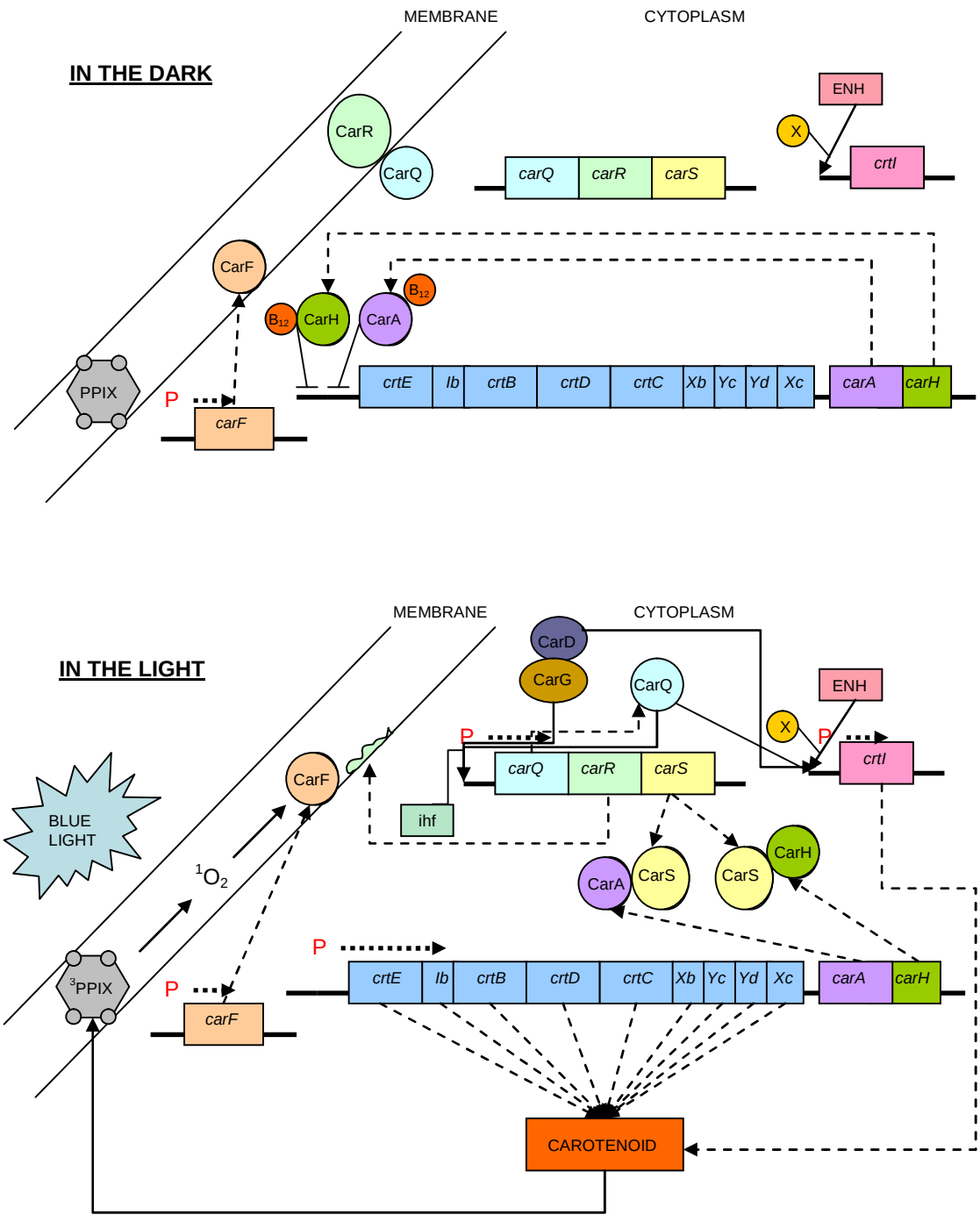


Figure 1.9 The known carotenogenesis protein cascade triggered in *Myxococcus xanthus* upon exposure to blue light (See text for detailed explanation). 'ENH' is an enhancer element and 'X' signifies catabolite alteration. 'P' indicates the promotion of gene transcription and the accompanying arrow the direction of transcription.

A novel third protein directly involved with this process is CarF, a recent protein addition to the light-induced carotenogenesis pathway. It is similar in structure to previously identified Kua proteins, which are well known in humans (Fontes *et al*., 2003). They are both membrane bound components containing similar histidine residues within their sequences. It is currently believed that CarF is the initial protein to react with $^1O_2$ in the protein cascade, a fact supported by its presence in the membrane alongside CarR. It appears that CarF inactivates CarR, binding with the protein and causing its resultant degradation, although the method by which this is achieved is currently unknown. The binding mechanism with CarR is similar to that observed between CarR and CarQ, indicating that CarF plays the role of an anti-anti sigma factor in the reaction (Galbis-Martinez *et al.*, 2008). This interaction prevents CarQ from binding with CarR and it is thus released from the membrane or prevented from being initially sequestered (Fontes *et al.*, 2003), increasing its level in the cell cytosol.

1.6.2. *crtEDBC* control

CarA is a negative regulator which forms a complex with vitamin B12 (Cervantes and Murillo, 2002) and binds to the promoter region of the *crtEDBC* operon, repressing overall operon activity (Lopez-Rubio *et al*., 2004). The operon contains eleven open reading frames (ORFs) (Figure 1.9) of which eight, *crtE, crtB, crtD, crtC crtXb, crtYc, crtYd* and *crtXc* encode proteins involved in carotenoid biosynthesis (Figure 1.7ab). The role of the *crtIb* product in carotenogenesis, if any, is unclear. Upstream of the *crtEBDC* operon is the promoter region that contains two operators that are recognised by CarA. CarA is able to bind to the upstream operator strongly, resulting in a change in CarA complex formation. Initial binding and protein

shape alteration enables CarA to undergo further secondary binding at the second downstream operator. Once the protein is bound to both operators, *crtEDBC* transcription is prevented and no carotenoids are produced. The activity of CarA is directly regulated through its interaction with the protein CarS.

The third protein that is produced from the *carQRS* operon, CarS, functions as an anti-repressor of CarA activity (Whitworth and Hodgson, 2001). It is unable to recognise or bind to either of the CarA operators, but instead binds directly to CarA (López-Rubio *et al*, 2002; Whitworth and Hodgson, 2001). Following exposure of *M. xanthus* to blue light and release of CarQ from the membrane, CarS is produced at an increased level. Upon contact, it is able to bind with the CarA protein N-terminus (Perez-Marin *et al.*, 2004), an interaction that has been studied in great structural depth (Navarro-Avilés *et al.*, 2007). This combination alters CarA shape, which prevents the CarA protein from binding to the CarA operators. If CarA is already partially bound to the DNA, CarS can still bind to the protein and prevent the subsequent recognition and binding at the secondary operator. This weakens the strength of the overall CarA and DNA interaction, resulting in the protein complex becoming detached from both operators (Lopez-Rubio *et al.*, 2004). Once repressor activity is negated, the operon is transcribed and the biosynthetic enzymes created. The feedback system, including the CarA and CarS interaction, also ensures that carotenoids are not being produced constantly within *M. xanthus*, thus preventing the potential waste of valuable cell resources and energy in their synthesis.

Situated downstream of *carA* is a gene that is co-transcribed with it, *carH*. The protein it produces, CarH, binds to the vitamin B12 (Pérez-Marín *et al.*, 2008) in a similar manner to that observed with CarA (Cervantes and Murillo, 2002). In a similar mechanism to CarA, the CarH-B12 complex is also capable of some degree of *crtEDBC* promoter repression in the dark. In conjunction CarS can also prevent

repression of the operon from the CarH-B12 complex as well as CarA. The precise role of B12 in this process is unclear and its binding to CarH does not seem to affect overall *crtEDBC* operon control resulting directly from CarA and CarS interaction (Pérez-Marín *et al.*, 2008). It is known that if vitamin B12 is present and the operator regions for both CarA and CarH are removed, then no visible operon repression occurs within the cell. CarH possesses an N-terminal end with 40% similarity to that of CarA, and as a result any effects caused by CarH presence are not observed upon CarA removal, as CarS is then instead able to bind to CarH, preventing operon repression. In the presence of CarA this reaction is not as noticeable as CarA displays a higher binding affinity for CarS than CarH does. Given these observed results the role of CarH is unclear, although it is still expressed in the cell when CarA is mutated as long as its operator is still intact (Pérez-Marín *et al.*, 2008), indicating that it may play some sort of role as a secondary *crtEDBC* regulator.

1.6.3. Additional key proteins

As CarQ directly controls CarS production, it is effectively the principle protein responsible for the control of carotenoid synthesis. This is supported by the fact that it also plays a role in the promotion of the separate *crtI* gene (Ruiz-Vazquez *et al.*, 1993). *crtI* encodes the biosynthetic enzyme phytoene dehydrogenase, which is responsible for two steps during primary carotenoid production (Figures 1.7a and 1.7b). *In vitro* studies of transcription activation have found that CarQ is incapable of *crtI* transcription alone, and the entire process also requires the host cells to have undergone a period of carbon starvation before it can occur (Browning *et al.*, 2003). It is hypothesised that additional regulator and catabolites and possibly enhancer sequences, may be involved in *crtI* control (personal communication, F. J. Murilllo).

An additional carotenoid protein, CarD, is also required for the successful transcription of both *crtI* and *carQRS* (Nicolas *et al*., 1996). Although the mode of action is still currently not fully understood, preliminary studies indicate that it functions as a transcription factor, essential in the activation of carotenogenesis. Besides its role in the reaction to blue light, it also plays a role in the cellular response to nutrient starvation, which may in part explain its involvement in *crtI* transcription (Galbis-Martínez *et al.*, 2004). It is the only known protein that is involved in both fruiting body formation and carotenogenesis in *M. xanthus* (Nicolas *et al*., 1994). A similar homologue has also been detected in the closely related *Stigmatella aurantiaca* (Cayuela *et al*., 2003).

CarG is another carotenogenesis transcription factor, which is zinc-associated, suggesting it plays some sort of structural role. The protein is unable to bind directly to DNA, but forms a reaction complex with CarD (Peñalver-Mellado *et al.*, 2006) Data indicate that when either of the two proteins is present in a cellular reaction then so is the other, and that binding between the two involves the N-terminus of CarD. The 1:1 ratio bind ratio observed for the two proteins in the cell is similar to that recorded for CarA and CarS, and it has been shown that CarD binding with CarG is essential for either to function effectively as a transcription factor.

Aside from the principal proteins controlling carotenogenesis, a number of additional elements have also been recorded to have a direct effect on the efficiency of the process. The best example of this is observed when increased levels of copper are present in the bacterial growth media, resulting in the stimulation of carotenoid synthesis (Moraleda-Munoz *et al.*, 2005). Copper, particularly in the Cu^{2+} form, is known to generate oxidising agents, hence the cellular response (Valko *et al.*, 2005). Studies indicate that each of the known genes involved directly in carotenogenesis reacts to the presence of copper ions, with the sole exception of *carF*. This suggests

that CarF is designed to specifically react to blue light and no other external stimuli. In turn this means that the copper ions, either directly or through an entirely different protein cascade, disrupt the CarR-CarQ complex in an alternative mechanism to carotenogenesis, still resulting in carotenoid formation.

1.7. CrtW

The final stage in the production of myxobacton, the more prevalent of the carotenoid esters in *M. xanthus*, is a ketonisation step. The enzyme believed to be responsible for this is a β-carotene ketolase. Similar enzymes that function at the equivalent stage of carotenogenesis have already been identified in a range of other bacteria, including *Bradyrhizobium* species (Hannibal *et al.*, 2000), *Synechocystis* species (Fernández-González *et al.*, 1997), *Synechococcus* species (Albrecht *et al.*, 2001) and *Brevundimonas* species (Tao *et al.*, 2006). They are usually annotated as either CrtO or CrtW, with the primary differences between the two ketolases being their lack of sequence similarity with each other and the former being generally twice the size of the latter (Takaichi and Mochimaru, 2007). Despite the final step in carotenoid production appearing to be a ketonisation, no candidate gene or enzyme for this function had previously been identified in *M. xanthus.*

Following preliminary analysis of the *M. xanthus* DK1622 genome sequence (D. Whitworth, unpublished), a candidate ketolase-encoding gene was identified. This was achieved through scanning the entire genome for potential open reading frames and identifying any homologues through online genome database searches. The product of one such gene, situated approximately 1.45kb from the *M. xanthus* origin of replication, displayed significant sequence similarity with the product of a *crtW* gene previously identified in *Gloeobacter violaceus*; a cyanobacterium and

carotenoid producer (Steiger *et al*., 2005). The enzyme encoded by the gene was responsible for part of the beta-carotene production pathway, functioning as a ketolase and was designated CrtW.

The preliminary sequence analysis of *M. xanthus crtW* (as the gene shall be referred to henceforth) and its associated flanking regions also identified three additional potential genes situated nearby, with the closest two less than 100bp either side of the sequence (Figure 1.10). Following further sequence analysis, the products of the pair of genes immediately downstream of *crtW* were identified as a putative oxidoreductase (product of *ord*) and an integral, inner membrane protein (product of *mmb*) respectively. Upstream of *crtW* a possible putative dipeptidylphosphatase encoding gene (*pdp*) was identified. In addition to the three genes, a gene encoding a potential heat-shock protein appeared to be located immediately upstream of the *pdp* gene, but transcribed in the opposite orientation to the *pdp crtW ord mmb* cluster. Sequence analysis suggests this encodes HtpG, a protein believed to be involved in the biosynthesis of tetrapyrroles in a range of cyanobacteria (Watanabe *et al*., 2006). Given the close proximity of the *pdp*, *ord* and *mmb* sequences to *crtW*, and the fact that all four of the genes are transcribed in the same orientation, it appears possible that they were arranged in an operon.

7484620 CGTCCACGCCAGCGGCGCTTCCCAGTCGTGCGTCAGGTGCTTGTAGAACTCCTGGTACTGCTCGTCCGTGATTTC
GGACTTGGAGCGCTGCCACAGGGCGCTGGCCTTGTTGACGACCTCCAACGACGTCTCCGTCTTCGCCTCGTCGCCGGTCCCCGTCG
TCTTGCTCACCTGGAGCTTGATGGGGTGGCCCACGTAGTCGGAGTACTGCGTGATGAGCGACCGCAGCCGCCACTCGCCCAGGAAC
TCCTTCTGGTCCTCCTTCAGGTGCAGGGTGATGGAGGTGCCGCGCGCGGCGCGCTCGGCGGGCTCCACCGTGAAGGAGCCCTTGGC
TTCCGACGTCCACCGCCAGGCGGACTGGCCCTGGCCGGCGGCGCGGCTGACCACCTCCACGCGGTCCGCGACCAGATAGGCGCTGT
AGAAGCCCACGCCGAACTGGCCGATGAGCTGCATGTCCTTCTGCTGGCCCTTCTGCGCCAGCGCCTCGATGAACTCGCGCGAGCCG
GAGTGGGCGATGGTGCCCAGGTTCTTCACCAGCTCGTCATGCGACATGCCGATGCCGGTGTCCTCGATGGTGAGGGTGCCCTTCGC
CTCGTCCGGGATGAGGCGCAGCTCCAGGGCCGGCTCGTCCGCCAGCAACTCCGGCTCCGTAATCGCGCGGAACCGGAGCTTGTCGA
GCGCGTCGGACGCGTTGGACACCAGCTCGCGGAGAAAAATCTCCTTGTGGCTGTAGAGCGAATTGATGACCAGGCTGAGGAGCTGA
TTGATTTCCGCCTGGAATGCGTGGGTCTCCCGCTGGGGGGCGTTTTCGACGGTCATGGTGACAGGGGCCTCCGGGGCGCCCGGTGG
GGCGCCAGCGTCGCTTTCCCTTAACCATGGAAAACCGCTTGTCGAGTCCGGACTGTCCGCCCGCCCGCTCAAGGGCGGTGTGCTG
GTGGTCCATTCCGGGTGCGCGCGAATGAGGGGGAAATGTGGGACGTTGGGTTTTCAGATGTGAGGACCCCTGGGAGGGTGGGCCCG
GGTTGGAGTAGGGAAGGGGACATGAACCGTACCCTCCTGTCCCTGCTCGGCGCGGCGATGCTGTCGGGCGCCGCTACCGCCGCGGA
GAAGACGCCCCCTGCCCGTTTCCCGGATGCCGCGGAGCTCCAGCGCCTGACGGCGCGCTTCGCGCCGGTGGAGCTCCGGGTGGACC
TGAAGGCGCTGCCGGAGTCCGAGCGCCGCGCCCTGGCCCGCATTGTCCAGGCCTCGAAGCTGATGGACACGCTCTTCCTGCGTCAG
CGGTGGGCGGGCAACGAGCCGCTGCTGCTGGACCTGGTCCAGGACACGACGCCGCTGGGCCGCGCGCGGCTGCAGGCGTTCCTGTT
GGACAAGGGGCCCTGGAACAGCCTCGACGAGGCGCGGCCCTTCATCCCGGGCGTGCCGGCCAAGCCCGCGTCCGCGAACTTCTATC
CGGCTGGCGCCACCCAGGCGGAGGTGGAGGCGTGGGTGAAGTCGCTGCCCGAGGCGCAGCAGAAGGAGGCCACCGGCTTCTACACC
ACCATCCGCCGCGGCACGGATGGCCGCTTCATCACGGTGCCCTACAGCGTGGAGTACCAGGGCGAGCTGGCCCTGGCCGCCGCGTT
GCTGCGTGAGGCCGCTGCGCTCACCCAGCAGCCCACGCTGAAGGCGTTCCTCACCTCGCGCGCGGACGCGTTCCTGTCCAACGACT
ACTACGCCAGCGAGGTGGCGTGGATGGAGCTGGACGCCAGCATCGAGCCCACCATCGGGCCCTACGAGGTCTACGAGGACGAGTGG
TTCAACTACAAGGCCGCCTTCGAGGCCTTCGTGGGCCTGCGCGACGACGCGGAGACGCAGAAGCTGGCGAAGTTCAGCGGGCAGCT
CCAGGGGCTGGAGAACAACCTCCCCATCGACCCGAAGCTGCGCAACCCGAAGCTGGGCGCGCTGGCGCCCATCCGCGTCATCAACA
GCCTGTTCTCCTCCGGTGACGGCAACCGGGGCGTGCAGACGGCCGCCTTCAACCTGCCCAACGACGAGCGGGTGTCGGAGAAGATG
GGCTCCAAGCGCGTGATGCTGAAGAACGTGCAGGAGGCCAAGTTCGAGCGCGTGCTGCTGCCCATCGCCAAGGTGGCCCTCACCCC
GGCGGACCAGAAGGACGTCTCCTTCGATGCCTTCTTCACGCACATCTTGATGCATGAGCTGATGCACGGCCTGGGGCCCAGCAACA
TCACCGTGGGTGGCAAGGCCACCACCGTGCGCAAGGAGCTCCAGTCGGCCTCCAGCGCCATCGAAGAGGCGAAGGCGGACATCTCC
GGCCTGTGGGCGCTCCAGCGCCTGGTGGACACCGGCGTCATCGACAAGTCGCTGGAGCGCACCATGTACACGACGTTCCTGGCCTC
CGCCTTCCGCTCCATCCGCTTCGGCGTGGACGAGGCGCACGGCAAGGGCATCGCCGTGCAGCTCAACTACTTCCTGGACACCGGCG
CGGTGAAGGTGAACGCGGACGGCACTTTCTCCGTGGTGCCGGCGAAGATGAAGAAGGCCGTCATCTCGCTGACGAAGCAGCTCATG
GAGATTCAGGGCCGCGGCGACCGGAAGGCCGCTGCCGAGGCGCTGCTGGCGAAGCTCGGCGTGGTGCGCCCGCCCGTGCAGCGCGTGCT
GGAGCGCCTCAAGGACGTGCCGGTGGACATCGAGCCGCGCTACGTCACCGCGGAGGAGCTGGTACGCGACGTGAAGAAGTAGCCAC
CGGCCCATTGCCCCGCCCCGCCGCCTGGGTATTTCCCATGCGCGGCGGGCCGGCGATGGCCAGAAAGCAGGTGAGATGGAGACTTC
CGCCCGCCAACTCCGTCCAGCGCCGCCCGGTCCCTGGGGCGTCGTCATCGCGCTCATCATCATGGGCGCGTGGGGTGGGCACCTTG
CCTGGGCGCTGACACGAGCGGAGCTGCCGTGGGTGGAGCCGCTCACCTGGCTGCACGTCGCTCTCCAGGCCTGGCTGTGCACGGGC
CTCTTCATCACCGGCCACGACGCCATGCACGGCACGGTGTCCGGCCGGCGCTGGGTGAACGAGGCCGTGGGCACGGTCGCCTGCTT
CCTCTTCGCGGGGCTGTCCTACCGTCGGCTGGTGGTGAACCACCGTGCCCACCATGCCCGACCCACGAGCGACGCGGACCCGGACT
TTTCCACCCACAGCCAGTCCTTCTGGCCGTGGCTGGGTACCTTCATGGCCCGCTACACCACGCTGCCCCAGCTTGGGGTGATGGCG
GCCAAGTTCAACGTGTTGCTCTTCCTGGGCGTCTCCCAGCCACACATCCTCGGCTATTGGGTGCTGCCCTCGGTGTTGGGCACGTT
GCAGCTCTTCTACTTCGGCACCTACCTGCCGCACCGGCGCCCGGAGACGCCGGACATGGCCCCTCACCACGCGCGCACGTTGCCGC
GCAATCACCTGTGGGCCCTGCTGTCGTGCTTCTTCTTCGGCTACCACTGGGAGCACCACGAGTCCCCGGGCACACCCTGGTGGCGG
CTGTGGCGCCTGAAGGACGCCCGGGCCCGTGAGGCCGCGCTGACGCAGAGCACCGGGACGCTCCCGGGACAGGAAGGCACCGCCCG
GTAACGCGCGGGCGGCCCGGGCCGTTATACAGGGGCCATGCGCGACAAGCCGCCCGCTGAGCCGCCCAGCTCCGAAGTCACCCCCG
AGAAGACGTACCTGCGCCGGCGCGAGCTGCTGAAGAACGCGGGCCTCTTCGCGGGCACGGCCGTCGCCGTCGCCGGAGGGCTGCAC
CTGCTGGGCCGCAAGCAGACGCGCCCCATGGAGCGCTTCGTCCCGGACGCGGGCCTGGTGGAGCAACCGGTGGCGCAGGCGATGGG
CCCCTTCGACACGGACGAGCCGCGCACGCCCTACGAGGACGTCACCACCTACAACAACTTCTACGAGTTCGGCTTCGACAAGAACG
ACCCGGCCCGCTTCGCGCACACGCTGAAGCCGAAGCCGTGGAGCGTCGTCATCGACGGCGAGGTGCATAAACCGCGGACGGTGGAC
GTGGAGCAGCTCACGTCCTGGTTCTCCCTGGAGGAGCGCGTCTACCGCATGCGCTGCGTGGAGGCCTGGTCCATGGTGATTCCGTG
GCTGGGCTTCCCGCTGGCGGCGCTGCTCCAGCGCGTGGAGCCCACCAGCCATGCGAAGTACGTCGCCTTCACCACGCTGCTGGACC
CGGAGCAGATGCCGGGCCAGCGCCGCGCCCTGTTGGATTGGCCGTACACGGAAGGACTGCGCCTGGACGAGGCGATGAACCCGCTC
ACGCTGCTGGCCACGGGGCTCTACGGCCGGCAACTGCCCAACCAGAACGGCGCGCCGCTGCGGCTCGTGGCTCCTTGGAAGTATGG
ATTCAAGGGCATCAAGTCCATTGTCCGCATCAGCCTCACGCGGGAGGAGCCCATGACGACGTGGCGCCTGTCCGCGCCGCGCGAGT
ATGGCTTCTACGCCAACGTGAATCCTTCCGTGCCCCATCCGCGCTGGAGCCAGGCCAGCGAGCGCCGCATCGGCGACTTCGAGCGC
CGCCCCACGCTGCCCTTCAATGGCTACGCGGAGCAGGTGGCCCACCTCTACACCGGCATGGACCTGCGCCGGTTCTACTGAGTCCC
CGCGCCATGGCCTCGTCTCCGTATCCCTGGCTCAACCCCGCGCTCGTCGTGGGTGGCCTGTCGCCGCTGCTGATGCTCGCCGTCCA
GGGGCCCCGGGGCGAGCTGGGGCCCAACGCGATTGAAGCCGCGCTCCACCAGACGGGGCTGCTGACGCTGGTGCTGCTGGTGGCCT
CGCTGACGTGCACGCCGCTGCGGCTGGTGGCGGGGTGGACGTGGCCCGCGCGCGTGCGCCGCACCCTGGGCCTCTTGGCCTTCACC
TACGCGGTGGCGCACTTCCTCGTGTACGCGGTGCTGGACCAGGGGCTGGCGTGGGGCGCGCTGTGGGCGGACGTCACCGAGCGCCC
CTTCATCACCGTGGGCTTCGCCGCGCTGGTGCTGCTGGTGCCCCTGGCCGTGACGTCGACGAACCGGTGGGTGCGGCGGCTGGGCT
TTCCACGCTGGCAGCGCCTGCACCGGCTGGCCTATGGGGCCGCGGCGCTGGGCGTGGTGCACTTCGTGTGGCGCGTGAAGAAGGAC
GTCACCGAGCCGCTCATCTACGGCGCGGTGCTGGCGCTCCTGATGGCCATTCGCGTGGGTGAAGCCATGCGAAAACGCCGGGCCCG
CGCCGCTGCGGCGGCCCGGAACCCGGCGTGAGTGAAGCGGTGCCCAGACGGGCGCGGCGCGCTACTTGCGCGGCAGCGGGAGGATG
GTGTCCACCAGCGTCATGAGCTGCGCGCAGCTCACCGGCTTGCGGACGAAGGCGCTGATGCCGGCCTTCTGACCCAGCGCGCGCAC
CTCCGCCACGTTCGGGTCGCCCGT 7490019

Green: Heat shock protein
Red: *pdp*
Orange: *crtW*
Blue: Oxidoreductase (*ord*)
Rose: Transmembrane protein (*mmb*)
Bold bases: Start/stop codons
Bold and underlined bases: Possible CarQ dependent promoter P1
Bold and underlined bases: Possible CarQ dependent promoter P2

Figure 1.10 The DNA sequence of the putative *crtW* operon, with gene regions labelled. The numbers represent sequence location in the *M. xanthus* genome.

1.8. Aims and objectives

The primary aim of this investigation was to study the putative *crtW* gene in *Myxococcus xanthus* through a range of experimental approaches in order to further the understanding of both its function and control. Previous sequence similarity studies suggest that the gene encodes a ketolase, which plays a role in the final stages of bacterial carotenoid synthesis. It was also identified as forming a potential operon with three other genes.

The role of the putative *crtW* gene product needs to be confirmed. Currently the proposed role relies on the basis of a single piece of bioinformatic information, so further evidence is required to prove that the gene product does play a role in carotenogenesis and that this role is as a carotenoid ketolase. Consequently it needs to be determined whether or not CrtW is responsible for the production of myxobacton, the primary protective carotenoid in the cell.

An important aim was to see if *crtW* did indeed form part of an operon. If the *crtW* gene does encode a biosynthetic enzyme involved in carotenoid production, then it could be expected that its activation is dependent upon cell exposure to a blue light source. The hypothesis is therefore that the gene is controlled directly by CarQ, the principal sigma factor involved in light-induced carotenogenesis. Given the similarity between *crtI* and putative *crtW* in their isolation from the remainder of the carotenoid enzyme encoding genes within the in *M. xanthus* genome, it is possible that the two were controlled in a similar manner and as a result that *crtW* also required a period of starvation in addition to light exposure prior to activation.

To test these hypotheses a number of studies were proposed. To determine the precise function of the *crtW* gene product, the intended method of study was to delete the gene through construction of a knock-out strain, and then to establish the carotenoids produced in that particular strain. This would verify whether CrtW is a ketolase involved in the final stages of carotenoid synthesis. Isolation of *crtW* and expression in a different organism, one that does not naturally produce carotenoids, should enable the *in vitro* determination that the gene does actually encode a carotenoid ketolase, using β-carotene. Using this purified protein it should also be possible to characterise CrtW in greater depth.

In order to determine control of *crtW* gene expression, the promoter region was fused to an *E. coli lacZ* gene and expression assessed through use of a LacZ assay. The promoter region was mapped through primer extension. The *lacZ* fusion strain should be instrumental in establishing the effects of exposure to blue light and carbon starvation upon *crtW* gene activation. To study the mode of regulation of *crtW* a series of *M. xanthus* mutant strains each deficient in some aspect of carotenogenesis control were proposed for use.

When attempting to study the potential operon it was proposed to use both practical and theoretical approaches. Initially primer extension and quantitative reverse transcriptase PCR were used to establish the existence of the operon and map the promoters. Through a range of bioinformatic techniques it was possible to increase understanding of the genes in the potential operon and their possible functions.

These were the fundamental aims, hypotheses and approaches intended to be used, although it was expected that other findings would be observed and other experimental methods would be required during the course of the investigation. This would prove particularly relevant if CrtW did not function in the carotenogenic pathway as was first anticipated.

Chapter 2

Identification of the *crtW* promoter(s)

2.1. Introduction

The putative product of the *crtW* gene, a ketolase, is believed to function in *Myxococcus xanthus* carotenogenesis, so a reasonable assumption is that it is controlled in a similar manner to other *crt* genes. Currently there are only three known proteins involved in carotenogenesis that directly bind to DNA sequences controlling gene activity, CarA, CarH and CarQ. The principal protein of those is CarQ, the sigma factor, known to directly promote both the expression of *crtI* (Ruiz-Vazquez *et al.*, 1993), as well as the *carQRS* operon (Whitworth *et al.*, 2004).

Unfortunately despite a number of studies of CarQ-controlled genes, the promoter sequence that it recognises is currently unknown. As a result it is difficult to identify whether it does indeed act to express the *crtW* gene in *M. xanthus* from sequence data alone. Initial studies by David Whitworth (unpublished data) suggested possible -35 and -10 recognition sequences of CCGA and GGCG respectively, for CarQ-containing RNA polymerase. Two such promoter regions were identified, the first (P1) located 145bp upstream of the *pdp* open reading frame (orf) of the putative *crtW* operon and the second (P2) 187bp upstream of *crtW* within the possible *pdp* orf (Figure 1.10).

The remaining two possible regulators, CarA and CarH, are both repressors of the *crtEDBC* operon (Lopez-Rubio *et al.*, 2004). The operator sequences were identified as the pairs of inverted palindromes AAGGTT/AACCTT and ACCCTT/AAGGGT for CarA and CarH respectively. Neither sequence was found in the *crtW* operon region or the surrounding area implying CarA and CarH are not regulators of *crtW*. *crtI* is directly regulated by CarQ but is also dependent on catabolic repression. The mechanism of catabolic repression of *crtI* is not known. The absence of a *carA* operator implies *crtW* is more likely to be regulated like *crtI* than *crtEDBC*.

If CarQ-containing RNA polymerase can recognise the potential promoter regions in the *crtW* region, it is possible both could be used to control the gene,

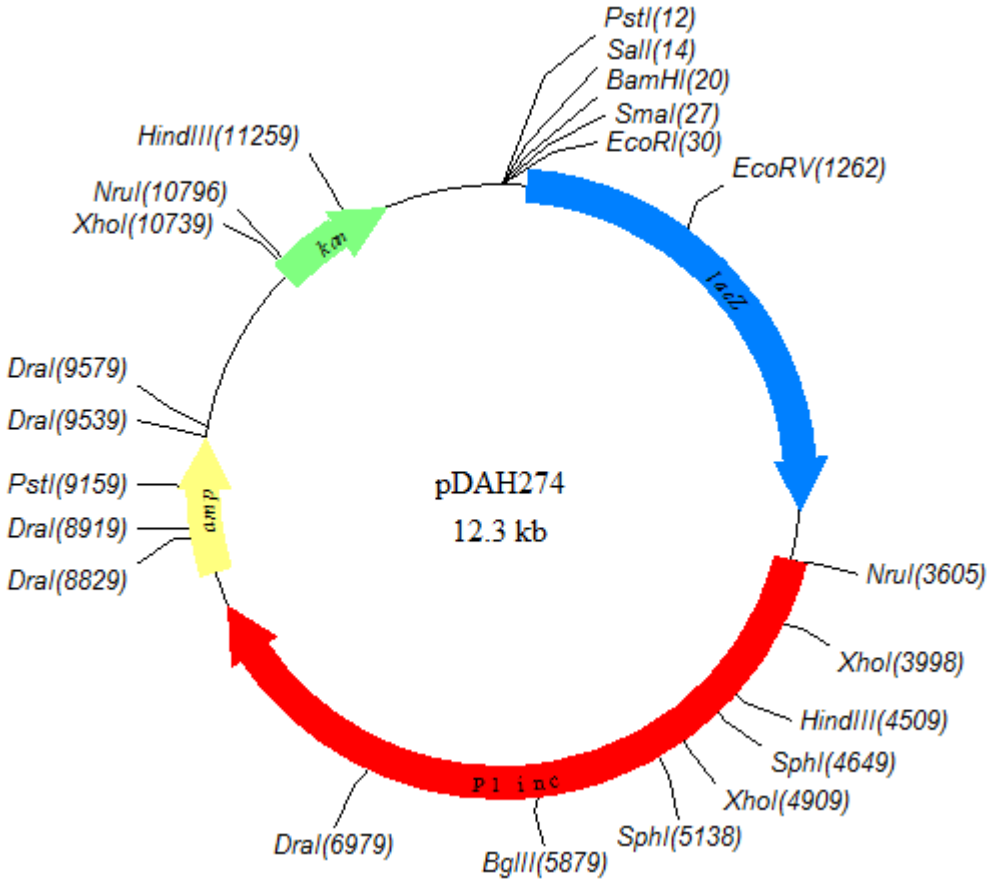
instead of just one. In this case one may be used preferentially over the other. However it is equally possible that neither could be recognised by CarQ RNA polymerase, and that the putative promoters are spurious.

In order to determine whether the potential *crtW* CarQ promoter regions (Figure 1.10) did indeed function as promoters within *M. xanthus*, they were tested using an integrative plasmid-based *lacZ* promoter probe. This involves inserting the potential promoter region of interest upstream of a promoterless *lacZ* gene. The promoter probes are then integrated into the *M. xanthus* genome. LacZ will be produced by the plasmid, if the integrated DNA contains active promoter regions. If the promoters are CarQ-dependent they will be light dependent, inactive in a *carQ* mutant and constitutive in a *carR* mutant.

To test the two *crtW* promoter regions for possible activity, the sequences were to be isolated separately and together. Bearing this in mind, two sets of primer pairs were designed to amplify out the regions of interest from the genome through the polymerase chain reaction (PCR). All of the primers were designed taking GC content, naturally high in *M. xanthus*, and melting temperature into consideration, with factors such as dimer formation and DNA looping identified through the use of a computer-based primer design software, DNA Star 5 – Primer Select. All primer sequences were of a similar length and melting temperature enabling the use of both the first forward primer and second reverse primer, PCRCRTWF1 and PCRCRTWR2, to synthesise a large genome fragment encompassing both promoter regions (‘Primers’ Methods).

2.2. pDAH274

As the promoter regions were to be tested using a *lacZ* promoterless gene, an appropriate plasmid was required containing the gene. A number of previous *M. xanthus* studies had utilised the plasmid pDAH274 for such a role (Galbis-Martínez *et al.*, 2004; Whitworth *et al.*, 2004; Moreno *et al.*, 2001). In this case a unique multiple cloning site (MCS) upstream of the *lacZ* gene allowed sequence insertion (Figure 2.1). The plasmid also contains antibiotic resistance genes for both ampicillin and kanamycin, plus a P1 *inc* region. The P1 incompatibility region allows the plasmid to be incorporated, packaged, into bacteriophage P1capsids. The plasmid can then be transferred by the phage and inserted into *M. xanthus*. Since the development of the



Promoterless *lacZ* region:	140-3209bp
P1 *inc* region:	3609-8409bp
amp:	8619-9479bp
kan:	10709-11519bp

Figure 2.1 The plasmid pDAH274, with key restriction enzyme sites labelled and the coding regions highlighted.

P1 system the process of electroporation has been developed to transfer plasmids to *M. xanthus* (Magrini *et al.*, 1998; Pérez-Marin *et al.*, 2004; Pham *et al.*, 2006). It was this method that was proposed for use following the insertion of the promoter regions into the plasmid. To ensure the sequences would insert correctly into the plasmid appropriate restriction enzyme (RE) sites were added to the ends of each primer, *Eco*RI and *Bam*HI, from the pDAH274 MCS (Figure 2.1). Following experimental verification that pDAH274 cut at the desired restriction sites ('RE Digestion' Methods), the primers were then used to amplify the promoter regions from the *M. xanthus* genome.

2.3. Generation of promoter fragments

Genomic DNA was isolated from a culture of DK1622 cells ('*M. xanthus* DNA Isolation' method) and used in the PCR mix. The 'Simple' PCR protocol was used along with homemade PCR master mix solution, using an annealing temperature of 48^{o}C. Working from the primer designs three fragments were expected from the PCR, designated S1-3 (Figure 2.2). S1 and S2 were the two regions containing the individual promoter regions, S1 before the *pdp* gene (P1) and S2 before the *crtW* gene (P2). S3 was the larger fragment incorporating both of the regions. In terms of size, working from the known genome sequence, the expected fragments were S1-1199bp, S2-1223bp and S3-2388bp in size. The large sizr with all fragments measuring over 1000bp is required to increase the chance of homologous recombination between the inserted gene construct and *M. xanthus* genomic DNA. All of the reactions produced bands following gel electrophoresis, with bands for S1 and S2 of the appropriate size generated (Figure 2.3). In the S3 reaction the band was less than 500bp in size, possibly caused by non-specific annealing of the primers, an interpretation supported by the observation of an additional smaller band in the S3 gel (Figure 2.3).

S1 Fragment (1199bp)

7485253 CCAGCAACTCCGGCTCCGTAATCGCGCGGAACCGGAGCTTGTCGAGCGCGTCGGACGCGTTGGACACCAGCTCGCG
GAGAAAAATCTCCTTGTGGCTGTAGAGCGAATTGATGACCAGGCTGAGGAGCTGATTGATTTCCGCCTGGAATGCGTGGGTCTCCC
GCTGGGGGGCGTTTTCGACGGTCATGGTGACAGGGGCCTCCGGGGCGCCCGGTGGGGCGCCAGCGTCGCTTTCCCTTAACCATGGA
AAACCGCTTGTCGAGTCCGGACTGTCCGCCCGCCCGCTCAAGGGCGGGTGTGCTGGTGGTCCATTCCGGGTGCGCGCGAATGAGGG
GGAAATGTGGGACGTTGGGTTTTCAGATGTGAGGACCCCTGGGAGGGTGGGCCCGGGTTGGAGTAGGGAAGGGGACATGAACCGTA
CCCTCCTGTCCCTGCTCGGCGCGGCGATGCTGTCGGGCGCCGCTACCGCCGCGGAGAAGACGCCCCCTGCCCGTTTCCCGGATGCC
GCGGAGCTCCAGCGCCTGACGGCGCGCTTCGCGCCGGTGGAGCTCCGGGTGGACCTGAAGGCGCTGCCGGAGTCCGAGCGCCGCGC
CCTGGCCCGCATTGTCCAGGCCTCGAAGCTGATGGACACGCTCTTCCTGCGTCAGCGGTGGGCGGGCAACGAGCCGCTGCTGCTGG
ACCTGGTCCAGGACACGACGCCGCTGGGCCGCGCGCGGCTGCAGGCGTTCCTGTTGGACAAGGGGCCCTGGAACAGCCTCGACGAG
GCGCGGCCCTTCATCCCGGGCGTGCCGGCCAAGCCCGCGTCCGCGAACTTCTATCCGGCTGGCGCCACCCAGGCGGAGGTGGAGGC
GTGGGTGAAGTCGCTGCCCGAGGCGCAGCAGAAGGAGGCCACCGGCTTCTACACCACCATCCGCCGCGGCACGGATGGCCGCTTCA
TCACGGTGCCCTACAGCGTGGAGTACCAGGGCGAGCTGGCCCTGGCCGCCGCGTTGCTGCGTGAGGCCGCTGCGCTCACCCAGCAG
CCCACGCTGAAGGCGTTCCTCACCTCGCGCGCGGACGCGTTCCTGTCCAACGACTACTACGCCAGCGAGGTGGCGTGGATGGAGCT
GGACGCCAGCATCGAGCCCACCATCGGGCCCTACGAGGTCTACGAGGACGAGTGGTTCAACTACAAGGCCGCCTTCGAGGC
7486452

S2 Fragment (1223bp)

7486418 CAACTACAAGGCCGCCTTCGAGGCCTTCGTGGGCCTGCGCGACGACGCGGAGACGCAGAAGCTGGCGAAGTTCAGC
GGGCAGCTCCAGGGGCTGGAGAACAACCTCCCCATCGACCCGAAGCTGCGCAACCCGAAGCTGGGCGCGCTGGCGCCCATCCGCGT
CATCAACAGCCTGTTCTCCTCCGGTGACGGCAACCGGGGCGTGCAGACGGCCGCCTTCAACCTGCCCAACGACGAGCGGGTGTCGG
AGAAGATGGGCTCCAAGCGCGTGATGCTGAAGAACGTGCAGGAGGCCAAGTTCGAGCGCGTGCTGCTGCCCATCGCCAAGGTGGCC
CTCACCCCGGCGGACCAGAAGGACGTCTCCTTCGATGCCTTCTTCACGCACATCTTGATGCATGAGCTGATGCACGGCCTGGGGCC
CAGCAACATCACCGTGGGTGGCAAGGCCACCACCGTGCGCAAGGAGCTCCAGTCGGCCTCCAGCGCCATCGAAGAGGCGAAGGCGG
ACATCTCCGGCCTGTGGGCGCTCCAGCGCCTGGTGGACACCGGCGTCATCGACAAGTCGCTGGAGCGCACCATGTACACGACGTTC
CTGGCCTCCGCCTTCCGCTCCATCCGCTTCGGCGTGGACGAGGCGCACGGCAAGGGCATCGCCGTGCAGCTCAACTACTTCCTGGA
CACCGGCGCGGTGAAGGTGAACGCGGACGGCACTTTCTCCGTGGTGCCGGCGAAGATGAAGAAGGCCGTCATCTCGCTGACGAAGC
AGCTCATGGAGATTCAGGGCCGCGGCGACCCGGAAGGCCGCGGCGAGGCGCTGCTGGCGAAGCTCGGCGTGGTGCGCCCGCCCGTGCAG
CGCGTGCTGGAGCGCCTCAAGGACGTGCCGGTGGACATCGAGCCGCGCTACGTCACCGCGGAGGAGCTGGTACGCGACGTGAAGAA
GTAGCCACCGGCCCATTGCCCCGCCCCGCCGCCTGGGTATTTCCCATGCGCGGCGGGCCGGCGATGGCCAGAAAGCAGGTGAGATG
GAGACTTCCGCCCGCCAACTCCGTCCAGCGCCGCCCGGTCCCTGGGGCGTCGTCATCGCGCTCATCATCATGGGCGCGTGGGGTGG
GCACCTTGCCTGGGCGCTGACACGAGCGGAGCTGCCGTGGGTGGAGCCGCTCACCTGGCTGCACGTCGCTCTCCAGGCCTGGCTGT
GCACGGGCC 7487641

S3 Fragment (2388bp)

7465253 CCAGCAACTCCGGCTCCGTAATCGCGCGGAACCGGAGCTTGTCGAGCGCGTCGGACGCGTTGGACACCAGCTCGCG
GAGAAAAATCTCCTTGTGGCTGTAGAGCGAATTGATGACCAGGCTGAGGAGCTGATTGATTTCCGCCTGGAATGCGTGGGTCTCCC
GCTGGGGGGCGTTTTCGACGGTCATGGTGACAGGGGCCTCCGGGGCGCCCGGTGGGGCGCCAGCGTCGCTTTCCCTTAACCATGGA
AAACCGCTTGTCGAGTCCGGACTGTCCGCCCGCCCGCTCAAGGGCGGGTGTGCTGGTGGTCCATTCCGGGTGCGCGCGAATGAGGG
GGAAATGTGGGACGTTGGGTTTTCAGATGTGAGGACCCCTGGGAGGGTGGGCCCGGGTTGGAGTAGGGAAGGGGACATGAACCGTA
CCCTCCTGTCCCTGCTCGGCGCGGCGATGCTGTCGGGCGCCGCTACCGCCGCGGAGAAGACGCCCCCTGCCCGTTTCCCGGATGCC
GCGGAGCTCCAGCGCCTGACGGCGCGCTTCGCGCCGGTGGAGCTCCGGGTGGACCTGAAGGCGCTGCCGGAGTCCGAGCGCCGCGC
CCTGGCCCGCATTGTCCAGGCCTCGAAGCTGATGGACACGCTCTTCCTGCGTCAGCGGTGGGCGGGCAACGAGCCGCTGCTGCTGG
ACCTGGTCCAGGACACGACGCCGCTGGGCCGCGCGCGGCTGCAGGCGTTCCTGTTGGACAAGGGGCCCTGGAACAGCCTCGACGAG
GCGCGGCCCTTCATCCCGGGCGTGCCGGCCAAGCCCGCGTCCGCGAACTTCTATCCGGCTGGCGCCACCCAGGCGGAGGTGGAGGC
GTGGGTGAAGTCGCTGCCCGAGGCGCAGCAGAAGGAGGCCACCGGCTTCTACACCACCATCCGCCGCGGCACGGATGGCCGCTTCA
TCACGGTGCCCTACAGCGTGGAGTACCAGGGCGAGCTGGCCCTGGCCGCCGCGTTGCTGCGTGAGGCCGCTGCGCTCACCCAGCAG
CCCACGCTGAAGGCGTTCCTCACCTCGCGCGCGGACGCGTTCCTGTCCAACGACTACTACGCCAGCGAGGTGGCGTGGATGGAGCT
GGACGCCAGCATCGAGCCCACCATCGGGCCCTACGAGGTCTACGAGGACGAGTGGTTCAACTACAAGGCCGCCTTCGAGGCCTTCG
TGGGCCTGCGCGACGACGCGGAGACGCAGAAGCTGGCGAAGTTCAGCGGGCAGCTCCAGGGGCTGGAGAACAACCTCCCCATCGAC
CCGAAGCTGCGCAACCCGAAGCTGGGCGCGCTGGCGCCCATCCGCGTCATCAACAGCCTGTTCTCCTCCGGTGACGGCAACCGGGG
CGTGCAGACGGCCGCCTTCAACCTGCCCAACGACGAGCGGGTGTCGGAGAAGATGGGCTCCAAGCGCGTGATGCTGAAGAACGTGC
AGGAGGCCAAGTTCGAGCGCGTGCTGCTGCCCATCGCCAAGGTGGCCCTCACCCCGGCGGACCAGAAGGACGTCTCCTTCGATGCC
TTCTTCACGCACATCTTGATGCATGAGCTGATGCACGGCCTGGGGCCCAGCAACATCACCGTGGGTGGCAAGGCCACCACCGTGCG
CAAGGAGCTCCAGTCGGCCTCCAGCGCCATCGAAGAGGCGAAGGCGGACATCTCCGGCCTGTGGGCGCTCCAGCGCCTGGTGGACA
CCGGCGTCATCGACAAGTCGCTGGAGCGCACCATGTACACGACGTTCCTGGCCTCCGCCTTCCGCTCCATCCGCTTCGGCGTGGAC
GAGGCGCACGGCAAGGGCATCGCCGTGCAGCTCAACTACTTCCTGGACACCGGCGCGGTGAAGGTGAACGCGGACGGCACTTTCTC
CGTGGTGCCGGCGAAGATGAAGAAGGCCGTCATCTCGCTGACGAAGCAGCTCATGGAGATTCAGGGCCGCGGCGACCCGGAAGGCC
CCGAGGCGCTGCTGGCGAAGCTCGGCGTGGTGCGCCCGCCCGTGCAGCGCGTGCTGGAGCGCCTCAAGGACGTGCCGGTGGACATC
GAGCCGCGCTACGTCACCGCGGAGGAGCTGGTACGCGACGTGAAGAAGTAGCCACCGGCCCATTGCCCCGCCCCGCCGCCTGGGTA
TTTCCCATGCGCGGCGGGCCGGCGATGGCCAGAAAGCAGGTGAGATGGAGACTTCCGCCCGCCAACTCCGTCCAGCGCCGCCCGGT
CCCTGGGGCGTCGTCATCGCGCTCATCATCATGGGCGCGTGGGGTGGGCACCTTGCCTGGGCGCTGACACGAGCGGAGCTGCCGTG
GGTGGAGCCGCTCACCTGGCTGCACGTCGCTCTCCAGGCCTGGCTGTGCACGGGCC 7487641

Green: Heat shock protein Red: *pdp*
Orange: *crtW* **Bold bases**: Start/stop codons
Bold and underlined bases: Possible CarQ dependent promoters P1/P2

Figure 2.2 The desired promotion region fragments from the *crtW* operon. The numbers represent sequence location in the *M. xanthus* genome.

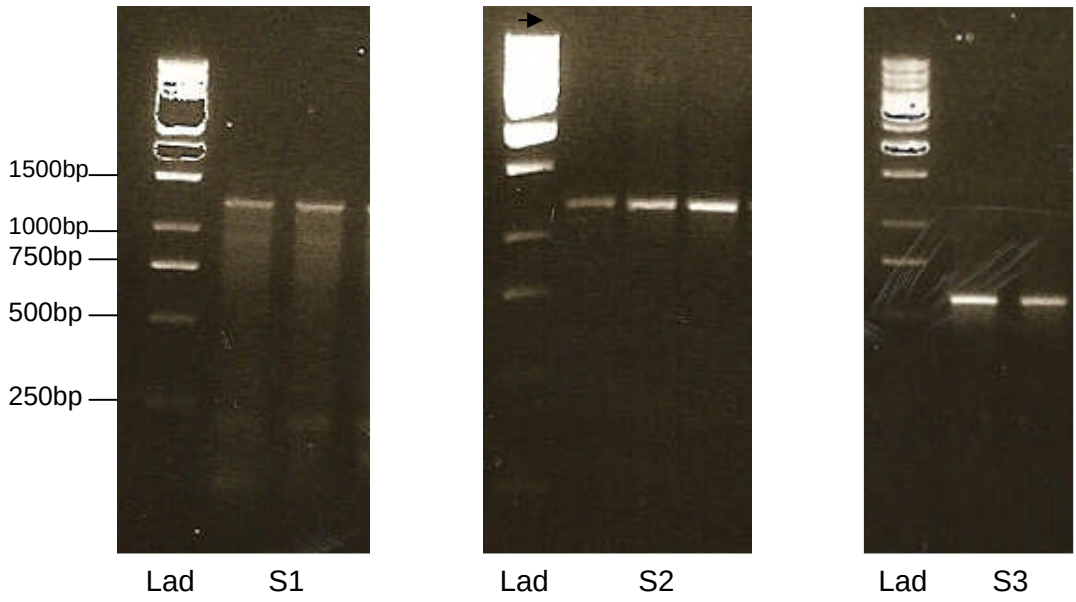


Figure 2.3 Three agarose gels of the PCR products. The products encompass the promotion regions (Figure 2.2).

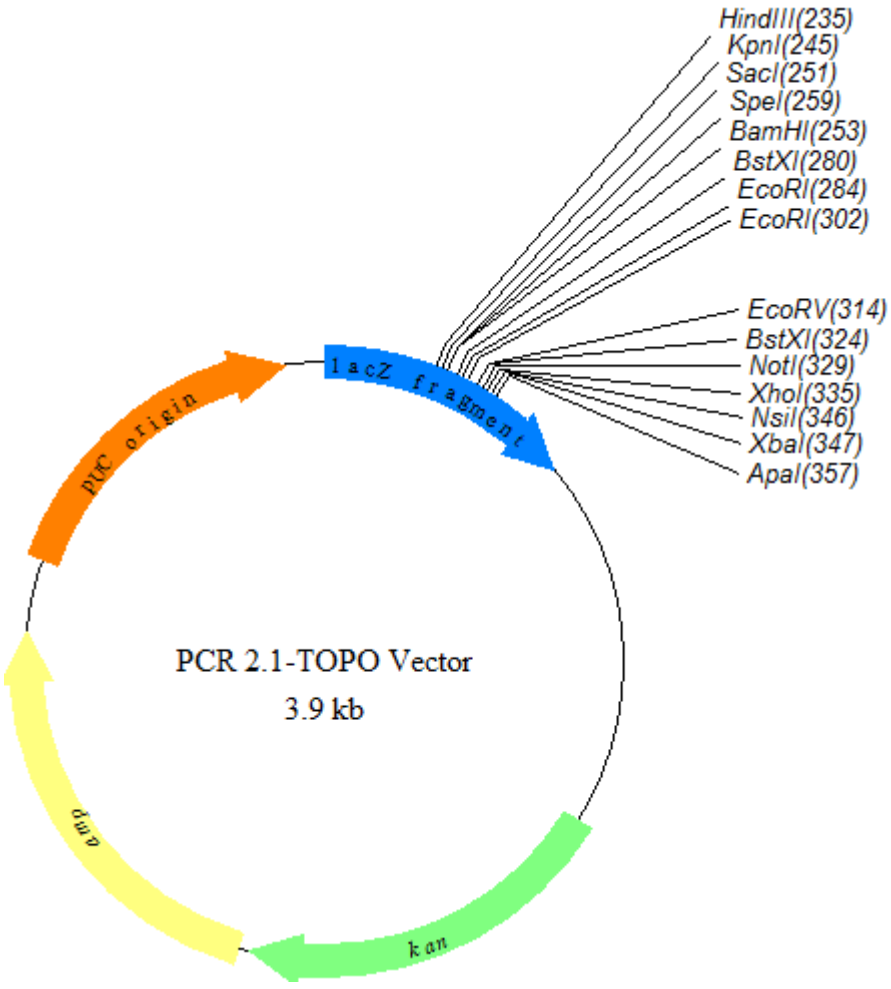
All four of the observed bands were excised and gel purified ('Gel Extraction' Methods) before being sent to the Biological Sciences in-house sequencing suite to determine whether or not the fragments were indeed the sequences desired ('Sequencing' Methods). The results confirmed that the S1 and S2 sequences had been successfully amplified, but the result for S3 indicated a small fragment that did not correspond to the desired DNA stretch for amplification. As S1 and S2 were correct the decision was made to press on with the insertion of those fragments upstream of the *lacZ* promoter in pDAH274. Both the plasmid and the two PCR fragments were digested with *Eco*RI and *Bam*HI, before being isolated via gel electrophoresis and purification. They were then mixed together in a series of ligation reactions and ligation controls ('Ligation' Methods) before being used to transform DH5α, a competent strain of *Escherichia coli*. No transformants were observed and the only growth witnessed initially was on a control plate for undigested pDAH274, and the competent cell control. This proved that although the cells were competent, there was a ligation problem that needed to be addressed.

2.4. Attempt to insert S1 and S2 into pDAH274

Following a number of ligation reaction alterations, principally involving changing the temperature at which the reaction was carried out from 4^{o}C to room temperature, amongst other things, a number of potential transformant colonies were obtained following ligation. This was a small sum, just managing to total double figures, but since only one colony was required that actually contained the insert it should have proved enough. To determine the content of each transformant the 'Simple' PCR protocol was used in conjunction with a pair of pDAH274 primers, amplifying from either side of the MCS, using a cell boilate. The resulting gel of the PCR samples was blank with the exception of ladder and primer dimers. Despite slightly altering primer and DNA concentrations there was no improvement. The colonies tested did not contain the vector with or without insert. Further repetition of the ligation procedure eventually led to more colonies being produced on most of the ligation plates, with high double figure colony numbers being recorded in each case. A number of the colonies, following PCR testing, yielded sequences that showed 99% similarity to the *lacZ* gene in pDAH274. This suggested vector religation or the insertion of a number of additional random bases that may have been present in one of the reagent mixtures. To eliminate any possible contamination, all of the fragments and vectors were digested and purified again, plus fresh buffers and ligation enzymes were used in a new set of ligation reactions. Once more banding was present on an agarose gel following the PCR of a number of transformants, but once again they were all too small and proved to be a result of vector religation. Following a number of similar results, still with the lack of any fragment insertion, a different approach was taken.

As the problem seemed to be the insertion of the fragment, and the fragments used were coming directly from a PCR, it was possible that there was not enough suitable quality or quantity of fragment for insertion. To avoid this problem it was proposed to insert the PCR fragments directly into an intermediate TA cloning vector, 2.1-TOPO (Figure 2.4). In TA cloning, the vector has a pre-cut site with T base overhangs, able to bind to the A-overhangs of the PCR synthesised product, thus incorporating it into the vector. The cut site is found within a *lacZ*α region containing a unique MCS. The principle of the vector is that a sequence is inserted within the

MCS disrupting the *lacZα* sequence, meaning the plasmid is unable to produce the required enzyme to digest X-Gal. To select for these plasmids, *E. coli* cells containing the ligated plasmids are grown on selective media with X-Gal, and identified through blue/white screening. If the colony turns blue it has an intact *lacZα* gene and contains a religated plasmid, but if it is white then there has potentially been some form of DNA insertion.



lacZα region:	1-547bp	M13 forward site:	341-406bp
T7 promoter region:	364-383bp	F1 origin:	548-985bp
M13 reverse site:	205-221bp	*kan*:	1319-2113bp
amp:	2131-2991bp	pUC origin:	3136-3809bp

Figure 2.4 The PCR 2.1-TOPO vector, with key restriction enzyme sites labelled and the coding regions highlighted.

Using freshly prepared PCR products, it proved fairly straightforward to obtain 2.1-TOPO vectors with the desired inserts. Frequency of insertion was still fairly low though, with many more blue colonies visible compared to white ones (a ratio of approximately 15:1). It also took almost 48 hours of growth at 37^{o}C to clearly see the colonies. From the colonies obtained 19 were selected for each insert and subjected to PCR using M13 primers ('Primers' Methods), that recognised sequences either side of the 2.1-TOPO MCS, and again the 'Simple' PCR protocol. Following the disappointing ratio of blue:white colonies, the PCR results gave only three possible inserts of appropriate size – two for S1 and one for S2. The fragments were excised and again sent away to be sequenced to confirm that they were the desired sequence. Fortunately they all were; however one of the S1 fragments had a number of bases incorrectly inserted. This is a consequence of not using proof-reading high fidelity Taq DNA polymerase during the PCR, with a standard Taq polymerase preferred for the generation of shorter sequences throughout this investigation. The latter incorrect PCR product was discarded leaving two vectors with the correct sequence, pTOPS1 and pTOPS2.

Despite a number of efforts to obtain the larger fragment, S3, basic PCR was not effective. This was entirely possible due to the size of the desired fragment and even though the PCR timings were modified, it proved to be to no avail. All fragments that were obtained by this method were of the wrong size and DNA sequence. As a final attempt to obtain S3 an approach known as splicing by overlap extension (SOE) was attempted (Horton *et al.*, 1989).

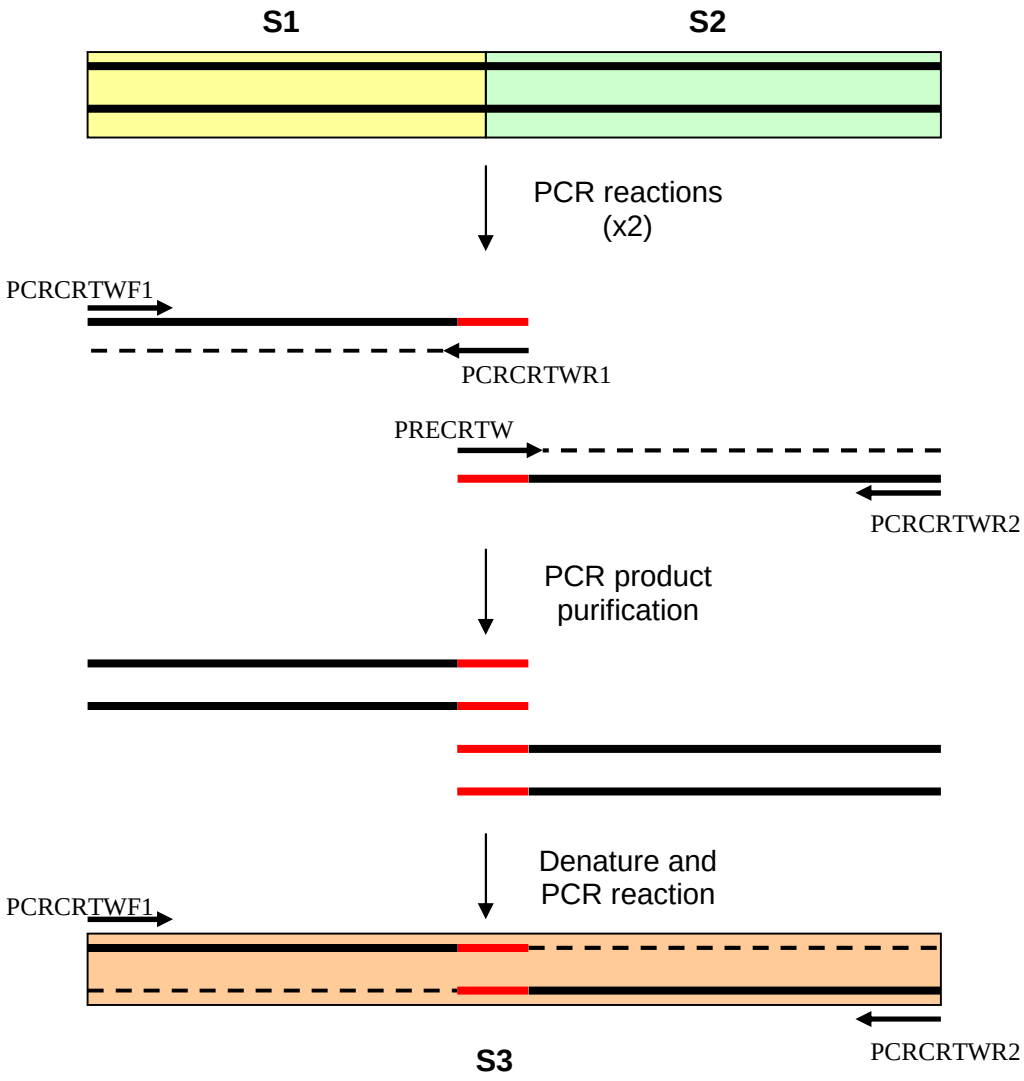


Figure 2.5 The principles behind the SOE approach to generate a fragment encompassing both possible *crtW* promoter regions. The overlapping region is highlighted in red.

The basic principle of SOE is that instead of amplifying one large fragment, two shorter fragments are produced via PCR working from either end of the S3 fragment, but displaying an overlap where they meet in the middle (40bp-60bp in length). The specific overlap allows the two pieces of DNA to become linked together or spliced in a separate round of PCR reactions (Figure 2.5). Following a number of new primer designs, to give rise to optimal overhang lengths, the process was attempted. It proved semi-successful in that obtaining the two secondary fragments was relatively straightforward, yet despite this, combining the two did not prove possible. Again this was despite using a number of protocols ('SOE' Methods) and reagents, including a high fidelity Taq polymerase. Consideration was given to annealing the two separate fragments together, but to do so would have involved

disrupting an area of the sequence. Given the potential role of much of the sequence, and that the location of any promotion region was currently unknown, this idea was rejected. Despite the amount of time and energy spent attempting to isolate the S3 fragment the experiment was halted. As both promoter regions had been successfully cloned into an intermediate vector, it would now be possible to test these regions individually. As the promoter probe integrates by homologous recombination, the use of the S2 fragment means that functionally both promoters are tested. However through the use of the S2 fragment *crtW*, *ord* and *mmb* will all become separated from the first putative promoter, P1. This would result in lack of expression of the three genes if P2 is not active, so S2 could potentially be mutagenic unlike S3 (Figure 2.6). Thus the only problems with not testing S3 is if S2 is mutagenic, or if *ord* or *mmb* are essential.

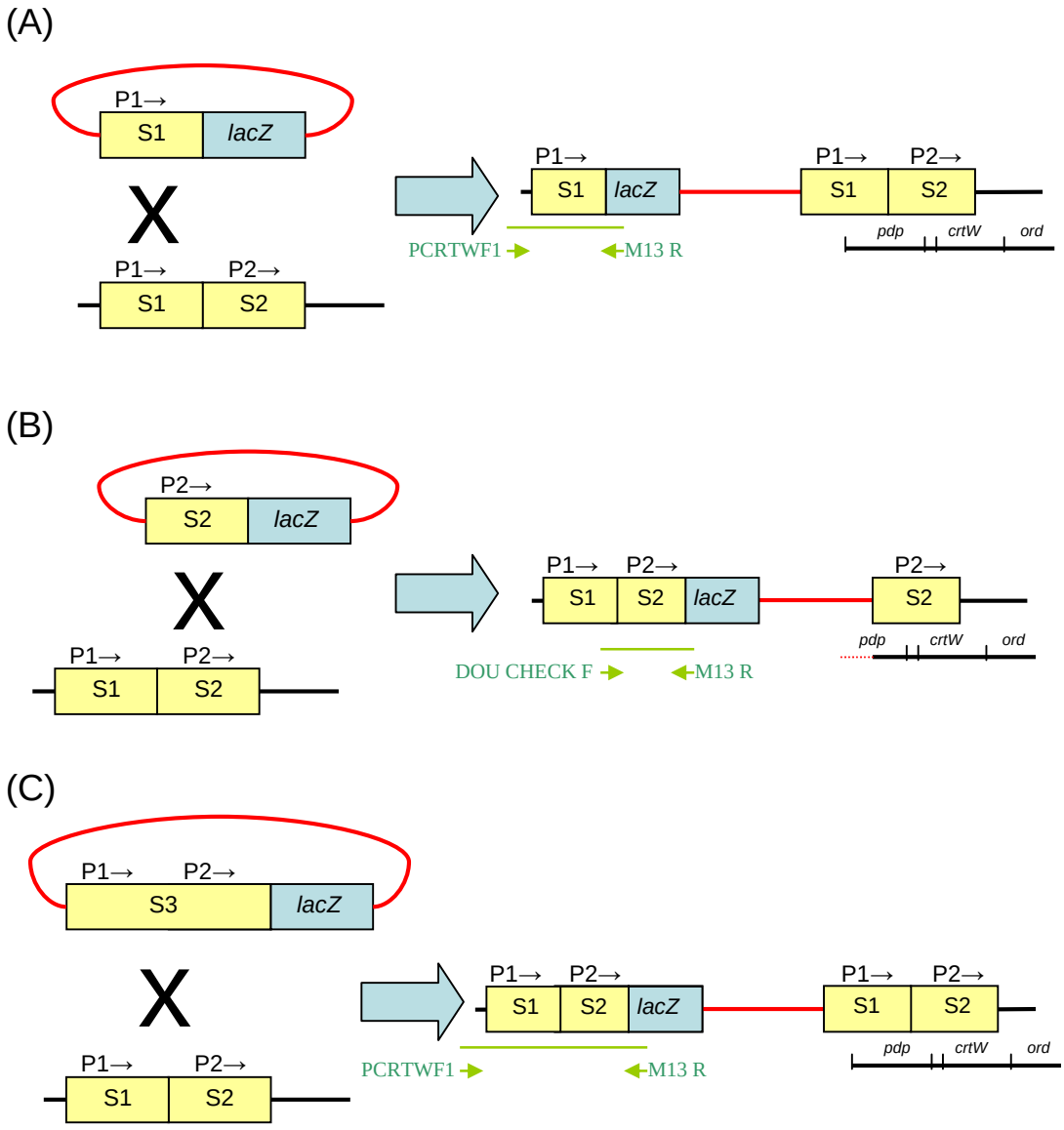


Figure 2.6 The expected genomic products following *M. xanthus* genome homologous recombination with: (A) an S1 vector, (B) an S2 vector and (C) an S3 vector. The red line indicates the vector DNA sequence. P1 and P2 indicate the promoters contained within the fragments. The green labelling shows the primers and region amplified in order to confirm successful homologous recombination between the genome and plasmid construct.

2.5. Subcloning of S1 and S2 into pDAH274

Initial attempts to subclone S1 and S2 from the pTOPS1 and pTOPS2 into pDAH274 failed due to an inability to efficiently isolate fragments from the agarose gel. To combat this problem a different extraction method was approached using

glass beads to bind the DNA fragments, thus extracting them from the gel ('Geneclean II' Methods). This method yielded a significant amount of purified fragment. Likewise the process was used on the digested pDAH274 vector to similar effect. It is unclear why there should be such an efficiency difference between the two techniques, although it appears that larger fragments are not effectively eluted from the QIAGEN filter columns. QIAGEN filter extraction was used for <1kb DNA fragment extraction. For larger fragments >1kb, the 'Geneclean II' purification method was henceforth adopted.

Once all of the fragments were successfully excised a new ligation between the reagents was attempted, following the same procedure as before. Finally this proved to be a great success, with an increased number of transformants and at the first attempt managing to obtain colonies that produced a band during gel electrophoresis of expected S1 size. This was confirmed from sequencing results on the region, creating the first desired insertion vector, pJRNS1. Following a couple of attempts with the S2 fragment, this was also correctly inserted into the vector creating the final desired vector, pJRNS2. Again sequencing was the key to ensuring the correct fragment had been inserted and that there had been no base alterations within the actual sequence.

2.6. Electroporation of pDAH274 vectors

To investigate the promotion regions and their associated controls further, the novel plasmid vectors were to be used to transform a number of *M. xanthus* strains via electroporation. Four strains were selected based on their mutant genotypes; DK101 (wild-type), UWM303 (*carQ*), DK717 (*carA*) and DK718 (*carR*) ('Strains' Methods). UWM303 (McGowan *et al.*, 1993), is unable to produce CarQ and allows us to test CarQ-dependence of any promoter activity. DK717 is a *carA* mutant; and would indicate any CarA-dependence. The DK718 strain cannot produce CarR, which means that CarQ is not sequestered to the cell membrane and so leads to constitutive expression of CarQ-dependent promoters.

Initially electroporation was used in preference to P1 packaging, to transfer the constructs pJRNS1 and pJRNS2. Following electroporation of the bacterial strain with the plasmid, the cells were subject to selection on kanamycin media. Any colonies that grew were then tested by PCR of a cell boilate, using pDAH274 primers,

to confirm that the plasmid had been taken up by the bacteria. Electroporation sometimes leads to growth of phenotypically kanamycin resistant colonies on DCY Kan. To eliminate this problem any colonies that did grow on the initial selection plate were picked from the soft agar onto a DCY Kan^{50} plate. Any bacterial contaminants became evident at this stage as they have a completely different appearance to *M. xanthus*.

The initial *M. xanthus* colonies that were isolated were low in overall number, only 2-3 clear colonies per plate. From these approximately a third did not grow once transferred to higher concentration kanamycin, whereas the ones that did showed no sign of the plasmid. These were most likely spontaneous Kan^R mutants, as had previously been recorded in similar investigations (Kuner and Kaiser, 1981). A number of attempts at plasmid insertion into the bacteria all proved unsuccessful, though 12-20 clearly identifiable *M. xanthus* colonies could be seen on each plate. As the actual technique was known to work and had been used at Warwick University (unpublished data), the only possible problems were with either the strains of *M. xanthus* being used or the plasmid vector. Again a number of the strains used had also been successfully transformed using the electroporation method, which only left issues with the newly created plasmids.

A possible explanation of the failure of electroporation was the size of the plasmids, bearing in mind that pDAH274 was large at 12.2kb. The addition of the S1 or S2 fragment increased the size of the plasmid to around 13.5kb, far larger than would be ideal for electroporation. One option was therefore to reduce the plasmid size, by excising any excess DNA sequences. Looking at the plasmid (Figure 2.1) the best option seemed to be the removal of as much of the P1 *inc* region as possible, since this took up almost 5kb of the vector and was not needed for electroporation. From the plasmid map, a number of potential restriction sites were present within the region, and the best option appeared to be a restriction enzyme (RE) digest using *NruI* and *DraI*. This would lead to the removal of all of the P1 *inc* region as well as *amp*, which again was a part of the vector that was not relevant to this experiment, as *M. xanthus* is naturally ampicillin resistant. In total this would lead to the plasmid being reduced to half its original size. Both enzymes gave a blunt ended cut site, so the ends should also religate without much difficulty.

The first RE digest, carried out using *Dra*I, proved very successful, leading to the reduction in size of the plasmid by 2.5kb. However further digestion using *Nru*I

proved problematic, as more bands were present than expected following gel electrophoresis. It was later discovered that there was an additional *Nru*I site in the Kan^R determinant. In an attempt to avoid this problem a different site within the same region was used, *Sph*I, and a digest was carried out in conjunction with *Dra*I. As *Sph*I leaves a 3' overhang, the overhang had to be removed prior to ligation with the *Dra*I digested end. Despite a number of attempts, obtaining a ligated plasmid proved impossible. It may be that the blunt ending process was not as effective as it requires an active 3'-5' exonuclease ('Blunt Ending' Methods). Regardless of the final reason, too much time had been spent in the attempt to alter the plasmid size and so electroporation was abandoned and the bacteriophage P1 packaging method was adopted to transfer the plasmids to *M. xanthus* ('P1 Packaging' Methods).

2.7. P1 packaging of pJRNS1 and pJRNS2

As previously mentioned the P1 packaging method, though fairly efficient, is exacting in its requirements. The initial problem was obtaining a stock of appropriate P1 phage (P1::Tn*9 clr100*) at the correct concentration for the procedure as we needed $5x10^8$ pfu. A number of P1 stocks were sourced, but only one out of six proved to contain phage. Eventually a final lysate with a titre of $1.3x10^{10}$ pfu/ml was obtained. A number of colonies of pJRNS1 transductants were obtained for each of the four *M. xanthus* strains (Table 2.1), and five transductant colonies of each strain were subjected to PCR to confirm pJRNS1 uptake. The PCR results confirmed that the gene construct had inserted in the correct orientation within the genome in each of the tested colonies (Figure 2.6). Following this success the procedure was repeated for pJRNS2. This was not as efficient as the pJRNS1 transduction but again yielded a number of colonies with the correct insert present. Four transductants from each plasmid transduction, one per strain, were saved DK101 [pJRNS1], UWM303 [pJRNS1], DK717 [pJRNS1], DK718 [pJRNS1], DK101 [pJRNS2], UWM303 [pJRNS2], DK717 [pJRNS2] and DK718 [pJRNS2]. The newly formed strains ('Strains' Methods) were now ready to be used in a LacZ assay.

M. xanthus strain	pJRNS1 transductants	pJRNS2 transductants
DK101	>100	42
UWM303	42	6
DK717	34	5
DK718	10	6

Table 2.1 A table showing the comparative totals of *M. xanthus* transductants detected following P1 phage packaging and transfer.

Bearing in mind that if the *crtW* promoter region is activated during carotenogenesis, and in particular by CarQ, it would be expected that the gene is activated only in the presence of a blue light source. To test this hypothesis the eight *M. xanthus* strains containing the promoter probes were streaked onto two separate sets of DCY X-Gal plates. One set was encased in aluminium foil, shielding the samples from light, and the second set was exposed. Both sets were incubated at 30^{o}C in the presence of a white light source for between 48-72 hours, until growth was clearly visible. Each sample produced growth as well as displaying a blue colour tinge to some degree. This indicates that not only do all the strains contain pDAH274-derived vectors, but also that the promoter(s) was active regardless of the light condition. These qualitative results seemed to eliminate CarQ from a role in the promoter expression.

2.8. LacZ assay

To study the level of promotion for each region a quantitative β-galactosidase assay (LacZ assay) was carried out. This experiment relies on growing a culture of each plasmid carrying *M. xanthus* strain to an exponential growth stage, then sampling over time performing quantitative β-galactosidase assays. The specific LacZ activity was determined per amount of cellular protein using a Lowry protein assay ('Lac Z Assay' Methods). The results of specific LacZ activity throughout the assay were measured as units/min/mg protein.

A preliminary test was carried out on pJRNS1 containing strains, taking samples that were grown in both dark and light conditions at the time points of 0 and 6 hours. It is around the latter time that CarQ is usually at its peak rate in production following activation during light-induced carotenogenesis (Hodgson, 1993). Although a number of the readings were unusually large, most likely due to experimental error, it was clear that there was *lacZ* activity at all time points. A

further preliminary test was carried out on each pJRNS2 strain promoter region with similar results, though overall activation appeared to be a lot lower. In each sample an increase in LacZ production was seen between 0 and 6 hours, and there was little difference between the cultures grown in the presence/absence of light. It also appeared that there was high initial *lacZ* activity at the initial time point.

Following the previous data findings, a full scale LacZ assay was conducted sampling over a 14 hour period. This would cover peak CarQ production and also extend beyond the stage at which the synthesis of carotenoids begins to reach a standard rate and plateau (12 hours). Samples were taken at the time points of 0, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12 and 14 hours. The majority of the samples were taken early in the experiment as this is the stage at which catrotenogenesis was most active. Each sample was reproduced in triplicate to account for any anomalous results, and due to the volume of samples that were generated, two separate time courses were conducted; one the pJRNS1 (Figures 2.7 and 2.8) and the other for pJRNS2 (Figure 2.9)[*]. Positive results arose from the LacZ assays for each of the two promoter regions. Firstly the overall trend of all the pJRNS1 strains studied was that the curves generally showed an increase in specific activity over the course of time independent of the light regime (Figures 2.7 and 2.8). This is displayed as a gradual change until around the 5 hour mark when a more dramatic increase was observed (Figures 2.7 and 2.8). When examining the LacZ assays of pJRNS2 strains (Figure 2.9) again there was an increase in specific activity over time, independent of light regime for DK101 [pJRNS2]. Again there was a gradual increase in activity until the 7 hour time-point when there was a more marked increase although less dramatic than for the pJRNS1 strains (Figure 2.9).

With both pJRNS1 and pJRNS2 strains there was evidence of constitutive promoter activity independent of light regime and genetic background.

[*] Please note that there are no error bars displayed on the graphical data as they prevent the clear visualisation of certain time points, particularly in the 4-8 hour region. An example of the graphical data with the standard y-error bars included can be found accompanying this chapter (Figure 2.10) and is a fair reflection of all the data recorded. The error bars were relatively small in the earlier time points although they increased in value for the latter couple of time points, at 12 and 14 hours.

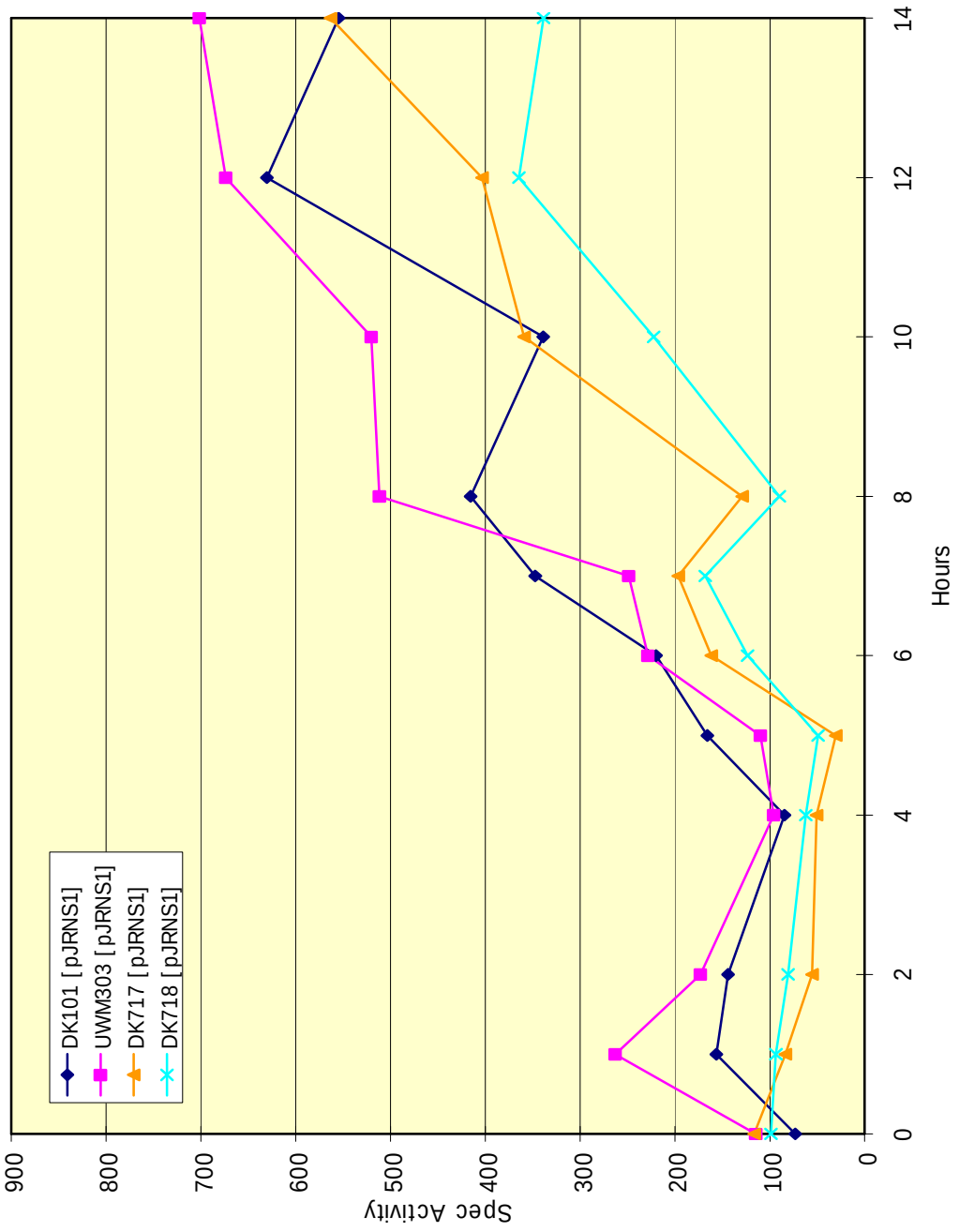


Figure 2.7 The specific LacZ activity (units/min/mg protein) expressed by four different *M. xanthus* strains containing pJRNS1 grown in the presence of light, against time (hours).

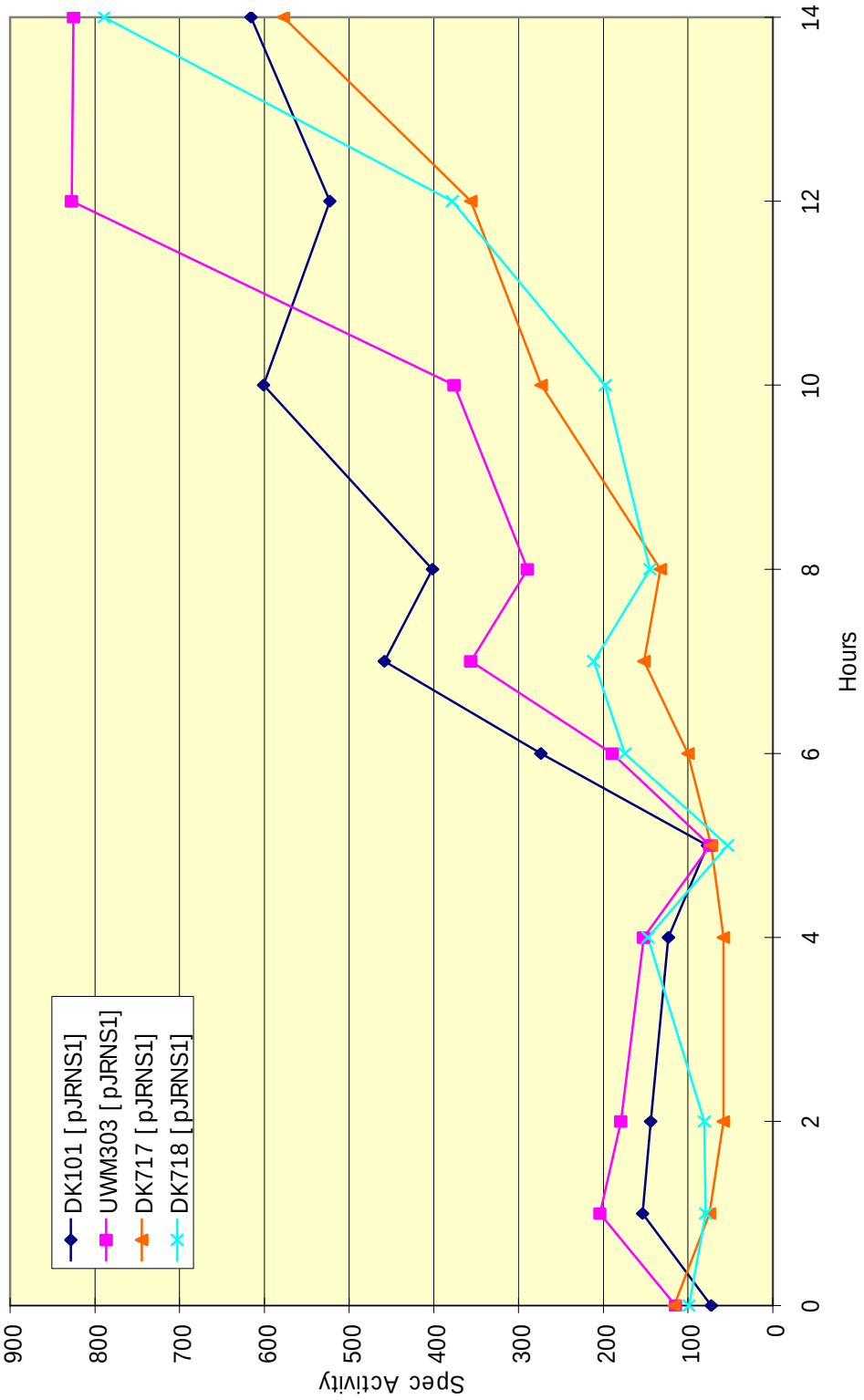


Figure 2.8 The specific LacZ activity (units/min/mg protein) expressed by four different *M. xanthus* strains containing pJRNS1 grown in the absence of light, against time (hours).

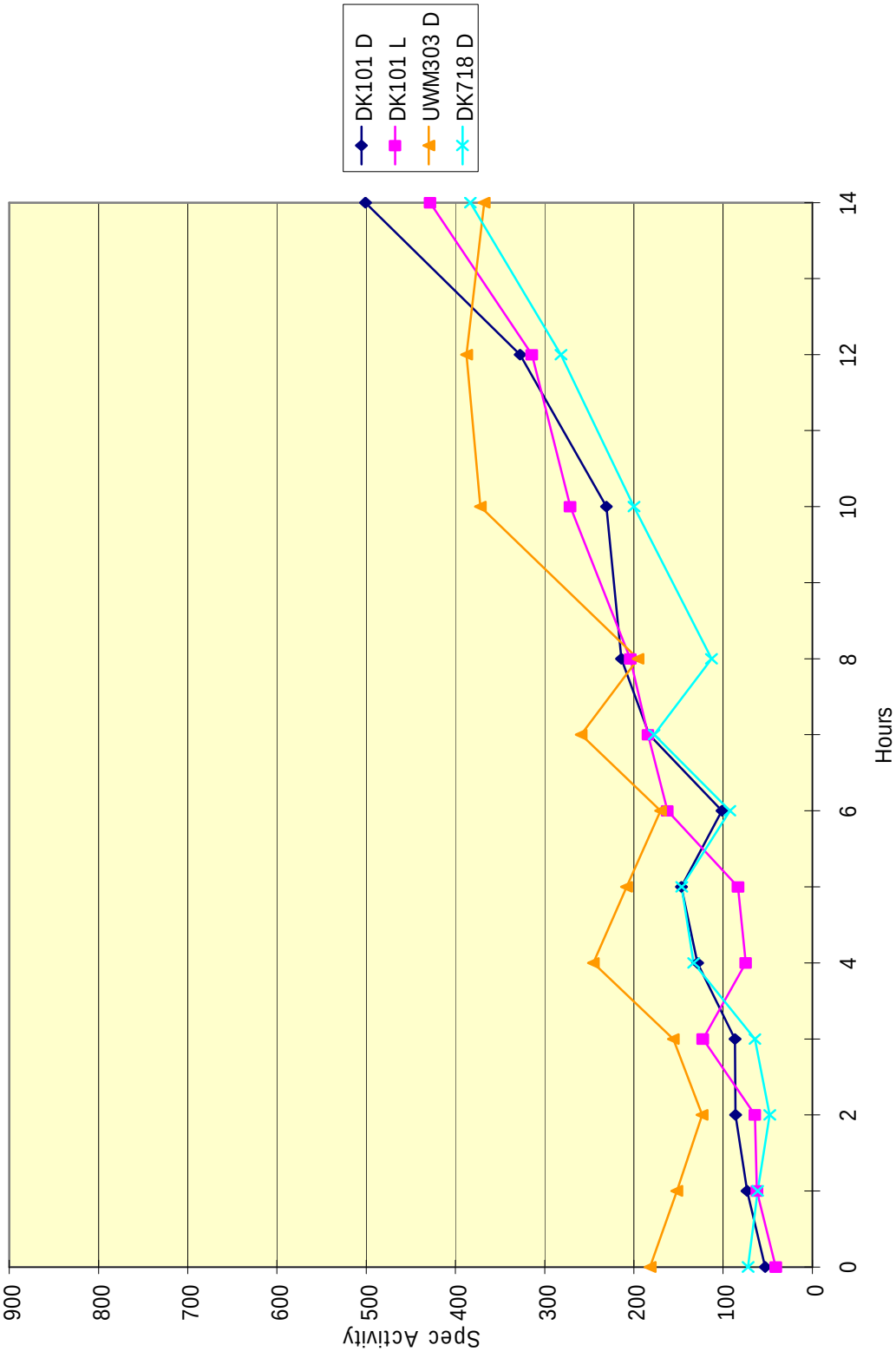


Figure 2.9 The specific LacZ activity (units/min/mg protein) expressed by three different *M. xanthus* strains containing pJRNS2, grown in the absence of light (D), against hours. The specific LacZ activity (units/min/mg protein) of DK101 [pJRNS2] grown in the presence of light (L) against time (hours)

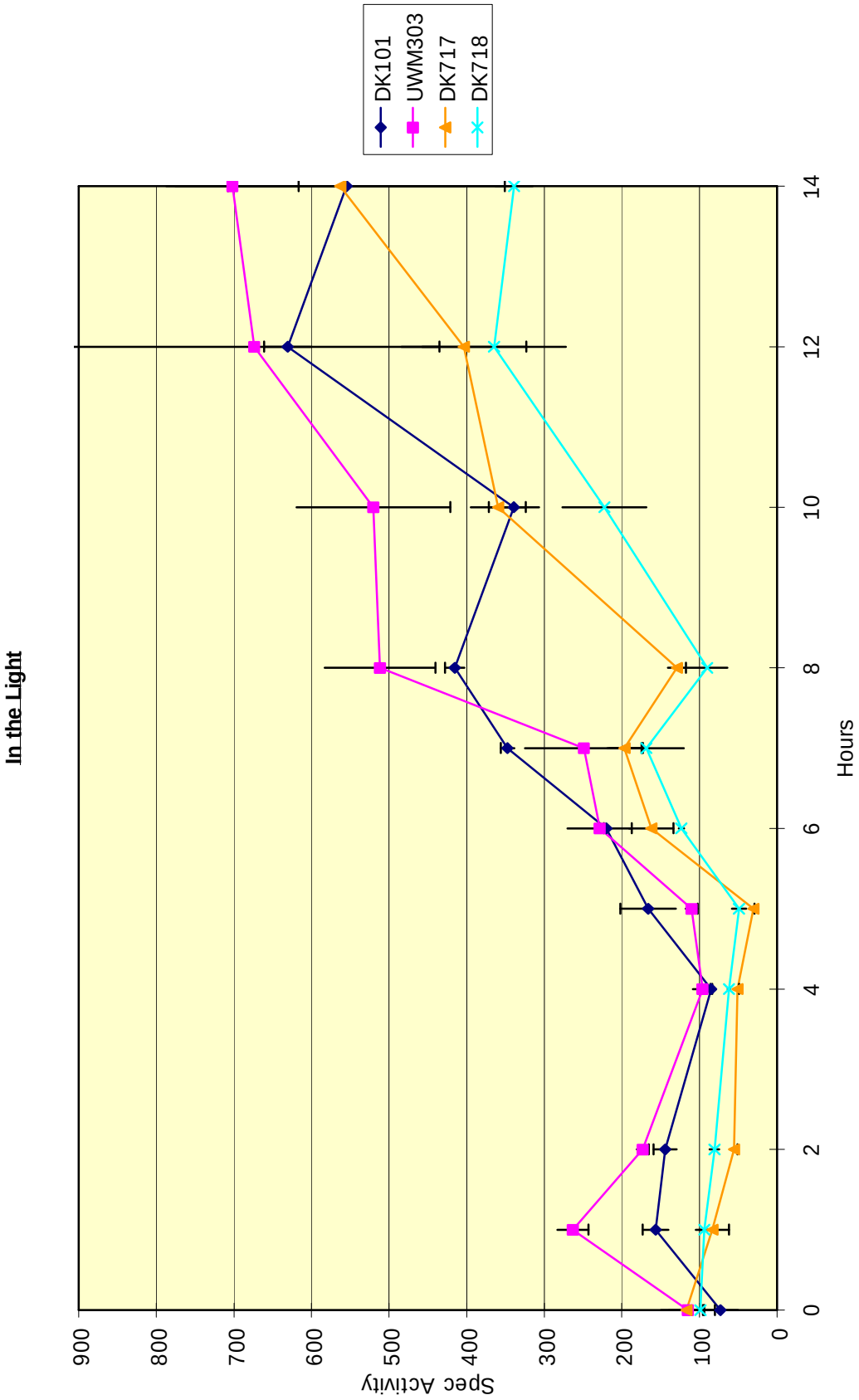


Figure 2.10 Figure 2.7 with the addition of standard y-error bars.

2.9. Promoter activity associated with fragment S1: the *pdp* promoter

The promoter activity seen with pJRNS1 is the promoter activity associated with the *pdp* gene as the S1 fragment ends in the middle of this orf (Figures 2.2 and 2.6). Looking at activity of the *pdp* promoter region in the light (Figure 2.7), there appears to be a fall in activity between the 3 and 5 hour mark in all of the samples. This follows an initial activity doubling during the first hour for the DK101 [pJRNS1] and UWM303 [pJRNS1] strains. Following this dip all four samples then increased sharply in activity, never reaching a plateau. A similar result can be seen when looking at the data for the strains grown in the dark (Figure 2.8), with again an initial doubling in activity observed for DK101 [pJRNS1] and UWM303 [pJRNS1] before a gradual decrease, with a visible dip at the 5 hour time point for all four samples. This is then once again followed by an increase in activity, finishing at similar levels to those seen in the samples exposed to light.

A key finding from the data was the actual value for LacZ specific activity at the 0 time point for the JRNS1 strains (Figures 2.7 and 2.8). All of the samples showed a value of 80-120 units/min/mg protein, which is an exceptionally high level of LacZ activity. To put this into context with regards to previous study of light-induced carotenogenic proteins in *M. xanthus*, during the activation of *carQRS* (Hodgson, 1993) initial *lacZ* specific activity recorded was 0 to 1 unit/min/mg protein and at its peak did not exceed values of 250 units/min/mg protein. In contrast the results for the *pdp* promoter gives initial activity values around 100 units/min/mg protein, and final values of up to 900 units/min/mg protein. As previously mentioned the graphical data suggest very little difference overall between any of the samples of pJRNS1 strains, in either dark or light growth conditions.

By carrying out statistical analysis of the samples, in this case a Student's t-test, it is possible to determine which results are different or similar to each other. In this case a paired t-test was carried out between the strains, with each time-point being compared with its equivalent. No time-points were excluded, so the probability (P) values calculated (Tables 2.2-2.4) use every recorded experimental value for each strain. For the final probability to indicate a result of significant difference it must be below 0.05 in value.

Strain	**DK101 [pJRNS1] *D***	**UWM303 [pJRNS1] *L***	**UWM303 [pJRNS1] D**	**DK717 [pJRNS1] *L***	**DK717 [pJRNS1] *D***	**DK718 [pJRNS1] *L***	**DK718 [pJRNS1] *D***
DK101 [pJRNS1] *L*	0.368469	0.093639	0.253484	0.016961	0.008681	0.00217	0.132526
DK101 [pJRNS1] *D*	~	0.578815	0.775454	0.003037	0.005215	0.002924	0.066369
UWM303 [pJRNS1] *L*	~	~	0.892275	0.002364	0.001874	0.002226	0.030027
UWM303 [pJRNS1] *D*	~	~	~	0.005869	0.003937	0.005165	0.015905
DK717 [pJRNS1] *L*	~	~	~	~	0.191131	0.096218	0.519077
DK717 [pJRNS1] *D*	~	~	~	~	~	0.287609	0.154033
DK718 [pJRNS1] *L*	~	~	~	~	~	~	0.165076

Table 2.2 A table showing the comparative Student's t-test P-values between *M. xanthus* strains containing pJRNS1. Text highlighted in red indicates results that display significant difference. The suffix '*L*' denotes a sample grown in the presence of light and '*D*' represents a sample grown under dark conditions.

In terms of the *pdp* promoter (Table 2.2), there is statistically no difference between any of the light/dark results within each strain. This indicates that there is no difference in promoter activity whether the promoter is exposed to blue light or not. Looking at results between the strains there is also no significant difference between the DK101 [pJRNS1] and UWM303 [pJRNS1] samples, and likewise the DK717 [pJRNS1] and DK718 [pJRNS1] samples. In contrast when comparing either DK717 [pJRNS1] or DK718 [pJRNS1] with UWM303 [pJRNS1], significant statistical differences are recorded. A similar result is seen when comparing the two samples with DK101 [pJRNS1], light and dark growth cultures, with the sole exception of the dark grown DK718 [pJRNS1] sample. These show no significant difference, and the actual values produced are still relatively small.

This supports the preliminary results gleaned from the DCY X-Gal plate cultures, where the promoters appear to function regardless of *M. xanthus* exposure to light or genetic background.

2.10. Promoter activity associated with fragment S2: the *crtW* promoter

The promoter activity seen with pJRNS2 is the promoter activity associated with the *crtW* gene as the S2 fragment ends in the 5' end of the *crtW* orf (Figures 2.2 and 2.6). Because of the geometry of the insertion promoter (Figure 2.6) there remains the possibility that promoter activity detected from fragment S2 is caused by the *crtW* promoter being read through from the *pdp* promoter in fragment S1. Following the *pdp* promoter, and bearing in mind that preliminary tests of the *crtW*

promoter indicated it also was not affected by light regime or genetic background fewer pJRNS2 *M. xanthus* strains were subjected to quantitative LacZ assays, and only DK101 [pJRNS2] was tested for the light effect.

When focussing on the second of the plasmids, pJRNS2, the increase in LacZ is less pronounced when compared to pJRNS1. Again for all of the samples measured in both light and dark conditions (Figure 2.9) a similar graphical pattern is seen. Unlike the S1 fragment though there is no clear point where all samples showed a sharp decrease in activity. All are expressed at a lower rate initially, with the majority being found below the LacZ specific activity of 100 units/min/mg protein. Despite being lower than the pJRNS1 containing strains, this again is still at a far greater initial level than witnessed during previous studies. Activity remains fairly constant over the first six hour period, only doubling over this time, at which point there is a more pronounced rise in the rate of activity. Even then this is only at about half the rate of increase recorded for the pJRNS1 samples, peaking in the DK101 [pJRNS2] dark sample at a specific activity value of 500 units/min/mg protein.

Strain	**DK101 [pJRNS2] *D***	**UWM303 [pJRNS2] *D***	**DK718 [pJRNS2] *D***
DK101 [pJRNS2] *L*	0.427075	0.004819	0.199336
DK101 [pJRNS2] *D*	~	0.020447	0.029274
UWM303 [pJRNS2] *D*	~	~	0.000021

Table 2.3 A table showing the comparative Student's t-test P-values between *M. xanthus* strains containing pJRNS2. Text highlighted in red indicates results that display significant difference. The suffix '*L*' denotes a sample grown in the presence of light and '*D*' represents a sample grown under dark conditions.

Looking at the statistical results for the *crtW* promoter, there are a number of points to note (Table 2.3). Firstly there is no significant difference between either the light or dark growth samples of DK101 [pJRNS2], which supports the findings of the *pdp* promoter, that there is no effect on promoter activity whether it is expressed in light or dark conditions. More interesting is the statistical data comparing the DK718 [pJRNS2] or UWM303 [pJRNS2] strains with DK101 [pJRNS2] cultures. With the exception of the DK718 [pJRNS2] and DK101 [pJRNS2] cultures grown in the presence of a light source, all show significant differences, though the differences between dark grown DK101 [pJRNS2] and the other strains are not highly significant. In contrast the UWM303 [pJRNS2] strain compared to a DK101 [pJRNS2] strain grown in light conditions, and also the DK718 [pJRNS2] culture, shows a very

significant difference. The reasons for this are unclear, since initially it was recorded, when studying the *pdp* promoter region, that UWM303 [pJRNS1] shows similar expression levels to that of DK101 [pJRNS1]. Looking at the graphical data (Figure 2.9) it is possible these findings may be a result of the higher initial expression levels of the UWM303 [pJRNS2] strain.

2.11. Comparison of the *pdp* promoter and the *crtW* promoter

Strain	**DK101 [pJRNS1] *D***	**DK101 [pJRNS1] *L***	**UWM303 [pJRNS1] *D***	**UWM303 [pJRNS1] *L***	**DK718 [pJRNS1] *D***	**DK718 [pJRNS1] *L***
DK101 [pJRNS2] *D*	0.001838	0.001661	0.005156	0.001768	0.022675	0.007421
DK101 [pJRNS2] *L*	0.000702	0.000603	0.002442	0.000795	0.008382	0.001142
UWM303 [pJRNS2] *D*	0.138764	0.308459	0.183005	0.110055	0.572139	0.000357
DK718 [pJRNS2] *D*	0.005237	0.005004	0.010005	0.004784	0.148376	0.932016

Table 2.4 A table showing the comparative Student's t-test P-values between *M. xanthus* strains containing pJRNS1 or pJRNS2. Text highlighted in red indicates results that display significant difference. The suffix '*L*' denotes a sample grown in the presence of light and '*D*' represents a sample grown under dark conditions.

By comparing the two promoter regions (Table 2.4) there are clearly two sets of results that display no significant difference. Firstly neither of the DK718 strains differ significantly between the promoter regions, and then almost all of the pJRNS1 strains show no significant differences when compared to the UWM303 [pJRNS2] strain. In this instance it appears that in the particular strain the promoter is behaving in a similar way to that of the *pdp* promoter, in contrast to the recorded results for the *crtW* promoter in other strains. Statistically the most similar result to UWM303 [pJRNS2] is that of the DK718 [pJRNS1] dark grown culture, whereas the DK718 [pJRNS1] light-cultivated strain shows a difference of high statistical significance.

2.12. Discussion

It appears from the results obtained that both the *pdp* and *crtW* promoters are constitutive. They are both unaffected by light regime and are therefore unlikely to be CarQ-dependent. This was confirmed by genetic background studies on a series of *M. xanthus* strains.

The experimental data show that in UWM303, both the *pdp* promoter and *crtW* promoter are still active, so neither are CarQ-dependent. If they were CarQ-dependent then it would be expected that the loss of *carR* would give rise to increased expression. In neither case, DK718 [pJRNS1] nor DK718 [pJRNS2], was there an increased level of expression to that observed for either DK101 [pJRNS1] or DK101 [pJRNS2]. Therefore neither promoter is CarQ-dependent and the putative consensus sequences that were initially detected in the regions of S1 and S2 are irrelevant.

The lack of any further effect following the loss of *carA* confirms that both the *pdp* promoter and *crtW* promoter are light-independent, constitutive promoters. From the results obtained and without a number of additional control reactions, it is impossible to say if promoter activity observed in strains containing pJRNS2 is as a result of read-through from the *pdp* promoter or not. If so then it is possible that pJRNS2 insertion is potentially mutagenic.

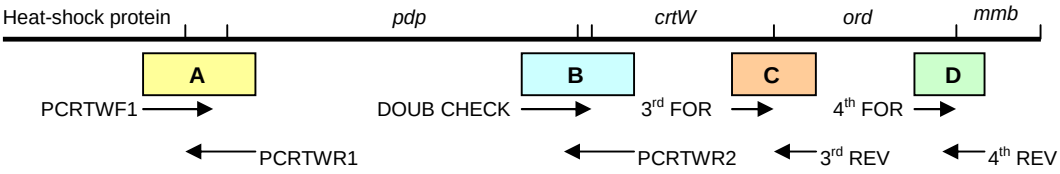
Chapter 3

The regulation of *crtW* transcription

3.1. Introduction

Two fragments with promoter activity had been positively identified within the *crtW* operon, the S1 and S2 fragments (Figure 2.2). Whether both of these function together or separately, control of *crtW* expression was still unclear. There is a possibility that the promoter activity previously detected in S2 was a result of read through from S1. To fully define the potential promoter(s) a key fact to ascertain was whether or not the four ORFs identified; *pdp, crtW,* a potential oxidoreductase (*ord*) and a trans-membrane protein encoding gene (*mmb*), form an operon. For direct genetic evidence the optimal way to determine this was through an RNA study. Once a promoter region is expressed, any ORF(s) downstream that form an operon will be co-transcribed as a polycistronic mRNA. Previous results had shown the promoter regions studied were constitutively expressed, and if all of the ORFs of the operon were transcribed together, then they should form a four ORF polycistronic mRNA. Using PCR and a suitable pair of primers covering each of the intergenic regions, it should thus be possible to determine whether the genes are co-transcribed (Figure 3.1a). A suitable set of primers was designed ('Primers' Methods) to cover the three primary intergenic regions. It was also decided to use the primers used for generating the S1 fragment to determine whether the operon promoter was located upstream of intergenic region 'A', upstream of *pdp* as presumed (Figure 3.1a).

a)



b)

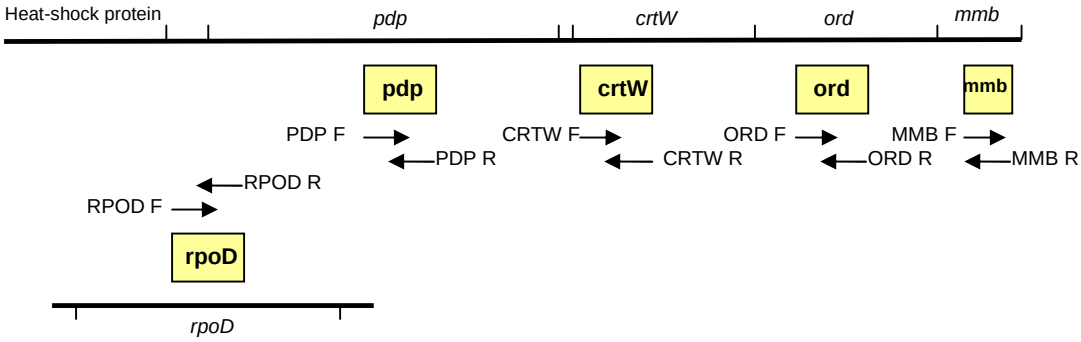


Figure 3.1 ***a*)** The four intergenic regions (A-D) that were sought using the PCR primers labelled in order to determine the extent of the *crtW* operon and ***b*)** the five DNA fragments amplified by QRT-PCR to quantify relative mRNA abundance.

3.2. Identification of the *crtW* operon

As each of the *M. xanthus* strains studied in the previous chapter displayed promoter activity, the DK101 (wild-type) strain was selected for RNA extraction. Also, as the exact time point at which the RNA was sampled did not appear to be critical, a 24-hour DCY liquid culture was prepared and the RNA extracted ('RNA Extraction' Methods). After ensuring the RNA was entirely DNA free, through the use of a DNase, it was then converted into cDNA. This was done using a combination of random hexamers, which lead to the majority of the RNA being converted into cDNA. In this form the material was more stable to work with during the PCR used to test the operon. Finally, as well as the four primer pairings being

tested on the cDNA, a control PCR reaction was also carried out using each of the pairs with purified DK101 chromosomal DNA.

There was a clear band of appropriate size in each of the lanes for the DK101 genomic DNA control, as was expected (Figure 3.2). The cDNA samples displayed bands encompassing the three intergenic regions B, C and D. This indicated that the four genes were indeed linked together in polycistronic mRNA and the four ORFs do form an operon. The lack of a band in the 'A' region confirmed that the heat shock protein gene was not part of the operon. As a band was visible in the genomic control, it also eliminated the possibility of any DNA contamination causing the results. This provides confirmation that the four genes are transcribed as a polycistronic mRNA and form an operon. It also indicates that the genes can all be transcribed from a single promoter region, which is almost certainly that previously identified in the S1 fragment, the *pdp* promoter. However, these data do not confirm or negate the possibility of internal promoters such as a *crtW* promoter in the S2 fragment.

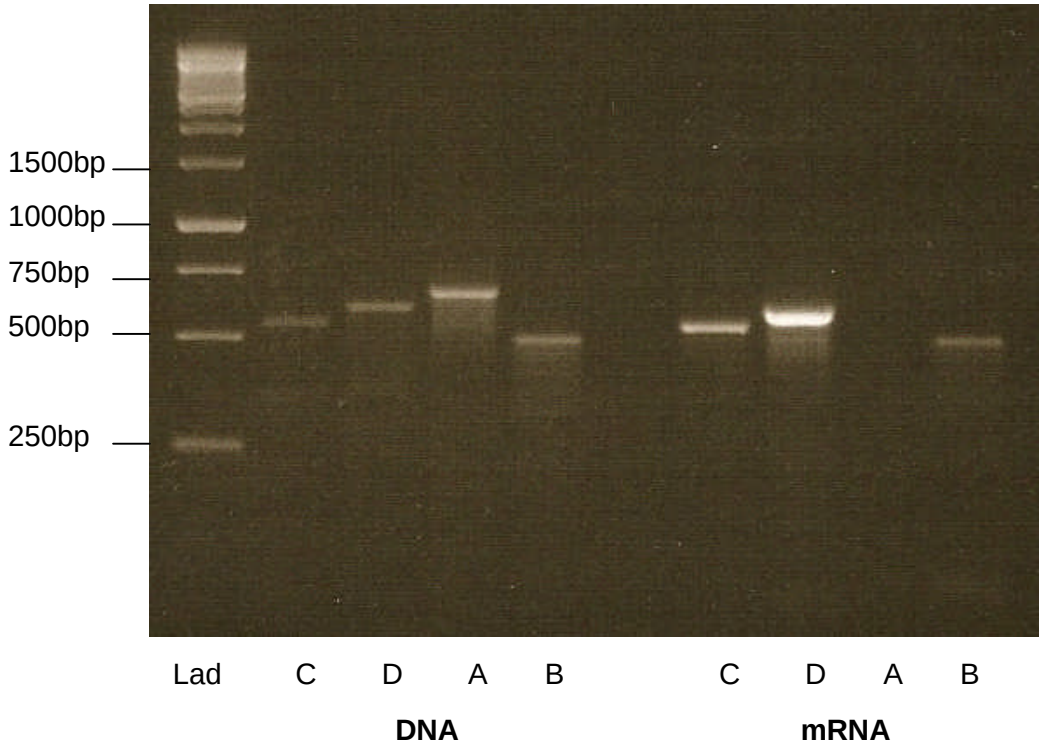


Figure 3.2 An agarose gel of the PCR products for the four regions of each gene in the *crtW* operon (Figure 3.1a) using a *M. xanthus* DK101 genomic DNA sample and a DK101 mRNA sample.

3.3. The effect of carbon starvation on *crtW* expression

As has been mentioned previously, there are a number of similarities between *crtW* and *crtI* in *M. xanthus,* including the fact that they both form separate loci in the genome distinct from the biosynthetic enzyme encoding genes found clustered together in the *crtEBDC* operon. For expression, *crtI* requires the cell to have undergone a period of carbon starvation in addition to the presence of light (Browning *et al.*, 2003). If there is similarity in the control of the two genes then *crtW* may also require carbon starvation to occur for optimal activation. As the vector pJRNS1 has already been created it is possible to test this hypothesis.

The four previously isolated *M. xanthus* strains, including pJRNS1, were cultured over a period of 48 hours in DCY until they reached the exponential growth phase. Each culture was then centrifuged and all excess media removed. The cells were then resuspended in minimal MC7 media, which was carbon and energy source free. The cultures were incubated with shaking at 30^{o}C for a three hour period, thus enabling the cells to undergo a period of starvation. They were then returned to fresh DCY medium and a LacZ assay carried out on each of the cultures. This time sampling was done at 0, 1, 3, 5 and 7 hour points, by which time in a similar *crtI* experiment the effect of carbon starvation was visible (Browning *et al.*, 2003). All LacZ assays were conducted similarly as when studying promoter activity ('LacZ assay' Methods) with samples grown in both the presence and absence of light.

Each of the strains displayed a similar pattern of results in the light and the dark (Figures 3.3 and 3.4). There is very little change seen in specific *lacZ* activity over the time period measured, and the initial readings were all relatively high. These data provide evidence that the promoter region is being constitutively expressed, yet the lack of a change in expression level over time rules out a positive effect caused as a result of carbon starvation. The starvation period instead led to the cells requiring a substantial period of recovery, during which levels of expression remain unaltered. The result confirms that the activation of the *crtW* operon is not reliant on carbon starvation or light regime, unlike *crtI*.

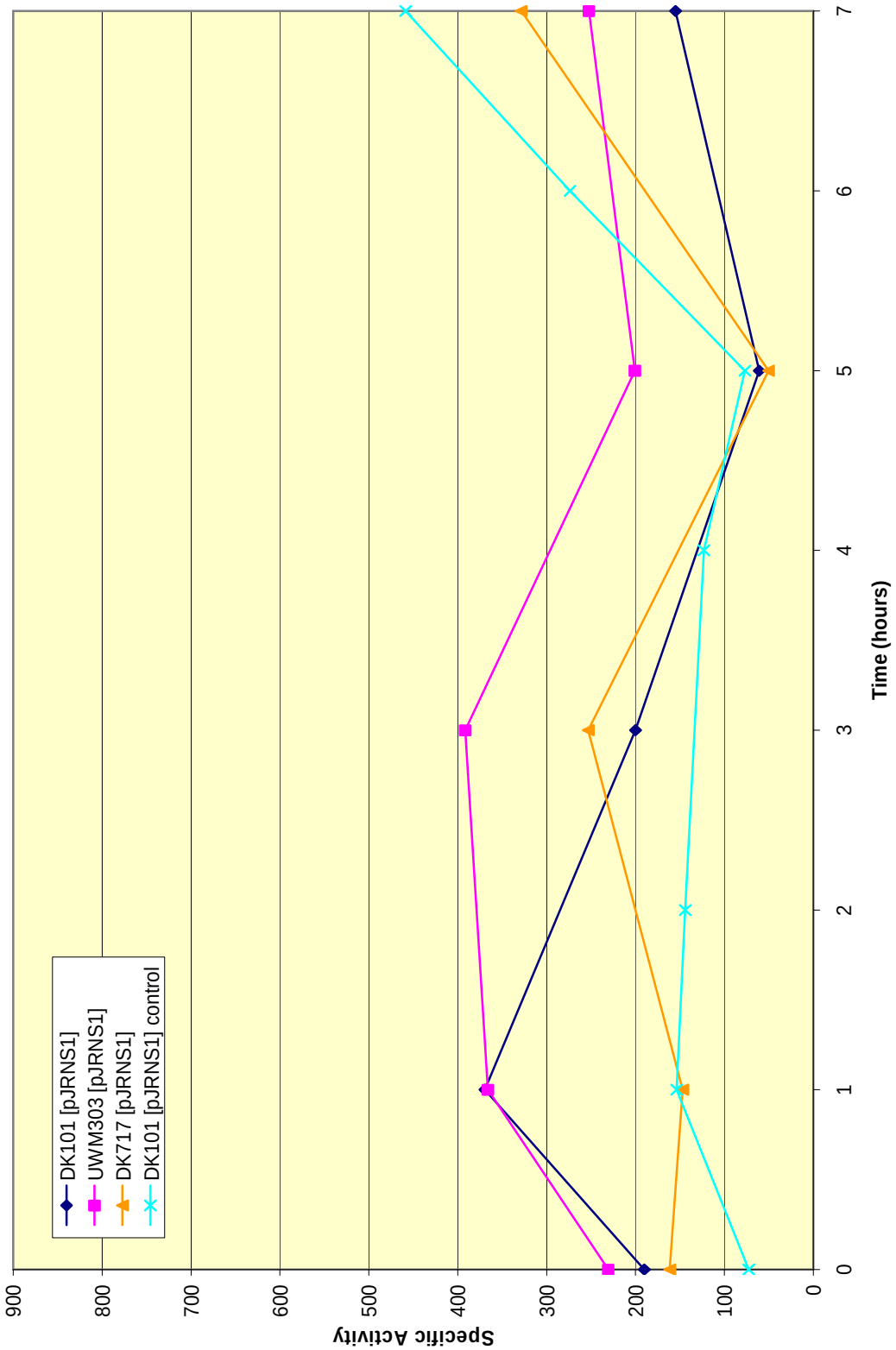


Figure 3.3 The specific *lacZ* activity against time (hours), expressed by three different *M. xanthus* bacteria strains grown in the absence of light under starvation conditions. A control strain that did not undergo starvation is also displayed.

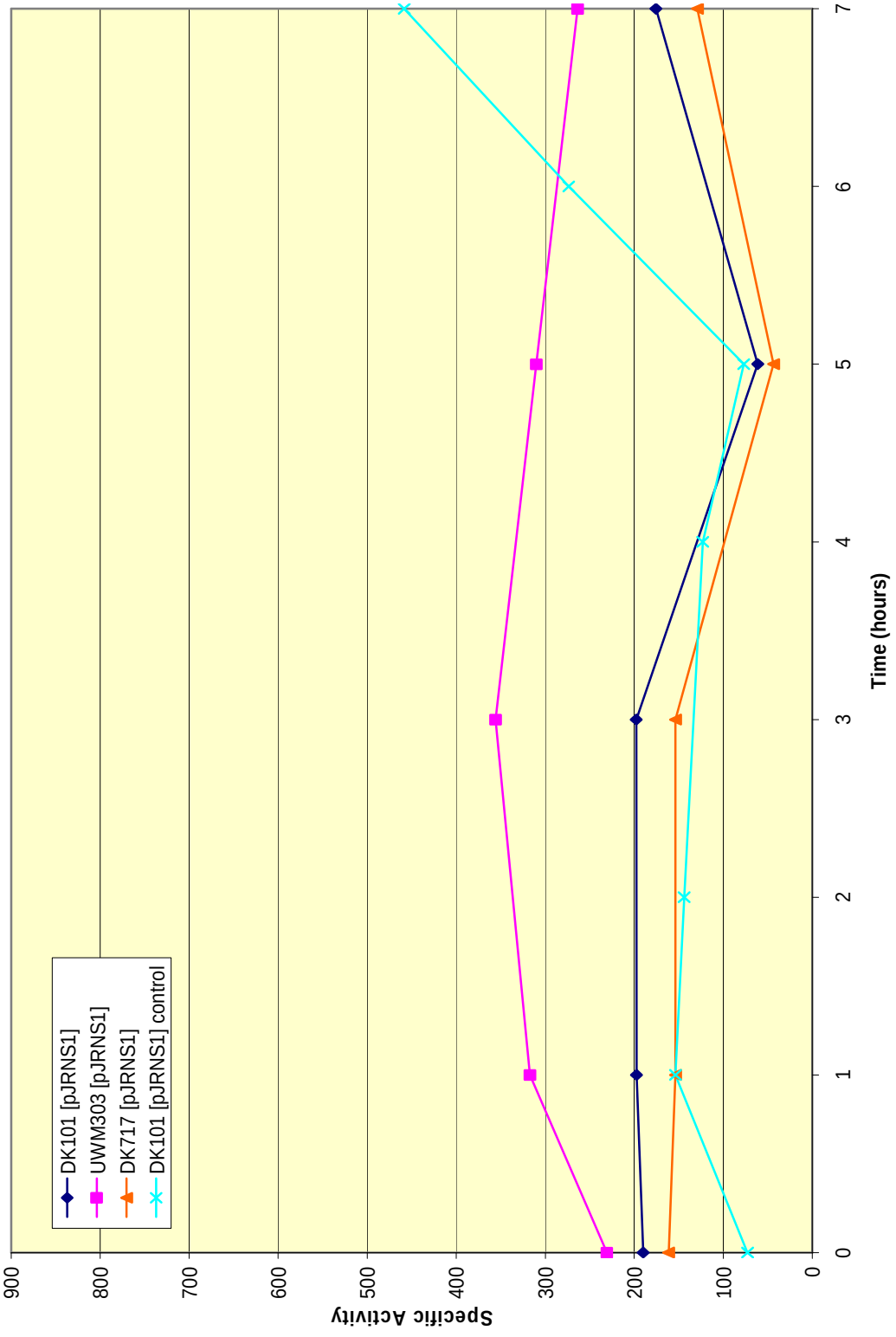


Figure 3.4 The specific *lacZ* activity against time (hours), expressed by three different *M. xanthus* bacteria strains grown in the presence of light under starvation conditions. A control strain that did not undergo starvation is also displayed.

Strain / Condition	DK101 [pJRNS1]	UWM303 [pJRNS1]	DK717 [pJRNS1]
Light to Dark	0.808035	0.208613	0.210406
Starvation to Replete	0.978655	0.303809	0.199269

Table 3.1 A table showing the comparative Student's t-test P-values between *M. xanthus* strains containing pJRNS1 when expressed under starvation conditions. The first row shows comparison between cultures grown in light and dark starvation conditions, and the second between starved and satiated cultures grown in the absence of light. Data for comparison in non-starvation conditions is taken directly from Chapter 2 (Figure 2.8)

Besides the graphical results, statistical analysis of the data also provides some insights (Table 3.1). A similar Student's t-test was carried out on the data as was utilised in Chapter 2. Looking firstly at the comparison between each individual strain and whether the starvation assay was carried out in either the light or the dark, there was no significant difference between any of the strains. Both UWM303 [pJRNS1] and DK717 [pJRNS2] had a P-value of approximately 0.2, which was higher than the majority of the P-values recorded during the preliminary Chapter 2 LacZ assays (Tables 2.2-2.4) . Under these conditions the most similar results were for DK101 [pJRNS1] with a value of 0.81. The more significant P-values were in the region of 1.0, which suggests that instead of displaying significant difference the data generated by the light and dark treated samples was indistinguishable. These results support the previous findings of the LacZ assays of the *pdp* promoter in Chapter 2, showing that there is no statistical difference between light or dark conditions in the level of *pdp* promoter expression.

As a second measurement, it is possible to directly compare the starvation results with those obtained from the *lacZ* promoter assay under replete conditions (Figure 2.8). In this instance, as no differences were evident between light or dark cultures, only the cultures grown in the absence of light were compared for the first seven hour period of each experiment. Once again the t-test results indicate no significant difference between the experimental data. The value for DK717 [pJRNS1] was again approximately 0.2, and the UWM303 [pJRNS1] strain showed even less significance with a value of 0.3. The final P-value of 0.98 from the comparison of the two conditions of the DK101 [pJRNS1] strain was once again more significant than any of the previous LacZ assay comparisons. This is the clearest evidence yet that there is no statistical difference between promoter expression in replete and starvation conditions. Starvation had no effect on the cells and neither did exposure to a light

source. The conclusion is that both light and starvation has no effect on the expression of the operon. Thus the *crtW* operon is unlike the *crtI* gene, *carQRS* operon or the *crtEDBC* operon.

3.4. Attempting to map *pdp* and *crtW* promoter regions through RACE

Having identified at least one promoter capable of controlling transcription within the *crtW* operon, it would be beneficial to determine an accurate location of this promoter and any other internal promoters. In order to determine exact locations of any promoter regions the preferred way would be to map them by identifying mRNA endpoints. In identifying mRNA endpoints it is possible to identify promoter start points or possibly, but less likely, mRNA processing points. There must be an mRNA endpoint situated upstream of the *pdp* start codon and any other endpoints within the operon could indicate internal promoters; such as upstream of *crtW*. To locate promoter regions, one of the most commonly used current methods for mapping DNA or RNA ends was adopted; RACE.

RACE (random amplification of cDNA ends) works using cDNA formed by the conversion of samples of mRNA, to generate PCR products which can be sequenced (Figure 3.5). In terms of mapping the *pdp* promoter region, the process worked by converting mRNA into cDNA as in the procedure adopted for operon detection. The sole difference was that in order to reduce the amount of non-specific cDNA strands created, a primer specific for the RNA sequence of interest was used instead of random hexamers. This led to the generation of gene specific cDNA. Terminal deoxynucleotidyl transferase (Tdt) was added to the cDNA along with dATP. The enzyme recognised the 5' end of the DNA sequence and added a poly dA tail. This in turn formed part of the specific marker for the RACE reaction. The 5' end sequence of the cDNA molecules was now known and as a result amplification of the entire region was carried out. To ensure maximum amplification efficiency, two 5' adapter primers were used in conjunction with a 3' primer with a sequence from within the ORF. One adapter primer (Adapter TTT) contained a unique sequence plus a sequence of seventeen thymines (Ts) that will hybridise to the poly-A tail of the cDNA. The second adapter primer (Adapter 2) contained the unique sequence without the seventeen dTs (See 'Primers' Methods). As a result the final amplified PCR product sequence should be made up of the adapter primer, followed by a series

of Ts, then the cDNA sequence up until the 3' ORF specific primer (Figure 3.5). It was then a case of sequencing the PCR product to map the mRNA endpoint.

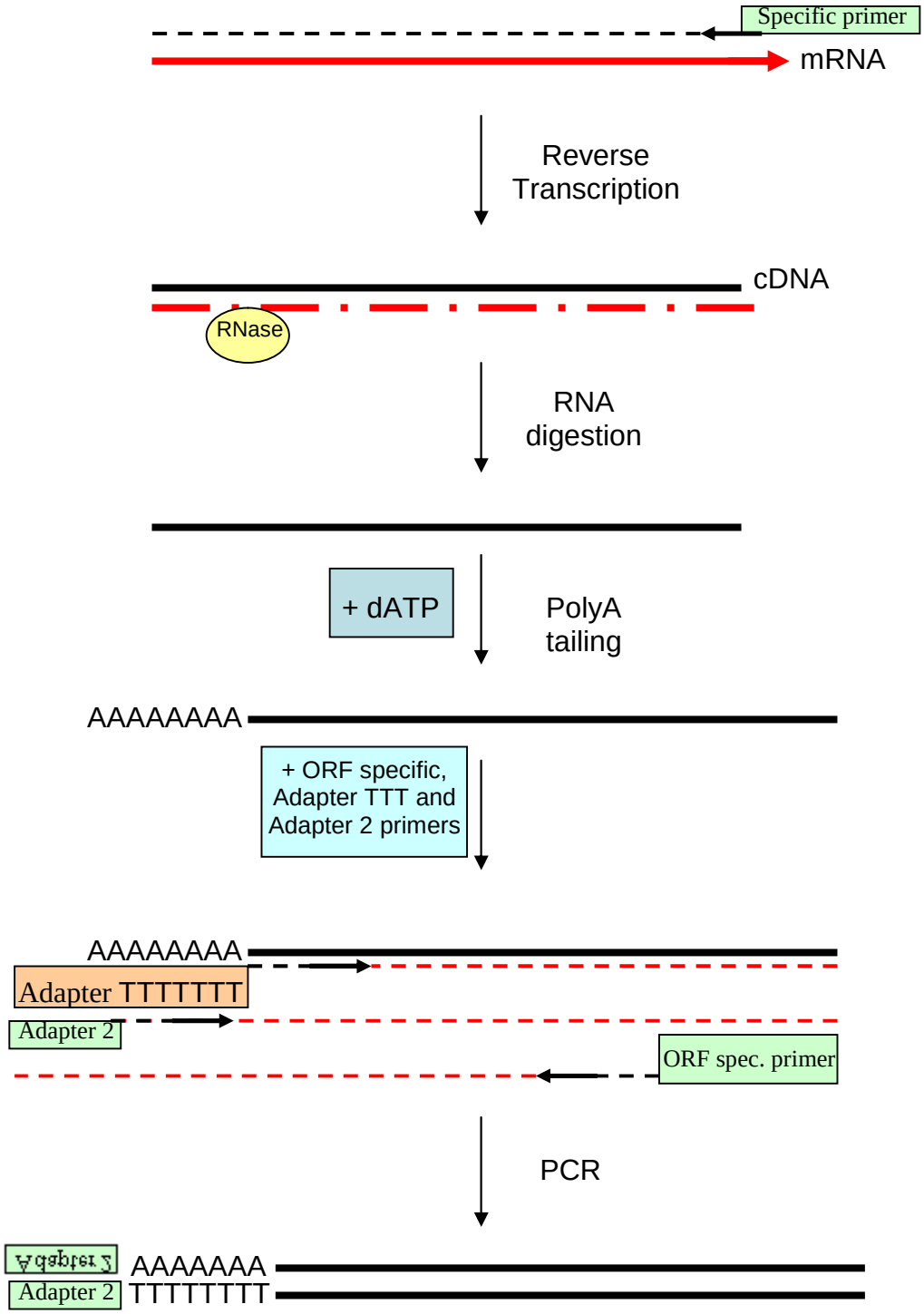


Figure 3.5 RACE analysis of mRNA endpoints (see accompanying text for a full description).

Initial attempts at the process ('RACE' Methods) led to the generation of either a large smeared gel band following gel electrophoresis, or no product at all. Obtaining a smeared band is not uncommon as during reverse transcription of mRNA to cDNA the enzyme may stop randomly before reaching the mRNA end. This would mean that a number of similar mRNA molecules would exist of differing lengths. In this instance the smeared band was excised and subjected to a further round of PCR, but still no clearer band pattern was produced. This proved the case regardless of a number of protocol adaptations; including the reduction and increase of initial dATP concentration, use of Adapter TTT without Adapter 2 during PCR, use of Adapter 2 alone in a separate step following initial Adapter TTT use, as well as altering the time and temperature for various stages of the PCR process. The only time a clear product was seen, it was either too smeared and produced no further bands when subjected to further PCR, or else was extremely faint and failed to generate a positive sequencing result (Figure 3.6). In order to identify the problem, and optimise the procedure in the process, a control gene was studied, *rpoD* (Inouye, 1990). This is a gene formerly known as *sigA*, that encodes the major housekeeping sigma factor of *M. xanthus*. Despite using an amplified region of DNA direct from the genome as a positive control – still no positive results were obtained. A final attempt was conducted using random hexamers instead of specific primers to initially convert RNA into cDNA. Although a large amount of random cDNA was created, it would eliminate any issues created through the use of the ORF specific primers.

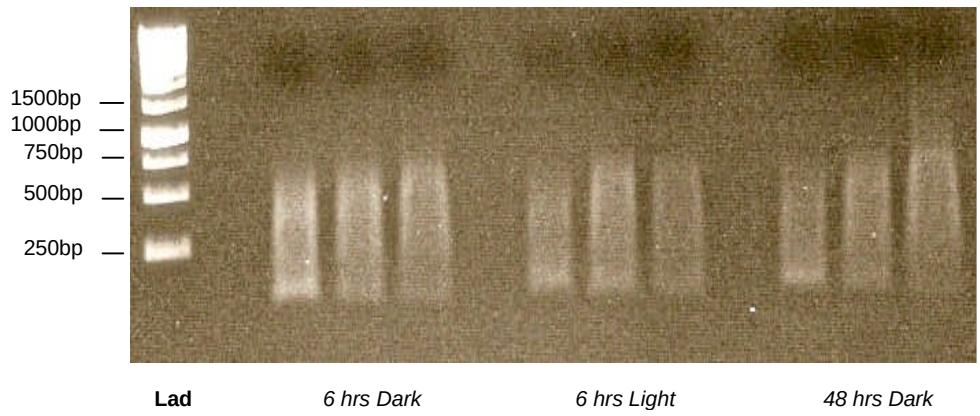


Figure 3.6 An agarose gel depicting three examples of the blurred bands visible following electrophoresis of the PCR products obtained when attempting to amplify cDNA samples. The cDNA was produced from mRNA extracted from *M. xanthus* DK101 growth cultures at the time and conditions labelled in italics.

This had a mixed effect upon the results, but led to a clear band being observed in the *rpoD* control reaction. Sequencing of the excised gel band confirmed that the *rpoD* gene had been amplified. Following this result the procedure was repeated for the upstream *pdp* and potential internal promoter regions. The first attempt using random hexamers yielded a pair of bands from PCR. However sequencing of the two bands showed that two different genes were amplified, one encoding a LysM domain protein and the other a peptidase from the protease Do (DegP) family. As the cDNA used was generated from RNA isolated from a DK101 strain, the result does show that the two genes are expressed constitutively during early exponential growth. No products were obtained from the *crtW* operon. Following a number of additional attempts and protocol amendments no other bands were obtained and due to time constraints this aspect of the investigation was concluded.

It is unclear why the protocol proved to be so unsuccessful. It could have been that the reverse transcriptase had a problem transcribing *crtW* operon mRNA. This may have been the result of insufficient primers or the poor quality of the RNA being used. Although all primers were designed to recognise close to the 5' end of the mRNA during cDNA construction (within a maximum of 150bp of the promoter region), it may be that the distance was still too great for RACE to be effective.

3.5. Quantification of *crtW* operon expression

At this stage expression of the *crtW* operon had been observed using the promoter probes pJRNS1 and pJRNS2, but the amounts of mRNA produced had not been quantified. One of the most effective ways of doing this quantifiably is through the use of QRT-PCR.

QRT-PCR (quantitative real time polymerase chain reaction) is a method of measuring the amount of transcription product that is created at each stage of a reaction. It directly measures the amount of mRNA present in the cell at any given time, and has already been used in conjunction with some myxobacteria species (Kegler *et al*., 2006). In the previous experiments, mRNA had not been quantified, but solely isolated for qualitative testing. During QRT-PCR a sample of mRNA or cDNA is used in a PCR, similar to a number of experiments already carried out in this investigation. The fundamental difference is that in this case after each PCR cycle the

amount of product is measured. Effectively the sample is being quantified as it is amplified, so rather than using PCR to obtain a product; the stages during creation provide information as to quantity. By comparing an unknown amount of amplified sample to a sample of known amount, it is then possible to determine the activity of each gene of interest at a particular point in a given sample.

For this investigation a number of RNA samples were isolated from a *M. xanthus* DK101 strain. Samples were taken from an overnight culture, transferred to fresh media to initiate a period of timed growth, in both light and dark conditions. Samples were extracted from the culture at 0, 2, 6 and 48 hour time points, covering the main stages at which carotenogenesis activity levels varied. Comparison of the samples should indicate whether all the genes in the operon are transcribed at the same level from the *pdp* promoter or if a secondary *crtW* promoter could be present generating higher levels of *crtW*. For consistent comparison each gene encoded by the operon was studied and primer pairs designed for each individual gene (Figure 3.1b and 'Primers' Methods). A specific primer designing program, DNA Star 5, was used to generate primers for use during the QRT-PCR reaction, as the functional requirements differed from standard PCR primers. In addition, a good control gene was required for testing; one known to be expressed constantly during cell growth and at a relatively constant level, a housekeeping gene. In this case *rpoD*, previously used as a control for the RACE experiments was selected and again an appropriate pair of primers designed for its amplification (Figure 3.1b and 'Primers' Methods).

For each gene and the control, triplicate technical samples at each time point were generated and studied ('QRT-PCR' Methods). The RNA was converted into cDNA, using random hexamers, for each time-point to ensure the primers functioned correctly. Initially for each gene a standard curve was generated, using a known amount of *M. xanthus* genomic DNA in a series of dilutions that was amplified using each primer pair. Negative controls, containing all reagents with the exception of a DNA source, were also run to confirm each sample was free of contamination. This proved invaluable following the first attempt at QRT-PCR, when the supposed nuclease-free water sample used was found to be contaminated. This problem was easily resolved through the use of a fresh batch of Qiagen nuclease-free water.

Following initial technical problems, data was obtained for all of the genes, using a range of 96-well plates ('QRT-PCR' Methods). Data were analysed using a relatively new method of real-time data comparison, that of Pfaffl (Pfaffl, 2001). This

was designed to provide an alternative to the standard SYBR Green comparative Δct analytical approach (Morrison *et al*., 1998). It worked on the basis of a calibration curve no longer being required for each reaction and instead relied on high primer specificity during the QRT-PCR reactions. In terms of this investigation each gene expression level for both light and dark samples was compared to a DNA dilution control, using a sample of *M. xanthus* chromosomal DNA. All results were subsequently standardised using the constitutively expressed *rpoD* housekeeping gene as a control. It was then possible to compare the growth in light and dark conditions within each sample and the control, thus identifying additional evidence for varied gene expression in these altered growth conditions.

3.6. Quantification of *crtW* operon genes in the absence and presence of light

Although results from the QRT-PCR were gained for each of the genes in the putative *crtW* operon, the ones for the fourth gene, *mmb*, could not be effectively analysed. This was primarily due to the fact that the dilution series for this gene failed to work. This was despite a number of attempts, and so as a result there was no suitable control reaction to compare to data obtained for each of the gene sample time-points. Initially comparison was made between each gene and its activity under both dark and light conditions using the Pfaffl calculation. The data for this comparison are shown in Figure 3.7, where the ratio is determined by comparing each sampled strain with the results from the dilution curve, and then standardising the result using data obtained for the control gene. Before progressing further it is worth recalling that the dilution series was made using *M. xanthus* DK101 chromosomal DNA, so the ratios produced a relative rather than an absolute measure of how active a gene was at a given time-point.

The first thing to note is the obvious variation of the gene levels in the samples. This was unexpected if the three ORFs were transcribed from a single promoter upstream of *pdp*. We would expect similar expression of *pdp*, *crtW* and *ord* or a decline from *pdp* to *ord*.

Firstly comparing light and dark for *pdp*, at time points 2, 6 and 48 hours there appears to be slightly more mRNA from dark-grown cells than light-grown cells. This is reversed at the 0 hour time point. In the case of *crtW* for time points 0, 2 and 48 hours there appears to be more mRNA from dark-grown cells than light-grown

cells. This is reversed at the 6 hour time point. In the case of *ord*, for the time points 0 and 48 hours the amount of mRNA seems about the same for light and dark, for the 6 hour time point there appears to be slightly more in the light and the 2 hour time point slightly more in the dark. I conclude, consistent with the promoter probe studies if anything there is a slight decrease in expression levels of the *pdp, crtW* and *ord* genes in the light. The *crtW* operon is definitely not light induced.

Comparing expression levels between the open reading frames, for the 0 hour time point and the 48 hour time point, independent of the light condition, there is consistently greater expression of *crtW* and *ord* than *pdp*. For time point of 2 hours, again there is consistently greater expression of *crtW* and *ord* than *pdp* in the dark. In the light expression of *crtW* is anomalous but there is greater expression of *ord* than *pdp*. For the time point 6 hours there is greater expression of *ord* than *pdp* in the light and the dark. Anomalously, there is less expression of *crtW* than *pdp* in the light and the dark. With these two exceptions there is greater expression of *crtW* and *ord* than *pdp*. This is the first evidence that there may be an internal promoter in the *crtW* operon situated upstream of *crtW*. The possibility does still exist however that the observed expression levels are still a result of the promoter upstream of *pdp*. *crtW* and *ord* expression may just be at a higher level compared to the transcribed *pdp*.

Given the apparent similarities between the light and dark cultures for each gene the two conditions can be directly compared using Pfaffl's method, whereby the dark culture is classified as the untreated sample and exposure to light is the cell treatment. By comparing the results graphically it may be possible to determine how similar gene expression is under each of the afore-mentioned conditions (Figure 3.8). If expression is the same, the ratio would be expected to be around 1.00 for all of the samples. Yet each displays some level of deviation, with the *ord* gene displaying most similarity consistently within 25% of equal ratio at each time point. *pdp* shows differences at the 2 hour and 48 hour time points of near 50%, despite displaying a 1:1 ratio at the other two points. *crtW* shows low similarity at the same two points, with ratios of 0.23 and 0.29 respectively.

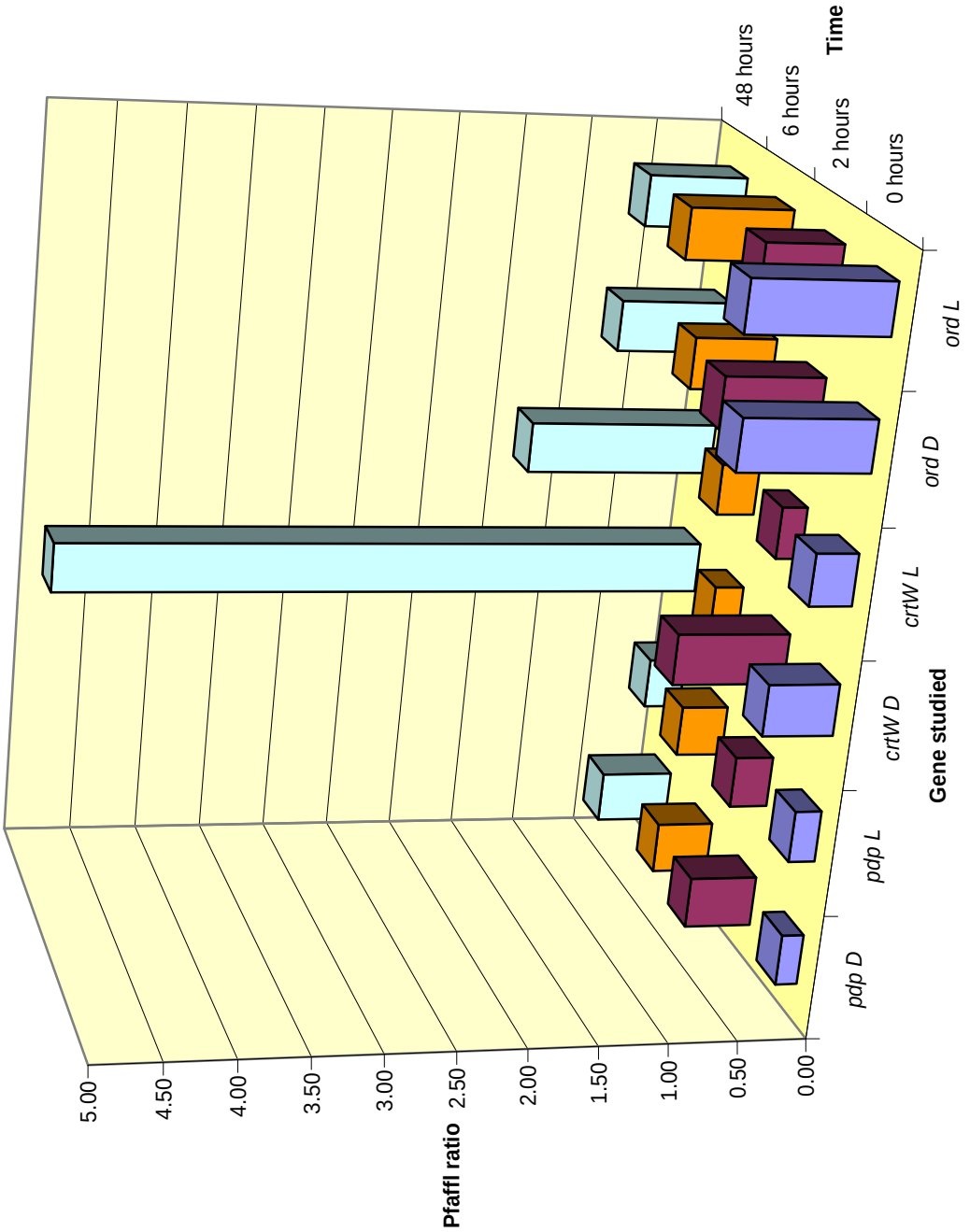


Figure 3.7 The Pfaffl QRT-PCR ratio of three genes expressed in a *M. xanthus* DK101 strain grown in the presence (L) and absence (D) of light.

Figure 3.8 The ratios of dark-grown to light-grown expression of *pdp*, *crtW* and *ord* as measured by Pfaffl QRT-PCR ratios.

Condition/gene \ Gene	***pdp***	***crtW***	***ord***
Between light and dark	0.3110	0.1514	0.2843
pdp	~	0.1732	0.5021
crtW	~	~	0.0765

Table 3.2 A table showing the comparative Student's t-test P-values between *M. xanthus* DK101 genes for QRT-PCR ratios obtained from DK101 strains grown in light or dark conditions. The first row shows comparison between cultures in the light and dark, and the following two rows are direct comparisons between the genes following a light versus dark Pfaffl ratio calculation.

To quantify observable differences, a Student's t-test statistical analysis was carried out on all of the samples studied in the same manner as before (Table 3.2). Firstly the ratios of the light and dark results for each gene were statistically tested. Despite apparent initial observable differences, particularly in the case of the 48-hour sample of *crtW*, none of the collected ratios proved to be statistically different overall within each sample. As a result it has been shown once again that there is little significant difference in the gene expression whether the culture is grown in either the presence or absence of light. The P-values also suggest two of the samples that appeared most similar from the graphical data (Figure 3.7), *pdp* and *ord*, are also statistically the most similarly expressed genes.

As a final t-test comparison the results from the dark and light ratio for each gene were compared to each of the others. Once more there was no significant statistical difference between any of the recorded values. The strongest statistical similarity was observed between *pdp* and *ord*. Although there was a difference recorded between *crtW* and the other genes at levels of 0.08 and 0.17, these were also not statistically significant. As a result the conclusion has to be made that there is no significant difference between the expression of any of the genes whether or not they are exposed to light during growth or sampled at any given tested time point.

3.7. Discussion

It is clear from the analysis of mRNA generated in the *M. xanthus* DK101 strain that an operon does exist containing *crtW*. The evidence indicates that as hypothesised it incorporates four genes; *pdp*, *crtW*, *ord* and *mmb*.

There was the possibility that *crtW* required a similar activation step as *crtI*, which needs the cell to have undergone a period of carbon starvation before it is expressed at significant levels. This was found to not be the case with *crtW* as

expression levels did not increase following starvation. If any difference was observed then it was that expression was actually fractionally reduced. Therefore *crtW* is controlled in a completely different way to *crtI*, in that it is not CarQ-dependent, subject to catabolite repression or light-inducible.

The quantitative analysis of mRNA production from the *crtW* operon ORFs revealed, consistent with the promoter probe studies (Chapter 2), that the operon was not light-induced and if anything slightly light-repressed. Comparing the quantities of the transcripts of the three ORFs tested; *pdp*, *crtW* and *ord*, there was consistently larger amounts of the transcripts of *crtW* and *ord* than that of *pdp*, independent of light regime, with a few anomalous exceptions. Transcript production of *crtW* was usually greater than *ord*, again independent of the light regime. These observations are consistent with the presence of an internal promoter within the *crtW* operon, upstream of *crtW*. The results however do not positively confirm an additional promoter presence.

I attempted to map the start sites of the *crtW* operon promoter upstream of *pdp* and the internal promoter upstream of *crtW* using RACE, but this was not successful. I was able to map other gene promoters using RACE or modifications of the standard protocol, but not the *crtW* operon promoters. This could be due to the low abundance of the *crtW* operon transcripts or due to difficulties with reverse transcribing the polycistronic mRNA. Alternative ways of mapping the transcription start sites could be primer extension or S1 nuclease analysis.

Chapter 4

Bioinformatics

4.1. Sequence analysis of the *crtW* operon

4.1.1. Introduction

Often the study of gene sequence data is a complex process that requires the bringing together of a large amount of knowledge. Studies of gene sequences and their related protein products have revealed a wealth of information, particularly during the biological boom of the last thirty years. During this time, methods of working with DNA have evolved dramatically; from the amplification of DNA through the PCR to the speed and accuracy of DNA sequencing. The continual development of techniques has led to an increased amount of genetic study and a surge in the practice of genome sequencing. Many suggest the pinnacle of such work was the eventual complete sequencing of the human genome (Venter *et al.*, 2001), but this is just a small example of what is currently possible. The overall cost of sequencing continues to fall, and the speed of the process increases, leading to the sequencing of more genomes and individual genes. It is now possible to analyse entire sequences of organisms ranging from bacteria like *E. coli* K-12 (Blattner *et al.*, 1997) and *Sorangium cellulosum* (Schneiker *et al.*, 2007), up to more complex plant species such as rice (Goff *et al.*, 2002) and possibly even a species of extinct mega fauna, the woolly mammoth (Miller *et al.*, 2008). Each newly produced sequence generates yet more data, further scientific interest and as a result an increased demand for access to this information. To cope with this, a number of genetic databases have been created worldwide in which the majority of sequence information generated is stored. A number of these are accessible via the internet enabling the rapid sharing of genetic information. The availability of data has led to the evolution of a new field of science, bioinformatics.

Bioinformatics is the use of biological information to compare or model various gene sequences and cell biological behaviour, usually through the use of a range of computer programs. Rather than conducting a large amount of work at the

practical level of biology, bioinformatics often uses data that have already been created and recorded to make predictions on the behaviour and activity of other similar systems. It is essentially a form of data analysis that utilises large sets of information of which an example is study through the use of the NCBI BLAST system. This enables access to the large NCBI database of DNA sequences, the majority of which have been studied in detail and then submitted to a central host. The database contains information on recorded genes, linked protein identities, phylogenetic relationships between species and the raw genetic sequence data. By comparing an unknown sequence with the database of known sequences, the origin of the genetic information studied can be determined. If previous analysis of a similar DNA sequence identified associated genes and also a function for those genes, then this will also be recorded on the database. Hence it is possible to identify a novel DNA sequence from experimental investigation and then through the use of the BLAST software establish what that sequence could possibly encode, the organism from whence it originated and any possible associated protein with the sequence. This is arguably the simplest and most common use of bioinformatics, but highlights how powerful a tool it can be in the study of novel genes within organisms. In this instance bioinformatics is primarily a prediction tool, whereby any data generated will still require practical confirmation

In terms of the study of *crtW* from *M. xanthus,* it was through bioinformatics that the potential *crtW* gene was first identified, based on its similarity with known *crtW* sequences in the NCBI database. Identities of the neighbouring three genes were also established, along with the possibility of the genes being located within an operon. However, despite initial success of such predictions, over time the database has dramatically increased in size and the information it contains has become much more detailed. As a result it was decided to repeat the preliminary bioinformatic work on the gene region in an attempt to increase our predictive power.

4.1.2. Putative identification of *crtW* operon gene function

To verify gene identity within the operon all four regions were submitted to fresh NCBI BLAST searches. Prior to searching, each of them was first translated into the equivalent amino acid sequence (Figure 4.1). The NCBI BLAST results are displayed in Figure 4.2. The ‘bit’ score is an alignment value and any value over 100

is considered significant, so anything over 400 would indicate a very strong relatedness between sequences. The e-value (expected value) is a measure of similarity of the sequences being compared together. For example a value of 0.0 would indicate that the sequences were identical and most likely originated from the same organism. The closer the value is to zero, the more similar the sequences compared, and the less likelihood that the similarity is a result of chance alone. The top six results for each amino acid chain are included in the table for analysis.

1st gene (*pdp*)

MLSGAATAAEKTPPARFPDAAELQRLTARFAPVELRVDLKALPESERRALARIVQASKLMDTL
FLRQRWAGNEPLLLDLVQDTTPLGRARLQAFLLDKGPWNSLDEARPFIPGVPAKPASANFYPA
GATQAEVEAWVKSLPEAQQKEATGFYTTIRRGTDGRFITVPYSVEYQGELALAAALLREAAA
LTQQPTLKAFLTSRADAFLSNDYYASEVAWMELDASIEPTIGPYEVYEDEWFNYKAAFEAFVG
LRDDAETQKLAKFSGQLQGLENNLPIDPKLRNPKLGALAPIRVINSLFSSGDGNRGVQTAAFNL
PNDERVSEKMGSKRVMLKNVQEAKFERVLLPIAKVALTPADQKDVSFDAFFTHILMHELMHG
LGPSNITVGGKATTVRKELQSASSAIEEAKADISGLWALQRLVDTGVIDKSLERTMYTTFLASA
FRSIRFGVDEAHGKGIAVQLNYFLDTGAVKVNADGTFSVVPAKMKKAVISLTKQLMEIQGRG
DRKAAEALLAKLGVVRPPVQRVLERLKDVPVDIEPRYVTAEELVRDVKK

2nd gene (*crtW*)

METSARQLRPAPPGPWGVVIALIIMGAWGGHLAWALTRAELPWVEPLTWLHVALQAWLCTG
LFITGHDAMHGTVSGRRWVNEAVGTVACFLFAGLSYRRLVVNHRAHHARPTSDADPDFSTHS
QSFWPWLGTFMARYTTLPQLGVMAAKFNVLLFLGVSQPHILGYWVLPSVLGTLQLFYFGTYL
PHRRPETPDMAPHHARTLPRNHLWALLSCFFFGYHWEHHESPGTPWWRLWRLKDARAREAA
LTQSTGTLPGQEGTAR

3rd gene (*ord*)

MRDKPPAEPPSSEVTPEKTYLRRRELLKNAGLFAGTAVAVAGGLHLLGRKQTRPMERFVPDA
GLVEQPVAQAMGPFDTDEPRTPYEDVTTYNNFYEFGFDKNDPARFAHTLKPKPWSVVIDGEV
HKPRTVDVEQLTSWFSLEERVYRMRCVEAWSMVIPWLGFPLAALLQRVEPTSHAKYVAFTTL
LDPEQMPGQRRALLDWPYTEGLRLDEAMNPLTLLATGLYGRQLPNQNGAPLRLVAPWKYGF
KGIKSIVRISLTREEPMTTWRLSAPREYGFYANVNPSVPHPRWSQASERRIGDFERRPTLPFNGY
AEQVAHLYTGMDLRRFY

4th gene (*mmb*)

MASSPYPWLNPALVVGGLSPLLMLAVQGPRGELGPNAIEAALHQTGLLTLVLLVASLTCTPLR
LVAGWTWPARVRRTLGLLAFTYAVAHFLVYAVLDQGLAWGALWADVTERPFITVGFAALV
LLVPLAVTSTNRWVRRLGFPRWQRLHRLAYGAAALGVVHFVWRVKKDVTEPLIYGAVLALL
MAIRVGEAMRKRRARAAAAARNPA

Figure 4.1 The amino acid sequence for each gene encoded by the *crtW* operon

Reference	Role/Function	Organism	BIT	e-value
1st gene (*pdp*) – Encoding a putative dipeptidyl peptidase in *M. xanthus*				
ZP_01463495.1	Hypothetical protein STIAU_8155	*Stigmatella aurantiaca* DW4/3-1	798	0.0
YP_001377726.1	MutT/NUDIX family protein	*Anaeromyxobacter* sp. FW109-5	665	0.0
YP_002490934.1	MutT/NUDIX family protein	*Anaeromyxobacter dehalogenans*	627	5e-178
YP_593300.1	MutT/NUDIX family protein	*Acidobacteria bacterium* Ellin 345	575	2e-162
YP_822815.1	MutT/NUDIX family protein	*Solibacter usitatus* Ellin 6076	556	9e-154
YP_565218.1	ATNUDT3 – NUDIX hydrolase	*Arabidopsis thaliana*	453	1e-125
2nd gene (*crtW*) – Encoding a carotenoid ketolase in *M. xanthus*				
ABB88952.1	CrtW	*Algoriphagus* sp. KK10202C	203	8e-51
ZP_01718386.1	Fatty acid desaturase	*Algoriphagus* sp. PR1	198	3e-49
YP_001869154.1	COG3239: Fatty acid desaturase	*Nostoc punctiforme* PCC 73102	160	5e-38
YP_322565.1	Fatty acid desaturase	*Anabaena variabilis* ATCC 29413	155	3e-36
YP_457553.1	β-carotene ketolase	*Erythrobacter litoralis* HTCC 2594	149	2e-34
NP_487229.1	β-carotene ketolase	*Nostoc* sp. PCC7120	146	1e-33
3rd gene (*ord*) – Encoding an oxidoreductase in *M. xanthus*				
ZP_01463488.1	Twin-arginine translocation signal	*Stigmatella aurantiaca* DW4/3-1	462	1e-128
YP_604348.1	Oxidoreductase, molybdopterin binding	*Deinococcus geothermalis*	402	1e-110
YP_001617206.1	Putative molybdopterin-binding protein	*Sorangium cellulosum*	367	5e-100
NP_296256.1	Putative sulfite oxidase subunit	*Deinococcus radiodurans* R1	362	1e-98
NP_900459.1	Hypothetical protein CV_0789	*Chromobacterium violaceum*	362	2e-98
YP_467178.1	Conserved hypothetical protein	*Anaeromyxobacter dehalogenans*	361	3e-98
4th gene (*mmb*) – Encoding a transmembrane protein in *M. xanthus*				
ZP_01463481.1	Permease of major facilitator superfamily	*Stigmatella aurantiaca* DW4/3-1	192	9e-48
NP_296257.1	Hypothetical protein DR_2537	*Deinococcus radiodurans* R1	150	5e-35
YP_590085.1	Ferric reductase transmembrane subunit	*Acidobacteria bacterium* Ellin 345	149	1e-34
YP_425685.1	Hypothetical protein Rru_A0594	*Rhodospirillum rubrum* ATC 11170	148	2e-34
YP_001101360.1	Hypothetical protein HEAR3130	*Herminiimonas arsenicoxydans*	148	2e-34
YP_113339.1	Hypothetical protein MCA2500	*Methylococcus capsulatus*	146	7e-34

Figure 4.2 The top six NCBI BLAST matches for the products of the four genes in the *crtW* operon. The match for *M. xanthus* has been excluded in each case. The reference number refers to the accession number on the NCBI database. All BIT scores and e-values are calculated using the default NCBI BLAST settings.

4.1.2.1. *crtW*

Looking at *crtW*, the primary gene of interest to this thesis, all listed proteins show great similarity with the *M. xanthus* sequence. Each protein encodes either a fatty acid desaturase or β-carotene ketolase gene, with the most similar being the *crtW* gene of an *Algoriphagus* species. Prior to this study *crtW* had not been positively identified in *M. xanthus*, hence it is not annotated as such in the database and current annotation is as a fatty acid desaturase family protein. All ketolases are members of this family, indicating the three other desaturase protein matches could also

potentially encode ketolases. All of the species recorded are bacterial carotenoid producers. With the sole exception of *M. xanthus*, all of the species are aquatic bacteria. Using the same NCBI BLAST search protocol, the next ten strongest matches (data not included) are proteins that have been identified in cyanobacteria or similar water-based species. The fact that they each create carotenoids, and the stronger matches are possible fatty acid desaturases or ketolases, increases the probability that the *M. xanthus* protein is CrtW, and that this is involved in carotenogenesis.

4.1.2.2. *pdp*

The first gene of the operon, the putative dipeptidyl peptidase (*pdp*), displays greater alignment with matching sequences than *crtW*, with the lowest 'bit score' still over 450 in value. The strongest match is with the protein from *Stigmatella aurantiacum*. This is a species very similar to *M. xanthus* that contains a large number of highly conserved sequences with *Myxococcus* (Rice and Lampson, 1995). It is currently under debate as to whether this similar soil-based spore producer should actually be reclassified as a *Myxococcus* sp. The following five matches all suggest MutT/Nudix proteins. Members of this family are hydrolases and contain a specific region known as the Nudix box (Bessman *et al*., 1996). This enables the hydrolysis of various diphosphates, although the specific action of each hydrolase does vary between species. The strong similarity suggests the previously annotated *pdp* gene may in fact be a MutT/Nudix family protein involved in the breakdown of specific molecule/s in myxobacteria. In contrast to *crtW*, all of the bacterial species with similar gene sequences are found in soil. For instance *Anaeromyxobacter* are, like *M. xanthus*, myxobacteria that grow in anaerobic conditions. All identified species display a form of environmental resistance, through either spore formation or the ability to adapt to extreme conditions, such as the *Acidobacteria*. The final strongest match in this comparison is that of the extensively studied thale cress, *Arabidopsis thaliana*. Although not recorded here, the next three strongest matches are also plants i.e. wheat and rice. Again a similar MutT/Nudix protein was found in each case.

4.1.2.3. *ord*

The third gene of the cluster encodes a possible oxidoreductase, Ord. As before with the *pdp* gene, all of the matches display strong similarity. Two of the matches do encode an oxidoreductase protein, but significantly they are both labelled as putative or hypothetical. The other sequence matches do not suggest anything beyond a general function for Ord. The matches that were found encompass a wide range of bacterial species that include: three Myxobacteria; *Deinococcus* species, organisms highly resistant to desiccation and radiation; and *Chromobacterium violaceum*, a common soil-based and aquatic facultative anaerobe.

The twin-arginine translocation system, or TAT, is a mechanism by which a bacterium can export fully folded proteins, including any co-factor, out of the cell. Thus the bacterium can insert the co-factor in the cytoplasm and then export the full active protein. The TAT system recognises an N-terminal signal consensus sequence: SRRXF(L)LK, including twin arginine residues (hence the name) in the protein to be exported (Stanley *et al*., 2000). Therefore the *Stigmatella aurantiaca* protein ZP_01463488.1 is a putative substrate for the TAT system and if you examine Ord in Figure 4.1 a similar TAT sequence is also present (Figure 4.3). As a result we can tentatively conclude Ord is also exported by the Tat system.

ZP_01463488.1

MSDKKAPQPPESEVTPETLYRRRREFIKNAGLFLGTAAAVGGGLYLLGPKRTRPGETPPPPEAG
GKAGPAEPPADEAPIAQKSLGPYDTDEAQTPYDDVTSYNNFYEFGLDKGDPVQNAHTLKTRP
WAVVVDGEVNKPQTVDIDQLLAWFPPEERIYRMRCVEAWSMVIPWLGIPLGELIKRVEPTSRA
KYVAFTTLKDPEQMPGQKRRVLDWPYVEGLRLDEAVHPLTLLATGLYGKALPPQNGAPLRL
VVPWKYGFKGIKSIVRITLTERQPPTTWNLAASNEYGFYANVNPAVDHPRWSQATERRIGEFR
RRPTLPFNGYAEQVASLYTGMDLRENY

Ord

MRDKPPAEPPSSEVTPEKTYLRRRELLKNAGLFAGTAVAVAGGLHLLGRKQTRPMERFVPDA
GLVEQPVAQAMGPFDTDEPRTPYEDVTTYNNFYEFGFDKNDPARFAHTLKPKPWSVVIDGEV
HKPRTVDVEQLTSWFSLEERVYRMRCVEAWSMVIPWLGFPLAALLQRVEPTSHAKYVAFTTL
LDPEQMPGQRRALLDWPYTEGLRLDEAMNPLTLLATGLYGRQLPNQNGAPLRLVAPWKYGF
KGIKSIVRISLTREEPMTTWRLSAPREYGFYANVNPSVPHPRWSQASERRIGDFERRPTLPFNGY
AEQVAHLYTGMDLRRFY

Figure 4.3 The amino acid sequences of *S. aurantiaca* ZP_01463488.1 and *M. xanthus* Ord with the TAT consensus sequence highlighted in red and underlined

From the *M. xanthus* genome annotation, Ord was annotated as having a molybdopterin co-factor. The molybdopterins are a series of cofactors, made up of pyranopterin and a pairing of thiolate groups (Rajagopalan and Johnson, 1992). Most of the enzymes that have molybdopterin co-factors are either reductases or oxidases. In Figure 4.2 the *Deinococcus geothermalis* protein and the *Sorangium cellulosum* proteins are also annotated as molybdopterin co-factor binding. From this we can tentatively conclude that Ord is an oxidoreductase that contains a molybdopterin co-factor, which is inserted into the apoprotein before being exported from the cell in a fully folded state by the TAT system.

4.1.2.4. *mmb*

The fourth and final gene of the operon, *mmb*, annotated as encoding a trans-membrane protein, also displays great sequence similarity with proteins from a range of other bacterial species. Once more a protein from *Deinococcus radiodurans* is one of the stronger matches, but also proteins from species such as *Methylococcus*, a methanotroph, and *Rhodospirillum* sp., a purple, photosynthetic freshwater species. The strength of sequence matches is reduced compared to that previously recorded for the *pdp* and *ord* genes of the cluster, although all the bit score values are above 100, so similarity is still very significant. The five most similar sequences do not enable the assignment of a role for the gene product. Only two of the matches suggest possible roles for the protein, and both are hypothetical. The strongest match is to a proposed permease and the second strongest is to a ferric-reductase. Of the next five matches (not shown in this study) a further three are classified as putative ferric-reductases. Given the similarly observed high e-values of all matches, it is possible that the gene does encode a metal ion reductase.

4.1.2.5. Summary

Looking at the information obtained from the preliminary NCBI BLAST search, it is evident that despite tentative labelling of the *M. xanthus crtW* operon actual gene product roles are still unclear. The *crtW* gene appears to encode a fatty acid desaturase, most likely a β-ketolase. *ord* encodes an exported molybdopterin-binding oxidoreductase, and *mmb* encodes a transmembrane protein that could be a

permease or a ferric reductase. The *pdp* gene product is homologous with hydrolases of the MutT/Nudix protein family. The problem is that if all of these genes are transcribed from the same promoter region and form an operon, then it might have been expected that their functions would be related. Yet there is no obvious link between the four. The *crtW* gene is present in a range of carotenoid producing species, whereas the oxidase and reductase genes have strong matches in completely separate species, though all are still bacterial. In contrast the hydrolase gene is found in a range of soil-based bacteria as well as a number of plant species. For *pdp*, *ord* and *mmb* there are no similarities with respect to *crtW* containing organisms, and the majority of the stronger matches were with species that were not even recorded in the top thirty matches of the other three genes. In order to test if there is any sort of linkage, a further analytical approach is required.

4.1.3. EMBL STRING

Identifying and then suggesting a role for a hypothetical protein is relatively straightforward when using NCBI BLAST searches. To fully understand the function of the gene though, it is beneficial to also look at the gene it is most commonly associated with i.e. the genetic context. This often makes it possible to establish the role played by a gene product by looking at the products of the gene(s) that surround it, especially if the surrounding genes form an operon. In the case of the *M. xanthus crtW* operon it would be of benefit to look at the strongest matched protein sequences and see if the genes involved are similarly linked together in other bacteria. It is possible to do this using a program known as EMBL STRING (search tool for the retrieval of interacting genes/proteins) which has access to a similar, but far more limited, database to that utilised by the BLAST searches. By conducting a search similar to that of the NCBI BLAST program it identifies protein sequences with great similarity to the compared sequence. The major difference is that for each sequence it also identifies the products of the neighbouring genes and uses a scoring system to indicate if they are also similar. This aids the positive determination of a role for each protein sequence of interest. A STRING study was carried out for each of the genes in the *crtW* operon.

4.1.3.1. *pdp* gene linkage

As the database is not extensive it contains only limited organism information. For instance, there is no information on *Myxococcus xanthus*. A *pdp* gene was identified belonging to *Anaeromyxobacter dehalogenans*. It was identified as encoding a MutT/Nudix protein, as per the NCBI BLAST results, and was found to be functionally coupled to a peptidase, with a score of 0.677. A score of 1.000 would indicate the two sequences were located next to each other, whereas a score of zero would indicate no linkage or relationship in the genome. A value over 0.5 is considered to represent significant linkage between the regions, so 0.677 shows strong linkage. The result means that in a different organism a similar MutT/Nudix protein gene is associated with a peptidase gene. This is not the case in *M. xanthus,* where the hydrolase is linked to a ketolase. As only one match was found for the *M. xanthus* gene it is still inconclusive as to how hydrolase activity relates to the rest of the *crtW* operon.

4.1.3.2. *crtW* gene linkage

Far more results were obtained when conducting a *M. xanthus crtW* gene STRING analysis as *Erythrobacter litoralis*, *Anabaena variabilis, Nostoc* sp. and *Gloeobacter violaceus* all gave strong sequence matches. *G. violaceus* is a bacterium found living under volcanic conditions; a hardy and resistant organism which has also been recorded to show carotenoid production (Tsuchiya *et al*., 2005). A similar gene is recorded in a *Synechococcus* species, a highly pigmented cyanobacterium. By looking at the sequences these genes are linked to, there appears to be two distinct functional roles identified. Firstly *E. litoralis* contains a β-carotene ketolase very strongly linked to a β-carotene hydroxylase, with a value of 0.816. In a number of carotenoid producers hydroxylases have been shown to interact with a carotenoid substrate, accompanying a ketonisation step (Misawa *et al.*, 1995; Lee and Kim, 2006), which may explain why the two genes are found linked together in *E. litoralis*. *Synechococcus* sp. WH8102 also encodes a β-carotene ketolase, which was instead found linked to a separate carotenoid binding protein (also with a value of 0.816), as well as two hypothetical proteins that displayed no homology with any other known protein sequence (0.723 and 0.634 respectively). Clearly the first gene product is involved in carotenogenesis, as is *crtW* but we can say nothing about the next two.

The second system identified is one linked to transport. *A. variabilis* shows strong linkage between an identified fatty acid desaturase and a putative iron transport system (0.56), then as with the *Synechococcus* results, two putative proteins with linkage values of 0.570 and 0.652. Similarly observed with the *Nostoc* sequence, a β-carotene ketolase has a strong link to an unknown protein, 0.705, and then also further strong links to a 50S ribosomal protein and a substrate binding protein of an ABC transporter (0.692 and 0.532 respectively). The ABC transporter system consists of a number of transporter proteins whose role is to transport a range of low molecular weight compounds across the cell membrane. Links to a possible transport system shows similarities to roles that were hypothesised for the *mmb* products during the NCBI BLAST searches, albeit of a completely different class. However the link here is based on a fatty acid desaturase rather than a positively identified β-carotene ketolase, as is the case for the gene associated with the ribosomal protein. A similar pattern is observed in the *G. violaceus* species, with links to a 50S ribosomal protein and also a ComA protein. The ComA protein is a response regulator expressed, amongst other times, when the bacterial host is exposed to extreme environmental stress (Friedrich *et al*., 2001). Despite being labelled as β-carotene ketolase it appears in these two instances the gene-linked encoding products are involved in substrate transport, protein synthesis and regulation. Again the genes that β-carotene ketolase is linked to in other organisms do not explain the roles of the genes of the *M. xanthus crtW* operon.

4.1.3.3. *ord* gene linkage

The *ord* gene product produces strong comparative values with similarities to a number of oxidoreductase proteins and putative molybdopterin-binding proteins. The strongest match is seen when compared to *Deinococcus geothermalis* which shows strong linkage to a ferric-reductase like protein (0.993) and then two further hypothetical proteins (0.669 and 0.644), suggesting another potential operon. The strong link to the possible ferric reductase is the first similarity to the *crtW* operon and the *mmb* gene product. This strong linkage is also expressed in the next match, *Deinococcus radiodurans*, which also has a 0.993 value for a link to a reductase-like protein. It is predicted to be linked to two other genes encoding a cytochrome C assembly protein and a ketol-acid reductoisomerase (0.816 and 0.862 respectively).

Cytochrome C is a heme protein that is an essential component of the electron transport chain. It can be oxidised or reduced while carrying electrons, indicating a possible link with a protein that has oxidoreductase activity. However, it is unclear what role the ketol-acid reductoisomerase may play in this process. The next strongest species matches are *Rhizobium etli,* with a link to a membrane protein (with a value of 0.993 again), and *A dehalogenans*. The link in this instance is to an unknown protein (0.992) and then, as with *D. radiodurans*, a cytochrome C assembly protein and ketol-acid reductoisomerase (0.734 and 0.680). In addition, there is also a suggested link with a zinc-binding alcohol dehydrogenase, 0.756. The Ord match of *Chromobacterium violaceum* in addition to links with a reductase (0.993), a cytochrome C protein (0.753) and a ketol-acid reductoisomerase (0.676), shows a link to a cytochrome C biogenesis trans-membrane protein (0.753). Thus it appears in bacteria other than *M. xanthus ord* homologues are linked to cytochrome C assembly proteins. In the *M. xanthus* genome *ord* is linked to the *mmb* gene, which may have ferric reductase activity as in *D. geothermalis*.

4.1.3.4. *mmb* gene linkage

When conducting an EMBL STRING search using *mmb*, the membrane protein and possible ferric reductase, the first two matches are identical species to those recorded for *ord*; *D. geothermalis* and *D. radiodurans*. Both display strong linkage to an oxidoreductase (0.993) and *D. geothermalis* shows links to three further hypothetical proteins. The latter species shows links once more to a ketol-acid reductoisomerase and a cytochrome subunit (0.654 and 0.715 values). A strong link of 0.811 is also recorded with a putative hydrolase. This is the first mention of a protein with a role similar to that encoded by *pdp,* first gene of the *crtW* operon, and in addition the link is stronger than that recorded for either the cytochrome or reductoisomerase proteins. However the hydrolase in question is not a MutT/Nudix protein, but a carbon-nitrogen (C-N) hydrolase. The third strongest match is with a species also identified during the NCBI BLAST searches, *Rhodospirillum rubrum*. The reductase is classed as a hypothetical protein associated with a second hypothetical protein (0.995). These are then in turn connected to an alcohol dehydrogenase (0.729) and a ketol-acid reductoisomerase (0.657). The final two species matches were for *A. dehalogenans* and *Methylococcus capsulatus*. The latter

showed a number of hypothetical protein links without positive role assignment to any of the proteins suggested. *A. dehalogenans* in contrast displayed a strong affinity to a hypothetical oxidoreductase (0.992) and then the previously recorded trio of zinc-binding alcohol dehydrogenase, cytochrome C assembly protein and ketol-acid reductoisomerase, with values of 0.716, 0.708 and 0.632. The *mmb* gene product is clearly an integral membrane protein and may function as a reductase as suggested by both NCBI BLAST and EMBL STRING searches. There also appears to be a strong link between *ord* and *mmb* in *D. geothermalis*.

4.1.3.5. *ord* and *mmb* transcription

Of the top species matches for *ord* and *mmb* products, three showed strong sequence similarity with both gene products; *D. radiodurans, D. geothermalis* and *A. dehalogenans*. All showed strong linkage to the second protein with scores of 0.992 or 0.993, regardless of which gene product was being studied at the time. Not only are the products of these two genes functionally linked in *D. radiodurans, D. geothermalis* and *A. dehalogenans*, but the genes are contiguous: MXAN_6048, *ord* and MXAN_6047, *mmb*, in *M. xanthus*; Dgeo_0877, the oxidoreductase gene and Dgeo_0878, the membrane gene in *D. geothermalis*; DR2536, the oxidoreductase gene and DR2537, the membrane gene in *D. radiodurans*; and Adeh_3976, the oxidoreductase gene and Adeh_3977, the membrane gene in *A. dehalogenans*. By looking at a number of the other species identified by the EMBL STRING search as displaying some degree of similarity to the sequences studied, the same pattern is clearly replicated (Table 4.1). Given the high incidence of the genes being located together, plus the strong linkage observed between the two, there is the real possibility they are transcribed together. This is the first sign of positive gene function linkage within the *crtW* operon and effectively splits it into three distinct sections – the hydrolase gene, *crtW* and then the *ord mmb* homologue. Given the number of sequences identified that matched both the membrane protein and oxidoreductase, they may have been studied in greater detail in one or more of the organisms. If a positive role identification could be assigned to the two genes then this may in turn aid in the understanding of their placement within the *M. xanthus* genome.

Organism	*ord*-like gene	*mmb*-like gene
Myxococcus xanthus	MXAN_6048 (*ord*)	MXAN_6047 (*mmb*)
Deinococcus geothermalis	Dgeo_0877	Dgeo_0878
Deinococcus radiodurans	DR2536	DR2537
Anaeromyxobacter dehalogenans	Adeh_3976	Adeh_3977
Chromobacterium violaceum	CV_0789	CV_0790
Caulobacter crescentus	CC2747	CC2748
Methylococcus capsulatus	MCA0843	MCA0844
Rhodospirillum rubrum	Rru_A0593	Rru_A0594
Escherichia coli K12	*yedY*	*yedZ*

Table 4.1 A table to show the *ord* and *mmb*-like genes in a number of organisms showing they are contiguous in all cases.

An NCBI BLAST search of the *E. coli* K12 genome was carried out using *ord* and *mmb,* which showed that *E. coli* K12 possessed two homologues given the notation *yedY* and *yedZ* (Table 4.1). Unfortunately neither has yet been assigned a definite function (Loschi *et al*., 2004). YedY was identified as a molybdoprotein containing a twin-arginine signal peptide, as suggested for Ord from NCBI BLAST and EMBL STRING results. It was also found to interact directly with YedZ forming a heterodimer (Loschi *et al*., 2004). YedZ was found to contain a haeme core and acts to anchor YedY, the catalytic subunit, to the cell membrane (Brokx *et al*., 2005). Crystallographic analysis of YedY found it to be dissimilar to known molybdenum containing enzymes, with the distinct molybdenum binding present giving rise to a unique YedY active site. Alteration of one part of the active site, Cys102, changing it to serine, resulted in YedY remaining associated with the cytoplasmic membrane rather than the periplasm as observed with wild-type YedY. In the absence of YedZ the Cys102Ser YedY was instead found in the periplasm, indicating that it must associate with YedZ at the cytoplasmic membrane (Brokx *et al*., 2005). A specific explanation for the change in activity following Cys102 replacement with serine is unclear, but does prove that YedY and YedZ interaction occurs. Detailed analysis of the heterodimer showed distinct reductase activity with a range of sulfoxide substrates, although it exhibited no sulphite oxidase activity (Loschi *et al*., 2004). Both genes were successfully knocked out in an attempt to identify their function (Brokx *et al*., 2005). The study found there was no observed phenotypic change in the resultant knockout strain. This either means that the functions of the two gene products were replaced by other proteins, or they were not essential for the culture

conditions chosen for the mutant. The two genes in the *E. coli* genome showed no link to any other neighbouring genes, appearing to be transcribed alone.

4.1.3.6. Summary

The EMBL STRING data as a whole supports the initial evidence that none of the studied organisms contain the four *crtW* operon genes located together. It has been demonstrated that the final two genes are contiguous in a wide number of bacteria and in *E.coli* are co-transcribed (Table 4.1). They encode a membrane-bound periplasmic molybdopterin-dependent, haeme-dependent oxidoreductase. The molybdopterin containing YedY homologue is attached to the membrane by the b-type haeme-containing YedZ (Brokx *et al*, 2005). The YedY subunit is exported into the periplasm in a fully folded co-factor containing state by the TAT protein export system. The *yedY* and *yedZ* homologues show no genetic linkage to *crtW* in any bacterium other than *M. xanthus*.

The *crtW* gene itself shows homology with two organisms that associate it with other carotenoid genes. The *crtW* gene was also found to be linked in a number of other organisms with genes encoding transport proteins and ribosomal subunits. The EMBL STRING results for the potential hydrolase, *pdp*, were limited, and only one positive match identified. This was to another myxobacterium; *A. dehalogenans*. In this case the MutT/Nudix family protein gene was linked to a peptidase gene, but still neither gene was found to be associated with the *yedY* and *yedZ* homologues. STRING results suggest an increased number of roles for each of the *crtW* operon genes and indicate that the final two genes are transcriptionally linked, and should be relabelled *yedY* and *yedZ.* The gene arrangement in the *crtW* operon thus appears to be a novel bacterial arrangement. There is still no clear explanation as to why the *crtW* operon genes should be linked. As the particular gene of interest is the one encoding the ketolase, *crtW*, focussing on this may aid understanding of the operon. In terms of bioinformatics, the best approach would be to compare and contrast known species of carotenoid producers. By studying their genetic control and gene arrangement with respect to carotenogenesis it may be possible to ascertain what is happening within *M. xanthus* and the *crtW* operon.

4.2. Genetics of carotenoid synthesis in other bacterial species

4.2.1. Introduction

The primary role of carotenoids in bacteria is to protect them from damaging species generated as a direct result of exposure to light. Hence they would initially have evolved in bacteria that were found in the presence of both O_2 and light. This group would have most likely consisted of the first evolved photosynthesisers, such as cyanobacteria. It is highly probable that carotenoids similarly arose in non-O_2 photosynthesisers, but in terms of groups such as the cyanobacteria they would have been essential. Following resultant oxygenation of the world by the O_2-evolved photosynthesisers, carotenoids would then have been of benefit to many non-photosynthetic anaerobes in the light, as a form of protection. Hence the first carotenoid generating cyanobacteria most likely gave rise to each of the bacterial, plant and fungal carotenoid-dependent species existing today.

In terms of the enzymes responsible for carotenogenesis, it is known that three are essential: CrtE, a hydroxyneurosporene synthase; CrtB, a phytoene synthase; and CrtI, a phytoene dehydrogenase. In addition, in the generation of cyclic carotenoids the lycopene cyclise, CrtY, is also essential. All other late-acting carotenoid biosynthetic enzymes, such as CrtC, CrtX, CrtZ and the ketolase CrtW, are considered non-essential, but improve the efficiency of carotenoid action under specific conditions.

4.2.2. Biosynthetic carotenogenesis gene arrangement

When conducting the initial NCBI BLAST search on the *M. xanthus crtW* gene product the strongest match was with a gene product in an *Algoriphagus* species, which had also been annotated as CrtW. By looking at the genes to which *crtW* is linked in this genome we may gain insight into the unique *crt* gene arrangement in *M. xanthus*. In the *Algoriphagus* sp. genome *crtW* is located with the majority of other biosynthetic genes arranged in a cluster: *crtEIBDCW*. All of the key carotenoid synthesis genes are contiguous and transcribed in the same direction. In *M. xanthus*, *crtI*, and possibly *crtW*, are located separately away from the other enzyme encoding genes.

In contrast, the closest related species to *M. xanthus* identified by the gene comparison is *Stigmatella aurantiaca*. In this organism the *crtW* homologue was annotated as a putative fatty acid desaturase. Given *St. aurantiaca*'s close phylogenetic relationship with *Myxococcus xanthus*, it is perhaps unsurprising to find its *crtW* homologue is separated from the main cluster of biosynthetic enzymes, in a similar arrangement to that of *M. xanthus*. The *St. aurantiaca crtW* homologue also appears to form part of a larger cluster. Given the similarity between the two species' sequences, the fact that *yedY* and *yedZ* homologues were not identified during STRING analysis was unexpected. However close analysis of the *St. aurantiaca* genome revealed *yedY* and *yedZ* homologues next to the *crtW* location. There was a strong match between the *yedY* and *yedZ* homologues of *M. xanthus* and *St. aurantiaca.*

Another organism previously identified as showing a close phylogenetic relationship with *M. xanthus* is the myxobacterium *Sorangium cellulosum*. Its complete genome has recently been sequenced (Schneiker *et al.*, 2007) and as a result a large number of genes have been identified and annotated. The carotenoid biosynthetic gene cluster displays a number of distinct differences to that of *M. xanthus*. The order of the genes is *crtI*, *crtB*, *crtZ*, *crtC*, *crtD* and *crtE*, where individual genes *crtZ*, *crtC* and *crtD* are transcribed in the opposite direction to the others. At some point during evolutionary divergence there appears to have been a rearrangement in gene order and gain and loss of genes, although the key genes have remained located together. In *S. cellulosum* there is no *crtW* gene, which is reflected in the final carotenoids produced: 1', 2'-dihydro-1'-hydroxy-torulene glucoside ester; 1', 2'-dihydro-3, 1'-dihydroxy-torulene glucoside ester; 1', 2'-dihydro-1'-hydroxy-torulene rhamnoside; and 1', 2'-dihydro-3, 1'-dihydroxy-torulene rhamnoside (Kleinig *et al*., 1971). There is instead a hydroxylase gene, *crtZ*, previously shown to encode a mono-oxygenase that inserts hydroxyl groups into carotenoid aliphatic rings and 1', 2'-dihydro-3, 1'-dihydroxy-torulene glucoside ester and 1', 2'-dihydro-3, 1'-dihydroxytorulene rhamnosuide both contain a hydroxyl group in the aliphatic ring structure.

In the *So. cellulosum* report (Schneiker *et al.*, 2007) a comparison was made to the carotenoid biosynthetic gene cluster of *Streptomyces coelicolor*, an organism that contains *crt* gene products that share sequence to *M. xanthus crt* gene products. *Streptomyces coelicol*or is a high GC ratio Gram-positive bacterium that is

phylogentically very distant from myxobacteria, although it shares an ecological niche, soil, and the high GC ratio. Its *crt* genes display completely different gene order and content to the myxobacteria. It contains two operons *crtEIBV* and *crtYTU*. Again, the *crtI* and *crtB* genes are contiguous and transcribed in the same direction. In this instance the second operon is inverted and adjacent to the first. In distinction to *St. aurantiaca*, which is very closely phylogentically related to *M. xanthus*, each of the other organisms studied have their biosynthetic genes located together, in distinct gene clusters. Between all of these, the actual gene arrangement does vary, as do the genes encoded, however, the genes that are different encode enzymes that carry out late, tailoring reactions in carotenoid biosynthesis.

The similarities are not confined to soil bacteria, shown in a study of a *Brevundimonas* species, a marine relative of *Caulobacter* species (Nishida *et al.*, 2005). *Brevundimonas* sp. strain SD212 possesses a number of clustered *crt* genes: the arrangement is *crtW, crtYIB, crtZE,* with *crtYIB* inverted with respect to the others. There are a number of other ORFs present in the cluster encoding sequences similar to *crtC* and *crtD*, along with genes encoding final step tailoring enzymes. The study also compared this species with five others – one soil bacterium and four marine species. Only two of the species possessed a *crtW* gene for the ketolase, *Paracoccus* species V81106 (previously *Agrobacterium aurantiacum*) and *Bradyrhizobium* sp. ORS278. In each case it was again found located with all of the other *crt* genes. In *Paracoccus* sp. the genes were all transcribed in the same direction in a *crtWZYIB* cluster and for the *Bradyrhizobium* sp. *crtE* was adjacent to *crtYIBW*, but divergently transcribed. No other species contained *crtW* though, but of *Paracoccus zeaxanthinifaciens, Pantoea ananatis, Pantoea agglomerans* and *Xanthobacter autotrophs* only the latter did not contain *crtZ.* In a number of the *crtZ* encoding organisms, including *S. cellulosum* and *Brevundimonas* sp. strain SD212, the gene is inverted and located immediately at the end of the biosynthetic cluster. This suggests it was a later evolutionary addition to the carotenoid producing genes which is unsurprising considering its product is active late in the pathway. *Bradyrhizobium* sp. ORS278 is the only organism in the study to possess *crtW* alone without a *crtZ* gene. Despite the varied genetic make-up of each cluster, the organism did have all of the principle biosynthetic genes clustered together. *Myxococcus* species are the only ones currently identified as having a potential *crtW* gene located separately in the genome from the other biosynthetic genes.

4.2.3. Ketolase and hydroxylase influence

CrtW is responsible for the final stage of myxobacton creation, the primary *M. xanthus* carotenoid product. Similarly in each of the previous organisms mentioned, with the sole exception of *S. coelicolo*r, which lacks both genes, the products of *crtW* or *crtZ* are believed to be responsible for the final carotenoid conversion step. In *Algoriphagus* sp. KK10202C, the organism that displays the strongest CrtW match to the *M. xanthus* protein, the gene is encoded at the end of the biosynthetic gene cluster. All of the genes in the cluster are transcribed in the same direction, though this has not proven universal in carotenoid producers. *X. autotrophicus* Py2 shows that if neither *crtW* nor *crtZ* are present, it is still possible to produce a functioning carotenoid. It has already been recorded that CrtZ is capable of acting on a substrate previously altered by CrtW, as is evident from the gene content of the *Paracoccus* species N81106. The *Bradyrhizobium* species ORS278, like *M. xanthus*, has no genetic equivalent of *crtZ*, so CrtW must also be able to operate exclusively and without subsequent substrate alteration by CrtZ. A number of organisms contain *crtZ* without *crtW*, in particular *So. cellulosum*. Therefore it must also be possible for CrtZ to function alone without CrtW present. Given the similar systems of carotenoid creation that exist in all of the organisms studied, the primary difference appears to be the presence of CrtW, CrtZ or neither, at the final biosynthetic stage. To establish reasons for this, the basic activity of both needs to be studied further.

It is possible to study the activity of both CrtZ and CrtW separately using a β-carotene substrate. When hydroxylated, β-carotene is converted into zeaxanthin, with a hydroxyl group added to the third carbon of each of the two carbon ring structures. If the carotenoid is exposed to a ketolase, a ketone group is added to the fourth carbon of each ring creating canthaxanthin. These modifications have a direct effect on the ability of the compounds to function in cell protection. A study has been carried out looking specifically at mechanisms of carotenoid antioxidant behaviour (Di Mascio *et al.*, 1990). One of the primary observations was that canthaxanthin was faster at quenching singlet oxygen species than zeaxanthin. The values recorded (in L mol^{-1} s^{-1}) were $2.1x10^{10}$ and $1.0x10^{10}$, respectively, so it was over twice as effective. This is a large difference when considering the actual size of the rate constants. The figures do differ when carotenoid activity in the presence of other specific radicals is studied.

Another study looked at the capability of carotenoids to react with nitrogen dioxide radicals, mercaptoethanol thiyl radicals, glutathione thiyl radicals and methanesulfonyl thiyl radicals (Mortensen *et al*., 1997). Zeaxanthin was far more capable of countering nitrogen-base radicals, being almost three times as effective as canthaxanthin. In contrast the opposite was true when exposed to mercaptoethanol thiyl radicals. The latter two listed radicals were controlled with similar effectiveness by each substrate. This shows that carotenoid structure has an effect on how effectively it can quench particular radical species. Highlighting this fact, a separate study (Woodall *et al.*, 1997) looked at the effectiveness of a set of carotenoids to react to the generation of peroxyl radicals. Canthaxanthin was least effective in standard comparisons, functioning at half the level of zeaxanthin. In a reaction where each substrate was exposed to a large amount of peroxyl oxidants at once, zeaxanthin was almost seven times more efficient at quenching than canthaxanthin. The distinct differences offer an insight into why different bacterial species from different environments would need to generate various carotenoids.

In terms of carbon bonding it has been recorded that the more alternate carbon-carbon double bonds present in a carotenoid, the more effective it is at quenching singlet oxygen species (Foote *et al*., 1970). In the examples of zeaxanthin and canthaxanthin the latter was far more effective, despite both possessing a similar carbon double bond number. In this case the addition of the two double bonds in the ketone group (which also fall into the alternate bonding pattern when combined with the other carbon double bonds) act to increase its effectiveness. This gives it thirteen double bonds in total, compared to the eleven present in zeaxanthin. The addition of a double bonded oxygen atom may also provide the molecule with added stability compared to that of zeaxanthin. This may explain why the two react differently to different radicals – clearly the presence of an OH group enables the carotenoid to quench peroxyl radicals easier and vice versa for glutathione thiyl radicals. In conclusion, the presence of both additional groups together in a carotenoid molecule should lead to the generation of an ester that is effective at quenching all of the reactive species from the canthaxanthin and zeaxanthin studies. When β-carotene is both hydroxylated and undergoes ketonisation, a structure is obtained which has a C_4 ketone and a C_3 hydroxyl group; astaxanthin.

Overall astaxanthin is a better quencher of singlet oxygen species, with a rate constant of 2.4×10^{10} (L mol^{-1} s^{-1}), than either hydroxylated or ketonised carotenoids.

The increased molecular stability provided by the hydroxyl groups appears to increase its effectiveness. Similarly with each of the previously tested nitrogen, glutamine thiyl and methanesulfonyl radicals; astaxanthin displays overall increased or at least similar maximum levels of quenching as for zeaxanthin and canthaxanthin. The sole exception is the mercaptoethanol thiyl radicals which are quenched at a reduced rate compared to either of the previously studied carotenoids. The reduction is not considerable, with approximately a 10% drop off in activity. When studying the peroxyl radicals, astaxanthin behaved in a similar way to canthaxanthin, with zeaxanthin still far more effective at quenching. This suggests that the presence of the ketone group leads to a decrease in the ability of the carotenoid to deal with these particular radicals. By looking at the three different carotenoids that can be produced from β-carotene using just hydroxylase and ketolase activity, it is obvious that each function has a significant effect. The most effective overall is astaxanthin, created through the combined activity of both CrtW and CrtZ, and the least effective zeaxanthin, formed by CrtZ alone, depending on which radical is being encountered by the carotenoid.

The myxobacton of *M. xanthus* has twelve alternating carbon double bonds and one additional oxygen double bond, as there is one cyclic carbon ring in the molecule. Without the ketone group the molecule should be a more effective quencher of singlet oxygen species than each of the previously mentioned β-carotene-derived carotenoids, due to its longer central twelve carbon double bond spine. Despite an increased stability associated with further hydroxylation of such a molecule, thus making it more effective at radical quenching, a CrtZ hydroxylase is not encoded within the *M. xanthus* genome. This result is also observed when looking at either the closely related *S. aurantiacum* or *Bradyrhizobium* sp. ORS278. Similarly when considering the increased instability and ineffectiveness of a hydroxylated species to deal with radical generation, there is no *crtW* gene in *So. cellulosum*, *Streptomyces coelicolor* or *Pantoea ananatis* (*Erwinia uredovora* 20D3).

4.2.4. *crtY* evolution

crtY encodes lycopene cyclase, the enzyme responsible for the conversion of lycopene to β-carotene . In all of the species studied it is closely linked to both *crtI* and *crtB*, with the three usually transcribed together in an operon which is part of the

main biosynthetic gene cluster. By searching the NCBI BLAST database it is possible to identify all organisms that have had all or part of the *crtY* gene sequence positively identified – 112 sequences in total. It is then possible to compare the sequences of the genes with each other using computer programs, ClustalW and ClustalX in this instance. The programs work by conducting a series of gene comparisons, whereby each amino acid sequence is compared for every gene. The process is done by calculating distance based algorithms for each individual sequence relationship. From this information a phylogenetic tree can be created, and then each relationship reassessed to see if the branch arrangement is correct, with similar sequences clustered together. If there are further differences the program will rearrange the branches and the sequences are compared with each other again. This process continues until the optimal tree is created whereby all similar branches are grouped together. The entire process was carried out for *crtY* and a tree can be seen overleaf in Figure 4.4. This was visualised using the TreeView 1.6.6 computer program.

The tree itself can be split into three primary branches, of which *M. xanthus* is on the smallest. The first one clusters the species into four distinct groups; the cyanobacteria, a group of plants and then two groups consisting of a selection of actinomycetes including mycobacteria. The second primary branch encompasses the actinomycetes again, including the myxobacteria, mycobacteria and clavibacteria. The final remaining distinct primary branch includes a number of pathogenic species as well as the soil-based *Xanthobacter* species and the streptomycetes. As the tree is constructed using information obtained from a single gene sequence it is beneficial to determine how consistent the predicted groupings are.

The cyanobacteria, *Synechococcus* and *Prochlorococcus* species are grouped together, suggesting that they all have closely conserved sequences. In this case more than one *crtY* gene has been identified in the majority of the *Prochlorococcus* individuals, hence why there are two distinct groupings of similar species. The program has successfully separated these species by the slight variations detected between the two CrtY products. Likewise as all of the plant species are clustered together, it is likely they all obtained *crtY* from the same origin. They appear to have a close association with the plant *Solanum* (tomato) and bacterial *Ostrococcus* species. This in turn suggests one of these may be the possible original gene source for the plants.

It has been previously shown that *M. xanthus* possesses two separate copies of the *crtY* gene, designated *crtYc* and *crtYd*, which are contiguous. Both of the two *M. xanthus crtY* halves are located together on the same branch. Over time *crtY* has either split or the two pieces have not yet fused in order to form a single *crtY* gene. This is also observed in one of the mycobacteria species, *Mycobacterium ulcerans* Agy99, which is grouped on the same primary branch as *M. xanthus*. Both individual sequences are clearly separated from each other on the branch, showing that despite the number of similarities between the organisms they are still very different from each other in terms of *CrtY* sequence. Species that appear to be most closely related include a *Clavibacte*r species; a gram-positive soil dweller, and a *Salinibacter*; an extreme salt-water bacterium. There are also two other species originally isolated from thermal vents, *Haloarcula* and *Haloquadratum*. Additionally an ammonia oxidiser, *Natronomonas*, also a marine-based organism, is classified in the same cluster. Each species inhabits a different ecological niche and as a result is clearly adapted to withstand various extremes, yet the tree suggests all are connected through the origin of the *crtY* gene. Also located on the same branch are a number of organisms that survive by breaking down chemical substrates like methane and chlorine, and are capable of living in conditions where these substrates are present in a high concentration, or where there is instead an extreme of temperature. All are recorded as producing carotenoid pigments.

The remaining species that are located on the third primary branch range from pathogenic species, including *Pseudomonas* and *Bdellovibrio*, to *Xanthobacter* and *Sphingomonas* species which are soil bacteria that produce carotenoid pigments. Similarly to the other two primary branches the organisms represented are from a large cross-section of bacterial phyla and classes, including Chloroflexi, α-proteobacteria and flavobacteria.

As the computer program has created the tree that it predicts best displays all of the gene comparisons, all of the groupings may still not be entirely consistent. Overall similar groups are found clustered together and distinct segregation still observed, particularly between large groups such as the cyanobacteria and all of the other carotenoid producers. Having proven that the program can generate consistent phylogenetic trees using the information from a single biosynthetic gene, it is possible to directly compare the *crtW* producing species.

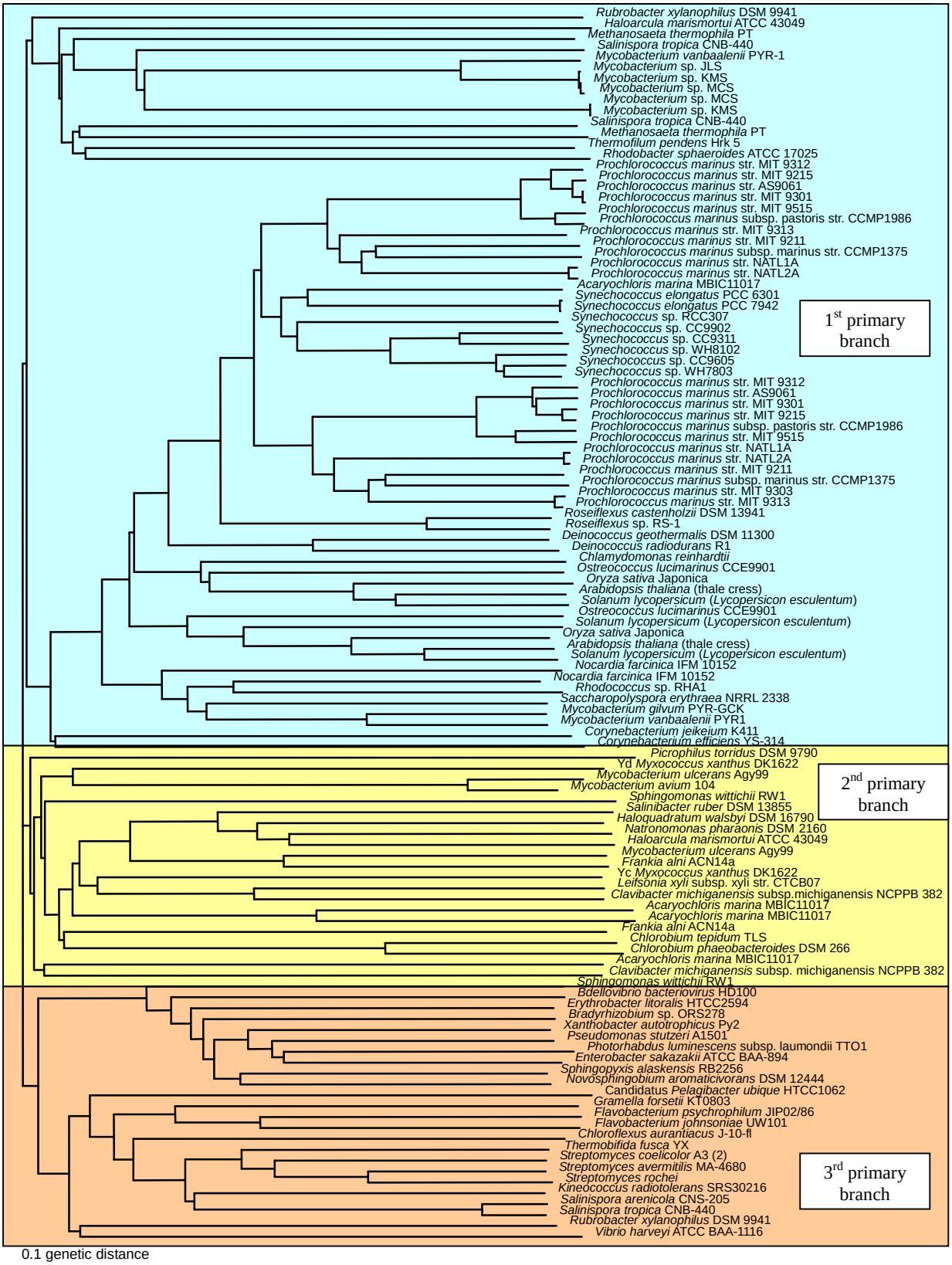


Figure 4.4 A rooted phylogenetic tree constructed by comparing CrtY protein sequences using a distance-matrix algorithm.

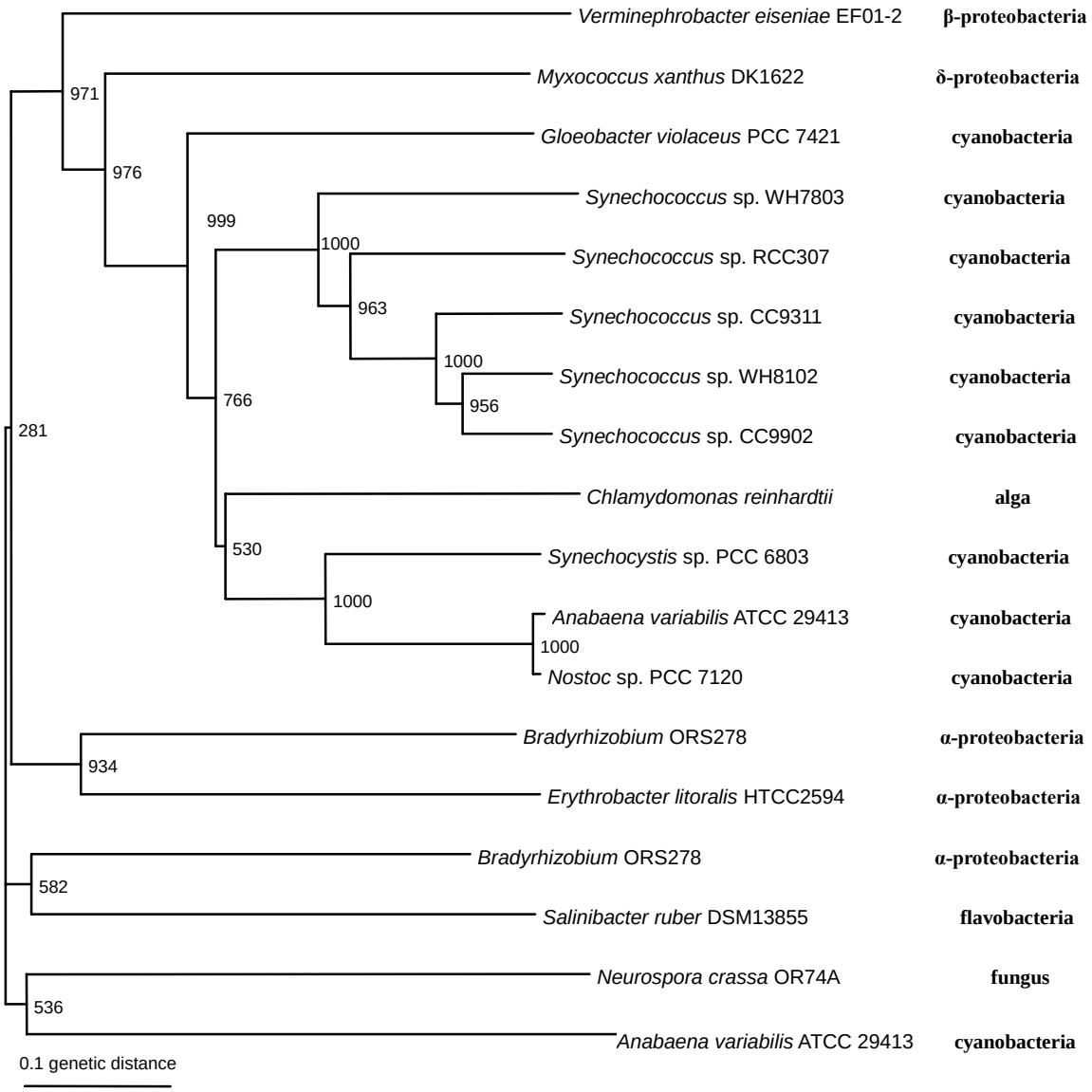


Figure 4.5 A rooted phylogenetic tree constructed by comparing CrtB protein sequences from organisms that also contain *crtW* using a distance-matrix algorithm. The number indicates the bootstrap value for each branch from 1000 replicates.

4.2.5. *crtB* evolution

A second control gene was used to produce an additional phylogenetic tree – CrtB. This is another highly conserved biosynthetic enzyme in carotenoid producing organisms. The tree was constructed using only organisms previously identified as containing *crtW*, or a carotene ketolase encoding gene. Thus it should

provide additional data for direct comparison with a tree created using CrtW generating organisms. This totalled 24 sequences, and from these 18 *crtB* genes were identified. The 18 CrtB sequences were used to construct the tree using an identical method to that previously adopted when studying the *crtY* gene. The results supported much of the data observed in the *crtY* tree, with most similar organisms grouped together (Figure 4.5).

Most of the cyanobacteria, and all of the *Synechococcus* species, were once again grouped together. The higher the bootstrap value, 1000 being the maximum, the more closely related a species appears to be, based on the sequence studied (Soltis and Soltis, 2003). Even within a closely related group of the same species there is still genetic variation, shown by *Synechococcus* bootstrap values below the maximum, i.e. 856 and 963. Species that were previously grouped furthest from the cyanobacteria in the *crtY* phylogenetic tree were again located on a separate branch when *crtB* was used as the predictor, i.e. *Bradyrhizobium* species, possessing two separate *crtB* genes, and *Erythrobacter*. These are two of the strongest related species with regards to CrtB sequence similarity, and *Erythrobacter* is also known to be a producer of red carotenoids. Linked with one of the *Bradyrhizobium* proteins is a *Salinibacter* species. Similar data were observed during *crtY* analysis, when *Salinibacter* was grouped with *M. xanthus* amongst other species. The final organisms on this branch are *Neurospora*, a fungal species, and *Anabaena*, which is a freshwater cyanobacterium.

Following this, there are two remaining branched sections in the same primary branch as the majority of the *Synechococcus* cyanobacteria. The larger of the two is split into two distinct segments; the smaller of which contains two cyanobacteria, a second *Anabaena* species and a *Nostoc* subspecies, which have a high bootstrap value. The remaining species on the branch are *Chlamydomonas*, an algal genus, and *Synechocystis*, another marine cyanobacterium.

The remaining tree branch, an offshoot of the larger one encompassing the groups discussed previously and *Synechococcus*, contains *M. xanthus*. It is grouped with a β-proteobacterium *Verminephrobacter* species, and the two display a strong relationship with regards to the bootstrap values. *Verminephrobacter eiseniae* is a bacterium isolated from the digestive tract of the earthworm. It is a species that lies dormant in the soil and then when ingested by an earthworm host, its life cycle is triggered. It essentially lives within the earthworm as a symbiont. As a species it has

become the subject of a number of recent scientific studies following its identification, principally attempting to determine its role within the gut (Pinel *et al.*, 2008).

The *crtB* phylogenetic tree supports the view of the effectiveness of the Clustal and TreeView programs. It produces similar results to *crtY*, but also suggests a number of novel species groupings. The fact that not all of the organisms encoding *crtW* also contained *crtB* is unexpected. As CrtB is highly conserved it suggests it has either been missed or there is an analogue of *crtB* in the *crtB*$^-$ species. As a result of this more species are present in the *crtW* phylogenetic tree.

4.2.6. *crtW* evolution

As has been alluded to previously, a number of species produce a β-carotene ketolase gene, *crtW*. It is not always given such notation, being also labelled as either *crtO* or *bkt*. *crtO* and *crtW* are genes that are not homologous, though they encode enzymes with a similar role. As a result there are two genes which have been identified that are able to produce a β-carotene ketolase (*bkt* is excluded as it displays great similarity to *crtW*). *crtW* is the more numerous of the two, although a number of species containing *crtW* do also contain a *crtO* gene e.g. *G. violaceus*. The benefit conveyed to the organism in this instance is unclear, as very little study has been conducted into comparative activity of the two genes. It is known that CrtW is the primary carotenoid ketolase in these examples, but it appears CrtO can also act simultaneously in a similar role, thus potentially increasing the final amount of carotenoid product. When constructing the CrtW tree both genes were considered and a number of CrtO sequences included for direct comparison. This may also help to identify a possible universal source for the two genes, whereby a single gene may have developed into two different forms over time. The majority of the tree was made using predominantly *crtW* sequences to allow easier comparison between organisms.

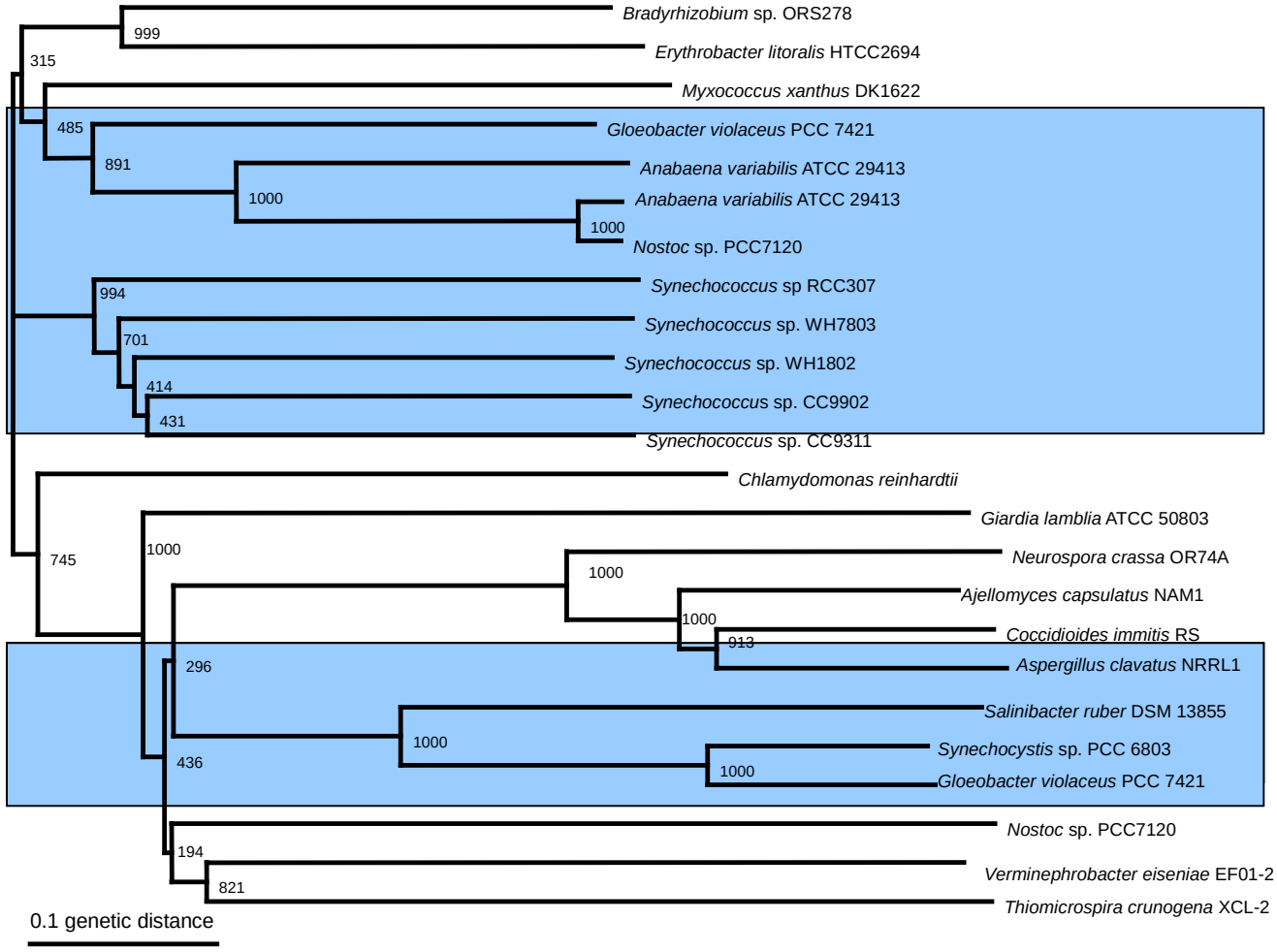


Figure 4.6 A rooted phylogenetic tree constructed by comparing CrtW and CrtO protein sequences using a distance-matrix algorithm. The number indicates the bootstrap value for each branch from 1000 replicates. All species contained within a blue box are grouped by comparing CrtO sequences.

The first point of note from the *crtW* phylogenetic tree (Figure 4.6) is the bootstrap values. There are a decreased number of branches reaching the maximum value of 1000 compared to the *crtB* tree. As a result a lot of the species appear to display stronger relationships with respect to the *crtB* gene than they do with *crtW*, suggesting decreased levels of species sequence conservation. One surprising result, in contrast to that initially observed during *crtB* analysis, is the weak linkage between the five *Synechococcus* species. They are again clustered together, but the bootstrap values linking them are as low as 414. This is further evidence of the variation that is present between individual members of the group.

Located on a nearby branch to the marine cyanobacteria species are again a *Nostoc* sp., *G. violaceus* and an *Anabaena* species. This is a similar arrangement to that generated when comparing *crtB* sequences. The difference is that once more the bootstrap values are lower, indicating a weaker relationship when comparing ketolases. In this instance all *Synechococcus* species and freshwater cyanobacterium *Nostoc* species were compared using CrtO. In the case of the *Anabaena* species the gene was classed as either *crtO* or *crtW* (Mochimaru *et al*., 2005). Effectively the CrtO data has been grouped separately from the majority of the CrtW sequences, on a separate branch. This is confirmation that the two genes display little similarity. This is not the sole CrtO containing branch of the tree as separately in the lower half is a grouping of three organisms; the flavobacterium *Salinibacter ruber*, and cyanobacteria *Synechocystis* and *Gloeobacter violaceus*. Once more these are arranged in a distinct separate grouping compared to neighbouring sequences and separated by a bootstrap value of only 296. This highlights the evolutionary difference is not only between CrtW and CrtO, but also within the CrtO proteins.

CrtO aside, the remaining ketolase data are split into five distinctive tree branches. The first of these encompasses four fungal species including *Aspergillus* and *Ajellomyces* species. The majority are linked together with a strong bootstrap value.

The next subset contains three isolated organisms, of which two are coupled together on the same branch. These are comprised of a β-proteobacterium, the *Verminephrobacter* species previously found to be strongly linked to *M. xanthus* when comparing *crtB* sequences, and a novel α-proteobacterium, *Thiomicrospira*. This species has strong similarities physically to *G. violaceus*. It too is isolated from a hydrothermal vent where it exists as a sulphur-utilising bacterium (Jannasch and Wirsen, 1981). The bootstrap value linking the two is below 1000, so not as strong as the majority of the surrounding species relationships. The third isolated species is *Nostoc*, the symbiotic nitrogen-fixing freshwater cyanobacterium, which is only weakly linked to *G. violaceus* and *Thiomicrospira*. The remaining species on the lower branches of the phylogenetic tree are the green algal protest, *Chlamydomonas reinhardtii* and *Giardia lamblia*; a flagellated intestinal protozoan parasite. It lives in the lumen of the gut and is spread via the ingestion of infected soil, following its initial transfer to the external environment through excretion. Although *Giardia* does create carotenoids it is unknown as to the role these may play.

The final branch is also split into two distinct segments – one of which contains *M. xanthus*. It appears on the same sub-branch as two *crtO* encoding species; although there is little sequence similarity (485 bootstrap value). This supports the hypothesis that *crtO* and *crtW* originated from the same gene source initially. Additionally it is also grouped alongside two cyanobacteria, *A. variabilis* and *G. violaceus*, mapped using the *crtW* gene, but still separated by a relatively low bootstrap value. It is separated from, but also classified as being most similar to the α-proteobacteria, *Bradyrhizobium* sp. ORS278 and *Erythrobacter*. As with the *Synechococcus* strains these two species have always been found located together on the phylogenetic tree regardless of whether the protein studied was CrtY, CrtB or CrtW. This is the first time *M. xanthus* has been strongly linked to either of these two species in any of the studies.

4.2.7. Phylogenetic tree summary

The data produced from the three trees for *M. xanthus* are conflicting. Whichever gene was studied, similar species were not always grouped with *M. xanthus*. With respect to *crtW* it was the α-proteobacteria, *Erythrobacter* and *Bradyrhizobium* subspecies; with *crtB* the earthworm symbiont *Verminephracter* and cyanobacterium *G. violaceus*; then when considering *crtY*, extremophiles. This suggests that it is still difficult to conclude anything based entirely on evolutionary relationships predicted from sequence comparison alone, especially if the gene is not highly conserved between organisms.

The *crtW* gene present in *M. xanthus* is not located along with the other biosynthetic genes in a cluster. To survey possible evolutionary origins for the biosynthetic enzymes within *M. xanthus*, the *crtE*, *crtD*, *crtC* and *crtI* gene products were compared in separate NCBI BLAST searches to determine their closest homologues. These gene products are responsible for synthesis of the core carotenoid.

4.2.8 *crt* gene homologues

Reference	Role/Function	Organism	BIT	e-value
crtE – Encoding geranylgeranyl pyrophosphate synthase in *M. xanthus*				
CAA79955.1	Geranylgeranyl pyrophosphate synthetase	*Myxococcus xanthus* DK1622	698	0.0
ZP_01460410.1	Geranylgeranyl pyrophosphate synthetase	*Stigmatella aurantiaca* DW4/3-1	469	1e-130
YP_023626.1	Geranylgeranyl pyrophosphate synthase	*Picrophillus Torridus* DSM 9790	172	7e-33
AAZ32465.1	Geranylgeranyl pyrophosphate synthase	Uncultured euryarchaeote	145	7e-33
NP_738677.1	Putative polyprenyl synthase	*Corynebacterium efficiens*	132	4e-29
crtD – Encoding hydroxyneurosporene dehydrogenase in *M. xanthus*				
ZP_01460400.1	Hydroxyneurosporene dehydrogenase	*Stigmatella aurantiaca* DW4/3-1	629	2e-178
YP_001995886.1	Phytoene desaturase	*Chloroherpeton thalassium*	250	2e-64
YP_001276151.1	Phytoene dehydrogenase-related protein	*Roseiflexus* sp. RS-1	244	1e-62
YP_601432632.1	Zeta-phytoene desaturase	*Roseiflexus castenholzii*	239	4e-61
YP_001543385.1	Zeta-phytoene desaturase	*Herpetosiphon aurantiacus*	238	1e-60
crtC – Encoding hydroxyneurosporene synthase in *M. xanthus*				
AAP59035.1	CrtC	*Thiocapsa roseopersicina*	110	2e-22
YP_780252.1	Hydroxyneurosporene synthase	*Rhodopseudomon palustris* BisA53	106	3e-21
NP_946865.1	Hydroxyneurosporene synthase	*Rhodopseudomon palustris* CG009	106	3e-21
EED35795.1	Hydroxyneurosporene synthase	*Gamma proteobacterium* NORS1-B	105	5e-21
AAR37857.1	Hydroxyneurosporene synthase	Uncultured marine bacterium	103	2e-20
crtI – Encoding phytoene dehydrogenase in *M. xanthus*				
ZP_01460393.1	Phytoene dehydrogenase	*Stigmatella aurantiaca* DW4/3-1	822	0.0
YP_643624.1	Amine oxidase	*Rubrobacter xylanophilus* DSM9441	425	6e-117
AAC84034.1	Diapophytoene dehydrogenase (CrtN)	*Heliobacillus mobilis*	421	1e-115
AA046894.1	Phytoene dehydrogenase	*Blakeslea trispora*	410	1e-112
AA046892.1	Phytoene dehydrogenase	*Blakeslea trispora*	410	1e-112

Figure 4.7 The top five NCBI BLAST matches of four carotenoid biosynthetic gene products. *M. xanthus* entries have been excluded in each case, except where two matches were recorded. The reference number refers to the accession number on the NCBI database. All BIT scores and e-values are calculated using the default NCBI BLAST settings.

4.2.8.1. *crtE* homologues

The first factor that is evident is the range of different species that display homologous sequences to each particular enzyme. In every case the strongest match is of the *M. xanthus* protein from whence the sequence originated. Similarly the second strongest match in most instances is *Stigmatella aurantiacum*, the closely related myxobacterium. Following these the best matches come from a wide variety of organisms (Figure 4.7).

Initially comparing the CrtE enzyme, the first enzyme encoded by the *M. xanthus* biosynthetic gene cluster, the next strongest match is that of the *Picrophilus*

torridus DSM 9790. This is an acidophile found at temperatures between 40^{o}C and 65^{o}C. It has been recorded growing at the lowest pH of any studied organism, where the pH is zero, albeit at a slower growth rate (Schleper *et al*., 1995). The next match on the list is from an uncultured bacterium, of which little is known as a result. Increasingly this kind of data is being added to the ever expanding sequence databases, inevitable given the trend towards sequencing DNA of environmental samples. The last strongest match listed is from *Corynebacterium efficiens*, a gram-positive non-sporulating actinomycete. The strongest matches to CrtE are all bacterial, though with few taxonomic relations.

4.2.8.2. *crtD* homologues

In *M. xanthus crtE* is transcribed with *crtD*, thus we might expect that the two are closely related in other species too, which should also be displayed in the sequence homology. *crtD* encodes a hydroxyneurosporene dehydrogenase in *M. xanthus*. The closest match is from *Chloroherpeton thalassium* ATCC 35110, a photosynthetic bacterium of the Chloroflexi phylum. It utilises sulphur as an energy source and was isolated from an oceanic water sample. It is also believed to function as a motile gliding bacterium. The next two database results are both *Roseiflexus* subspecies, one isolated from Yellowstone National Park (Boomer *et al*., 2002), and both highly pigmented with a deep red colouration. As in *Chloroherpeton thalassium* they utilise sulphur ions as electron donors in the photosynthetic process. Both individual subspecies again display gliding behaviour. The next match from the NCBI BLAST database is that of *Herpetosiphon aurantiacus* strain ATC23779, again belonging to the Chloroflexi phylum along with *Roseiflexus*. This strain is isolated from the slime coat cover of a freshwater alga (Holt and Lewin, 1968). It is a yellow filamented organism that has again been recorded moving by a gliding mechanism. It is highly similar to *M. xanthus* with respect to its predatory behaviour. The organism is capable of hunting using a similar 'wolf-pack' method, attacking bacterial prey in a swarm-like mass of cells. The cells are then digested using hydrolytic enzymes that are exuded by the bacteria (Holt and Lewin, 1968). Unexpectedly each of the matches recorded for CrtD are from organisms with a similar behavioural pattern to *M. xanthus*, each having gliding motility with one displaying a similar predation method.

4.2.8.3. *crtC* homologues

The gene *crtC*, encodes hydroxyneurosporene synthase. Of all genes studied, this displays the least similarities with the homologous sequences. The best CrtC match is with a gene from *Thiocapsa roseopersicina*, which produces the carotenoid spirilloxanthin. This plays a photoprotective role in the marine purple sulphur α-proteobacterium. It is similar to the next strongest matches i.e. two *Rhodopseudomonas palustris* strains. The two strains are fresh-water purple non-sulphur photosynthetic α-proteobacteria (Harwood and Gibson, 1988). Hence they are phylogenetically related to *Thiocapsa*. The final two strongest sequence matches are from organisms that again have not been studied in depth; an α-proteobacterium and an uncultured marine bacterium. All of the known species containing CrtC homologues are from the same taxon, α-proteobacteria, although this is not the taxon of *M. xanthus*.

4.2.8.4. *crtI* homologues

The gene *crtI* encodes phytoene dehydrogenase, a core enzyme of carotenoid biosynthesis. As a result the sequence is expected to be highly conserved between organisms. It is found located separately from the primary biosynthetic gene cluster in *M. xanthus*, though a similar sequence is encoded within the cluster, but its product does not function as a phytoene dehydrogenase. *M. xanthus* CrtI does display great similarity with other CrtIs, with no protein recorded in the top ten matches showing a weaker alignment score than 350. Once again the strongest match is from *Stigmatella aurantiaca*. Following this, the actinomycete *Rubrobacter xylanophilus* CrtI displays the highest degree of similarity. This species was initially discovered growing in areas of high radiation. It has been likened to the phylogenetically unrelated highly radioactivity-resistant *Deinococcus radiodurans* in its extreme survival. The next match is from the Gram-positive photosynthetic bacterium *Heliobacillus mobilis*. The protein match in *H. mobilis* is actually to CrtN, which is a diaphytoene dehydrogenase. This is a prime example of an organism possessing a different carotenoid gene that can play a similar role to that of *crtI* during carotenogenesis. The two probably have the same evolutionary ancestor, but subsequent development has slightly altered the gene function. The final two matches are from the same species,

Blakeslea trispora; there are two copies of the same gene present. *B. trispora* is a fungal plant pathogen and is the species from which the majority of β-carotene is extracted for commercial use. The occurrence of two similar carotenoid biosynthetic genes in the same organism is not uncommon e.g. the two *crtI*-like genes in *M. xanthus.* It has been estimated that 15% of the *M. xanthus* genome may be made up of duplicated gene regions (Goldman *et al*., 2006).

4.2.8.5. Homologue summary

Study of the *M. xanthus* key carotenoid biosynthetic genes has confirmed that all show distinct differences when comparing the species that display the strongest matches to those found for *crtW*. The strongest matches were recorded for the enzyme CrtI. This is to be expected as the enzyme is one of the central proteins in carotenoid synthesis, and hence will have stronger matches than enzymes such as CrtW which is only a peripheral biosynthetic enzyme. This also accounts for the low number of species possessing a CrtW-like enzyme compared to CrtI, CrtB and CrtY. The most unexpected observation from the phylogenetic tree construction was that CrtB did not appear to be present in all of the carotenoid producers. This is also an essential biosynthetic enzyme in carotenogenesis, so it is most likely an analogue of the enzyme is instead present in those carotenoid-producing species lacking CrtB. It was also evident that on a number of occasions, biosynthetic enzymes were highly conserved within a similar taxon e.g. CrtC in the α-proteobacteria. In these instances, despite the similarity in sequence to the original *M. xanthus* protein source, the taxon was often different to δ-proteobacteria.

It is clear the essential enzymes recorded during carotenogenesis CrtE, CrtI and CrtB, in conjunction with CrtY in the case of *M. xanthus*, are as expected the most highly conserved biosynthetic enzymes. This is evident in part by the occurrence in a number of the organisms in this investigation of the gene cluster *crtYIB.* Being essential, these genes and their related products would have been the first to evolve in carotenoid-producing species. The enzymes involved in the latter stages of carotenoid production such as CrtC, CrtW and CrtZ would therefore have been the final enzymes to evolve and hence be the last adopted. This would also explain why *crtW* and *crtZ* are the least commonly recorded biosynthetic genes, and also possibly why *crtW* is located separately to the remainder of the biosynthetic

enzyme encoding genes in *M. xanthus*. The separate *crtI* gene within *M. xanthus* was most likely obtained through the incorporation of additional DNA into the host genome. This was then either used in preference to the *crtI* that was already encoded by the biosynthetic gene cluster, or because that gene was unable to function effectively for some reason. The inheritance of the majority of genes in this way would also account for the range of gene clusters observed in the organisms studied and why some are transcribed in a completely different direction to nearby biosynthetic genes. The genetic isolation of *crtW* may be a result of it only recently being acquired in terms of the evolutionary time scale. As such it may not yet have had sufficient time to become completely associated with the other biosynthetic genes.

4.2.9. *St. aurantiaca crtW* operon arrangement

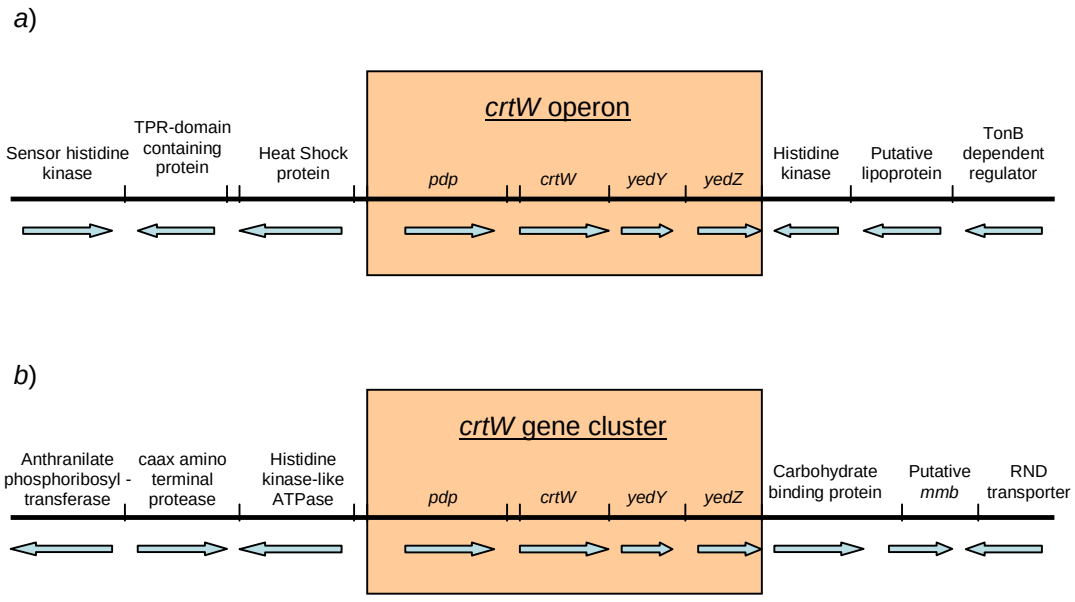


Figure 4.8 Comparison of *a*) *M. xanthus* and *b*) *St. aurantiaca crtW* gene clusters and neighbouring genes. Arrows indicate the direction in which the gene is transcribed. All labelled genes for *St. aurantiaca* if not stated are putative.

Due to its strong similarity with *M. xanthus* and as a result the little information that could be obtained, the bacterium *Stigmatella aurantiaca* has not been considered in detail throughout this study. In terms of the *crtW* operon this is true, as a similar structure exists within *St. aurantiaca*, leading to identical carotenoids being

produced by the bacterium (Figure 4.8). The operon is however located in a different place to the one from *M. xanthus*. It is not immediately preceded by a heat shock protein transcribed in the opposite direction, thus ruling out a link between this gene and the operon. This includes the possibility of a promoter region for the operon being encoded in part of the gene region. At the opposite end of the operon two different genes are encoded after it, transcribed in the opposite direction which is also observed in *M. xanthus*. These appear to encode a putative lipoprotein and a histidine kinase. In *M. xanthus* the *crtW* operon is followed by a series of unlinked genes that are not found in *St. aurantiaca.* High conservation of the gene sequences in *St. aurantiaca* and *M. xanthus* when the surrounding genes are altered further indicates that the four genes form an operon.

4.3. Discussion

By repeating initial gene comparison studies on NCBI BLAST, results indicate *M. xanthus crtW* is most similar to other genes encoding a CrtW carotenoid ketolase. The two genes downstream of *crtW* in the *crtW* operon, previously annotated as *ord* and *mmb*, encode a periplasmic molybdopterin and haeme-dependent oxidoreductase, YedYZ. These are intrinsically linked in a host of bacterial organisms, as they are in the operon. The remaining operon gene, *pdp*, also appears to encode a protein linked to the MutT/Nudix family. In this respect it is likely the enzyme plays a role within the cell as a hydrolase, involved in the formation or destruction of a cell substrate. Principally, with the exception of *M. xanthus* and *St. aurantiaca*, the putative MutT/Nudix gene, *crtW* and the *yedYZ* pair were never found transcribed together in any of the species identified during NCBI BLAST analysis.

EMBL STRING studies also indicate that none of the operon genes are usually found linked together in any other species, with the exception of *yedYZ*. Similar *crtW* gene sequences are however identified usually linked to other carotenoid biosynthetic enzymes. The search provided no additional information on the MutT/Nudix protein as there were not enough matches identified in the EMBL database. In contrast the strength of the *yedY* and *yedZ* link was indicated by large linkage values for the two in a number of bacterial species. For the *crtW* gene in particular links were identified in a number of organisms to genes that encoded ribosomal subunits and transport proteins, although again there is no evidence for similar linkage in *M. xanthus*.

Conversely, the *yedY* and *yedZ* homologues displayed no genetic linkage to *crtW* in any bacteria other than *M. xanthus* and *St. aurantiaca*. Overall the STRING results suggest an increased number of roles for each of the genes encoded by the *crtW* operon and that the gene arrangement in the *crtW* operon is a novel bacterial arrangement.

Further study of carotenoid producing organisms suggests that the biosynthetic enzyme encoding genes are usually found located together in a cluster. However in the case of *M. xanthus* there is a major cluster and two separate genes, *crtI* and *crtW*. This is unusual as *crtI* is an essential carotenoid biosynthetic gene, found clustered with the other essential biosynthetic genes in the majority of carotenoid-producing bacteria. When *crtW* is present in a larger gene cluster it is sometimes found associated with a hydroxylase gene, *crtZ*, but more usually it is either *crtW* or *crtZ* alone. Joint activity of the two can give rise to a carotenoid that possesses both a ketone and a hydroxyl group. This is highly stable and very effective at dealing with a number of reactive, cell damaging reactive species. In *M. xanthus crtW* is not linked to a *crtZ*, and nowhere in the genome is a carotenoid hydroxylase encoded. More in depth study of the effects of ketonisation and hydroxylation on a β-carotene substrate indicates that in most cases the former produces a carotenoid that is more effective at quenching reactive species. The primary exception to this rule is when negating the effect of reactive peroxyl species, where a solely hydroxylated carotenoid is far more effective.

The evolutionary relationship between *M. xanthus* and other carotenoid producing species is found to vary depending on which gene sequences are being compared. By initially comparing the *crtY* sequence it was clear that the phylogenetic comparisons were often intermingled. When repeating the process with *crtW* and also *crtB*, similar organisms grouped together once more, although *M. xanthus* is paired with different individuals in each case. Overall comparison of the biosynthetic gene sequences revealed very little phylogenetic correlation. Considering both the phylogenetic results and the data obtained from direct comparison of additional biosynthetic enzymes involved in *M. xanthus* carotenogenesis, it is implied that the genes responsible for their creation have originated from a number of sources prior to *M. xanthus* acquisition. The obtained results show a distinct difference between the origin of the essential genes *crtE, crtB* and *crtI*, semi-essential gene *crtY*, and the non-essential *crt* gene *crtW*, of *M. xanthus.*

The bioinformatic work suggests that with the exception of *Stigmatella aurantiaca* the four operon genes are not found located together in any other organisms. The MutT/Nudix protein is not even encoded in most of the organisms that contain homologues of the other three genes. As a result it is unclear the role that the operon plays in *M. xanthus*, although a number of separate roles for the genes and the pathways in which they function have been suggested. No evidence as of yet indicates a benefit of possessing the *crtW* operon and the four genes being grouped.

Chapter 5

CrtW structure and function

5.1. Introduction

Despite the positive identification of an active *pdp* promoter and confirmation that the gene itself forms part of a larger four-gene operon, there remains no experimental evidence for the actual role CrtW plays in *M. xanthus*. Preliminary bioinformatic studies seem to suggest that it functions as a ketolase involved in the final stage of carotenogenesis. This is supported further by the predicted secondary structure of the CrtW protein.

Without biochemical analysis of the protein, it is still possible to predict its likely secondary structure. This is achieved through the use of the protein amino acid sequence and comparing it with other known protein sequences and structures. From this information, it should be possible to determine what individual regions of the sequence encode and thus the possible function and location of the protein in the cell. In order to make such a study possible, a number of computer software packages are widely available enabling amino acid sequence analysis and protein secondary structure prediction. A program known as TMpred (prediction of transmembrane regions and orientation) was initially used in the study of CrtW. TMpred assigns scores which are a numerical representation of helix strength and the likelihood of its actual occurrence; any values over 500 are considered to be significant.

5.2. The secondary structure of CrtW

CrtW is predicted to play a role in the process of light-induced carotenogenesis, and as a result is most likely to be located at the cell membrane. This is where the majority of carotenoid biosynthetic enzymes are already known to be located. This suggestion is supported by TMpred, which predicts that the protein possesses a number of transmembrane regions. Two potential helical arrangements were suggested – Model 1 and Model 2, with the most likely of the two being Model 1 due to its higher score (Table 5.1). For Model 1 each of five helices span the

membrane, and the protein has an N-terminal extension located inside the cell and a C-terminal extension located outside of the cell membrane. In terms of helix length, each is similar in size, although the overall scores for each do vary greatly, ranging from one at 2177 to another at just 870. Model 2 is constructed of four helices, each of more varying size than in Model 1. In this instance though, the individual helix scores are much closer together in value and the final predicted protein would have both the N-terminal and C-terminal extensions on the inner surface of the cell membrane (Table 5.1).

Model 1	**Helix length (a.a.)**	**Score**	**Orientation (in cell)**
1	22	2177	Inside – Outside
2	19	870	Outside – Inside
3	21	1008	Inside – Outside
4	20	1022	Outside – Inside
5	20	1477	Inside – Outside
Model 2			
1	22	1865	Inside – Outside
2	18	1495	Outside – Inside
3	24	1194	Inside – Outside
4	20	1477	Outside - Inside

Table 5.1 A table depicting the two predicted helical arrangements of CrtW by TMpred.

This is not the first time that the protein structure has been modelled of a member of the β-ketolase family. A study was recently conducted of the CrtW produced by the α-proteobacterium *Paracoccus* sp. Strain N81106 (Ye *et al.*, 2006). Part of the study included initial prediction of the secondary protein structure, which determined that four transmembrane helices were encoded by the amino acid sequence. This is the same as predicted in Model 2 for *M. xanthus* CrtW, which leaves both termini on the inside of the cell. Using this precedent it appears most likely that Model 2 is correct, rather than Model 1. Also in the same study a direct sequence alignment was carried out comparing the *Paracoccus* sp. Strain N81106 protein sequence with those of *Brevundimonas* sp. SD212 (α-proteobacterium) and an *Alcaligenes* species (β-proteobacterium). This was done in an attempt to identify any similarities that existed between the three distantly related organisms and to determine whether there were other additional key elements encoded, as predicted from the amino acid sequence. For the purpose of this study the alignment was repeated, but also incorporating recent sequence data from *Myxococcus xanthus* (δ-proteobacterium) DK1622, *Algoriphagus* species (Bacteroidetes/Chlorobi group) and

Bradyrhizobium (α-proteobacterium) ORS278 (Figure 5.1). Key features were found to be highly conserved in all of the proteins.

One of the features that was identified during the original comparison between the three species was the presence of three histidine-rich (His) motifs which were conserved in each CrtW protein (Ye *et al*., 2006). These His features were spread along the peptide chain, but it was predicted each was still located on the inner surface of the membrane. Two of the motifs are located between the second and third helices, and the other immediately following the fourth helix at the C-terminus. Further analysis of each individual motif, whereby a selection of the histidine residues were mutated, established that the majority were essential for effective CrtW ketolase activity in *Paracoccus* species. In any instances where enzymatic activity was still observed following mutation, it was always found to be greatly reduced. Identical His motifs were present in each species, including *M. xanthus* (Figure 5.1). They are all highly conserved, implying their importance, although the second histidine residue in His3 is substituted with an aromatic residue in both *M. xanthus* and *Algoriphagus* species. Mutation studies of this residue in *Paracoccus* species resulted in reduced overall levels of substrate interaction, but the motif itself still remained functional (Ye *et al*., 2006). All of the identified His motifs were previously proven to be involved in iron coordination, enabling the enzyme access to electron donors.

In addition to the three His motifs, a further conserved amino acid sequence was located between His2 and the third transmembrane helix. This motif was conserved in all of the six species and consists of DPDF. The motif is believed to play a similar role in aiding iron coordination to that of the His motifs. In addition there is a separate histidine residue located immediately after the final helix, which although not forming part of a larger motif may still play a significant role during catalysis. This is supported by the fact that in its absence ketolase activity is greatly reduced. Mutation of a single phenylalanine (F) amino acid residue in the *Paracoccus* sequence reduced ketolase activity (Ye *et al*., 2006). This residue was found to form part of a far larger motif encoded by the sequence W(*xxx*)F(*xxx*)Y (where *x* represents any amino acid residue). Once again it was a highly conserved protein region present in each of the six CrtW peptides and was predicted to be found on the cytosolic side of the cell membrane (Figure 5.1).

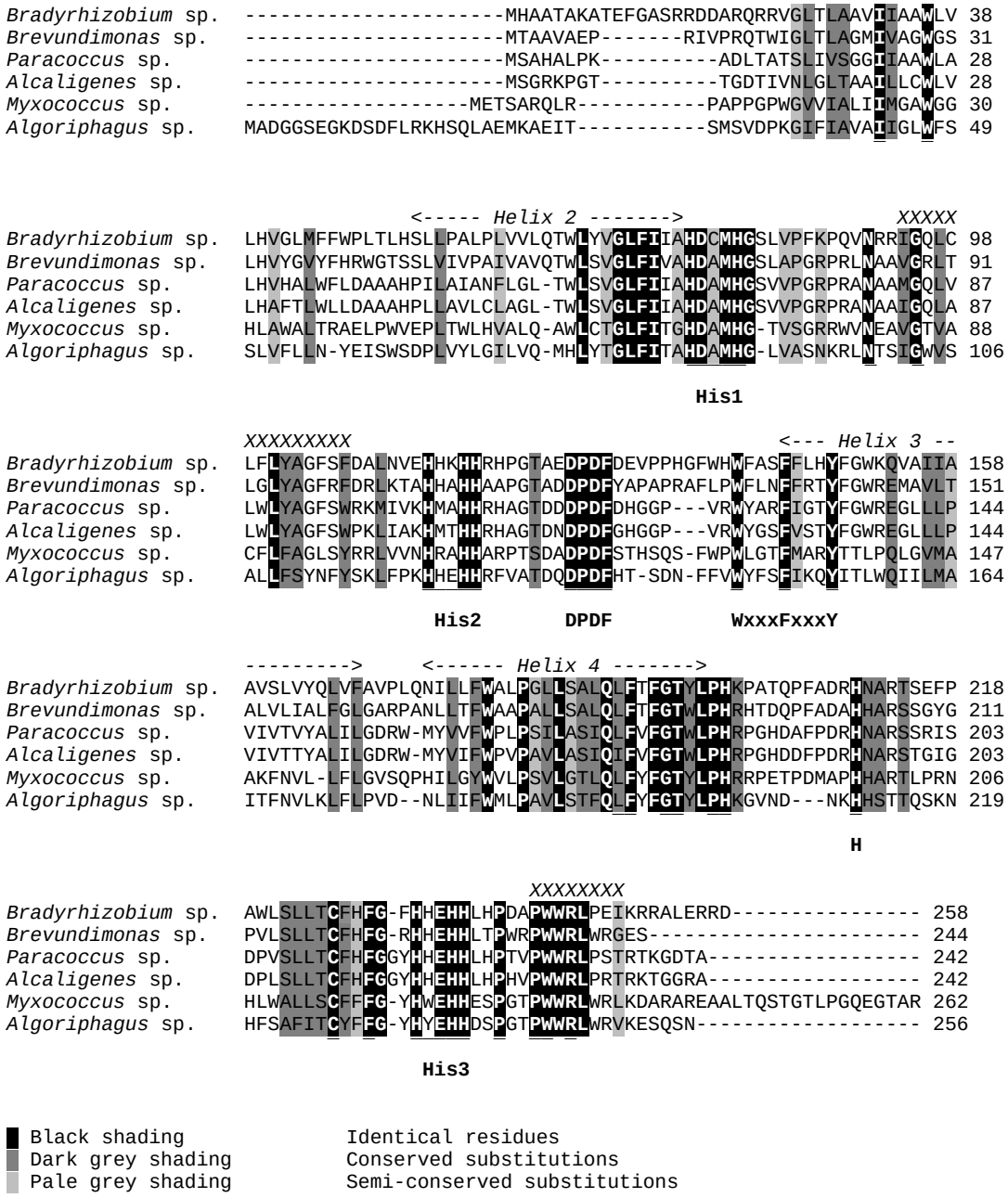
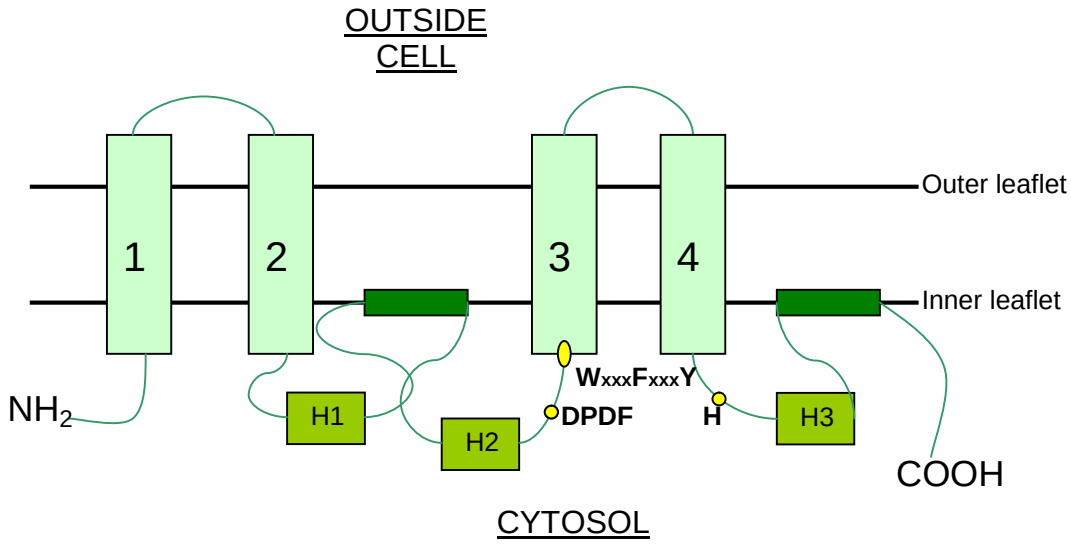
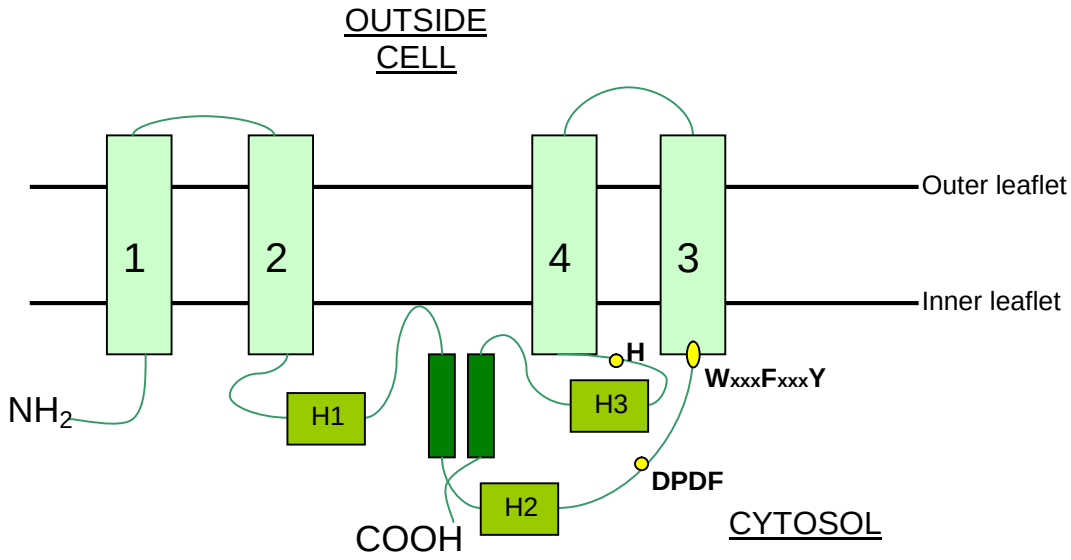


Figure 5.1 A ClustalW alignment of CrtW sequences for six bacterial species; *Bradyrhizobium* sp. ORS278 (GenBank accession number AAF78203), *Brevundimonas* sp. SD212 (GenBank accession number AB181388), *Paracoccus* sp. N81106 (GenBank accession number ABL09497), *Alcaligenes* sp. (NCBI accession number BAA09596), *Myxococcus xanthus* DK1622 (GenBank accession number ABF90074) and *Algoriphagus* sp. (GenBank accession number ABB88952). Any unlabelled underlined bases are previously identified from a comparative CrtW study as being highly conserved (Ye *et al.*, 2006). A string of *X*s indicate the area in which hydrophobic regions have been identified. The conserved motifs His1, His2, His3, DPDF, H, and WxxxFxxxY are marked and are believed to be involved in metal coordination (See text).





Helix domain 1:	Amino acids 17-30
Helix domain 2:	Amino acids 44-66
Helix domain 3:	Amino acids 137-156
Helix domain 4:	Amino acids 162-184

Figure 5.2 Two topographic models of the predicted secondary structure for CrtW from *M. xanthus*. The helices are coloured pale green (1-4), His motifs lime green (H1-3), the hydrophobic regions a deep sea green and conserved motifs labelled yellow. The difference in appearance is dependent upon whether the hydrophobic regions interact with each other or the membrane.

The final elements identified during the initial ketolase modelling process were two highly hydrophobic regions. These consisted of various hydrophobic residues, but all were located in the same regions in each CrtW (highlighted in Figure 5.1). Simplified versions of the predicted CrtW topographic model can be seen in Figure 5.2. The difference between the two models is whether the hydrophobic regions interact with the membrane or each other. To fully understand the role CrtW plays in *M. xanthus* carotenogenesis, experimental evidence was still required. In terms of carotenoid production the preferred method of study is usually through the creation of a gene knockout strain and examination of the resultant phenotype.

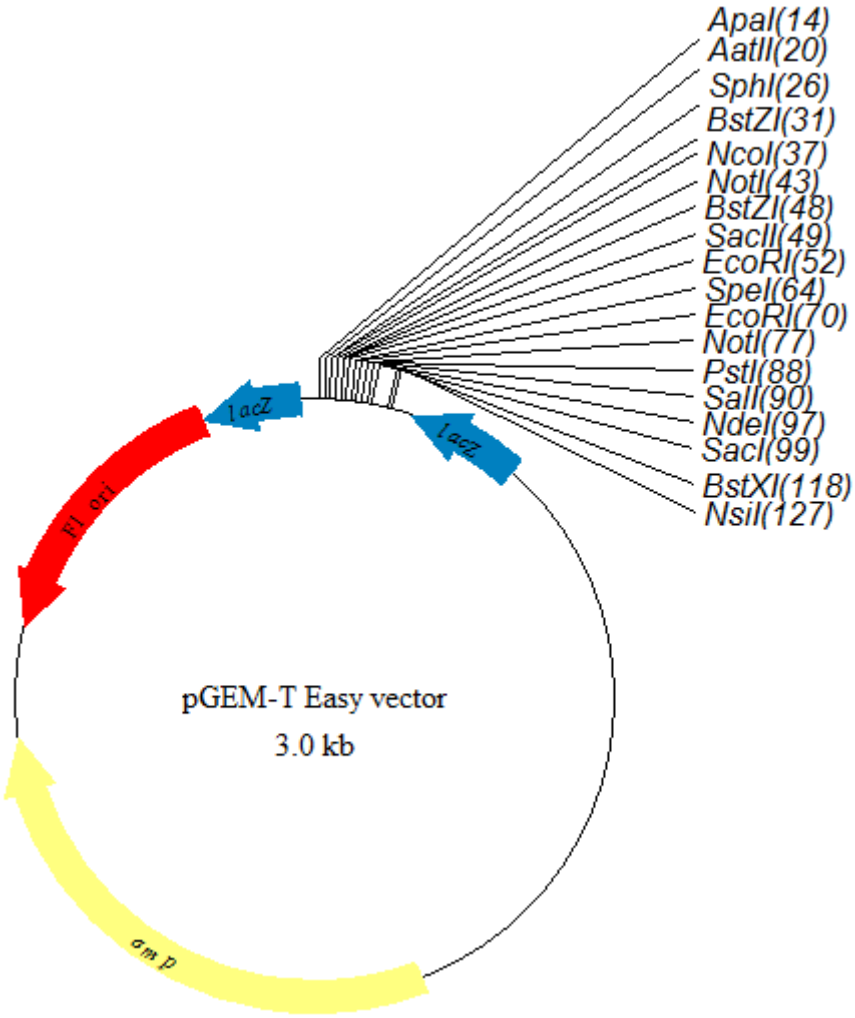
5.3. Construction of a knockout construct

A knockout strain is one in which the activity of a gene or a number of genes of interest have been removed. Through study of the resultant phenotype it should then be possible to determine the function of the missing gene. In terms of this study *crtW* was predicted to encode a ketolase that was involved in carotenoid modification. Working on the basis that the proposed carotenoid biosynthetic enzyme pathway for *M. xanthus* is correct, the gene should prove to be non-essential to the cell and carotenoid biosynthesis. If CrtW was absent then the secondary carotenoid, myxobactin (Kleinig and Reichenbach, 1970), should still be created through the pathway, but at an increased level. A second issue to consider is regarding the neighbouring genes of the *crtW* operon. Although the removal of *crtW* activity is desired, the surrounding operon genes must remain unaffected; otherwise additional functions may also be altered or lost. Considering the high levels of transcription previously recorded in the operon (Figures 2.7 and 2.8), it is possible that essential genes are encoded within the sequence. If these were disrupted during construction of the knockout strain then it would prove lethal to the cell. Principally this means the open reading frames downstream of *crtW*, *ord* and *mmb*, must remain the same following sequence alteration, and no other part of the operon should be altered except for *crtW*. There is currently no evidence of translational coupling of *crtW*, *ord* and *mmb*.

To successfully generate gene knockouts a range of experimental approaches have been developed. One of the most common is to insert a short sequence of random DNA into the middle of the gene of interest, thus disrupting its coding

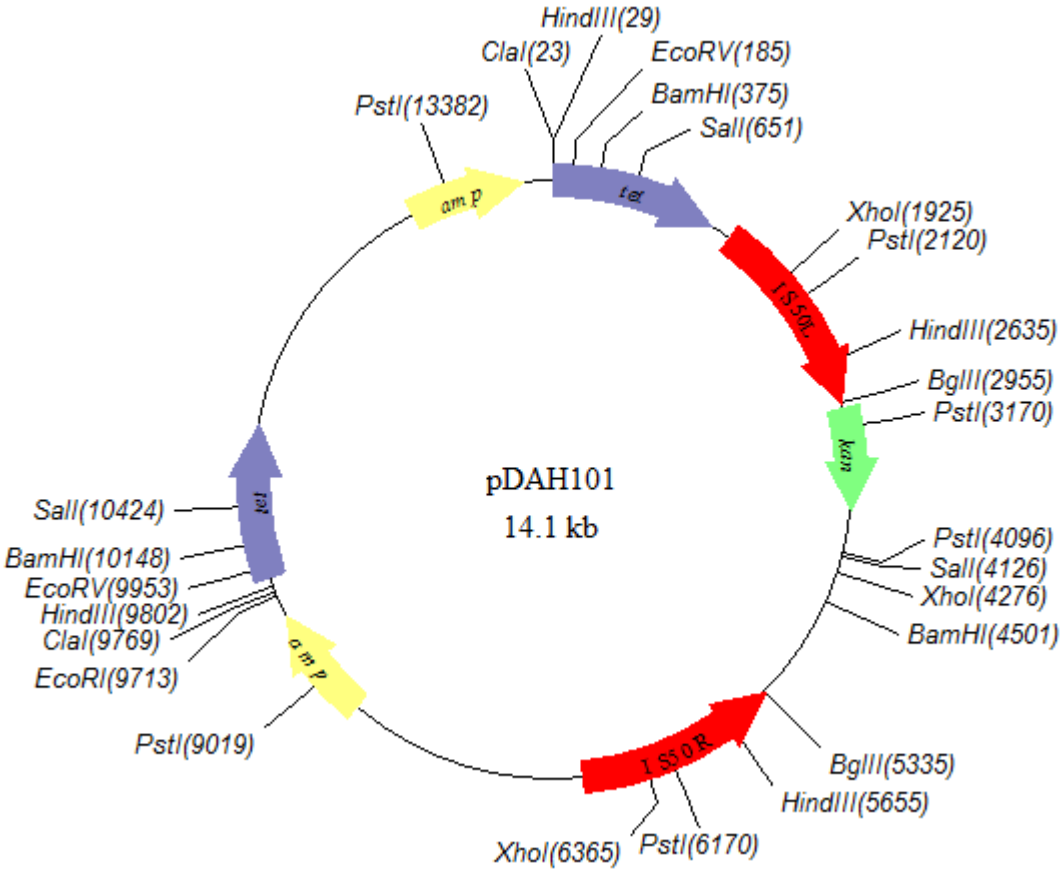
sequence. This is often achieved through the use of some form of transposon which creates a genetic 'scar' within the sequence. However if this method was applied to remove *crtW* activity it would also be likely to affect the downstream genes, as it would result in polar effects if it contains transcription and translation termination signals. The most sensible option is to completely remove the gene, but in doing so ensure that the downstream genes are unaltered. To aid in the final detection of the knockout, the *crtW* was replaced with an appropriate antibiotic resistance gene. When the operon is transcribed the inserted resistance gene should be expressed as well and we know *crtW* is constitutively expressed. This also offers a simple method for successful knockout mutant selection. *M. xanthus* displays a natural resistance to a range of antibiotics including ampicillin (Karwowski *et al.*, 1996), but is sensitive to kanamycin, which was chosen as the resistance gene for insertion.

In order to completely remove the gene and subsequently replace it without disrupting the operon, crossing over of two homologous regions in a *kan*-carrying plasmid and the bacterial genome is required. In order to achieve this, a plasmid firstly needs to be created containing *kan* flanked on either side by *pdp* and *ord*, the sequences naturally flanking *crtW* in *M. xanthus.* Through the processes of crossing over and homologous recombination between the plasmid and the gene, *crtW* should then be substituted by *kan*. Large flanking regions for the gene were selected, over 1kb in size, in order to increase the likelihood of recombination. The primary intention was to use the experimental approach of electroporation to transfer the plasmid into the bacteria, so vector size was an important consideration. It was discovered that in previous electroporation procedures with *M. xanthus* (unpublished data) a relatively small pGEM vector had been used to transfer genetic material into similar strains (Figure 5.3). A second plasmid was used as a source of the *kan* gene, pDAH101, which consists of Tn*5* inserted into a pBR322 dimer (Figure 5.4). The *kan* gene itself is approximately 800bp long, so is of a size that can be amplified from the plasmid and then used for insertion into the pGEM vector. To enable insertion of the flanking regions RE sites were added to either end of the sequence, before the start of each flanking region and the *kan* gene. This would result in the addition of six bases either side of both the gene and flanking regions within *M. xanthus*, which would have no overall effect on operon transcription.



F1*ori*:	2380-2835bp
lacZ (in two pieces):	166-345bp & 2836-2996bp
amp:	1337-2197bp
T7 promoter region:	2999-5bp
T7 RNA polymerase transcription site:	1bp

Figure 5.3 The pGEM®-T Easy vector, with key restriction enzyme sites labelled and the coding regions highlighted.



tet:	36-1273bp
IS50L:	1440-2974bp
kan:	2991-3782bp
IS50R:	5316-6856bp
amp:	8704-9564bp
tet:	9859-11046bp
amp:	13067-13927bp

Figure 5.4 The pDAH101 (2xpBR322::Tn*5*) vector, with key restriction enzyme sites labelled and the coding regions highlighted.

Name	Sequence	Description
KOI F	CCGCGGCTACAAGGCCGCCTTCGAGGCCTT	PDP fragment generation
KOI R	GCTAGCCTCACCTGCTTTCTGGCCATCGCC	PDP fragment generation
KOII F	GCTAGCCGCGCGGGCGGCCCGGGCCGTTAT	ORD fragment generation
KOII R	GAATTCTCAGTAGAACCGGCGCAGGTCCAT	ORD fragment generation
KOKAN F	GCTAGCATGATTGAACAAGATGGATTGCAC	*kan* KO fragment creation
KOKAN R	GCTAGCGAAGAACTCGTCAAGAAGCCGATA	*kan* KO fragment creation

Table 5.2 The primers used to obtain the two flanking regions PDP and ORD, and the *kan* insertion region for construction of the KO construct. *Nhe*I ends are highlighted in blue, *Sac*II ends red and *Eco*RI ends green.

Considering the unique RE sites already present in the plasmid vector, PCR primers were designed in order to obtain two 1kb gene flanking regions, PDP and ORD (Figure 5.5), and for the isolation of the *kan* gene (Table 5.2). The restriction sites used were *Nhe*I for either end of the *kan* region, and *Eco*RI and *Nhe*I, or *Sac*II and *Nhe*I respectively, for each of the flanking regions. The approach was taken to firstly insert one of the flanking regions into the pGEM vector and then ligate the second one alongside it. Further digestion of the plasmid using *Nhe*I should then enable direct *kan* insertion. Following a slight adaptation of the PCR protocol to compensate for the larger flanking region size, PDP and ORD were both successfully amplified from a *M. xanthus* genomic DNA sample. As was the case with the S1 and S2 fragments previously, they were then inserted into an intermediate 2.1-TOPO vector. This enabled production of a high copy number of each of the flanking regions (Figure 5.5). Insertion of the PDP region into the 2.1-TOPO vector and selection through blue/white screening proved straightforward, but addition of ORD to the vector proved more problematic. The cause of this was eventually established as being a combination of issues regarding the intermediate 2.1-TOPO vector; including the introduction of additional random restriction sites to the fragment and also the poor quality of the initial ORD PCR product, with incomplete RE sites present. Repetition of the entire process, fresh intermediate insertion and subsequent sequencing rectified this problem. The fragment was ligated with PDP in the 2.1-TOPO vector through digestion of the restriction sites *Nhe*I and *Xho*I, ensuring the *Eco*RI regions already present in the vector were not disrupted (Figure 5.6). A combination of *Eco*RI and *Sac*II RE digestion sites was then used to insert PDP and ORD into pGEM, which generated a plasmid that contained the two flanking regions ligated together within the pGEM MCS (Figure 5.7a).

PDP Fragment (1016bp)

7486421 CTACAAGGCCGCCTTCGAGGCCTTCGTGGGCCTGCGCGACGACGCGGAGACGCAGAAGCTGGCGAAGTTCAGCGGG
CAGCTCCAGGGGCTGGAGAACAACCTCCCCATCGACCCGAAGCTGCGCAACCCGAAGCTGGGCGCGCTGGCGCCCATCCGCGTCAT
CAACAGCCTGTTCTCCTCCGGTGACGGCAACCGGGGCGTGCAGACGGCCGCCTTCAACCTGCCCAACGACGAGCGGGTGTCGGAGA
AGATGGGCTCCAAGCGCGTGATGCTGAAGAACGTGCAGGAGGCCAAGTTCGAGCGCGTGCTGCTGCCCATCGCCAAGGTGGCCCTC
ACCCCGGCGGACCAGAAGGACGTCTCCTTCGATGCCTTCTTCACGCACATCTTGATGCATGAGCTGATGCACGGCCTGGGGCCCAG
CAACATCACCGTGGGTGGCAAGGCCACCACCGTGCGCAAGGAGCTCCAGTCGGCCTCCAGCGCCATCGAAGAGGCGAAGGCGGACA
TCTCCGGCCTGTGGGCGCTCCAGCGCCTGGTGGACACCGGCGTCATCGACAAGTCGCTGGAGCGCACCATGTACACGACGTTCCTG
GCCTCCGCCTTCCGCTCCATCCGCTTCGGCGTGGACGAGGCGCACGGCAAGGGCATCGCCGTGCAGCTCAACTACTTCCTGGACAC
CGGCGCGGTGAAGGTGAACGCGGACGGCACTTTCTCCGTGGTGCCGGCGAAGATGAAGAAGGCCGTCATCTCGCTGACGAAGCAGC
TCATGGAGATTCAGGGCCGCGGCGA**CCGGA**AG**G**C**C**GCCGAGGCGCTG**CTGGCG**AAGCTCG**G**C**GT**GGTGCGCCCGCCCGTGCAGCGC
GTGCTGGAGCGCCTCAAGGACGTGCCGGTGGACATCGAGCCGCGCTACGTCACCGCGGAGGAGCTGGTACGCGACGTGAAGAAG**TA**
GCCACCGGCCCATTGCCCCGCCCCGCCGCCTGGGTATTTCCCATGCGCGGCGGGCCGGCGATGGCCAGAAAGCAGGTGAG
7487437

ORD Fragment (1023bp)

7488226 CGCGCGGGCGGCCCGGGCCGTTATACAGGGGCC**ATG**CGCGACAAGCCGCCCGCTGAGCCGCCCAGCTCCGAAGTCA
CCCCCGAGAAGACGTACCTGCGCCGGCGCGAGCTGCTGAAGAACGCGGGCCTCTTCGCGGGCACGGCCGTCGCCGTCGCCGGAGGG
CTGCACCTGCTGGGCCGCAAGCAGACGCGCCCCATGGAGCGCTTCGTCCCGGACGCGGGCCTGGTGGAGCAACCGGTGGCGCAGGC
GATGGGCCCCTTCGACACGGACGAGCCGCGCACGCCCTACGAGGACGTCACCACCTACAACAACTTCTACGAGTTCGGCTTCGACA
AGAACGACCCGGCCCGCTTCGCGCACACGCTGAAGCCGAAGCCGTGGAGCGTCGTCATCGACGGCGAGGTGCATAAACCGCGGACG
GTGGACGTGGAGCAGCTCACGTCCTGGTTCTCCCTGGAGGAGCGCGTCTACCGCATGCGCTGCGTGGAGGCCTGGTCCATGGTGAT
TCCGTGGCTGGGCTTCCCGCTGGCGGCGCTGCTCCAGCGCGTGGAGCCCACCAGCCATGCGAAGTACGTCGCCTTCACCACGCTGC
TGGACCCGGAGCAGATGCCGGGCCAGCGCCGCGCCCTGTTGGATTGGCCGTACACGGAAGGACTGCGCCTGGACGAGGCGATGAAC
CCGCTCACGCTGCTGGCCACGGGGCTCTACGGCCGGCAACTGCCCAACCAGAACGGCGCGCCGCTGCGGCTCGTGGCTCCTTGGAA
GTATGGATTCAAGGGCATCAAGTCCATTGTCCGCATCAGCCTCACGCGGGAGGAGCCCATGACGACGTGGCGCCTGTCCGCGCCGC
GCGAGTATGGCTTCTACGCCAACGTGAATCCTTCCGTGCCCCATCCGCGCTGGAGCCAGGCCAGCGAGCGCCGCATCGGCGACTTC
GAGCGCCGCCCCACGCTGCCCTTCAATGGCTACGCGGAGCAGGTGGCCCACCTCTACACCGGCATGGACCTGCGCCGGTTCTAC**TG**
A 7489249

Red: *pdp*
Blue: Oxidoreductase (*ord*)
Bold bases: Start/stop codons
Bold and underlined bases: Possible CarQ dependent promoter P2

Figure 5.5 The knockout flanking region fragments from the *crtW* operon. The numbers represent sequence location in the *M. xanthus* genome.

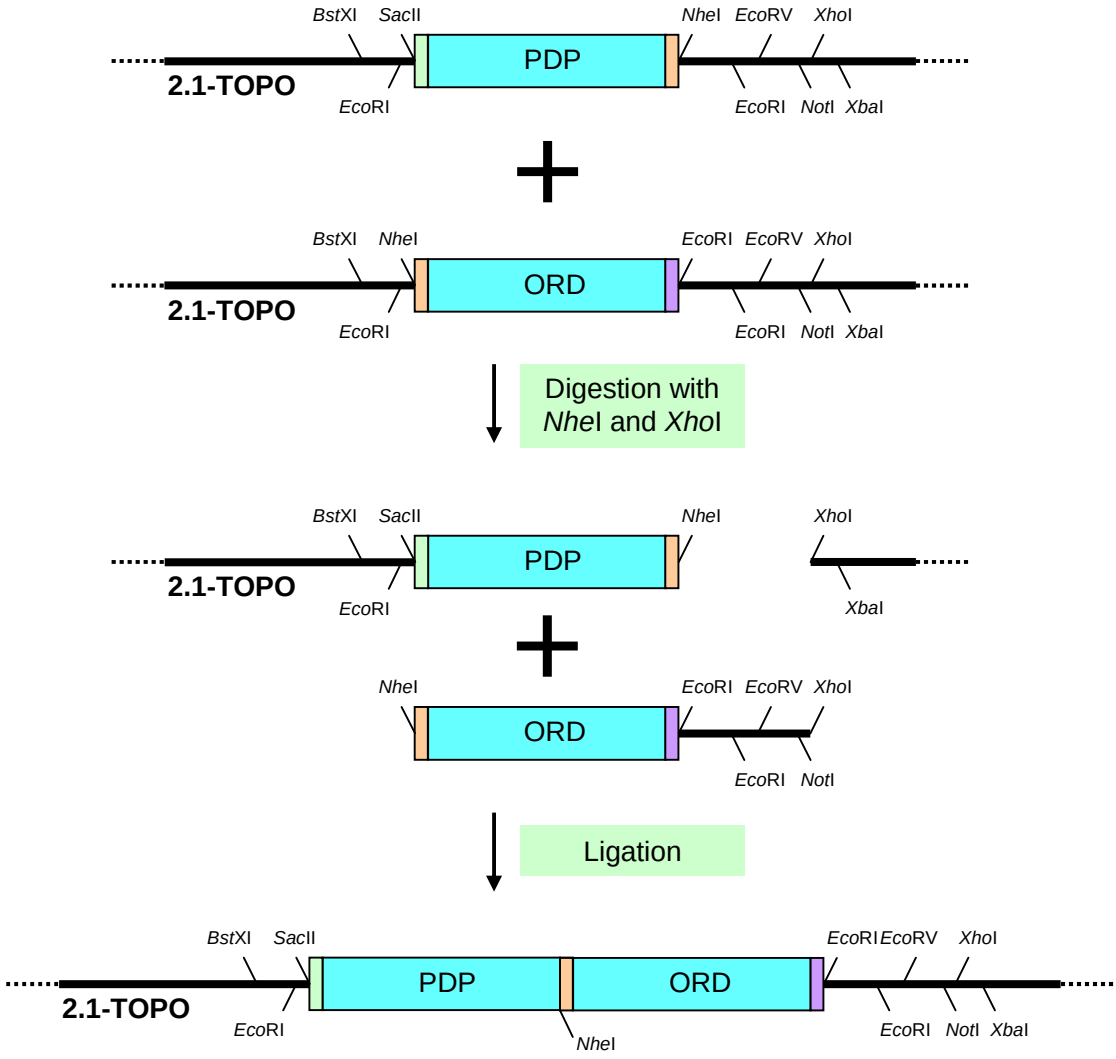
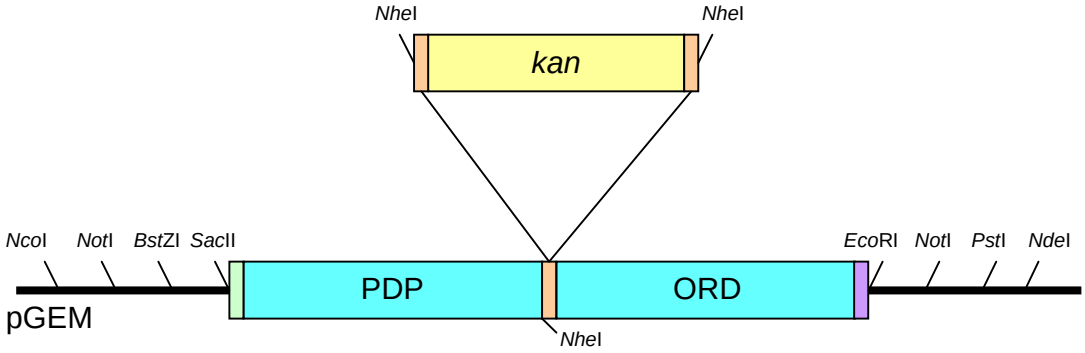
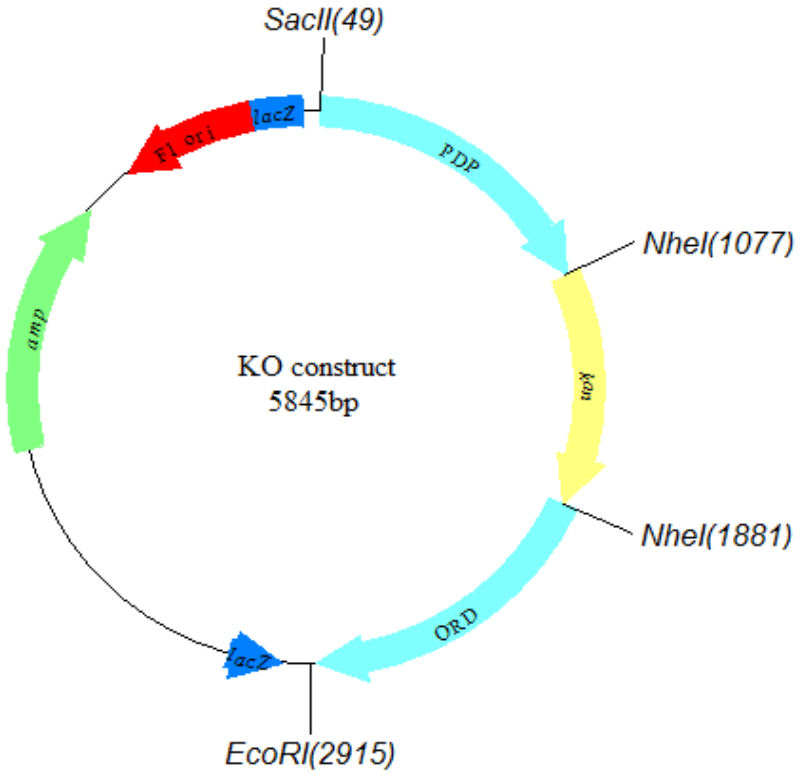


Figure 5.6 The ligation of PDP to ORD in the intermediate 2.1-TOPO vector. RE site location is not shown to scale.

a)



b)



PDP:	60-1076bp
kan:	1077-1881bp
ORD:	1882-2905bp
F1*ori*:	5225-5680bp
lacZ (in two pieces):	3011-3190bp & 5681-5841bp
amp:	4182-5042bp
T7 promoter region:	5844-5bp
T7 RNA polymerase transcription site:	1bp

Figure 5.7 *a)* The intended insertion of *kan* between the PDP and ORD regions using added *Nhe*I sites *b)* The novel 'KO construct' constructed through the combination of a pGEM®-T Easy vector and three PCR fragments. The key restriction enzyme sites and the coding regions are highlighted.

A similar procedure was then carried out on pDAH101 to amplify the *kan* gene. It too was then transferred to the intermediate 2.1-TOPO vector, but greater care was taken to ensure an accurate insert was generated by sequencing the products. This verified that all of the RE sites were intact, unaltered in the plasmid, although a number of attempts were required to obtain an appropriate sequence. Initial results following sequencing indicated that either a small percentage of bases were incorrectly inserted throughout the gene or the sequenced regions were incompletely amplified. The final step in knockout plasmid creation was the ligation of *kan* between the PDP and ORD flanking regions (Figure 5.7a). A result of the approach designed to achieve this meant insertion of *kan* with two *Nhe*I sites, which proved problematic. This was primarily because after the ligation procedure, the only way of verifying that the plasmid did contain the inserted gene was through the PCR and then sequencing. Despite containing *kan*, with no promoter present to initiate gene transcription, kanamycin resistance was not conferred to the KO vector. The promoter was not present to ensure there was minimal disruption to other promoter regions in the *crtW* operon. Consequently it was not possible to identify transformants through growth on kanamycin containing media. All transformants were tested and proved to contain only the religated plasmid. To reduce the background levels of religation, a dephosphorylation step was carried out following digestion of the intermediate KO construct with the PDP and ORD inserts ('Dephosphorylation' Methods). This effectively removed the phosphate groups from the two cleaved vector ends, preventing the two from religating back together. To ligate they must now instead bind to a DNA strand that possesses phosphate groups, in this instance the *kan* insert. Of the resulting transformants produced, which were much reduced in number, a significant proportion were still religation products, most probably a result of failing to completely dephosphorylate all construct ends. This time, in addition, there were also a number of transformants that proved to contain the insert. Although the initial ones tested contained an inverted copy of the gene (due to identical RE sites at both ends), a construct was also identified that encoded the correct sequence in the right orientation. The construct created was named 'KO construct' (Figure 5.7b).

5.4. *crtW* knockout

To observe the effects caused by the knockout, it is only strictly necessary to remove the gene activity from a wild-type *M. xanthus* strain. However, attempts were made to generate knockouts in four of the strains used in this investigation; DK101, DK717, DK718 and UWM 303, though no carotenoids should be produced in the latter strain. They could then act as controls for each other generating comparable results. A number of results were possible following gene knockout. Firstly gene removal may lead to a clearly visible phenotypic effect, which would be primarily observed as a colour change. As the principal carotenoid is no longer being produced the intensity of colour may either decrease or increase. A completely different colour may even be produced, caused by alternative carotenoid creation. It is also possible that there may be no visible alteration in colour. In this case proof would require the chemical identification of the carotenoids produced. Any other physical cell alteration detected would imply that the gene does not function as originally expected in carotenogenesis. The knockout created may potentially have no observable phenotypic effect on the cell at all. This would suggest that the gene was not functioning as a CrtW protein. The final possible scenario is that *crtW* removal gives rise to a lethal mutation. In this case no detectable colony growth would be observed following recombination, indicating that *crtW* was an essential gene in *M. xanthus*.

To confirm homologous recombination has taken place between the KO construct and *M. xanthus* DNA (Figure 5.8), the primary selection criterion is resistance to kanamycin. If following electroporation with the KO construct the tested *M. xanthus* strains produce colonies on DCY Kan^{50} plates, then it is highly probable that kan has been incorporated into the genome. Through a PCR of the resistant colonies the extent of recombination can be determined. In the case of only a single crossover successfully occurring the resultant PCR should be positive for both *kan* and *crtW* presence. If a double crossover has occurred then the PCR will show positive *kan* banding but no *crtW* present (Figure 5.8).

Using the previously adopted protocol for electroporation, the procedure was carried out using the four strains and the KO construct with selection on kanamycin media. All initial attempts resulted in plate contamination and very few *M. xanthus* colonies for either UWM303 or DK101. A number of colonies were produced on the

DK717 and DK718 plates though, numbering between 30 and 40 on each. Of these 20 were selected in total from each strain and the colonies replated onto separate DCY Kan^{50} plates. Repeat of the electroporation procedure using the UWM303 and DK101 strains, resulted in similar colony numbers being observed, although slightly reduced, with 50-60 produced in total for each strain. 20 of each were again streaked onto further selective media to verify that the colonies produced were kanamycin resistant *M. xanthus*. All tested colonies did prove to be resistant *M. xanthus*, although they grew in two distinct different colourations in the case of the DK101 strain; tan or a bright yellow. This phenomenon has been previously recorded (Kuner and Kaiser, 1981), and it is believed that the colour phase variance may influence spore development (Laue and Gill, 1995). It is clear this variance has no effect on non-sporulating colonies (personal communication, D. A. Hodgson), but both tan and yellow colonies were still sampled if they were present.

As every colony produced was kanamycin resistant, the PCR was carried out on all of the strains to verify whether the plasmid was present or flanking region crossover had occurred. Two novel primer pairs were designed that encompassed a region either end of *kan* and then a small region from the *M. xanthus* genome flanking regions. Following a PCR for each colony, none resulted in the production of agarose gel banding after electrophoresis. A separate check for *kan* gene presence using the original primers designed for amplification of *kan* from pDAH101 was conducted, and each gave a positive result. This confirms that all the colonies tested contained *kan* that was being successfully expressed and all cells were kan^R. The samples were additionally tested for plasmid presence using the 'KO primers' originally used during the creation of the KO vector, and all the returned results were negative. If any crossover is occurring, then it is likely that plasmid and genome intermediates would exist, and be detectable at this stage (Fig 5.8). As an additional control the samples were all subjected to further PCR to establish whether *crtW* was still encoded, each producing a band following gel electrophoresis. The results indicate that there has been no removal of the *crtW* gene, but all generated colonies following electroporation are still kanamycin resistant. Despite repetition of the experiment identical results were once again observed.

Following these results a range of additional PCR controls were carried out on the original isolated colonies using existing primers and primer pairs including: 'DOU CHECK F' with 'KAN^R REV' and '3^{rd} REV'; '3^{rd} REV' with '3^{rd} FOR';

'KAN^R FOR' with '3rd REV'; 'KAN^R FOR' with '4th REV' (Figure 5.9 and 'Primers' Methods). The only visible bands obtained after gel electrophoresis were for sequences normally present in *Myxococcus xanthus* genomic DNA and none from the inserted plasmid, with the exception of the *kan* gene alone. Comparison of a control DCY Kan^{50} plate of electroporated unaltered DK101 cells to a plate of an identical strain electroporated with the KO construct showed the latter to have three times as many resistant colonies growing. This suggests the results are not due to spontaneous kanamycin resitance alone. A second identical KO generation was attempted for the final gene in the *crtW* operon, *mmb*, which also proved to be unsuccessful.

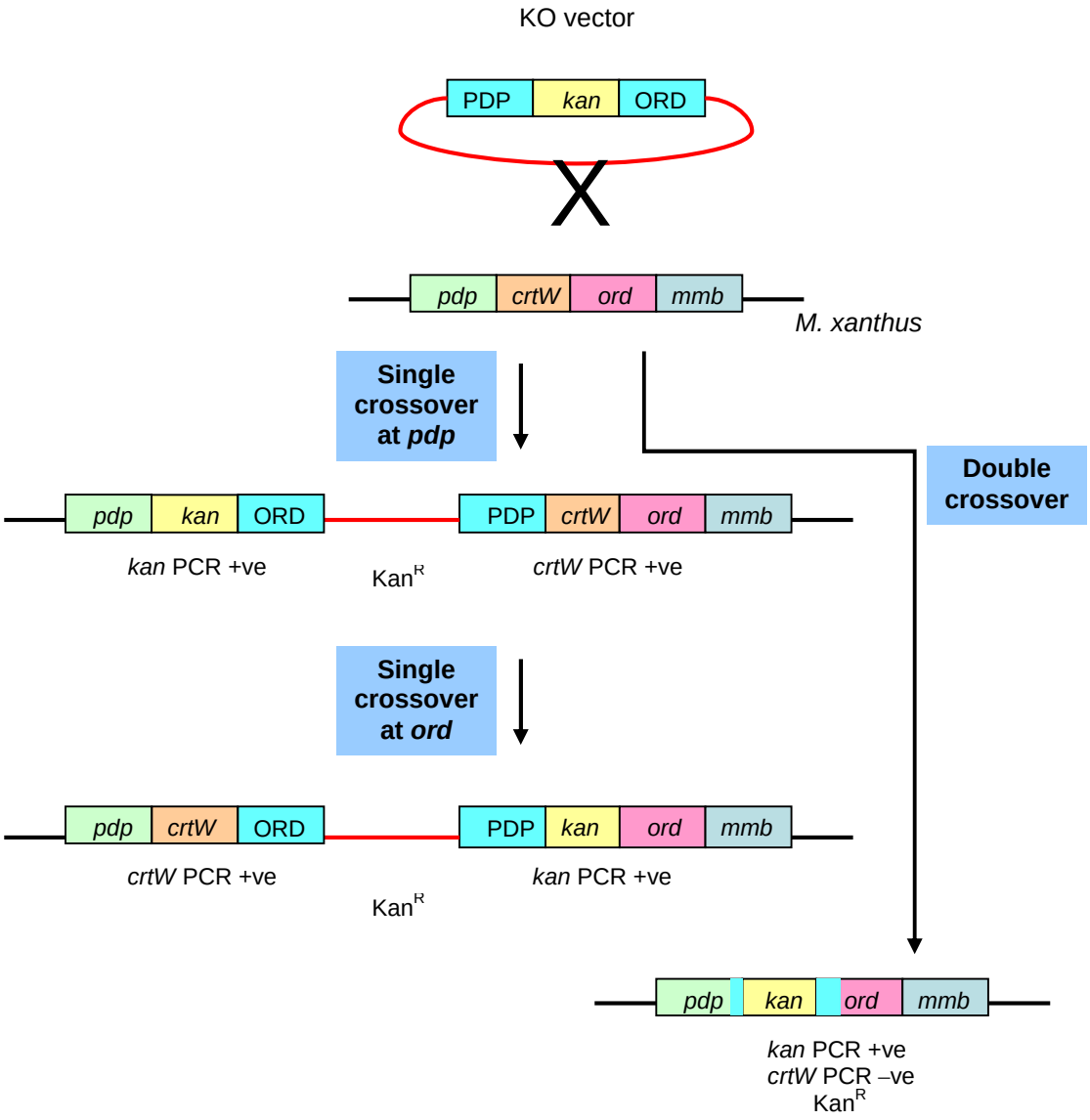


Figure 5.8 The three possible crossover events following transformation of *M. xanthus* strains with the KO construct

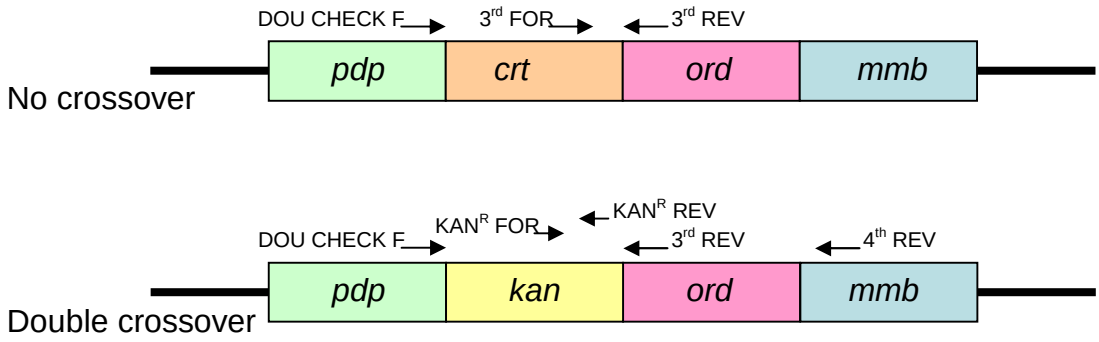


Figure 5.9 The location of the primers used to determine whether a double-crossover had occurred between *M. xanthus* genomic DNA and the KO construct. The primer sequences can be found in 'Materials and Methods'.

5.5. Knockout result analysis

Once the *kan* gene is taken up by the cell, any level of expression will confer some degree of kanamycin resistance to the cell. This could provide a safe period whereby spontaneous kanamycin resistance mutations could arise. As a result transformation with any resistance determinant will increase the spontaneous resistance mutant frequency to that drug, above that of the background level.

The probable reason for the failure of the KO construction approach is that in order for the *kan* gene to be expressed, recombination between the PDP fragment and its homologue in the chromosome is required before a promoter is attached to the *kan* sequence. As a result not only does the DNA have to enter the cell, but also be fully integrated into the genome through homologous recombination before enzymatic resistance can arise. It is entirely possible that spontaneous kanamycin resistance frequency, boosted as it is by the *kan* fragment transformation effect mentioned above, is greater than the frequency of transformation plus homologous recombination. However this effect does still not explain the *kan* gene detected in *M. xanthus* resistant colonies following KO construct electroporation. Besides relying solely on *kan* it would be beneficial to include an additional selective marker elsewhere in the KO construct, such as tetracycline. Using counter-selection would have made identification of single crossovers and recombination events easier. In this case a form of replica plating could have been used to aid the identification of transductants in which a desired double-crossover had taken place.

An alternative strategy is to insert a kanamycin resistance determinant with promoter downstream of the *mmb* gene which belongs to a *crtW* operon from which the *crtW* gene has been deleted. It should then be possible to replace the wild-type *crtW* operon with the *crtW*⁻ operon by selecting for kanamycin resistance and using PCR to identify coinheritance of the mutant operon (Figure 5.10). A similar approach was adopted during analysis of the *carQRS* region of *M. xanthus* (McGowan *et al.*, 1993). Due to time constraints this approach could not be taken.

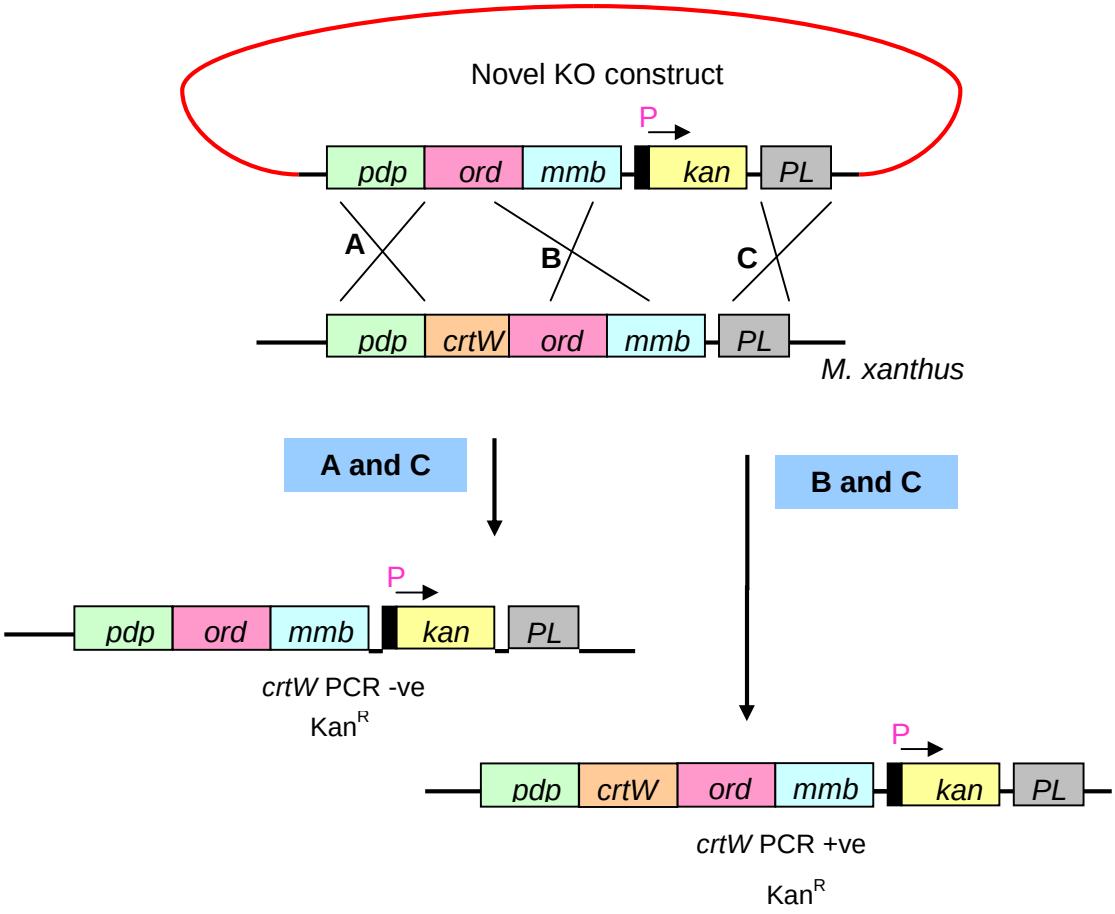


Figure 5.10 The knockout of *crtW* via homologous recombination between a plasmid carrying *kan* and its promoter and the *M. xanthus* genome. *PL* is the next gene in the genome after the *crtW* operon, a putative lipoprotein, and **A**-*C* represent three possible crossover events.

An alternative strategy was provided by Zhao *et al.* (2008), which instead of knocking out *crtW* allowed the 'knocking in' of the gene, i.e. over-expression of *crtW* in its natural host. This followed the discovery of novel autonomously replicating

plasmids in *M. fulvus*, shown to successfully transform *M. xanthus* DZ1 (See 'Chapter 6').

5.6. Discussion

The results obtained from secondary modelling studies indicate that CrtW is a membrane bound protein, most likely spanning the cell membrane four times. The presence of both hydrophobic regions along with the highly conserved histidine motifs and individual residues, create protein loops for substrate interaction within the cell. The similarity between *M. xanthus* CrtW predicted structure and a number of similar proteins produced in a range of bacteria indicate that the *crtW* gene investigated does indeed encode a ketolase.

The strategy adopted for gene replacement did not appear to be successful, with what appears to be a large number of kanamycin resistant mutants being generated. It is still possible that a double crossover event between the KO vector and the *M. xanthus* genome could result in a lethal mutation, but with no evidence of homologous recombination between the two, no conclusions can be made. In order to verify the effects of *crtW* removal an alternative approach is required, such as gene knockout via the insertion of *kan* with its own promoter (Figure 5.10). Alternatively *crtW* gene activity could be investigated through the over-expression of *crtW* within *M. xanthus* (Zhao et al., 2008), which would not require any alteration of the bacterial genome (See 'Chapter 6' for further details).

Chapter 6

CrtW protein characterisation

6.1. Introduction

Myxococcus xanthus DK1622, and its progenitors DK100 and DK101, does not contain any autonomously replicating episomes and the vast majority of exogenous episomes are unable to replicate autonomously of the chromosome when introduced into *M. xanthus*. Thus it has been impossible to over-express proteins in the bacterium using multi-copy plasmid vectors, which rely on increased gene dosage and highly-expressed regulated promoters to be effective. Previous genetic manipulation of *M. xanthus* has relied on integration of vectors into the chromosome and so gene dosage is limited to a single copy.

Despite concerted efforts to locate an autonomous episome for *M. xanthus*, none were identified, which remained the case until very recently. An attempt to locate self-replicating plasmids within a *Myxococcus* species using a different experimental approach proved successful (Zhao *et al.*, 2008). Individual cell cultures of myxobacteria produce a significant amount of extra-cellular slime which make investigation of plasmid content awkward. In this study, a plasmid isolation method was adopted that had previously been used with *Streptomyces* (Kieser, 1984), and applied to a strain of *Myxococcus fulvus*, a close relative of *M. xanthus*. This resulted in the eventual discovery of a unique 18.5kb plasmid, designated pMF1. Characterisation of the plasmid showed it to have a similar G+C content to the host, 23 predicted ORFs and approximately 86% sequence similarity to the host genome. An important piece of information absent from the initial study was how the plasmid replicated. No replicon was identified during bioinformatic analysis of the plasmid sequence.

Information on plasmid replication is essential, as without these details any alteration of the plasmid could disrupt plasmid replication. In order to determine whether the plasmid could successfully self-replicate, a 14kb fragment was inserted into an unrelated kanamycin resistant plasmid, pZJY1 (Figure 6.1).

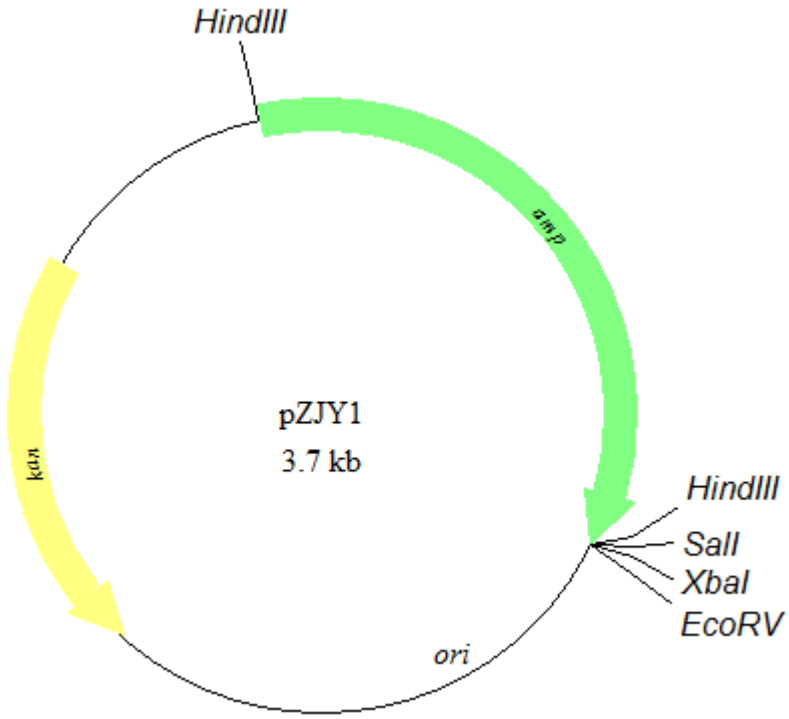


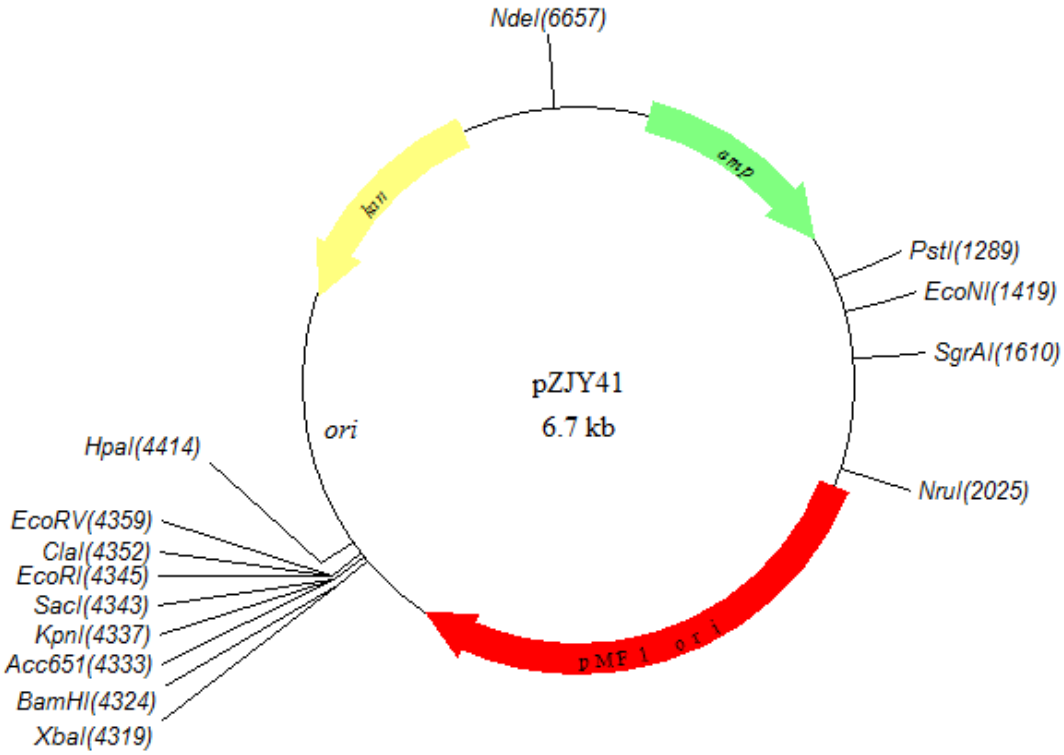
Figure 6.1 A schematic map of the plasmid pZJY1 (Zhao *et al*., 2008), with key restriction enzyme sites labelled and the coding regions highlighted.

Previously this had been shown to be unable to self-replicate in *M. xanthus*. The resultant plasmid, designated pZJY2, when used to electroporate a *Myxococcus xanthus* strain produced kanamycin resistant colonies (at a frequency of $4x10^2$ cfu/µg DK1622 DNA). Analysis of plasmids in the kanamycin resistant cultures found they were much smaller than the original construct, being only 5-8kb in size, and so they had undergone deletion. The smaller plasmids were found to contain regions from the original pZJY2 vector, and through individual plasmid sequencing, the key plasmid replicon was identified as pMF1.14 (Zhao *et al.*, 2008). A stable *M. xanthus* plasmid had now been effectively created which was stably inherited. To successfully utilise the smaller plasmid for the transfer of genetic information into a *M. xanthus* host, it also needed to be capable of replication in *E. coli*. However the pZJY2 deletant plasmids extracted from *M. xanthus*, pZJY7 and pZJY15, when tested were unable to transform *E. coli* cells. They were detected at frequencies of $1x10^5$ and $1x10^4$ cfu/µg DK1622 DNA respectively. Analysis of the two plasmids revealed they lacked a portion of the *E. coli* origin of replication, plus part of the *amp* gene encoded in pZJY1. Further screening of a larger number of *M. xanthus* transformants containing

pZJY2 deletants yielded a pair of plasmids whose size was unaltered following transformation and replication. The two different forms, designated pZJY41 (Figure 6.2) and pZJY156 (Figure 6.3) were both stably inherited and contained an entire copy of the *E. coli* origin region, so both were able to transform *E. coli* and *M. xanthus* cells to kanamycin resistance. The two were recorded at similar frequencies to the unmodified pZJY7 and pZJY15 in *M. xanthus* DK1622, estimasted to number between 10-20 copies per individual cell.

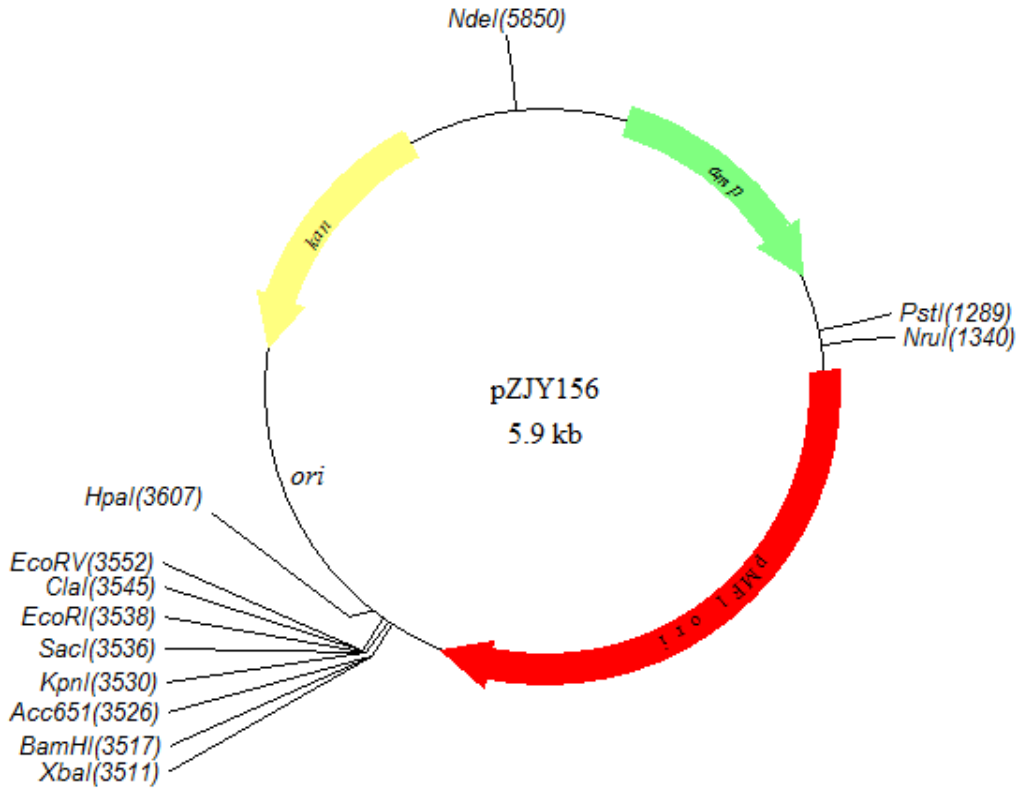
The two self-replicating plasmids (Figures 6.2 and 6.3) differ from each other in size by approximately 800bp. This is the result of the presence of an additional DNA sequence between the *kan* gene and pMF1 *ori* region in pZJY41. This does however provide the larger plasmid with a greater number of unique RE sites. Studies of each plasmid within *M. xanthus* recorded that both were capable of autonomous replication, although for stable inheritance they still need to be continually selected for by the kanamycin resistance conferred by the plasmid. Despite the plasmids proven to be capable of self-replication this selective pressure is still required, which may explain why it proved so difficult to identify similar plasmids previously. Without the use of antibiotic selection, after 42 generations 92% of *M. xanthus* cells sampled were found to have lost the plasmid (Zhao *et al.*, 2008).

In order to verify the effectiveness of the plasmid under experimental conditions it was used in conjunction with a *M. xanthus pilA* mutant strain. The mutant individuals are unable to create pili, so the intention was to compliment the mutation through the insertion of an unmutated *pilA* gene along with its promoter into the strain using the two novel shuttle vectors. The *pilA* strain was transformed by electroporation with the $pilA^{+}$ constructs. The kanamycin resistant transformants were able to produce pili and were thus capable of social motility (Zhao *et al*., 2008). The creation of a self-replicating vector provides the opportunity for complementation studies and development of protein expression vectors.



amp:	296-1090bp
pMF1 *ori*:	2114-4018bp
kan:	5413-6273bp

Figure 6.2 The pZJY41 vector, with key restriction enzyme sites labelled and the coding regions highlighted.



amp:	296-1090bp
pMF1 *ori*:	1429-3333bp
kan:	4606-5466bp

Figure 6.3 The pZJY156 vector, with key restriction enzyme sites labelled and the coding regions highlighted.

6.2. Over-expression of *crtW*

Previous attempts at construction of a gene knockout for *crtW* proved to be unsuccessful (Chapter 5). The creation of vectors, pZJY41 (Figure 6.2) and pZJY156 (Figure 6.3), allows the possibility of over-expression of genes within a *M. xanthus* strain, thus offering an alternative method to confirm the function of the product of *crtW*. This would primarily enable verification of *crtW*'s role during carotenogenesis. It was planned to insert the *crtW* gene in a similar manner to *pilA*, along with a promoter region into one of the vectors. Transformation of wild-type *M. xanthus* DK1622 with the resultant plasmid was expected to result in the over-expression of CrtW. If the protein did play a role in carotenogenesis then there could be a phenotypic change with respect to the final carotenoids produced.

The only problem in adopting the approach is the uncharacterised *crtW* promoter. It was proposed to use the c*arQRS* promoter (Whitworth *et al.*, 2004). This promoter has been well documented and was known to be controlled by the sigma factor CarQ, which has been proven not to control *crtW* transcription (Chapter 2). The *carQRS* promoter is large (Whitworth et al., 2004) so a region of 140bp was selected. Primers were designed to enable amplification of both *crtW* and the *carQRS* promoter region, and attachment of suitable RE sites ('Primers' Methods). Using the 2.1-TOPO vector, the intention was to firstly insert *crtW* and the promoter region into separate intermediate vectors; the former ligated using *Nhe*I and *Eco*RI and the latter *Bam*HI and *Nhe*I. The identical *Nhe*I sites would then be the points at which the two fragments were subsequently ligated together. *Bam*HI and *Eco*RI, were to be used as they were unique in both the 2.1-TOPO vector and also the shuttle vector.

The *crtW* fragment was generated and cloned, followed by a solitary band of appropriate size that once sequenced proved to be that of the *carQRS* promoter. The *carQRS* promoter was then inserted upstream of the *crtW* gene in an intermediate 2.1-TOPO vector.

Vector pZJY156 (Figure 6.3) was selected for use in the investigation of *crtW*, solely because it was the smaller in size than pZJY41. Insertion of the *carQRS* promoter and *crtW* fragment into pZJY156 worked with an extremely high success rate – 95% of resultant colonies tested containing the insert. Insertion was confirmed through sequencing of the vector, which was designated pJAM156. The pJAM156 vector was purified from *E.coli* for transfer into *M. xanthus*.

6.3. Shuttle vector electroporation

Three strains of *Myxococcus xanthus*, DK101 (wild-type), DK1622 (wild-type) and UWM303 ($CarQ^-$), were selected for transformation with pJAM156. DK101 and DK1622 were intended as the primary strains of study and UWM303 would function as a control, in which the gene would not be expressed following strain transformation. The sole difference between the two 'wild-type' strains of DK101 and DK1622 is a reduced slime-coat covering of the latter strain. There is no evidence recorded of vector readthrough when testing *pilA* expression in pZJY156 (Zhao et al., 2008). As a result when designing the *carQRS* promoter/*crtW* construct this was not considered. If readthrough does occur it will be evident in the *M. xanthus*

control strains – particularly UWM303, and if occurring will act to increase the concentration of CrtW produced overall.

It proved impossible to isolate pJAM156 transformants of DK101 and UWM303 despite a number of attempts. A number of putative DK1622 [pJAM156] transformants were however obtained. Ten of these transformants were then tested using PCR of boilates to detect the presence of the desired plasmid. The PCR used the primers 'PCRCRTW R' and 'CARQ PRO F', which should generate a fragment of 350bp containing both the inserted *carQRS* promoter plus an additional 194 bases of the *crtW* gene. Of those tested, seven produced a positive band of appropriate size. For additional confirmation, three of the gel bands were selected at random (samples a, c and g) and the fragments sequenced. All proved to encode the desired fragment, confirming successful transformation with pJAM156. These three transformants were then studied further. The plasmid was originally designed such that *crtW* transcription was controlled by CarQ, and so as a result it should be over-expressed in the DK1622 strain when exposed to a blue light source. A phenotypic change was expected with respect to the final carotenoids produced, at least in their overall final ratio when compared to wild-type strains. A simple visual test was carried out for verification, whereby the three DK1622 [pJAM156] colonies along with a DK1622 control strain were re-plated onto two pairs of DCY Kan^{20} and DCY only plates respectively. One pair of paltes (DCY and DCY Kan^{20}) was then incubated in the presence of an artificial light source and the other in complete darkness. Following a 4-day incubation period both plates were examined, comparing the colouration of the samples. There was no colour differentiation between the DK1622 [pJAM156] samples and controls upon inspection, with all of the dark incubated cultures appearing yellow and those in the light, orange. Each of the samples that was incubated in the dark did however grow slightly better. This is regularly observed with the lack of blue-light enabling *M. xanthus* cells to grow more readily. Although the carotenoid levels are expected to change, this does not result in a distinct visible colour change, possibly due to the subtle carotenoid difference. As a result any subsequent colour change that does occur may be undetectable to the human eye.

6.4. High-performance liquid chromatography analysis

To examine carotenoid production of DK1622 and DK1622 [pJAM156] high-performance liquid chromatography (HPLC) was necessary. In order to gain both access to and use of a high performance liquid chromatography unit relevant skills in carotenoid analysis are required. Following contact with a research group based at Royal Holloway University, London, permission was obtained for both the use of their HPLC equipment and expertise in order to aid in the identification of the carotenoids produced in the newly created DK1622 [pJAM156] strains. For direct comparison unaltered DK1622 and DK718 strains were used as controls, one displaying the standard carotenoid phenotype and the latter producing an excess of carotenoids.

The results obtained from the HPLC were very disappointing. This was primarily due to the fact that the data produced were very messy. In terms of the identification of carotenoid changes in the newly synthesised *M. xanthus* DK1622 [pJAM156] strain compared to unaltered DK1622 and DK718 strains, it was impossible to identify any differences between the samples. This was not caused through the production of similar proteins in each of the strains, but by the fact that none of the samples produced clear analysable results following HPLC. It appears that the carotenoid products from the reaction are too ionic for this particular HPLC approach, making it ineffective in identifying the differing carotenoids produced. It does suggest though that each of the strains is still producing complex carotenoids, even following genetic manipulation. An alternative explanation for these results would be that the samples were somehow contaminated during initial isolation resulting in the generation of erroneous results.

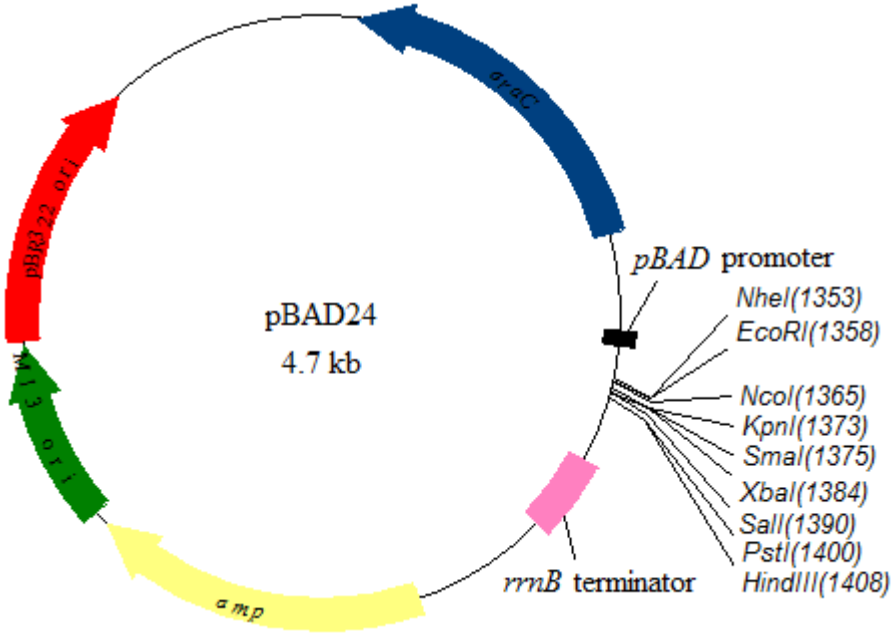
6.5. *M. xanthus crtW* expression in *E. coli*

To test if *crtW* encoded a carotenoid ketolase, an attempt was made to express the protein in *E. coli* and examine ketolase activity *in vitro*. Two potential systems were identified in order to achieve this; the pBAD and pET expression systems.

6.5.1. The pBAD expression system

The pBAD vectors (Guzman *et al.*, 1995) were developed for controllable expression of proteins in *E. coli*. The vectors were primarily designed for the study of

membrane proteins in *E. coli.* The control system of the plasmids consists of a P_{BAD} promoter region taken directly from the arabinose operon and the *araC* activator gene (Figure 6.4). When an *E. coli* strain containing the plasmid is grown in the presence of L-arabinose, transcription occurs from the P_{BAD} promoter. Without the addition of L-arabinose, transcription can still occur but at a much reduced level. Experimentally, the gene of interest is inserted downstream of the P_{BAD} region and exposure of the transformed strain to high levels of L-arabinose results in the expression of the newly inserted gene.



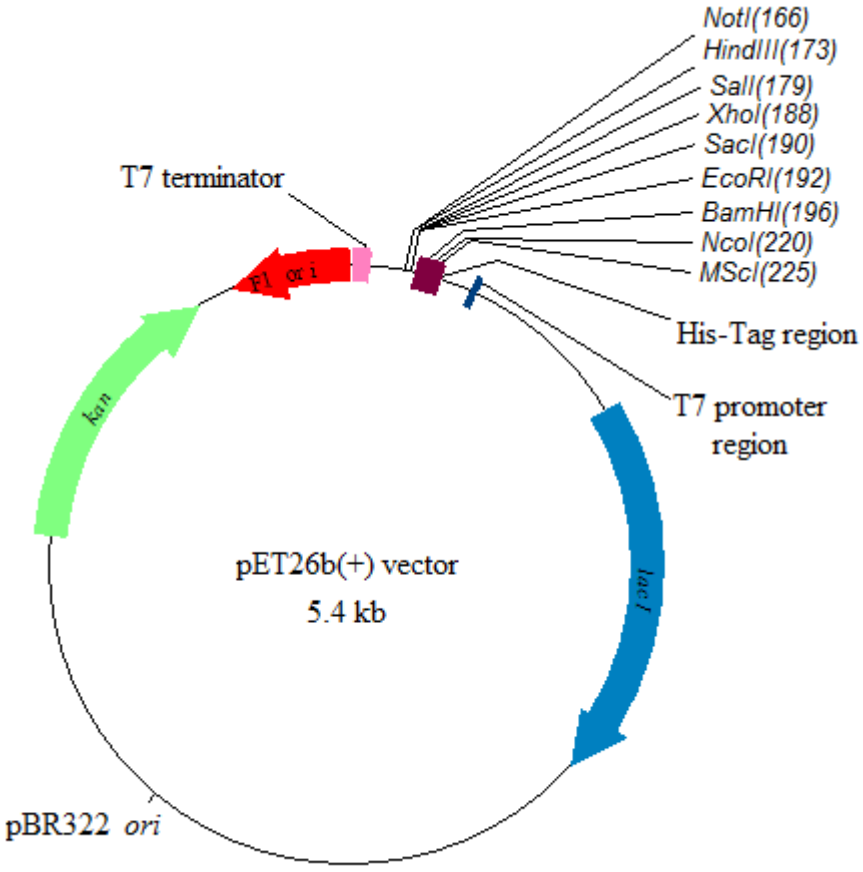
M13 *ori*:	2784-3242bp	pBR322 *ori:*	3248-3945bp
araC:	96-978bp	*amp*:	1885-2745bp
pBAD promoter:	1230-1277bp	Kozak region:	1317-1319bp
rrnB terminator:	1577-1775bp		

Figure 6.4 The pBAD24 vector, with key restriction enzyme sites labelled and the coding regions highlighted.

6.5.2. The pET expression system

The pET expression system was developed initially to be used as an RNA expression system, being highly selective for T7 RNA polymerase (Studier and Moffatt, 1986). Principally this relied on one of two different methods of T7 RNA

polymerase maintenance within the host, either through the incorporation of the polymerase gene under *lacI* control into a plasmid, or by its insertion into the chromosome via a prophage. The pET vector consists of a T7 promoter, an origin of replication (similar to pBAD), *amp*, plus a *lacI* gene (encoding a *lac* repressor) and a *lac* operator that acts to prevent transcription from the T7 promoter (Figure 6.5). Downstream of the T7 promoter is the MCS, enabling easy gene insertion. Expression was activated by the presence of T7 RNA polymerase in the host cell induced using IPTG. As long as the *lac* operator is not being repressed, which does not occur in the presence of IPTG, the desired protein is then created.



lacI:	764-1842bp	*kan*:	3986-4798bp
pBR322 *ori*:	3277bp	F1 *ori*:	4894-5349bp
T7 promoter region:	361-377bp	T7 transcription start:	360bp
T7 terminator:	26-72bp	His-Tag region:	158-225bp

Figure 6.5 The pET26b(+) vector, with key restriction enzyme sites labelled and the coding regions highlighted.

6.6. Construction of a *crtW* expression construct

During the creation of both of the expression constructs, the orientation of the newly inserted gene was paramount. The vectors used were pBAD24 (Figure 6.4) and pET26b (Figure 6.5). Two primers were designed to create a *crtW* gene fragment for insertion, with RE ends of *Bam*HI and *Hin*dIII respectively ('Primers' Methods), which enable a fragment insertion 20bp downstream of the Shine-Dalgarno sequence in the pBAD24. A *crtW* fragment was created by PCR with *Nde*I and *Nco*I ends to be inserted into the pET26b vector (Figure 6.5). Use of the pET26b vector results in the addition of a histidine tag to the end of the expressed protein. This enables simple purification of the protein.

During the generation of each vector, the desired gene region was firstly amplified through a PCR and then inserted into the intermediate TA cloning vector 2.1-TOPO. The sequence could then be excised and directly inserted into the expression vector of choice. Both PCR products were inserted into vector 2.1-TOPO and verified by insert sequencing. However, subsequent analysis of the *Nde*I-*Nco*I insert for transfer to the pET26b vector revealed that the final base of the *Nco*I sequence was not present, most likely a result of an inefficient PCR. Hence the attempt to create a *crtW* pET26b expression vector was suspended at this point.

Insertion of the appropriate fragment into the pBAD24 vector was attempted and also failed. Further analysis of the pBAD24 sequence revealed the presence of two previously unnoticed *Bam*HI restriction sites in the MCS site within the plasmid. As a result digestion of the plasmid would generate two *Bam*HI ends and no *Hin*dIII end, as the site was excised. To rectify the problem the *crtW* PCR fragment was redesigned, using instead a *Hin*dIII site and an *Nco*I site in preference to *Bam*HI. The new fragment was created by PCR and inserted into the 2.1-TOPO vector.

The fragment was excised from the 2.1-TOPO vector using *Hin*dIII and *Nco*I and inserted into an appropriately digested pBAD24 vector. Only one transformant contained an insert, but it proved to have inserted at the *Hin*dIII site alone, and in the wrong orientation. Fresh digestion reactions of both vector and insert were carried out and the ligation reactions repeated. One transformant of 19 tested proved to contain an insert of appropriate size, which was confirmed as *crtW* in the correct orientation by sequencing. The plasmid was designated pBADJAY.

6.7. CrtW synthesis using pBADJAY

A concentration level of 0.2% L-arabinose was identified as a standard amount adopted during pBAD studies that was known to result in gene expression (Bost *et al.*, 2007). As a result four separate *E. coli* DH5α [pBADJAY] samples were cultured overnight at 37^{o}C in LB containing 0, 0.1, 0.2 and 0.4% L-arabinose. The cells were harvested by centrifugation and snap frozen and allowed to thaw. This freeze-thaw process was repeated three times in order to break the cells up and release the proteins.

In order to detect CrtW, the four samples were each subjected to general SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis), which separated the proteins in each sample ('SDS-PAGE Analysis' Methods). The SDS-PAGE gel of the complete cell extracts revealed no significant difference except in the case of a possible protein band at 30kDa (Figure 6.6). However a band of 66kDa is expected for *M. xanthus* CrtW, based on the prediction of *crtW* protein product size. A control strain of *E. coli* DH5α [pDAH274] would be of benefit here to detect a unique 66kDa band in the samples. However given the data produced (Figure 6.6) it is not essential at this stage. A negative control however, of *E. coli* DH5α, would confirm the efficiency of L-arabinose control, verifying the system functions as intended. This is required as the banding observed at 0% L-arabinose concentration (Figure 6.6) is likely to be a result of low level construct transcription, observed previously in the pBAD expression system despite the absence of L-arabinose (Guzman *et al*., 1995).

From Chapter 4 CrtW is predicted to be a membrane protein. The membrane forms a tiny portion of the total cell protein content. Therefore it may not be possible to see the membrane protein in the initial protein gel. The cell membranes instead need to be isolated and their protein content analysed separately.

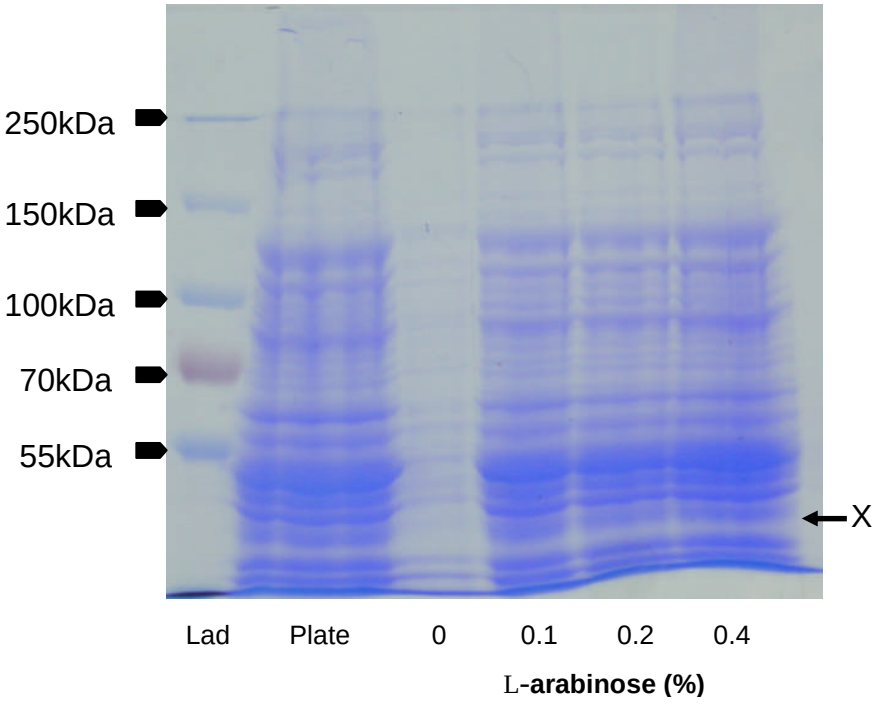


Figure 6.6 A protein gel of the total soluble protein extract of *E.coli* DH5α [pBADJAY] grown in different L-arabinose concentrations. The 'Plate' lane is a sample of cells removed directly from 0.4% L-arabinose plate. The arrow on the right hand side indicates the 30kDa region at which the 0.4% L-arabinose band appears a lot stronger than the 0% control concentration.

To achieve this similar cell cultures were prepared as for use in the complete cell extract SDS-PAGE, up until the point at which the samples were snap-frozen. The cells were then broken up through bead-beating, effectively breaking open the cell membranes and inner components, releasing proteins. Following centrifugation, the protein extraction process was then repeated with the resultant cell pellet, with the addition of a low concentration of sarkosyl detergent ('Sarkosyl extraction' Methods). This aids hydrophobic protein resuspension the subsequent supernatant, and by using a low detergent concentration should result in the resuspension of the majority of inner membrane proteins. Upon supernatant removal the process was repeated on the cell pellet using a higher concentration of detergent, resulting in the resuspension of the remaining proteins, principally those from the cell outer membrane. CrtW is expected to be present in the first sarkosyl wash, having been predicted as bound to the inner membrane in *M. xanthus*.

Following a repeat SDS-PAGE of the novel cell supernatant samples, no additional band patterns were observed. A greater number of proteins appeared present in the latter higher concentration detergent washes, but no difference was

observed when comparing the 0 control to the maximum 0.4% L-arabinose samples. Ideally at this stage, protein presence would now be verified using an antibody specific to CrtW. As the protein has not been previously studied in such depth this is not possible as an antibody has not been obtained that is specific to the ketolase. Hence it was not possible to isolate or observe CrtW through the use of SDS-PAGE.

If L-arabinose presence results in the increased synthesis of CrtW, the isolated cell samples can still be tested. This is possible by the addition of a carotenoid substrate that will readily undergo ketonisation by CrtW, to the *E. coli* DH5α [pBADJAY] strain. This approach has been previously exploited to great effect (Tao and Cheng, 2004), whereby it was found that once *E. coli* cells expressing *crtW* were broken up, releasing membrane-bound proteins, and the proteins used directly in a reaction with β-carotene, substrate ketonisation was observed. This was despite β-carotene not being the usual carotenoid precursor for the CrtW tested. In contrast a control reaction including a strain unable to express the gene showed no detectable carotenoid change. The study confirmed that not only was CrtW produced, but that it also functioned as a ketolase, capable of ketone group addition to β-carotene (Tao and Cheng, 2004).

All four of the previous *E. coli* [pBADJAY] cultures grown in the presence of 0, 0.1, 0.2 and 0.4% L-arabinose were assayed for enzyme activity ('Ketolase Assay' Methods). Upon experiment completion, four small carotenoid pellets were obtained, each less than $0.1mm^3$ in size. To analyse their composition each was then subjected to HPLC following resuspension in 0.5ml methanol to identify the specific carotenoid substrates present. If the *crtW* gene does encode a ketolase that interacts with β-carotene, then the reactions should result in the production of an amount of canthaxanthin or similar carotenoid molecules. Any additional observed carotenoid change would suggest an alternative gene and protein function. This was an identical analytical approach to that employed during previous bacterial CrtW ketolase assays (Tao and Cheng, 2004).

6.8. Ketolase assay results

The identification of the carotenoids produced following β-carotene exposure to CrtW using HPLC resulted in inconclusive results being obtained. Analysis of the samples was unable to identify any of the carotenoids present. It was expected that a

proportion of the β-carotene would be converted to canthaxanthin, but with neither positively identified following HPLC analysis, no conclusions are possible from this assay.

6.9. Discussion

Following the discovery of autonomous replicating plasmids in *M. xanthus* it was practically possible to over-express proteins in the bacterium using multi-copy plasmids. *crtW* was successfully inserted along with the *carQRS* promoter region into one such shuttle vector, pZJY156 (Figure 6.3), with the novel plasmid created designated pJAM156. Attempts to transform *M. xanthus* DK1622 (wild-type) with the plasmid were successful, although similar attempts using DK101 and UWM303 strains failed. The problem would most likely have been addressed through repetition of the transformation procedure, but was unnecessary as *M. xanthus* DK1622 [pJAM156] had already been created. When comparing the strain to control cultures grown in both the presence and absence of light, no phenotypic difference was observed with respect to colour. Given the potentially small change in produced carotenoid levels, this was not unexpected. In order to identify any carotenoid differences *M. xanthus* DK1622 [pJAM156] was instead analysed through HPLC. The results obtained were messy and inconclusive, with no carotenoid positively identified. The samples appear to be too ionic for the specific HPLC approach adopted, and so as a result no conclusive data was obtained.

Through use of the vector pBAD24 it was possible to construct a strain in which to study the over-expression of the *crtW* gene, *E. coli* DH5α [pBADJAY]. This did not prove possible when attempting a similar procedure with the pET26b vector. This was principally a result of poor conservation of restriction sites during intermediate creation. Despite successful pBADJAY generation, attempted observation of *crtW* over-expression in the strain *E. coli* DH5α [pBADJAY] through SDS-PAGE was unsuccessful. This was most likely a result of an inadequate protein isolation method. The problem could be solved through refinement of the sarkosyl extraction method and the testing of alternative *E. coli* strians, neither of which was possible in this investigation due to time constraints. Further attempts at a ketolase assay using the strain, testing the effect of the generated CrtW on β-carotene also failed, with no conclusive results from HPLC analysis.

Through the alteration of various experimental components of the HPLC procedure, including both the column and mobile phase solution adopted, it is believed that the final carotenoids present in both the ketolase assay and DK1622 [pJAM156] samples could be positively identified. However in order to determine the changes required, in addition to repetition of the experiments, it is estimated that a period upward of six months is required. Given the time constraints and expertise required to achieve this goal, the approach is no longer feasible as part of this investigation. An alterantive may be to consider using a mass spectrometry approach to identify individual carotenoids. This has been previously attempted when studying myxobacton and myxobactin, resulting in poor quality final results (personal communication, D. E. Whitworth). Therefore, although time-consuming, in order to obtain the best possible carotenoid data, HPLC should be the preferred option.

Chapter 7

Discussion

7.1 Conclusions

Repetition of initial gene homology searches provided evidence that the gene that was the subject of study in this investigation encoded a form of fatty acid desaturase, most similar to CrtW, a carotenoid ketolase involved in bacterial carotenogenesis (Chapter 4). The gene was also believed to form part of a larger operon (D. E. Whitworth, unpublished data). Experimental evidence showed this to be the case and *crtW* is the second gene in a four gene operon consisting of *pdp*, *crtW*, *ord* and *mmb*. All four appear to be transcribed together from a single promoter located upstream of *pdp*. As the four genes are all linked together it is possible they function in a similar pathway. In addition to the operon promoter, a second putative promoter region was identified upstream of *crtW*, most likely within the *pdp* gene. Through study of *M. xanthus* mRNA, larger amounts of *crtW* and *ord* transcripts were recorded compared to that of *pdp*. Additionally *crtW* transcript production appeared to be greater than that of *ord*, although not by a significant amount. This further suggests the presence of a second internal *crtW* promoter within the operon. It does not however confirm its existence as the varied gene transcript levels recorded may still be a result of promotion from the *pdp* promoter. Attempts to map either of the promoters using RACE were unsuccessful, so the precise location of either is currently unknown (Chapter 3).

It has been confirmed that, contrary to the initial expectation the sigma factor responsible for light-induced carotenogenesis CarQ, was not responsible for initiating gene transcription of either promoter of *crtW*. This was shown by inserting the two putative promoter regions upstream of a promoterless *lacZ* gene in the pDAH274 promoter probe plasmid. These constructs were used to transform four *M. xanthus* strains; DK101 (wild-type), UWM303 (*carQ*), DK717 (*carA*) and DK718 (*carR*), and *lacZ* expression was monitored in the presence and absence of light (Chapter 2). Consistently, there was no significant difference observed between any of the strains and in each strain LacZ was definitely produced. As a result it is evident that CarQ and the other known carotenogenesis biosynthetic gene repressor CarA did not control

crtW transcription. The strains were also found to react no differently whether grown in the presence or absence of light. Essentially if the *crtW* gene is active during carotenogenesis it is not light-induced as was hypothesised. This raises the possibility that it may not operate solely during carotenogenesis and may be involved in additional *M. xanthus* pathways.

Initially parallels were drawn between *crtW* and *crtI*, particularly with respect to their isolation from the majority of the biosynthetic enzyme genes. Following conclusive evidence that *crtW* was neither light nor CarQ-dependent, it was also found that carbon starvation had no effect on its transcription (Chapter 3). This is opposite of *crtI*, which requires the cell to have undergone a period of carbon starvation in addition to the presence of light before it is transcribed. As a result *crtW* is clearly highly dissimilar in its control to that of *crtI* and the *crtEDBC* operon. It is also unclear as to which sigma factor is responsible for *crtW*'s transcription. Despite being transcribed from two promoter regions, there does not appear to be any significant sequence similarity between the two regions and no known sigma factor binding sites were present. It appears *crtW* expression is controlled by a sigma factor different to CarQ but it is not known which one.

Through detailed analysis of the amino acid sequence of CrtW it proved possible to model the structure of the protein. It appears to span the cell membrane four times in total, as has also been suggested in previous CrtW modelling studies (Ye *et al.*, 2006). The protein itself is also made up of two hydrophobic regions and three highly conserved histidine motifs, creating the unique protein loops required for substrate interaction within the cell. The predicted protein is also very similar to a number of other CrtW proteins from a range of bacteria, supporting the fact that the gene studied does indeed encode a ketolase. Likewise from the predicted topographic models of CrtW (Figure 5.2), it is most likely that the hydrophobic regions are embedded in the cell membrane, if the protein is similar to those ketolases previously modelled. This would provide the protein with added stability and uniform protein loops (Chapter 5).

It did not prove possible during this investigation to remove CrtW activity by deletion of the gene (Chapter 5). This may have been the result of a number of factors, with the most likely appearing to be the increased rate of spontaneous kanamycin resistant mutants. This made it extremely difficult to detect any successful knockout strains that were generated. Another major problem was that the drug

resistance that replaced *crtW* could only be expressed once the construct had been fully integrated into the genome. Alternatively, although indeterminable from the results; it may be that the knocking out of *crtW* gave rise to a lethal mutation. This is possible given the high levels of constitutive gene expression observed with the LacZ assays. It is however unlikely that *crtW* itself would be the essential gene in this instance, with one of the other genes encoded within the *crtW* operon being a more likely candidate.

Both over-expression studies using autonomous replicating plasmid vectors in M. *xanthus* (Zhao et al., 2008) and *in vitro* carotenoid ketolase assays, using an *E. coli* DH5α [pBADJAY] strain capable of *crtW* expression, proved unsuccessful. This is a direct result of the inability to successfully identify the carotenoids and the levels at which they were produced using HPLC. It appears that the procedure adopted is unable to differentiate between the carotenoids due to their high ionic charge. To counteract this, the HPLC procedure needs to be adapted and a number of experimental parameters changed (Chapter 6).

Analysis of the *crtW* operon in more detail using NCBI BLAST and EMBL STRING searches revealed that the first gene *pdp* encoded a protein closely related to the MutT/Nudix family (Chapter 4). This suggests it plays a role as a hydrolase within the cell. It was also determined that the final two genes in the operon, *ord* and *mmb*, together encode a periplasmic molybdopterin and haeme-dependent oxidoreductase, YedYZ. The two genes were found to be highly conserved in a wide range of bacteria and were always contiguous. These genes were, annotated as *yedY* and *yedZ* respectively in *E. coli*. As a result it is suggested that the genes be annotated as such in *M. xanthus* from now on. Despite occurrence of the putative hydrolase gene, *crtW* and *yedYZ* in a number of organisms, they were never found together in a cluster with the exception of *M. xanthus* and *Stigmatella aurantiaca*. This makes the arrangement a novel one. It is still unclear as to the purpose the operon serves, or as to whether the gene products are capable of functioning in a similar pathway, although given their diverse roles this appears unlikely.

Considering the evolutionary relationship between *M. xanthus* and other carotenoid producing species, it is found to vary depending on the gene sequences being compared. Similar organisms tended to group together in each phylogenetic tree, but overall comparison of biosynthetic gene sequences reveals little phylogenetic correlation. It is clear that the results indicate a distinct difference in carotenogenesis

biosynthetic gene origin in *M. xanthus* when comparing the essential, semi-essential and non-essential *crt* genes (Chapter 4).

Overall this investigation has furthered the understanding of *M. xanthus* carotenogenesis with respect to the production of myxobacton via ketonisation. The gene responsible for the generation of CrtW, *crtW*, has been identified, with experimental and bioinformatic evidence suggesting it encodes a carotenoid ketolase. The role of the genes within the *crtW* operon is also suggested, although the function of the operon is still unclear. Despite failure to successfully isolate and characterise CrtW, a number of strains were created from which this is possible, given more time.

7.2 Future work

The investigation leaves a lot of potential future work relating to *crtW*. The *in vitro* study of the ketolase, CrtW, should be one of the primary objectives. Adoption of a similar over-expression method using a pET vector would enable the production of a His-tagged CrtW, aiding protein purification. It would then be possible to determine the protein structure, which has not yet been achieved for a ketolase, and also conduct a host of additional proteomic studies. In conjunction with this it would be beneficial to conclude the HPLC studies attempted during this investigation. The strain *E. coli* DH5α [pBADJAY] has already been constructed, so only HPLC procedure refinement is required, though this may be fairly time consuming. A successful ketolase-assay should also be a priority, as this would confirm that CrtW does function as a carotenoid ketolase. Likewise concluding the HPLC work when using the *M. xanthus* DK1622 [pJAM156] *crtW* over-expression strain would confirm the enzyme function *in vivo*. In respect to this it may be worthwhile conducting all of the afore-mentioned experiments on a larger scale, using bigger cell preparations. This would ensure a significant amount of material was being utilised to draw conclusions from and in terms of the CrtW protein, a large amount was being successfully generated for future analysis.

Repetition of a number of experiments from this investigation may also be of benefit, such as the Lacz assays. As well as conducting these on a larger scale, variations could be made to the intensity of the light to which the growing cultures are exposed and the time of exposure, to see if this makes any significant difference to promoter activity. Similarly, alterations could be made to the basic carbon starvation

assay; varying the time of exposure of *M. xanthus* cells to the starvation media to see whether a change in promoter activity is observed.

It would be particularly interesting to determine the roles of the other *crtW* operon gene products in *M. xanthus*. Although roles are suggested for all of them from bioinformatic studies, experimental evidence is still required and may indicate in which system(s) they function. This may be achievable through the knock-out of the genes and then observation of resultant cell phenotype, or alternatively over-expression in the cell as was attempted for *crtW*. Given the high transcription levels observed for the genes, they appear to play an important role in the cell, so understanding their function may be the key to understanding *crtW* transcription and the overall role of the *crtW* operon. Identification of the sigma factor(s) responsible for *crtW* transcription could be a priority. Accurate mapping of the *crtW* operon promoter regions through a process such as RACE, primer extension or S1 nuclease mapping should assist in this aim.

Often studies of *M. xanthus* carotenoid biosynthesis have been based on the identification of intreresting mutant strains following the random insertion of Tn5 into the host genome. It must therefore be worth considering repeating this type of study, and looking for any additional pigment mutants created. This may determine additional genes involved in carotenoid biosynthesis, or possibly even result in a mutation in the *crtW* operon. Similar experimental studies could also be conducted using other myxobacteria, such as *St. aurantiaca* and *So. Cellulosum*, which have not had their carotenoid synthesis pathways studied to the same extent as *M. xanthus*. This woul help to identify any significant differences within the family and also provide a clearer understanding of the role carotenoids play in each individual family member.

It is clear that carotenoid production is both of medical and economical importance. Novel carotenoids are frequently being discovered, and often the process of synthesis for these molecules is unknown. Increased understanding of carotenogenesis in a model system, such as *M. xanthus*, will aid in the study of carotenoids. Although this study has focussed on *crtW*, there are a number of other potential carotenoid biosynthetic enzymes in a whole host of organisms that still require characterisation, including in *M. xanthus*. Until similar studies are conducted on these it seems we will never truly understand the complex process of carotenogenesis.

Chapter 8

Materials and Methods

8.1. Media Recipes

LB

Liquid (1L): 10g tryptone 5g yeast
10g NaCl ddH_2O up to 1L

Agar Plates (LBA): As above plus 15g Sigma-Aldrich purified agar (15g/l)

DCY

Liquid (1L): 20g casitone 2g yeast
8ml 1M $MgSO_4$ 5ml 1M TrisHCl (pH 8.0)
ddH_2O up to 1L

Agar plates (DCYA): As above plus 15g Sigma-Aldrich purified agar (15g/l)

Soft agar (DCY SA): As above plus 7.5g Sigma-Aldrich purified agar (7.5g/l)

LC

As LB with the addition $CaCl_2$ to a final concentration of 5mM

LGC

As LC with the addition of 2ml 50% (w/v) glucose per litre

MC7 (starvation media)

Liquid: 10mM MOPS
1mM $CaCl_2$
ddH_2O to a final pH of 7.0

8.2. Plasmid Miniprep

Each plasmid miniprep was achieved by taking a 10-20ml overnight liquid culture of the desired vector containing host (usually an *E.coli* strain) and pelleting the bacterial cells. This was done through a 20 minute centrifugation at 16,100g. A 'Qiagen Spin Miniprep Kit' was then used to isolate the plasmid, following the 'QIAprep Spin Miniprep Kit using a Microcentrifuge' procedure accompanying the kit. The only adaptation made to the protocol was the increase of spin time following the initial 10 minute centrifugation. All 60 second spin steps were doubled in time and the final ethanol buffer removal step conducted for 3 minutes in total. This was to ensure that all of the supernatant passed through the filter, leaving a final clean plasmid concentrate.

8.3. Gel extraction procedure

When extracting DNA fragments from an agarose gel, following gel electrophoresis, a 'Qiagen Qiaquick Spin Kit' was used. The microcentrifuge 'Qiaquick Gel Extraction Protocol' was then adopted, unless the fragment was 1kb or greater in size, when a 'Geneclean II' kit was used in preference. Following the initial gel dissolving step, each subsequent 60 second spin was increased to 2 minutes, and the final buffer removal step to 3 minutes. This was to ensure the removal of all of the supernatant and prevent any potential ethanol contamination in the final isolated sample.

8.3.1. Geneclean II

This is a protocol using an older bead purification method in a separate kit, so as a result the entire procedure with the kit is detailed below. It can be easily replicated by using a standard batch of silica bead (<0.5mm in diameter) and an effective elution buffer.

The DNA fragment of choice was initially excised from an agarose gel, transferred to a microcentrifuge tube and weighed. Three times volume of NaI solution was then added to the gel; so 300µl for 100mg, 150µl for 50mg etc. The tube

was incubated at 50^{o}C for approximately 10 minutes or until the gel was completely dissolved. Glassmilk beads were then added to the mixture (5µl for a sample up to 500µl and 15µl for any higher volume). The mixture was vortexed for a period of 60 seconds allowing the beads to thoroughly mix and become suspended within the sample. The tube was then incubated at room temperature for around 5 minutes, gently mixing every 1-2 minutes. For the best results an incubation period of 15 minutes is optimum, allowing maximal DNA to bead binding. The beads were then washed 4 times with NEW Wash buffer (10x the volume of the existing microcentrifuge tube mixture). After each resuspension in the buffer, the beads were again centrifuged for 4-5 seconds at 16,100g. Following the final centrifugation all remaining liquid was pipetted off the beads. The pellet was then left to air dry for 5-10 minutes at room temperature. To elute the DNA, an equal volume of ddH_2O to pellet was used to resuspend the beads. The mixture was then centrifuged for an additional 30 seconds and the supernatant removed to a clean tube for storage and use.

8.4. Ligation reactions

After much experimentation and variations of the actual ligation reaction set up, two different ligation mixes were used for each reaction to ensure there was an appropriate level of insert in at least one of them. All reactions were carried out by initially preparing the ligation mixture, gently tapping the microcentrifuge tube to mix all of the reagents and then incubating it at room temperature overnight (a period of approximately 20 hours):

Ligation Mixture	**No. 1**	**No. 2**
Vector Digest (50ng/µl)	2µl	2µl
Insert Digest (100ng/µl)	2µl	5µl
Ligase (T4)	1µl	1µl
Ligation Buffer	1µl	1µl
ddH_2O	4µl	1µl

8.4.1. DNA blunt end ligation

When attempting to make pDAH274 smaller in size, the ‘sticky end’ overhangs created during plasmid digestion did not always match up to each other.

As a result, in order to effectively ligate the ends together they were both blunted, before ligation was attempted. To do this a standard reaction mixture was used:

1µl 10x T4 DNA polymerase buffer
2µl purified digested plasmid (50ng/µl)
6µl ddH_2O

The mixture reaction was incubated at 70^oC for 5 minutes. 1µl of T4 DNA polymerase was then added and the entire mixture gently pipetted, mixing all of the reagents. It was then incubated at 37^oC for a further 7 minutes. Upon completion the tube was transferred immediately to ice for 30 minutes. This ensured the deactivation of the polymerase. The mixture could then be run on an agarose gel and the freshly produced fragments excised and purified. This can be used directly in a ligation reaction. Most of the products will be religated plasmid, due to the size of the fragment removed, so it is advisable to dephosphorylate the 5' end of the plasmid, thus increasing the proportion of desired ligation product generated.

8.4.2. Dephosphorylation

The simple procedure below was conducted using NEB calf intestinal alkaline phosphatase (CIP)

The genomic or plasmid DNA was resuspended in 1x NEB buffer to a final concentration of 0.5µg per 10µl. For each µg of DNA half a unit of CIP was added and the entire mixture gently pipetted. The reaction was then incubated at 37^oC for a period of 1 hour. Standard purification protocols were then followed to ensure that only the DNA is collected and all enzymes are removed.

8.5. Competent cell preparation

For each preparation of competent cells an identical protocol was used regardless of strain. This involved growing the *E.coli* strain (See '*E.coli* Strains') in a universal tube containing 10ml LB, inoculated with a large loopful of cells from a fresh *E. coli* lawn. The tubes were incubated at 37^oC overnight, and then the following morning a 150µl suspension from each was transferred to 4 further universals along with an additional 10ml LB. These were in turn incubated at 37^oC

while being shaken, until the cells had grown significantly (OD_{600} ~ 0.4-0.6). The tubes were then centrifuged at 16,100g and all resultant pellets resuspended in a total volume of 1ml ice-cold 0.1M $MgCl_2$, transferred to a singlemicrocentrifuge tube. This was then centrifuged at 16,100g for 10 minutes and the pellet resuspended in 1ml of ice-cold 0.1M $CaCl_2$. A further centrifugation step and resuspension in $CaCl_2$ followed, with the process being repeated for a total of three times. The final suspension of cells was then stored on ice for approximately 4 hours or in $^2/_3$ volume of glycerol at -80^{o}C for later experimental use.

8.5.1. Cell transformation

Following the 4-hour ice incubation, or using a previously prepared competent cell strain, transformation is then possible. Successful transformation of the newly prepared competent cells requires 3-5µl of plasmid to be inserted, added to 100µl of the competent cell mix. Each tube is then gently mixed before being left on ice for 30 minutes. The tubes are then heat-shocked in a water bath at 42^{o}C for 2 minutes, before immediate transfer back to ice for an additional 2 minutes. 1ml of LB is added to each tube and then all samples incubated for 1 hour at 37^{o}C whilst being shaken. Following incubation all samples are then centrifuged for 10 minutes at 16,100g and the resultant pellets resuspended in 30µl of the remaining supernatant. This can be directly spread onto selective plates and incubated overnight at 37^{o}C of whichever temperature offers optimum growth conditions for the desired plasmid.

8.6. *Myxococcus xanthus* DNA extraction

For a number of procedures, isolated *M. xanthus* DNA has been utilised, usually in some form of PCR. To ensure there was an ample supply of DNA, some was initially isolated using an adapted method from R. Gill (unpublished). This has been designed specifically for *M. xanthus* DNA extraction and although relatively complex compared to a number of current isolation methods, proved highly effective.

Materials

Sucrose-Tris: 25% sucrose; 10mM Tris (pH 8)
Lytic mix: 5% SDS; 0.125M EDTA; 0.5M Tris (pH 9.4)
Pronase (Roche) 10mg/ml in TE buffer
TE Buffer: 10mM Tris, 1mM EDTA (pH 7.6)
Phenol – saturated with TE buffer (pH 7.6-8.0)
Chloroform
1M NaOAc
100% Ethanol
Ice-cold 70% ethanol
RNase H
Ice-cold isopropanol

Method

5ml of a 48-hour incubated *M. xanthus* DK101 culture (grown up in DCY) was spun down at 16,100g in a bench micro-centrifuge for 10 minutes. The supernatant was removed and the remaining pellet resuspended in 3ml of Sucrose-Tris. To this 600µl of lytic mix was added and the mixture incubated for 45-60 minutes at 65^{o}C. Simultaneously 1ml of pronase was incubated at 37^{o}C for 60 minutes. The resuspended pellet mixture was cooled to 37^{o}C and pronase added to a final concentration of 1mg/ml. The samples were then incubated overnight at 37^{o}C (a minimum two-hour incubation period is advised).

The sample was cooled to room temperature (18-20^{o}C) and an equal volume of phenol added. To this 10% volume of chloroform was added and the solution mixed carefully with a 15 minute period of continuous inversion. The tube was then spun at 16,100g for 15 minutes, in order to separate the phases. The aqueous layer was removed, placed in a clean universal tube and the phenol and chloroform extraction process repeated using the removed layer. This was repeated thrice, resulting in no visible interface layer in the final solution. NaOAc was added to the freshly removed aqueous layer until at a final 0.3M concentration. 2.5 volumes of 100% ethanol were added and the tube gently mixed. This resulted in precipitation of the DNA, which was spun for another 10 seconds at 16,100g before being washed using 0.5ml of ice-cold 70% ethanol. Upon completion of the wash, the DNA was finally dissolved in 0.5ml TE buffer with additional RNase (10µg/ml). The isolated DNA was then safely stored at -20^{o}C.

Although the process yields a considerable amount of DNA there is still some RNA in the mixture. To eliminate this problem the RNA was removed. This was done by taking the previously isolated sample and adding RNase H up to a final concentration of 50µg/ml. The solution was then incubated at 37^{o}C for 30 minutes. NaOAc was added to a final concentration of 0.3M and on top of this 0.625 volumes of ice-cold isopropanol, which resulted in the precipitation of the DNA. The DNA was removed by spooling it around a pipette tip and then transferring it to a separate 1.5ml microfuge tube, where it was washed in ice-cold 70% ethanol. Finally it was once again dissolved in 0.5ml TE buffer, leaving a viscous, clear liquid, which was then stored at -20^{o}C.

8.6.1. *M. xanthus* hot Phenol RNA extraction

The protocol for *M. xanthus* RNA extraction was adapted from J. Jakobsen (unpublished), who in turn adapted the protocol from an Applied Biosystems RNA isolation kit. Prior to attempting the method it is advisable to have all of the materials ready to use, as a number of the solutions require a period of preheating.

Materials

Phenol – saturated with 0.02M NaOAc (pH 4.5) containing 0.1% 8-hydroxyquinoline, an antioxidant (This should then be split into aliquots of 600µl in 1.5ml microfuge tubes and pre-heated to 65^{o}C in a water bath).

Solution A: 0.3M sucrose; 0.01M NaOAc (pH 4.5)
Solution B: 2% SDS; 0.01M NaOAc (pH 4.5) – preheated to 65^{o}C in a water bath
Chloroform
RNase-free ddH_2O
Ethanol and dry ice for snap-freezing purposes
Water bath at 65^{o}C
16,100g centrifuge (which will need to be cooled to 4^{o}C)
EtOH

Throughout the following process RNase AWAY spray was used ensuring all equipment used and work surfaces were RNA free. Any steps involving hot phenol were carried out in a fume hood.

Frozen cell pellets, obtained from various *M. xanthus* cell samples originally grown in 20ml DCY, were resuspended in 300µl of ice-cold Solution A. The mixture was then transferred into a 1.5ml microfuge tube. The sample was then ready for phenol extraction.

To each tube 400µl of pre-heated phenol was added and the samples then incubated at 65^{o}C for approximately 5 minutes. A minimum of 3 minutes is advised for this step. When all the tubes were sufficiently heated they were snap-frozen using an ethanol and dry ice bath. All frozen tubes were then centrifuged for 5 minutes, again at 16,100g. The tubes displayed a clear layering pattern, and the uppermost aqueous layer was transferred to a fresh centrifuge tube that was already prepared containing 300µl phenol and 300µl chloroform. The samples were mixed, by inversion, and centrifuged for 10 minutes at 16,100g. Once more following centrifugation a clear layering was visible in the mixture. The aqueous layer was again removed and transferred to a fresh 1.5ml microfuge tube containing 600µl chloroform. The tubes were mixed by inversion and centrifuged for a further 10 minutes at 16,100g. Finally following the resultant layer formation, the aqueous layer was transferred to another separate 1.5ml microfuge tube containing 40µl NaOAc (pH 4.5) and 900µl of 96% (v/v) ethanol. The samples were then refrigerated at -20^{o}C overnight, although the next step should be possible following a minimum 15 minute freezing period.

The frozen tubes were taken and centrifuged at 4^{o}C and 16,100g for a duration of 20 minutes. The supernatant was then removed and disposed, while the pellet was washed with 200µl ice-cold 70% ethanol. The tubes were subjected to a further 4^{o}C micro-centrifugation for a period of 10 minutes. Supernatant from the tubes was removed before the pellets were dried in a laminar flow cabinet for 45 minutes. Preferably a vacuum centrifuge should be used for 5 minutes, but I found this to be just as effective, if a little longer. Once dried the pellet was resuspended in 50µl of RNase free dd$H_2$0, and centrifuged for just 10 seconds at 16,100g. The newly isolated RNA sample was then transferred to another fresh 1.5ml tube and snap-frozen again using ethanol and dry ice before storage in a -80^{o}C freezer.

8.6.2. cDNA synthesis

Before being able to conduct any QRT-PCR reactions, previously isolated RNA samples were firstly required to be converted into the more stable cDNA form. Each RNA sample was taken and the amount of RNA contained in each tube determined through the use of a Nanodrop Spectrophotometer machine. This functions through the analysis of a single droplet of the RNA solution (5µl) and calculates how much RNA is present through an absorbance reading. It then extrapolates the result to establish the total amount of RNA present in the microcentrifuge tube from which the 5µl droplet was initially removed. Using a known amount of RNA, a reaction could then be carried out to copy the desired mRNA and generate a single cDNA strand:

0.1µM gene specific primers/random hexamers
5µg RNA
1µl 10mM dNTP mix
ddH_2O up to a total 13µl

The entire reaction mixture was ligated at 65^{o}C for a period of 5 minutes and then stored on ice for a further minute. This led to the generation of a number of DNA and RNA hybrid sequences. At this stage for most cDNA conversions 2µl random hexamer sequences were used, as they ultimately produced a better cDNA product. The tubes were then centrifuged at 16,100g for 30 seconds, in order to collect together all of the contents of the tube. To this was then added:

4µl 5x RT buffer
1µl 0.1M DTT
1µl RNase
1µl Superscript III RT

This was completely mixed by gentle pipetting of the tube contents. The tubes were then incubated for an hour at a temperature of 50^{o}C. If specific primers were initially used instead of random hexamers then this was increased to 55^{o}C. In order to deactivate the reverse transcriptase in the reaction a further incubation period of 15 minutes at 70^{o}C was also required. The reverse transcriptase recognises and acts on the newly synthesised cDNA strand forming a complete double-stranded molecule. Simultaneously the RNase acts to break down any remaining pieces of RNA, effectively leaving a sample of entirely cDNA. Invitrogen Superscript III RT was

used as it was adopted in previous protocols for *M. xanthus* cDNA production (Jakobsen, unpublished data). It is also a highly accurate enzyme in terms of inserting the correct bases during double-strand creation. The newly synthesised cDNA samples were then stored at -80^{o}C until required.

8.7. Polymerase Chain Reaction (PCR)

Reagents

The reagents used for each individual PCR rarely varied, with the exception of the original DNA sample used. Varying individual DNA sample concentrations resulted in various water volumes being used, although all other reagent measurements remained constant. When DNA was prepared directly from a cell extract, a cell boilate was used. The boilate was prepared by using a toothpick to pick a single cell colony from a plate, into 50µl of ddH_2O. This was then boiled for 5 minutes and 5µl of the resulting mixture used in the PCR reaction.

Usually a ready-made mastermix solution was used for PCR, obtained from Promega (2x PCR Mastermix), but an alternative homemade one was also created, the recipe for which is detailed below. This was used each time the pre-made mix ran out.

PCR Mastermix: 1x Taq Polymerase buffer
0.2mM each dNTP
1.25U Taq DNA Polymerase (per 50µl)
25mM $MgCl_2$

PCR Reagent mix: 15µl master mix solution
2-5µl isolated DNA
0.5µl each forward/reverse primer (conc. of 0.1µg/µl)
ddH_2O to total volume of 40µl

Protocols

For the majority of this investigation three PCR protocols were adopted for basic fragment amplification, with 'Hotspur' preferentially used unless stated. Each reaction was carried out on a Biometra T3000 Thermocycler machine.

‘Simple’: 2 mins 95°C
15 secs 48°C
30 secs 72°C
30 secs 95°C (Repeat from second step 30 times)
4 mins 72°C

‘Hotspur’: 3 mins 94°C
90 secs 95°C
90 secs 48°C
100 secs 72°C (Repeat from second step 35 times)
5 mins 72°C

‘HotJay’: 3 mins 98°C
15 secs 98°C
15 secs 48°C
30 secs 72°C (Repeat from second step 30 times)
5 mins 72°C

8.7.1. Splicing by Overlap Extension (SOE)

This procedure is usually used in order to remove a sequence from between two genes, thus leaving the two genes joined together. For this investigation it was used in a slightly different way, to amplify a large DNA fragment (S3) in two halves and then to ligate these pieces together through the PCR. The principle protocol used for the reaction was adapted from one designed by Leonides Calvo-Bado (unpublished data), and is proven to work with bacterial DNA. It utilises three separate PCR steps. The first one generates half of the desired larger fragment. The second step then encourages this half to bind to genomic template DNA, creating double strands of DNA with a distinctive primer sequence at the end of the S1 fragment. The final step is then to amplify the entire S3 region, which should be possible as there is now more desired template to recognise. This step is longer than the previous two due to the increase in the length of DNA being formed.

	Reaction mix	*Cycle*
Step 1	100ng template DNA	2 mins 94°C
	50pmol Primers (*PCRTWF1*/*PCRTWR1*)	45 secs 94°C
	200µM dNTPs	45 secs 50°C
	2.5U High Fidelity Taq Polymerase	90 secs 72°C (Repeat from 2nd step 10x)
	1x Buffer	45 secs 94°C
	5% DMSO	45 secs 55°C
	ddH_2O up to 40µl	90 secs 72°C (Repeat from 5th step 10x)
		5 mins 72°C

Step 2	50ng template DNA	2 mins 94°C
	1µl Step 1 product	15 secs 94°C
	1x Buffer	30 secs 45°C
	2.5U High Fidelity Taq Polymerase	2 mins 72°C
	ddH_2O up to 40µl	5 mins 72°C

Step 3	10µl Step 2 product	2 mins 94°C
	2µl Step 1 product	45 secs 94°C
	50pmol Primer (*PCRTWR2*)	45 secs 50°C
	200µM dNTP	150 secs 72°C (Repeat from 2nd step 10x)
	2.5U High Fidelity Taq Polymerase	45 secs 94°C
	1x Buffer	45 secs 55°C
	5% DMSO	150 secs 72°C (Repeat from 5th step 10x)
	ddH_2O up to 40µl	5 mins 72°C

8.7.2. QRT-PCR

When running a quantitative real-time PCR reaction a 96-well plate was used, which enables the simultaneous analysis of 96 reactions. As a result it was possible to run all of the desired reactions for a single gene of interest on one plate and still include a control housekeeping gene for comparison. For this investigation the housekeeping gene control was *rpoD*, formerly known as *sigA*. As a comparative DNA standard, pure *Myxococcus xanthus* DNA was used in a dilution series, producing final measurements of 0, 0.001, 0.01, 0.1, 1, 10 and 20ng DNA per reaction. These figures could then be used to create a standard dilution curve if required. Each reaction was identically treated, with the same volume of template DNA solution used – 0.5µl of each appropriate concentrated solution. When studying the specific genes, the template mixture was designed so that the final DNA concentration was 0.5ng/µl in each reaction well.

Reaction mix: 1µl Primers (0.5µl of each 0.5µM primer)
0.5µl template DNA
0.2µl 50x ROX reference dye
10µl 2x Mastermix
8.3µl ddH_2O

Each reaction was conducted in triplicate in order to reduce the overall likelihood of experimental error. This was done for both control and individual genes

studied. An example of one of the plates used is detailed below, clearly showing that the majority of the wells were being used.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	NTC	0	0.001	0.01	0.1	1	10	20	D0*a*	D0*b*	D0*c*	L0*a*
B	L0*b*	L0*c*	D2*a*	D2*b*	D2*c*	L2*a*	L2*b*	L2*c*	D6*a*	D6*b*	D6*c*	L6*a*
C	L6*b*	L6*c*	D48*a*	D48*b*	D48*c*	L48*a*	L48*b*	L48*c*				
D	101*a*	101*b*	101*c*	303*a*	303*b*	303*c*	717*a*	717*b*	717*c*	718*a*	718*b*	718*c*
E	NTC	0	0.001	0.01	0.1	1	10	20	D0*a*	D0*b*	D0*c*	L0*a*
F	L0*b*	L0*c*	D2*a*	D2*b*	D2*c*	L2*a*	L2*b*	L3*c*	D6*a*	D6*b*	D6*c*	L6*a*
G	L6*b*	L6*c*	D48*a*	D48*b*	D48*c*	L48*a*	L48*b*	L48*c*				
H	101*a*	101*b*	101*c*	303*a*	303*b*	303*c*	717*a*	717*b*	717*c*	718*a*	718*b*	718*c*

The first four rows are for the control gene, *rpoC*, and the latter four for *crtW*. Individual letters (*a*, *b* and *c*) indicate the three triplicate samples studied. NTC represents a negative template control, with no RNA or DNA included in the reaction. The other *M. xanthus* strains studied (DK101, UWM303, DK717 and DK718) were used as additional controls and measures of gene expression across the range of *M. xanthus* carotenoid mutants.

8.8. Sequencing

The Biological Sciences in-house sequencing facility at the University of Warwick used an ABI Prism 3130xl Genetic Analyser. This operated through the combination of a submitted DNA sample and a standard dye terminator sequencing ladder. The sample was run with POP7, an acrylamide-based liquid polymer on a 36cm capillary array. This means a single sequencing run takes only 45 minutes for a given sample. The results produced were between 600 and 800bp in length; the machine is not designed to sequence larger fragments. The sample itself was submitted in a total volume of 10µl in a 0.5ml microcentrifuge tube. Of this final volume, the majority of the sample was ddH_2O, added to 5.5pmol primer and an appropriate amount of DNA. This varied depending on the type of DNA used and the initial purification method adopted, but as general guidance the table overleaf was followed.

DNA size	Amount required
100-200bp	2-5ng
200-500bp	5-17ng
500-1000bp	8-33ng
1000-2000bp	17-67ng
Over 2000bp	33-85ng
Plasmids up to 15kb	250-500ng

Sequencing has been found to be hampered when attempting to sequence GC rich samples, over 65% GC content, which includes the majority of *M. xanthus* samples. The advice was to add 4% DMSO to the final sequencing reaction. However with the exception of a couple of instances, the standard sequencing reaction worked fine on *M. xanthus* DNA and DMSO addition was not required.

8.9. P1 Phage work

TM Buffer: 50mM TrisHCl (pH 7.5)
10mM $MgSO_4$

8.9.1. P1 culture preparation

Before attempting P1 packaging, the P1 phage had to be cultured so that a stock solution of at least 5 x 10^8 pfu was achieved. To do this an *E.coli* lawn of strain MC1061 (See 'P1 Packaging') was grown on an LC agar plate. To this 50µl of P1::Tn*9 clr*100 phage solution of unknown concentration was added. The plate was then incubated at 33^{o}C for two days, until plaques were visible. The plaques were visible as small, round clear circles on an otherwise unblemished cell lawn. A number of plaques were picked from the plate, using a glass pipette, removed in the form of a small plug. The plugs were then placed in 3 volumes of TM buffer and given a short vortex to resuspend all of the phage. The solution was mixed with 250µl *E. coli* MC1061 in LGC (grown to 0.21 at OD_{600}), and 200µl of the resultant mixture spread on an LGC agar plate. The plate was again incubated at 33^{o}C for 24-48 hours until plaques were visible. Any plates that displayed signs of lysis, now evident at a much higher level than from the initial P1 solution, were soaked with 6ml TM buffer. The buffer was poured off and stored with a drop of chloroform in the

refrigerator ready for use during P1 packaging. The novel phage concentration was calculated as being $1.3x10^{10}$ pfu.

8.9.2. P1 packaging

(The following process was carried out using *E.coli* MC1061 cells, as the process requires a Rec^+ *E. coli* strain. This is due to the fact that the plasmid requires a rec system in order to integrate into the phage as well as the fact P1 does not grow well with Rec^- strains.)

E. coli MC1061 strains containing the previously created plasmids, pJRNS1 and pJRNS2, were grown overnight at 37°C in 10ml $LBKan^{40}$ (Kanamycin at 40µg/ml). The following morning 250µl of each cell suspension was mixed with 5ml $LBKan^{40}$, and $MgCl_2$ added to a final concentration of 10mM. Again the tube was incubated at 37°C, until the culture had a reading of 0.21 at OD_{600}, which was approximately 80 minutes. To this 0.5ml of the pre-prepared P1::Tn*9* *clr*100 was added (See 'P1 culture preparation') and the mixture left for 20 minutes at room temperature with no form of shaking. The mixture was further diluted into 5ml $LBKan^{20}Cam^{12.5}$ and 10mM $MgCl_2$ again, before being left shaking at 30°C overnight. The following day 100µl of the overnight culture was diluted in 5ml $LBKan^{20}Cam^{12.5}$ and 10mM $MgCl_2$. The tube was once more incubated at 30°C, until the culture produced a reading of 0.21 at OD_{600}, again after approximately 80 minutes. At this point the culture was transferred to a 42°C water bath and incubated with vigorous shaking for a period of 35 minutes. The tube was then incubated for a further 2 hours at 37°C whilst being gently shaken. During this time the mixture could be seen to turn grainy in appearance as cell lysis occurred. Upon completion of the incubation, 50µl of chloroform was added and the mixture returned to the 37°C incubator without shaking, for an additional 15 minutes. The entire mixture was then centrifuged for a further 15 minutes at 16,100g. The final solution was then stored under a small amount of chloroform and refrigerated. It will remain viable for a number of months using this method.

To use the P1 phage in conjunction with *E. coli* strain MC1061 and also ensure a number of colonies grew, a series of tubes were prepared, with varying lysate and TrisMg buffer volumes:

Ingredient	TUBE 1	TUBE 2	TUBE 3
M. xanthus (0.21@600nm)	500µl	500µl	500µl
50mM $CaCl_2$	100µl	100µl	100µl
1M TrisMg Buffer	300µl	200µl	-
Lysate mix	100µl	200µl	400µl

The three tubes were incubated at room temperature for 40 minutes, without shaking. To each tube 3ml of molten DCYSA was added, mixed and then poured onto a DCY Kan^{20} plate. The plate was left to set before being incubated at 33°C overnight. The following day a further 2.5ml of molten DCYSA Kan^{130} was poured on top of the plates, giving an overall kanamycin concentration throughout the plate of 75µg/ml. All of the plates were then incubated for 4-5 days, again at 33°C, until colony growth was visible.

8.10. Electroporation

This protocol is adapted specifically for *M. xanthus* using a method designed by Youderian & Hartzell (1992) as the basis.

Materials

In addition to a *M. xanthus* culture, for each individual reaction:

Universal containing 1.5ml DCY
Microcentrifuge tube containing 5µl desired plasmid (at 50µg/µl) for electroporation
3ml DCYSA (freshly molten)
1ml DCY
2ml sterile distilled water
DCY Kan^{50} plate with Kanamycin (50µg/ml)

M. xanthus cultures were grown up in 20ml DCY until they reached an exponential growth stage (OD_{600} ~ 0.5). This was usually achieved after 48 hours; following initial inoculation using a fresh loopful of *M. xanthus* cells from a DCYA plate. The cells were pelleted through centrifugation of the tubes at 16,100g for 10 minutes. This was carried out at room temperature, with the pellets produced then resuspended in 0.5ml of sterile distilled water. The centrifugation and wash step was repeated three times, ensuring any traces of DCY had been completely removed. From the final suspension, 200µl of cell mixture was transferred to a pre-prepared

microcentrifuge tube containing 5µl of the plasmid DNA for electroporation. This was mixed by gentle pipetting and then transferred to a fresh electroporation cuvette. Using a 'Biorad Gene Pulser 2' the cuvette was electroporated at 400 Ω, 25 µFD and 0.65kV (if done accurately, a time constant reading of >9.0 was observed). Immediately following electroporation, 1ml of DCY was added to the cuvette, again mixed by pipetting and then the mixture transferred to an awaiting universal tube containing 1.5ml DCY. Speed is of the essence at this stage, as the *M. xanthus* cells tend to clump together if not resuspended soon after treatment. Each universal tube was then transferred to a shaking incubator at 30^{o}C for 4-8 hours.

Following incubation, 3ml of molten DCYSA was added to each universal tube and the resultant mixture poured onto the DCY Kan^{50} plates. Once set, the plates were incubated at 30^{o}C for 4-7 days or until colony growth was clearly visible. Colonies could then be screened through replating and using the PCR to establish whether the plasmid had been successfully electroporated.

8.11. LacZ and Lowry protein assay

The protocol below is adapted from D. A. Hodgson (1993) and was utilised for each of the beta-galactosidase assays carried out. All OD readings were taken using an Ultrospec 3000 Pro machine.

Materials

Z Buffer:	60mM Na_2HPO_4	40mM NaH_2PO_4
	10mM KCl	1mM $MgSO_4$
Z Buffer/ONPG/Beta-mercaptoethanol:		Z Buffer (pre-prepared and autoclaved)
		ONPG 1mg/ml
		Beta-mercaptoethanol (2.5µl/ml)

1M Na_2CO_3
5% Na_2CO_3
1M NaOH
Folin-Ciocalteu reagent
0.1% SDS
2% Na/K Tartrate
1% $CuSO_4.5H_2O$
ddH_2O

Sampling

Sampling in the assays was carried out on growing cells, so cell density often varied between tested time points. To account for this, prior to sampling each individual culture density was determined at OD_{600} and an appropriate volume of cells taken based on this result. This followed a linear pattern:

10ml of suspension if $OD_{600} = 0.1$

5ml of suspension if $OD_{600} = 0.5$

1ml of suspension if $OD_{600} = 1.0$

Etc.

Each sample was centrifuged at 16,100g for ten minutes and then the cell pellet resuspended in 500µl Z buffer. A second centrifugation was then carried out under the same conditions, followed by a further resuspension step in 500µl Z buffer. The sample was then split into three aliquots of 150µl. One of these was further processed through a third centrifugation and pellet resuspension in 150µl ddH_2O. 150µl 1M NaOH was then added and mixed through tube inversion, causing the cells to lyse (visible as the cloudy mixture turns colourless). The three tubes were then snap-frozen for use in both the LacZ assay and Lowry Protein assay.

8.11.1. LacZ assay

Prior to conducting the assay two blank samples were prepared as reference points when taking OD measurements. Blank B was incubated with all of the other tested samples at 37°C, for a period of three hours, in order to generate a measurement for the spontaneous rate of ONPG hydrolysis. All of the samples could then be directly referenced against Blank A and the reading for Blank B deducted from each individual reading in order to produce an accurate assay value.

Blank A:	50µl 1% SDS
	0.5ml Z buffer
	0.5ml 1M Na_2CO_3
Blank B:	50µl 1% SDS
	0.5ml Z Buffer/ONPG/Beta-mercaptoethanol
	0.5ml 1M Na_2CO_3

The two unprocessed sample tubes (without additional NaOH) were thawed and split into 4 further microcentrifuge tube samples, 2 of 10µl (1X samples) and 2 of 50µl (5X samples). Each was made up to a total volume of 100µl through the addition of Z buffer and then a further 50µl 1% SDS added. To begin the assay, 0.4ml of ice cold Z Buffer/ONPG/Beta-mercaptoethanol was added to each tube and the tubes then immediately transferred to a 37^{o}C water bath. The tubes were monitored until a colour change was observed, whereby the colourless solution turned a distinctive pale yellow. This time point was noted and 0.5ml 1M Na_2CO_3 added to the tube, providing colour enhancement. The mixture then had an absorbance reading measured at OD_{420} using the two blanks for comparison. Any samples that took longer than 3 hours to change in colour were discarded. All times were tabulated and the data processed in Microsoft Excel 2003.

8.11.2. Lowry protein assay

Prior to carrying out the assay, as with the 'LacZ assay', a suitable blank was created for sample comparison. This was processed identically to each of the other samples to provide a control reference. In addition, two solutions ('Reagent C' and 'Ciocalteu mix') were prepared fresh at the beginning of each assay, immediately before they were required in the experiment.

Blank C: 300µl 1M NaOH
Reagent C: 5% Na_2CO_3 2% Na/K Tartare 1% $CuSO_4.5H_2O$ (**50:1:1 ratio**)
Ciocalteu mix: Equal amount of Folin-Ciocalteu reagent and ddH_2O

Initially a standard protein curve was constructed for use in protein calculations. This was done by the preparation of a dilution series of BSA (Bovine Serum Albumin) in 300µl of 0.5M NaOH. An identical assay procedure to the one detailed below was then carried out for each dilution. From the values recorded a standard curve was created and from this the amount of protein in each sample could be estimated.

The protein assay was conducted using the remaining original sample tube, in which the cells had already undergone lysis. The tube was further separated into two

aliquots of 50µl and 250µl, and each sample made up to a total volume of 300µl through the addition of 1M NaOH. The tubes were transferred to a boiling water bath for a 5 minute period and then allowed to cool to room temperature. Once cool, 750µl fresh Reagent C was added to each tube and the samples incubated for a further 10 minutes at room temperature. 150µl of 'Ciocalteu mix' was added to the tubes and they were incubated at room temperature again, for an additional 30 minutes. Samples were then finally transferred to cuvettes and an absorbance reading at OD_{750} taken, standardising the value against that of Blank C. The OD value could then be read from the standard curve and the amount of protein present in each sample estimated.

8.12. Standard Polyacrylamide Gel Electrophoresis

1 X TBE Buffer: 10.781g Tris-base
0.58g EDTA
5.5g Boric acid
ddH_2O up to 1l

When running DNA samples on a gel following RE digestions or a PCR, a 1% agarose gel was used e.g. for a small gel, 60ml of 1xTBE buffer and 0.6g agarose.

All gels were run in Biorad Gel Electrophoresis trays at 100V for a period of 60 minutes. If larger gels were used, over $150ml^3$, then this was increased to 150V for 90 minutes in duration, ensuring adequate band separation. The resultant gels were then visualised using a GeneFlash Syngene Bio-Imaging gel documentation system.

8.12.1. SDS-PAGE analysis

A 9% standard SDS-PAGE (sodium dodecyl sulphate - polyacrylamide gel electrophoresis) method was used in order to observe proteins during this investigation. All gels were made and run using a Biorad Gel Electrophoresis Kit.

Materials

Loading Gel: 2ml ddH_2O
1ml 30% acrylamide
1.875ml Tris-HCl (pH 8.8)
50µl 10% SDS
15µl TEMED
10µl 10% APS (Ammonium persulphate)

Stacker Gel: 3.675ml ddH_2O
0.65ml 30% acrylamide
0.625ml Tris-HCl (pH 6.8)
50µl 10% SDS
15µl TEMED
25µl 10% APS

1x Running buffer (Laemmli buffer): 3g Tris-base
14.4g Glycine
1g SDS
All in 1l ddH_2O

Coomassie Blue staining solution: 0.1% (w/v) Coomassie Blue
20% (v/v) methanol
10% (v/v) acetic acid.

Destain solution: 50% (v/v) methanol
10% (v/v) acetic acid.

Method

The loading gel was poured into the gel tray until approximately 10mm below the tip of where the well comb will finish, and then when set the stacker gel was poured on top of the gel and the comb inserted. It takes approximately 15 minutes for the gel to set completely. To ensure the gel sets evenly on the top, a small amount of isopropanol was pipetted on top of the loading gel following pouring. This results in a clear two layer separation keeping the top of the gel flat. It also made it easier to determine exactly when the gel has set, as an additional layer became visible between the gel and alcoholic covering layer. The isopropanol was removed once the gel was set by using a piece of filter paper to absorb the excess alcohol, then the stacker gel was added. The gel itself was run for 50 minutes at 200V in 1x running buffer, by which time the smallest bands had reached the very bottom of the gel. 25µl of each sample was loaded into a well, as any excess tended to lead to overflow into adjoining

wells. To each 5µl of sample used, 1µl of SDS Protein Dye was added, and as a comparison 5µl of Fermentas Pageruler^{TM} Plus Prestained Protein Ladder was run on the gel alongside the samples.

When complete the gel was removed from the plates by carefully separating the plates using a scalpel blade, taking care not to break either the plates or the gel. The gel remained attached to one of the plates was easily removed by using the scalpel blade in conjunction with a small amount of water pipetted between the gel and plate. The gel was then transferred to a small plastic box along with 100ml of Coomassie Blue staining solution, submerging the gel. This was gently agitated on a shaker for 15 minutes, after which time the excess staining solution is poured away. Three further washes using the Destain solution were then carried out for 30 minutes each, which was enough time to remove the majority of excess dye from the gel. The gel was then visualised under a bright white light source for optimal results.

8.12.2. Sarkosyl extraction

***Lysis buffer* (pH 7.0-8.0):** 50mM Tris-HCl (pH 7.8)
5% Glycerol
50mM NaCl

100µl of a previously harvested *E. coli* strain cell culture was pelleted through centrifugation at 16,100g for 15 minutes. All supernatant was removed and the remaining pellet weight established. The pellet was then freeze thawed twice, by snap-freezing and allowing it to thaw gradually at room temperature. It was then resuspended in lysis buffer, with 7ml buffer used per gram of cell pellet produced. Additionally to this mixture 7.5kU lysozyme was added for each 1ml of lysis buffer used. Sarkosyl detergent was added to a total volume of 0.05%, acting to solubilise any mildly-hydrophobic protein molecules present within the sample. Finally a quarter-volume of silica beads, 0.07-0.08mm in diameter, were included in the mixture to enable the break-up of the cells. The entire suspension was shaken vigorously for 2 hours at 4^{o}C. The insoluble cell debris was separated via centrifugation and the supernatant removed for use in an SDS-PAGE reaction. The resuspension process was then repeated with the latest cell pellet using an increased concentration of sarkosyl detergent, 0.2%, in order to solubilise the more highly-

hydrophobic proteins associated with the outer cell membrane. No further bead-beating was carried out at this stage, solely a resuspension step.

8.13. RACE

The following procedure is adapted from Chapters 8.54-8.60 of Molecular cloning – A Laboratory Manual, 3rd edition, Sambrook and Russell, Spring Harbor Laboratory Press (2001).

10µl of previously synthesised cDNA was taken and made up to a final volume of 20µl through the addition of: 4µl 5x Tdt Buffer

4µl 1mM dATP

10U Tdt

ddH_2O up to 20µl

The mixture was incubated at 37°C for a period of 15 minutes allowing the addition of polyA DNA tails. A further 3 minute incubation at 80°C was then required to inactivate the Tdt. This negated the need for enzyme removal from the reaction at this stage. For the amplification step the mixture was as follows:

10µl cDNA mixture (from previous step)
5µl 10x amplification buffer *
5µl 20mM solution dNTPs *
1.6µl 25mM Adapter TTT Primer (17dTs)
3.2µl 25mM Adapter Primer
3.2µl Gene specific primer
1µl DNA Polymerase *
21µl ddH_2O

**Alternatively these ingredients may be substituted for 11µl of the ready-made Promega 2x Mastermix.*

The amplification mixture was then used in a PCR:

5 mins 94°C
5 mins 48°C
40 mins 72°C
40 secs 94°C
1 min 48°C
3 mins 72°C
40 secs 94°C (Repeat from 4th step 40x)
1 min 48°C
15 mins 72°C

8.14. Ketolase Assay

(*The method below is adapted from Tao and Cheng, 2004)*

Materials

50mM Tris-HCl (pH7.5)
50mM Tris-HCl (pH7.8)
Triton x100
β-carotene substrate
Acetone
Ethanol and dry ice (or alternatively liquid nitrogen)
Methanol
Diethyl ether
NaCl

Method

E.coli DH5α [pBADJAY] strains that can express *crtW* were grown up to an OD_{600} in 50ml liquid media. The sample was then centrifuged for 20 minutes at 16,100g to pellet the cultured cells. Using ethanol and dry ice the cell pellet was subsequently frozen and then thawed. The freeze-thawing process was repeated three times. The pellet was then resuspended with 2ml of ice-cold 50mM Tris-HCl (pH 7.5) and 0.25% Triton x100. 10μg β-carotene substrate mixed with 50μl acetone was then added to the suspension and the entire sample mixed thoroughly by pipetting. The resultant mixture was then evenly split into two microcentrifuge tubes and once again frozen and thawed. Both tubes were centrifuged at 16,100g for a further 20 minutes and the two lots of supernatant combined before being transferred into a fresh microcentrifuge tube. Any remaining cell debris was discarded and the supernatant removed diluted with 3ml of 50mM Tris-HCl (pH 7.8) in a 50ml flask. The mixture was then incubated at 30^{o}C, whilst being shaken for a period of 2 hours. 5ml methanol was added to stop the reaction, along with 5ml diethyl ether and a small amount of NaCl which acted to separate the mixture into two phases within the flask. The upper layer was the one required and contains the carotenoids, so was pipetted off and transferred to an awaiting bijou and air-dried. A small carotenoid pellet remained for further analysis.

8.14.1. HPLC

Following the preparation of a carotenoid sample from *M. xanthus* or *E.coli* (as in 'Ketolase Assay') high performance liquid chromatography (HPLC) can be carried out. A Waters Nova-Pak, 6µm C_{18}, 3.9 x 500mm reversed phase C_{18} column was used in conjunction with a mobile phase of acetonitrile/methanol/isopropyl (90:6:40, v/v/v). Any results produced following the passing of the sample through the column were then analysed by comparing them to known standards of various carotenoid compounds measured previously under identical conditions. This process can detect a minimum of 0.0003 absorbance units (0.5pmol/h/mg enzyme activity). The resultant data are then displayed in graphical peak format.

8. 15. Primers

The list below is of all the key primers used in this investigation, their base sequence and the primary experiment/s for which they were used. The majority were designed using the computer program DNA Star 5 – Primer Select.

Name	Sequence (5'- 3')	Role
PCRTWF1	CAGGATCCATCCAGCAACTCCGGCTCCGTAATCG	S1/S3 fragment amplification 'A' operon fragment amplification
PCRTWR1	CTGAATTCGTGCCTCGAAGGCGGCCTTGTAGTTG	S1/S3 fragment amplification 'A' operon fragment amplification
PCRTWF2	CAGGATCCATCAACTACAAGGCCGCCTTCGAGGC	S2 fragment amplification 'B' operon fragment amplification
PCRTWR2	CTGAATTCGTGTGATGAAGAGGCCCGTGCACAGC	S2/S3 fragment amplification 'B' operon fragment amplification
274 FOR	CATCGCCCACGCATAACCGA	Confirmation of pDAH274 insert
274 REV	GTTGTAAAACGACGGGCAGT	Confirmation of pDAH274 insert
M13 F	GTAAAACGACGGCCAG	Confirmation of 2.1-TOPO insert
M13 R	CAGGAAACAGCTATGAC	Confirmation of 2.1-TOPO insert
DOU CHECK F	ATGAAGAAGGCCGTCATCTC	'B' operon fragment amplification
3rd FOR	CGTTGCAGCTCTTCTACTTC	'C' operon fragment amplification
3rd REV	CGAACTCGTAGAAGTTGTTG	'C' operon fragment amplification
4th FOR	GCTCCTTGGAAGTATGGATT	'D' operon fragment amplification
4th REV	CGTAGGTGAAGGCCAAGAGG	'D' operon fragment amplification
Adapter 2	TACTCAAGCTATGCATC	RACE specific primer
Adapter TTT	TACTCAAGCTATGCATCTTTTTTTTTTTTTTTT	RACE specific primer + 17 dTs
KOI F	CCGCGGCTACAAGGCCGCCTTCGAGGCCTT	PDP knockout fragment generation
KOI R	GCTAGCCTCACCTGCTTTCTGGCCATCGCC	PDP knockout fragment generation
KOII F	GCTAGCCGCGCGGGCGGCCCGGGCCGTTAT	ORD knockout fragment generation
KOII R	GAATTCTCAGTAGAACCGGCGCAGGTCCAT	ORD knockout fragment generation
KOKAN F	GCTAGCATGATTGAACAAGATGGATTGCAC	*kan* knockout fragment creation
KOKAN R	GCTAGCGAAGAACTCGTCAAGAAGCCGATA	*kan* knockout fragment creation
KANR FOR	TTGGCTACCCGTGATATTGCTGAA	Testing *kan* gene insertion
KANR REV	TCTACCTTGCTCCTGCCGAGA	Testing *kan* gene insertion
PRE CRTW (F)	ACGCGTTCCTGTCCAACGACTACT	*crtW* knockout detection
KO YEDZ F1	CTGCAGATGCGCGACAAGCCGCCCGCTGA	*yedZ* knockout creation
KO YEDZ R1	GCTAGCGGCGCGGGGACTCAGTAGAA	*yedZ* knockout creation
KO YEDZ F2	GCTAGCGTGAAGCGGTGCCCAGACGG	*yedZ* knockout creation
KO YEDZ R2	ACTAGTCCGTGCCGCTGCGTGAGCTG	*yedZ* knockout creation
CARQ PRO F	GGATCCGCGCCCAACACTAACGCCCTC	*carQRS* promoter region creation
CARQ PRO R	GCTAGCCGACTCCTCTACGGGCCACCT	*carQRS* promoter region creation
NEW CRTW F	GCTAGCATGGAGACTTCCGCCCGCCAA	*crtW* fragment for shuttle vector
NEW CRTW R	GAATTCTTACCGGGCGGTGCCTTCCTG	*crtW* fragment for shuttle vector
pBAD24 F	GGATCCATGGAGACTTCCGCCCGCCAA	pBAD24 insert creation

pBAD24 R	AAGCTTTTACCGGGCGGTGCCTTCCTG	pBAD24 insert creation
pET26b F	CATATGATGGAGACTTCCGCCCGCCAA	pET26b insert creation
pET26b R	GCATGGTTACCGGGCGGTGCCTTCCTG	pET26b insert creation
pBAD24 F 2	CCATGGATGGAGACTTCCGCCCGCCAA	New pBAD24 insert creation
PDP F	CTGAAGGCGTTCCTCACCTC	*pdp* detection in RT-PCR
PDP R	CGTCGTTGGGCAGGTTGAAG	*pdp* detection in RT-PCR
CRTW F	TCGTCATCGCGCTCATCATC	*crtW* detection in RT-PCR
CRTW R	TTGAACTTGGCCGCCATCAC	*crtW* detection in RT-PCR
ORD F	CAGCTCACGTCCTGGTTCTC	*ord* detection in RT-PCR
ORD R	CCTTCCGTGTACGCCCAATC	*ord* detection in RT-PCR
MMB F	CCTCTTGGCCTTCACCTACG	*mmb* detection in RT-PCR
MMB R	TGGCTTCACCCACGCGAATG	*mmb* detection in RT-PCR
RPOD F	TGCGGCTCAACAAGAAGCAG	*rpoD* detection in RT-PCR
RPOD R	AAATCACGCAGCTCGTCCTC	*rpoD* detection in RT-PCR

8.16. Strains and plasmids

Listed below are the plasmids, and *M. xanthus* and *E. coli* strains, used during this investigation, along with their primary phenotype or genotype. All plasmids used were developed during this investigation unless an alternative source is given.

M. xanthus

DK101 (D. A. Hodgson, 1993) Wild type
DK1622 (D. A. Hodgson) Wild type
DK717 (D. A. Hodgson) Car A^-
DK718 (D. A. Hodgson, 1993) $CarR^-$
UWM303 (McGowan *et al*., 1993) $CarQ^-$

UWM303 [pJRNS1] $CarQ^-$, *crtW* promoter region (i) (positions 20-1219) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
DK718 [pJRNS1] $CarR^-$, *crtW* promoter region (i) (positions 20-1219) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
DK101 [pJRNS1] *crtW* promoter region (i) (positions 20-1219) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
DK717 [pJRNS1] $CarA^-$, *crtW* promoter region (i) (positions 20-1219) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
UWM303 [pJRNS2] $CarQ^-$, *crtW* promoter region (ii) (positions 20-1243) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
DK718 [pJRNS2] $CarR^-$, *crtW* promoter region (ii) (positions 20-1243) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*

DK101 [pJRNS2]	*crtW* promoter region (ii) (positions 20-1243) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
DK717 [pJRNS2]	CarA$^-$, *crtW* promoter region (ii) (positions 20-1243) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*

E. coli

(For the majority of the work carried out in this investigation the E. coli strain used was DH5α unless stated otherwise)

DH5α (Invitrogen)	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
MC1061 (Invitrogen)	*araD139 Δ(araA-leu)7697 ΔlacX74 galK16 galE15(GalS) λ$^-$ e14$^-$ mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2*

Plasmids

pDAH101 (D. A. Hodgson, 1993)	*amp*, *kan*, *tet*, IS50L, IS50R
pDAH274 (Scanlan *et al*., 1990)	*lacZ*, P1 incΔ, *amp*, p15A *ori*
pGEM®-T Easy (Promega)	f1 *ori*, *lacZ*, *amp*
PCR®2.1-TOPO® (Invitrogen)	f1 *ori*, pUC *ori*, *amp*, *kan*, *lacZα*
TOP S1	f1 *ori*, pUC *ori*, *amp*, *kan*, *lacZα*, *crtW* promoter region (i) (positions 312-1511)
TOP S2	f1 *ori*, pUC *ori*, *amp*, *kan*, *lacZα*, *crtW* promoter region (ii) (positions 312-1535)
pJRNS1	*crtW* promoter region (i) (positions 20-1219) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
pJRNS2	*crtW* promoter region (ii) (positions 20-1243) controlling *lacZ*, P1 incΔ, *amp*, p15A *ori*
pBAD24 (Guzman *et al*., 1995)	*araC*, *amp*, pBR *ori*, pBAD promoter, pMB *ori*
pBADJAY	*araC*, *amp*, pBR *ori*, pBAD promoter controlling *crtW* (positions 1318-2107)
pET26b (Promega)	*kan*, *lacI*, f1 *ori*
pZJY41 (Zhao *et al*., 2008)	*amp*, *kan*, pMF1 *ori*
pZJY156 (Zhao *et al*., 2008)	*amp*, *kan*, pMF1 *ori*
PJAM156	*amp*, *kan*, pMF1 *ori*, *carQRS* promoter region controlling *crtW* (positions 3517-4448)

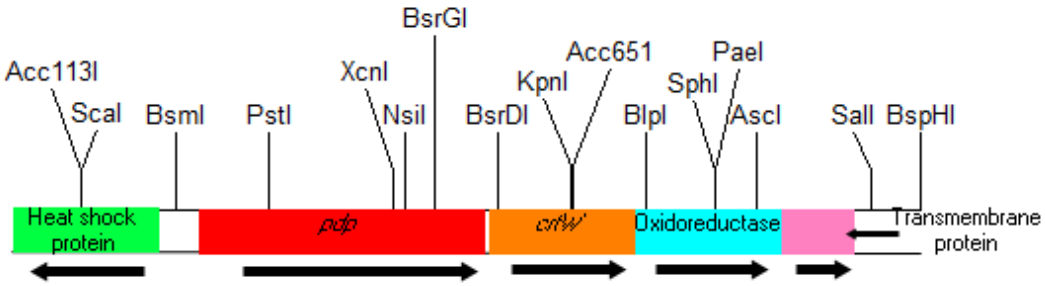
Appendices

Appendix 1

Listed below are the abbreviations for each amino acid that encodes the CrtW protein,

A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

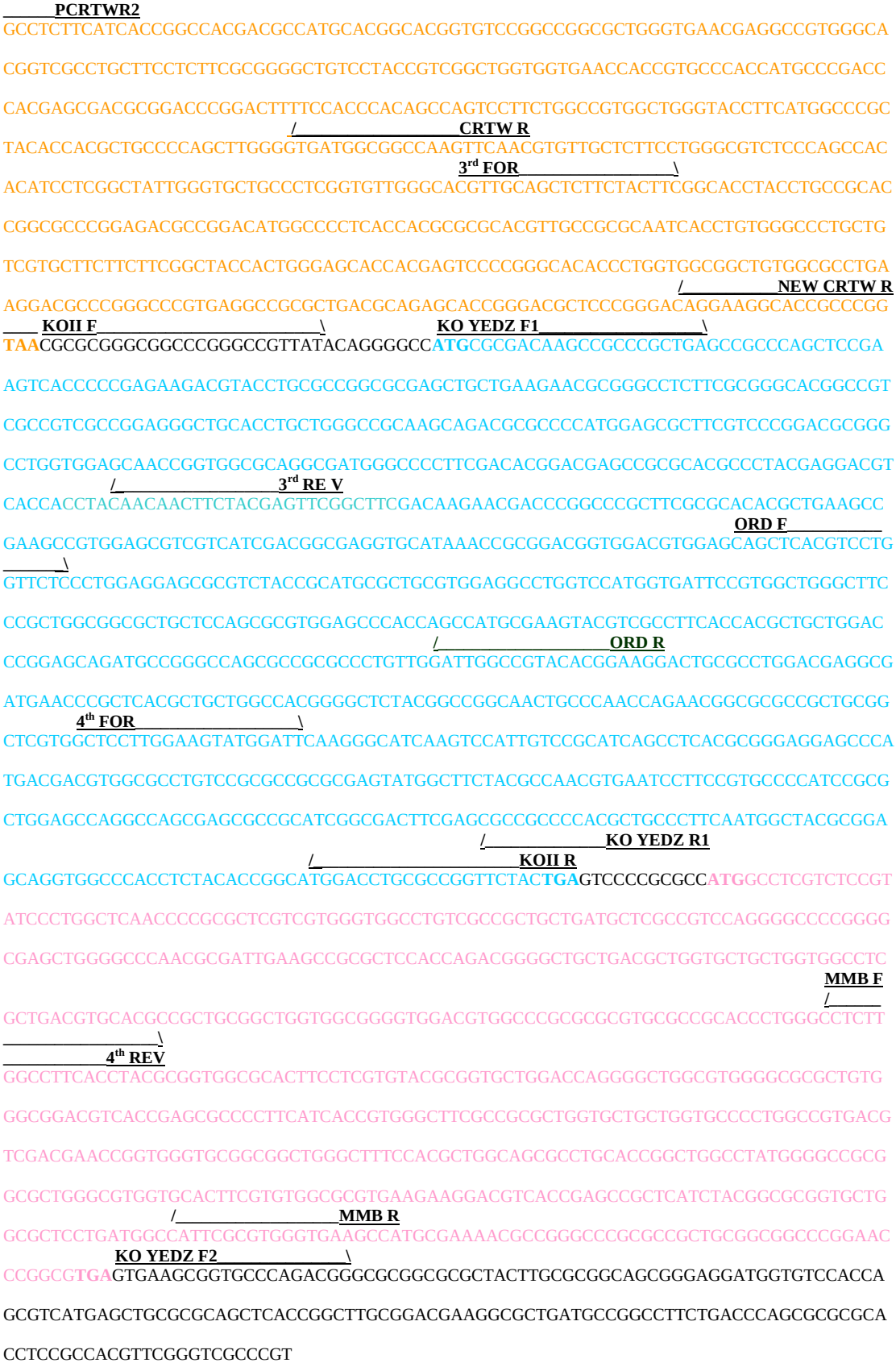
Appendix 2



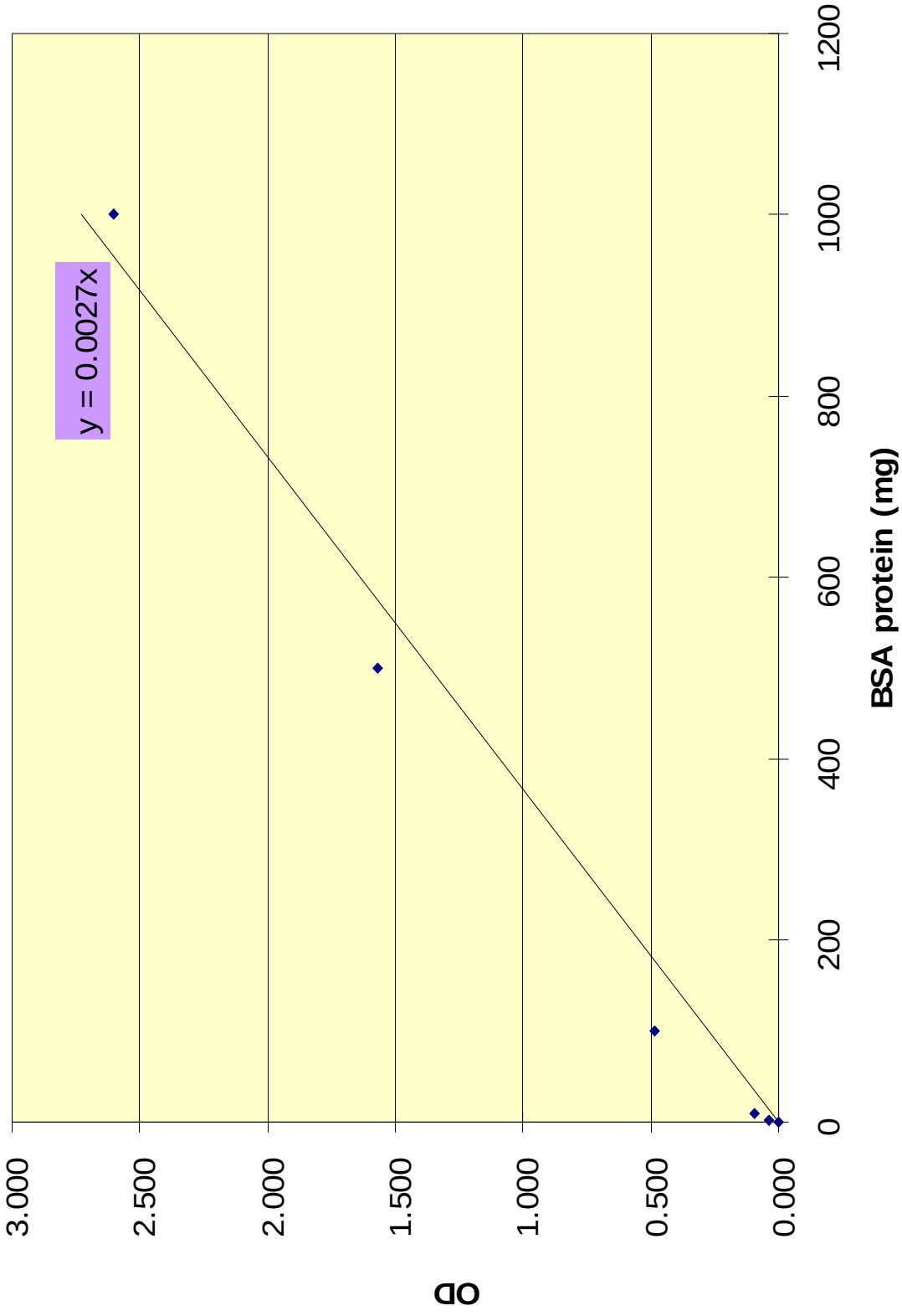
The original *crtW* operon region with a selection of unique restriction enzyme sites mapped. The bold arrows indicate the direction of gene transcription.

Appendix 3

A simple map of the primers used on the *crtW* operon during this investigation. All marked primers refer to the strand located beneath.

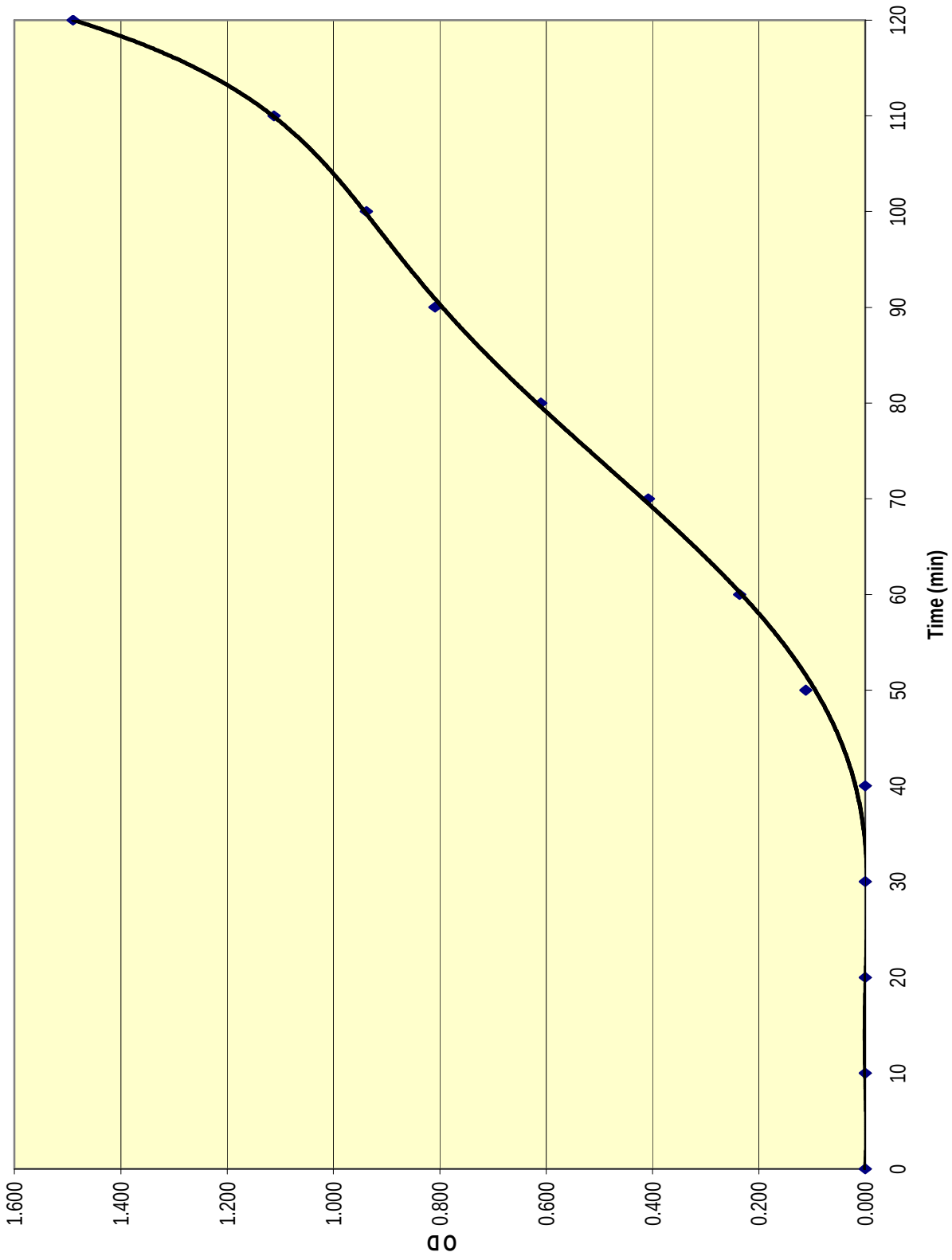

PCRTWR2
GCCTCTTCATCACCGGCCACGACGCCATGCACGGCACGGTGTCCGGCCGGCGCTGGGTGAACGAGGCCGTGGGCA
CGGTCGCCTGCTTCCTCTTCGCGGGGCTGTCCTACCGTCGGCTGGTGGTGAACCACCGTGCCCACCATGCCCGACC
CACGAGCGACGCGGACCCGGACTTTTCCACCCACAGCCAGTCCTTCTGGCCGTGGCTGGGTACCTTCATGGCCCGC
CRTW R
TACACCACGCTGCCCCAGCTTGGGGTGATGGCGGCCAAGTTCAACGTGTTGCTCTTCCTGGGCGTCTCCCAGCCAC
3rd FOR
ACATCCTCGGCTATTGGGTGCTGCCCTCGGTGTTGGGCACGTTGCAGCTCTTCTACTTCGGCACCTACCTGCCGCAC
CGGCGCCCGGAGACGCCGGACATGGCCCCTCACCACGCGCGCACGTTGCCGCGCAATCACCTGTGGGCCCTGCTG
TCGTGCTTCTTCTTCGGCTACCACTGGGAGCACCACGAGTCCCCGGGCACACCCTGGTGGCGGCTGTGGCGCCTGA
NEW CRTW R
AGGACGCCCGGGCCCGTGAGGCCGCGCTGACGCAGAGCACCGGGACGCTCCCGGGACAGGAAGGCACCGCCCGG
KOII F
KO YEDZ F1
TAACGCGCGGGCGGCCCGGGCCGTTATACAGGGGCCATGCGCGACAAGCCGCCCGCTGAGCCGCCCAGCTCCGA
AGTCACCCCCGAGAAGACGTACCTGCGCCGGCGCGAGCTGCTGAAGAACGCGGGCCTCTTCGCGGGCACGGCCGT
CGCCGTCGCCGGAGGGCTGCACCTGCTGGGCCGCAAGCAGACGCGCCCCATGGAGCGCTTCGTCCCGGACGCGGG
CCTGGTGGAGCAACCGGTGGCGCAGGCGATGGGCCCCTTCGACACGGACGAGCCGCGCACGCCCTACGAGGACGT
3rd RE V
CACCACCTACAACAACTTCTACGAGTTCGGCTTCGACAAGAACGACCCGGCCCGCTTCGCGCACACGCTGAAGCC
ORD F
GAAGCCGTGGAGCGTCGTCATCGACGGCGAGGTGCATAAACCGCGGACGGTGGACGTGGAGCAGCTCACGTCCTG
GTTCTCCCTGGAGGAGCGCGTCTACCGCATGCGCTGCGTGGAGGCCTGGTCCATGGTGATTCCGTGGCTGGGCTTC
CCGCTGGCGGCGCTGCTCCAGCGCGTGGAGCCCACCAGCCATGCGAAGTACGTCGCCTTCACCACGCTGCTGGAC
ORD R
CCGGAGCAGATGCCGGGCCAGCGCCGCGCCCTGTTGGATTGGCCGTACACGGAAGGACTGCGCCTGGACGAGGCG
ATGAACCCGCTCACGCTGCTGGCCACGGGGCTCTACGGCCGGCAACTGCCCAACCAGAACGGCGCGCCGCTGCGG
4th FOR
CTCGTGGCTCCTTGGAAGTATGGATTCAAGGGCATCAAGTCCATTGTCCGCATCAGCCTCACGCGGGAGGAGCCCA
TGACGACGTGGCGCCTGTCCGCGCCGCGCGAGTATGGCTTCTACGCCAACGTGAATCCTTCCGTGCCCCATCCGCG
CTGGAGCCAGGCCAGCGAGCGCCGCATCGGCGACTTCGAGCGCCGCCCCACGCTGCCCTTCAATGGCTACGCGGA
KO YEDZ R1
KOII R
GCAGGTGGCCCACCTCTACACCGGCATGGACCTGCGCCGGTTCTACTGAGTCCCCGCGCCATGGCCTCGTCTCCGT
ATCCCTGGCTCAACCCCGCGCTCGTCGTGGGTGGCCTGTCGCCGCTGCTGATGCTCGCCGTCCAGGGGCCCCGGGG
CGAGCTGGGGCCCAACGCGATTGAAGCCGCGCTCCACCAGACGGGGCTGCTGACGCTGGTGCTGCTGGTGGCCTC
MMB F
GCTGACGTGCACGCCGCTGCGGCTGGTGGCGGGGTGGACGTGGCCCGCGCGCGTGCGCCGCACCCTGGGCCTCTT
4th REV
GGCCTTCACCTACGCGGTGGCGCACTTCCTCGTGTACGCGGTGCTGGACCAGGGGCTGGCGTGGGGCGCGCTGTG
GGCGGACGTCACCGAGCGCCCCTTCATCACCGTGGGCTTCGCCGCGCTGGTGCTGCTGGTGCCCCTGGCCGTGACG
TCGACGAACCGGTGGGTGCGGCGGCTGGGCTTTCCACGCTGGCAGCGCCTGCACCGGCTGGCCTATGGGGCCGCG
GCGCTGGGCGTGGTGCACTTCGTGTGGCGCGTGAAGAAGGACGTCACCGAGCCGCTCATCTACGGCGCGGTGCTG
MMB R
GCGCTCCTGATGGCCATTCGCGTGGGTGAAGCCATGCGAAAACGCCGGGCCCGCGCCGCTGCGGCGGCCCGGAAC
KO YEDZ F2
CCGGCGTGAGTGAAGCGGTGCCCAGACGGGCGCGGCGCGCTACTTGCGCGGCAGCGGGAGGATGGTGTCCACCA
GCGTCATGAGCTGCGCGCAGCTCACCGGCTTGCGGACGAAGGCGCTGATGCCGGCCTTCTGACCCAGCGCGCGCA
CCTCCGCCACGTTCGGGTCGCCCGT

Appendix 4



A graph to show the optical density readings obtained from a dilution series of BSA, for use in LacZ assay calculations.

Appendix 5



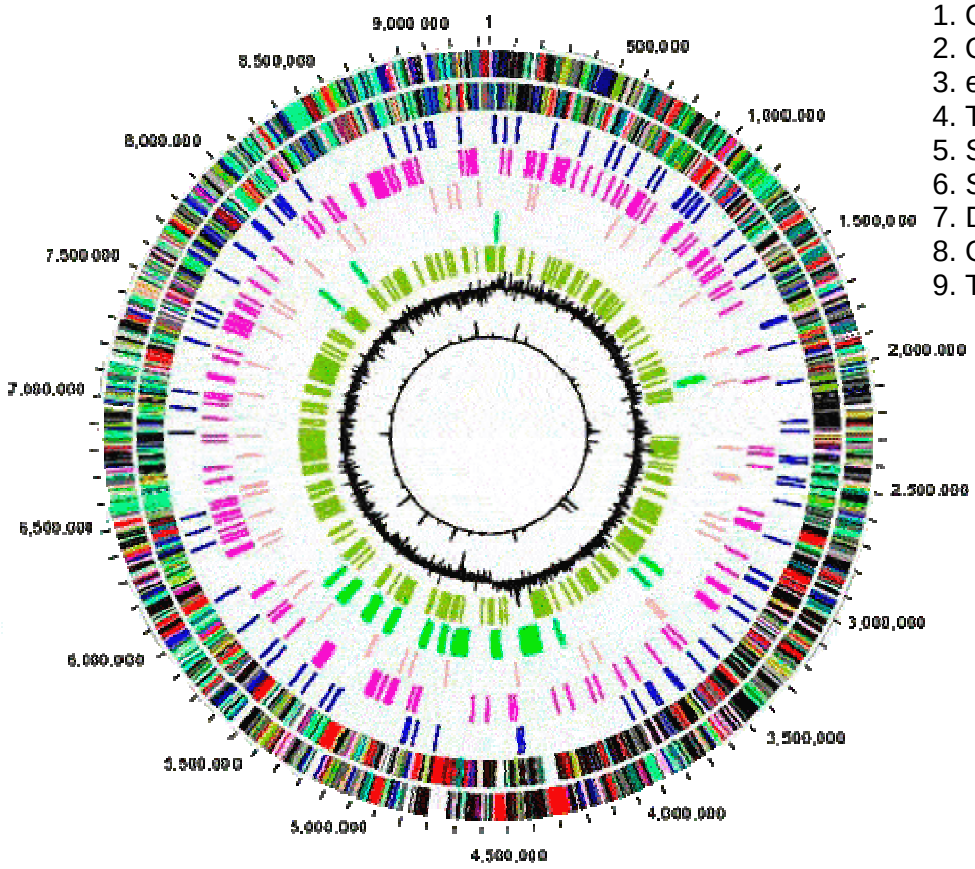
A graph to show the optical density readings obtained for an *E. coli* DH5α cell culture following fresh inoculation of an LB sample with a 24 hour old *E. coli* culture.

Appendix 6

A detailed *M. xanthus* DK1622 genome map with similar gene families colour coded and labelled.
This is taken directly from the *M. xanthus* website (www.xanthus.wikimods.org/myxopedia).

Layers from outside

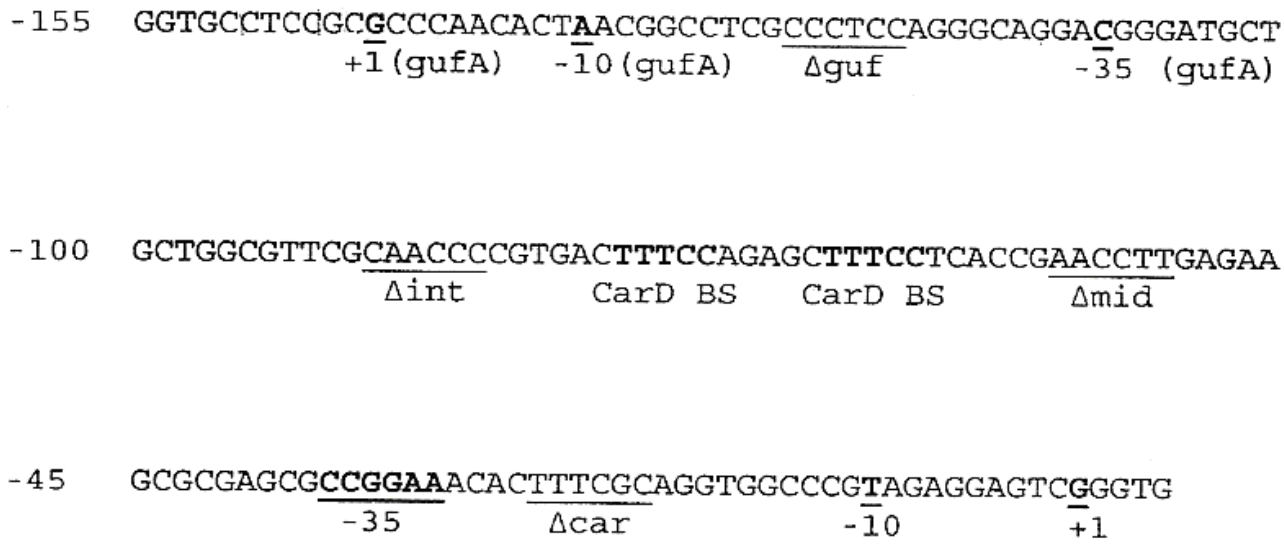
1. ORF (clockwise)
2. ORF (counter-clockwise)
3. ePKs
4. TCSTS
5. Sigma factors
6. Secondary metabolism
7. Digestive enzymes
8. GC-skew
9. Trinucleotide signature



Colour	Gene Role	# of Genes	%/7388 Genes
	Fatty acid and phospholipid metabolism	137	1.85%
	Biosynthesis of co-factors, prosthetic groups, and carriers	129	1.75%
	Cell envelope	734	9.94%
	Transport and binding proteins	407	5.51%
	Amino acid and biosynthesis	79	1.07%
	DNA metabolism	125	1.69%
	Transcription	100	1.35%
	Mobile and extra-chromosomal element functions	116	1.57%
	Protein synthesis	162	2.19%
	Protein fate	314	4.25%
	Cellular processes	410	5.55%
	Purines, pyrimidines, nucleosides and nucleotides	59	0.80%
	Conserved hypothetical proteins	838	11.34%
	Central intermediary metabolism	58	0.79%
	Regulatory functions	581	7.86%
	Signal transduction	267	3.61%
	Energy metabolism	343	4.64%
	Unknown function	1097	14.85%

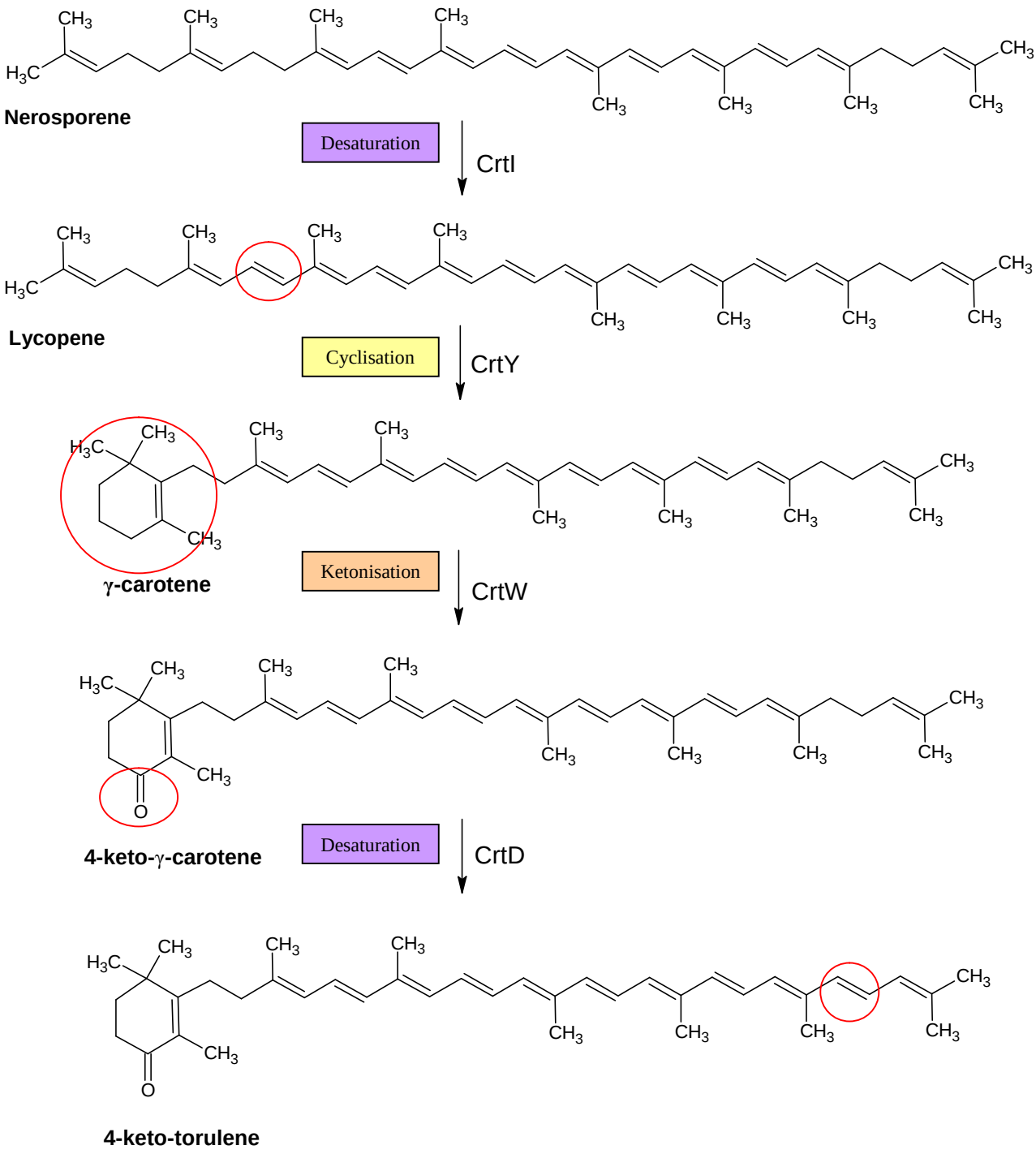
Appendix 7

The sequence below has been kindly reproduced from an original diagram (Whitworth *et al*., 2004) and then adapted. It depicts the *carQRS* promoter region starting from the +1 on the uppermost line until the final GTG on the third line.



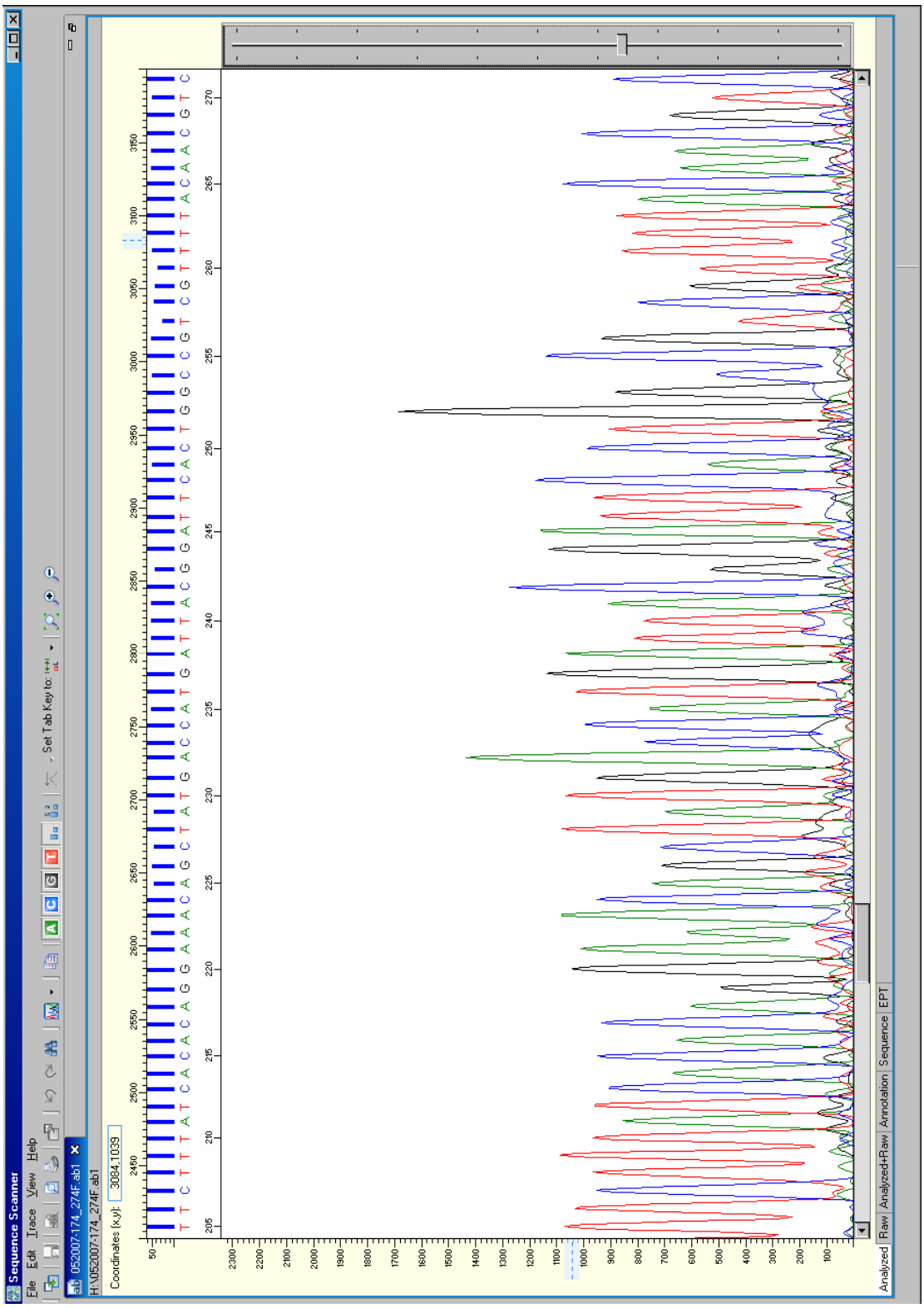
Appendix 8

The alternative part of the carotenoid pathway of *M. xanthus*, from the generated neurosporene up to the production of 4-keto-torulene. Each arrow is labelled with the enzyme that is believed to be responsible for that stage of the process, and the basic enzymatic process that occurs. The structural change at each stage is circled.



Appendix 9

An example of the sequencing data output obtained from an ABI Prism 3130xl Genetic Analyser as visualised using Applied Biosystems Sequence Scanner V 1.0.



Bibliography

Journals

Albrecht, M., Steiger, S., and Sandmann, G. (2001) Expression of a ketolase gene mediates the synthesis of canthaxanthin in *Synechococcus* leading to tolerance against photoinhibition, pigment degradation and UV-B sensitivity of photosynthesis. *Photochem. Photo biol*. **73:** 551-555

Anthony, J. R., Warczak, K. L., and Donohue, T. J. (2005) A transcriptional response to singlet oxygen, a toxic byproduct of photosynthesis. *Proc. Natl. Acad. Sci. USA*. **102:** 6502-6507

Armstrong, G.A. (1997) Genetics of Eubacterial carotenoid biosynthesis: A colourful tale. *Annu. Rev. Microbiol*. **51:** 629-659

Beebe, J. M. (1941) The morphology and cytology of *Myxococcus xanthus*, N. Sp. *J. Bacteriol*. **42:** 193-223

Berleman, J. E., Chumley, T., Cheung, P., and Kirby, J. R. (2006) Rippling is a predatory behaviour in *Myxococcus xanthus*. *J. Bacteriol*. **188:** 5888-5895

Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem*. **271:** 25059-25062

Beveridge, T. J. (2001) Use of the gram stain in microbiology. *Biotech. Histochem*. **76:** 111-118

Blattner, F. R., Plunkett III, G., Bloch, C. A., Perma, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A.., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*. **277:** 1453-1462

Bode, H. B., Meiser, P., Kiefisch, T., Cortina, N. S., Krug, D., Göhring, A., Schwär, G., Mahmud, T., Einakady, A., and Müller, R. (2007) Mutasynthesis-derived myxalamids and origin of the isobutyryl-CoA starter unit of myxalamid B. *Chem. Biochem*. **8:** 2139-2144

Boomer, S. M., Lodge, D. P., Dutton, B. E., and Pierson, B. (2002) Molecular characterization of novel red green nonsulfur bacteria from five distinct hot spring communities in Yellowstone National Park. *Appl. Environ. Microbiol*. **68:** 346-355

Bost, S., Silva, F., and Belin, D. (1999) Transcriptional activation of *ydeA*, which encodes a member of the major facilitator superfamily, interferes with arabinose accumulation and induction of the *Escherichia coli* arabinose P_{BAD} promoter. *J. Bacteriol.* **181:** 2185-2191

Botella, J. A., Murillo, F. J., and Ruiz-Vazquez, R. (1995) A cluster of structural and regulatory genes for light induced carotenogenesis in *Myxococcus xanthus*. *Eur. J. Biochem*. **233:** 238-248

Brokx, S. J., Rothery, R. A., Zhang, G., Ng, D. P., and Weiner, J. H. (2005) Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the role of a conserved active site cysteine in assembly and function. *Biochemistry*. **44:** 10339-10348

Browning, D. F., Whitworth, D. E., and Hodgson, D. A. (2003) Light-induced carotenogenesis in *Myxococcus xanthus*: Functional characterization of the ECF sigma factor CarQ and anti-sigma factor CarR. *Mol. Microbiol***. 48:** 237-251

Burchard, R. P., and Dworkin, M. (1966a) Light-induced lysis and carotenogenesis in *Myxococcus xanthus*. *J. Bacteriol*. **91**: 535-545

Burchard, R. P., and Dworkin, M. (1966b) Action spectrum for the photolysis of *Myxococcus xanthus*. *J. Bacteriol*. **91:** 896-897

Burchard, R. P., and Hendricks, S. B. (1969) Action spectrum for carotenogenesis in *Myxococcus xanthus*. *J. Bacteriol*. **97:** 1165-1168

Cayuela, M. L., Elías-Arnanz, M., Peñalver-Mellado, M., Padmanabhan, S., and Murillo, F. J. (2003) The *Stigmatella aurantiacum* homolog of *Myxococcus xanthus* high-mobility-group A-type transcription factor CarD: Insights into the functional modules of CarD and their distribution in bacteria. *J. Bacteriol*. **185:** 3527-3537

Cervantes, M., and Murillo, F. J. (2002) Role for vitamin B12 in light induction of gene expression in the bacterium *Myxococcus xanthus*. *J. Bacteriol*. **184:** 2215-2224

Conn, P. F., Schalch, W., and Truscott T. G. (1991). The singlet oxygen and carotenoid interaction. *Photochem. Photobiol. Sci.* **11:** 41-47

Datta, S. K., Datta, K., Parkhi, V., Rai, M., Baisakh, N., Sahoo, G., Rehana, S., Bandyopadhyay, A., Alamgir, M., Ali, M. S., Abrigo, E., Oliva, N., and Torrizo, L. (2007) golden rice: introgression, breeding, and field evaluation. *Euphytica.* **154:** 271-278

Dawid, W. (2000) Biology and global distribution of Myxobacteria in soils. *FEMS Microbiol. Rev*. **24:** 403-427

Di Mascio, P., Devasagayam, T. P., Kaiser, S., and Sies, H. (1990) Carotenoids, tocopherols and thiols as biological singlet molecular oxygen quenchers. *Biochem. Soc. Trans*. **18:** 1054-1056

Dworkin, M. (1996) Recent advances in the social and developmental biology of Myxobacteria. *Microbiol. Rev*. **60:** 70-102

Erkoç, Ş., and Erkoç, F (2002) Structural and electronic properties of porphyrin skeleton of chlorophyll. *J. Mol. Struct. THEOCHEM*. **579:** 41-44

Fernández-González, B., Sandmann, G., and Vioque, A. (1997) a new type of asymmetrically acting beta-carotene ketolase is required for the synthesis of echineone in the cyanobacterium *Synechocystis sp.* PCC 6803. *J. Biol. Chem*. **272:** 9728-9733

Fontes, M., Galbis-Martínez, L., and Murillo, F. J. (2003) A novel regulatory gene for light-induced carotenoid synthesis in the bacterium *Myxococcus xanthus*. *Mol. Microbiol*. **47:** 561-571

Foote, C. S., Chang, Y. C., and Denny, R. W. (1970) Chemistry of singlet oxygen. X. Carotenoid quenching parallels biological protection. *J. Am. Chem. Soc*. **92:** 5216-5218

Fraser, P. D., Miura, Y., and Misawa, N. (1996) In vitro characterization of astaxanthin biosynthetic enzymes. *J. Biol. Chem*. **272**: 6128-6135

Friedrich, A., Hartsch, T., and Averhoff, B. (2001) Natural transformation in mesophilic and thermophilic bacteria: Identification and characterization of novel, closely related competence genes in *Acinetobacter* sp. Strain BD413 and *Thermus thermophilus* HB27. *Appl. Env. Microbiol*. **67:** 3140-3148

Galbis-Martinez, M., Fontes, M., and Murillo, F. J. (2004) The high-mobility group A-type protein CarD of the bacterium *Myxococcus xanthus* as a transcription factor for several distinct vegetative genes. *Genetics*. **167:** 1585-1595

Galbis-Martínez, L., Galbis-Martínez, M., Murillo, F. J., and Fontes, M. (2008) An anti-antisigma factor in the response of the bacterium *Myxococcus xanthus* to blue light. *Microbiol*. **154:** 895-904

Gass, J., Ehren, J., Strohmeier, G., Isaacs, I., and Khosla, C. (2005) Fermentation, purification, formulation, and pharmacological evaluation of a prolyl endopeptidase from *Myxococcus xanthus:* Implications for Celiac sprue therapy. *Biotechnol. Bioeng***. 92:** 674-684

Gerth, K.., IrschiK H., Reichenbach, H., and Höfle, G. (1982) The myxovirescins, a family of antibiotics from *Myxococcus virescens* (*Myxobacterales*). *J. Antibiot*. **35:** 1454-1459

Gerth, K., Jansen, R., Reifenstahl, G., Höfle, G., Irschik, H., Kunze, B., Reichenbach, H., and Thierbach, G. (1983) The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacterales). I. Production, physico-chemical and biological properties, and mechanism of action. *J. Antibiot*. **36:** 1150-1156

Glaeser, J., and Klug, G. (2005) Photo-oxidative stress in *Rhodobacter sphaeroides*: protective role of carotenoids and expression of selected genes. *Microbiol*. **151:** 1927-1938

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). *Science*. **296:** 79-92

Goldman, B. S., Nierman, W. C., Kaiser, D., Slater, S. C., Durkin, A. S., Eisen, J., Ronning, C. M., Barbazuk, W. B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H. S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S. A., Vaudin, M., Wiegand, R., and Kaplan, H. B. (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci. USA*. **103:** 15200-15205

Goldman, B., Bhat, S., and Shimkets, L. J. (2007) Genome evolution and the emergence of fruiting body development in *Myxococcus xanthus*. *PLoS ONE*. **26:** e1329

Goodwin, T. W. (1986) Metabolism, nutrition and function of carotenoids. *Annu. Rev. Nutr*. **6:** 273-297

Gorham, H. C., McGowan, S. J., Robson, P. R. H., and Hodgson, D. A. (1996) Light-induced carotenogenesis in *Myxococcus xanthus*: Light-dependent membrane sequestration of ECF sigma factor CarQ by anti-sigma factor CarR. *Mol. Microbiol*. **19:** 171-186

Guzman, L-M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol*. **177:** 4121-4130

Hannibal, L., Lorquin, J., D'Orteli, N. A., Garcia, N., Chaintreuil, C., Masson-Boivin, C., Dreyfus, B., and Giraud, E. (2000) Isolation and characterisation of canthaxanthin biosynthesis genes from the photosynthetic bacterium *Bradyrhizobium sp*. Strain ORS278. *J. Bacteriol*. **182:** 3850-3853

Harwood, C. S., and Gibson J. (1988) Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol*. **54:** 712-717

Hillesland, K. L., Lenski, R. E., and Velicer, G. J. (2007) Ecological variables affecting predatory success in *Myxococcus xanthus*. *Microb. Eco*. **53:** 571-578

Hodgson, D. A. (1993) Light-induced carotenogenesis in *Myxococcus xanthus*: genetic analysis of the CarR region. *Mol. Microbiol***. 7:** 471-488

Hodgson, D. A., and Berry, A. E. (1998) Light regulation of carotenoid synthesis in *Myxococcus xanthus*. **In:** Microbial responses to light and time. Society for General

Microbiology Symposium. Vol 56 (Caddick, M. X., Baumberg, S., Hodgson, D. A., and Phillips-Jones, M. K. Eds.) Cambridge University Press, Cambridge. pp. 185-211

Holt, J. G., and Lewin R. A. (1968) *Herpetosiphon aurantiacus* gen. *et* sp. n., a new filamentous gliding organism. *J. Bacteriol*. **95:** 2407-2408

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene.* **77:** 61-68

Inouye, S. (1990) Cloning and DNA sequence of the gene coding for the major sigma factor from *Myxococcus xanthus*. *J. Bacteriol*. **172:** 80-85

Jannasch, H. W., and Wirsen, C. O. (1981) Morphological survey of microbial mats near deep-sea thermal vents. *Appl. Environ. Microbiol*. **41:** 528-538

Jelsbak, L., Givskov, M., and Kaiser, D. (2005) Enhancer-binding proteins with a forkhead-associated domain and the sigma(54) regulon in *Myxococcus xanthus* fruiting body development. *Proc. Natl. Acad. Sci. USA.* **102:** 3010-3015

Johnson, E. A. (2003) *Phaffia rhodozyma*: colorful odyssey. *Int. Microbiol* **3:** 169-174

Karwowski, J. P., Sunga, G. N., Kadam, S., and Mcalpine, J. B. (1996) A method for the selective isolation of *Myxococcus* directly from soil. *J. Ind. Micriobiol. Biotechnol*. **16:** 230-236

Kegler, C., Gerth, K., and Müller, R. (2006) establishment of a real-time PCR protocol for expression studies of secondary metabolite biosynthetic gene clusters in the G/C-rich myxobacterium *Sorangium cellulosum* So ce56. *J. Biotech*. **121:** 201-212

Kieser, T. (1984) Factors affecting the isolation of CCC DNA from *Streptomyces lividans* and *Escherichia coli*. *Plasmid*. **12:** 19-36

Kleinig, H., and Reichenbach, H. (1970) A new type of carotenoid pigment isolated from myxobacteria. *Naturwissenschaften*. **57:** 92-93

Kleinig, H., Reichenbach, H., Achenbach, H., and Stadler, J. (1971) Carotenoid pigments of *Sorangium compositum* (Myxobacterales) including two new carotenoid glucoside esters and two new carotenoid rhamnosides. *Arch. Mikrobiol*. **78:** 224-233

Kraft, B. J., Masuda, S., Kikuchi, J., Dragnea, V., Tallin, G., Zaleski, J. M., and Bauer, C. E. (2003) Spectroscopic and mutational analysis of the blue-light photoreceptor AppA: A novel photocycle involving flavin stacking with an aromatic amino acid. *Biochemistry*. **42:** 6726-6734

Krinsky, N. I., and Johnson E. J. (2005) Carotenoid actions and their relation to health and disease. *Mol. Aspects. Med*. **6:** 459-516

Krug, D., Zurek, G., Revermann, O., Vos, M., Velicer, G. J., and Müller, R. (2008) discovering the hidden secondary metabolome of *Myxococcus xanthus*: a study of intraspecific diversity. *Appl. Environ. Microbiol*. **74:** 3058-3068

Kuner, J. M.., and Kaiser, D. (1981) Introduction of transposon Tn5 into *Myxococcus* for analysis of developmental and other nonselective mutants. *Proc. Natl. Acad. Sci. USA.* **78:** 425-429

Kunze, B., Bedorf, N., Kohl, W., Hofle, G., and Reichenbach, H. (1989) Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (*Myxobacterales*). *J. Antibiot*. **42:** 14-17

Kuspa, A., Kroos, L., and Kaiser, D. (1986) Intercellular signalling is required for developmental gene expression in *Myxococcus xanthus*. *Dev. Biol*. **117:** 267-276

Laue, B., and Gill, R. E. (1995) Using a phase-locked mutant of *Myxococcus xanthus* to study the role of phase variation in development. *J. Bacteriol*. **177:** 4089-4096

Lee, F. Y., Borzilleri, R., Fairchild, C. R., Kamath, A., Smykla, R., Kramer, R., and Vite, G. (2008) Preclinical discovery of ixabepilone, a highly active antineoplastic agent. *Cancer Chemother. Pharmacol*. **63:** 157-166

Lee, J. H., and Kim, Y. T. (2006) Cloning and characterization of the astaxanthin biosynthesis gene cluster from the marine bacterium *Paracoccus haeundaensis*. *Gene*. **370:** 86-95

López-Rubio, J. J., Elías-Arnanz, M., Padmanabhan, S., and Murillo, F. J. (2002) A repressor-antirepressor pair links two loci controlling light-induced carotenogenesis in *Myxococcus xanthus*. *J. Biol. Chem*. **277:** 7262-7270

López-Rubio, J. J., Padmanabhan, S., Lazaro, J. M., Salas, M., Murillo, F. J., and Elías-Arnanz, M. (2004) Operator design and mechanism for CarA repressor-mediated down-regulation of the photoinducible carB operon in *Myxococcus xanthus*. *J. Biol. Chem*. **279:** 28945-28953

Loschi, L., Brokx, S. J., Hills, T. L., Zhang, G., Bertero, M. G., Lovering, A. L., Weiner, J. H., and Strynadka, N. C. J. (2004) Structural and biochemical identification of a novel bacterial oxidoreductase. *J. Biol. Chem*. **48:** 50391-50400

Lozano, G. A (2001) Carotenoids, immunity, and sexual selection: comparing apples and oranges? *Am. Nat*. **158:** 200-203

Maan, M. E., Van Rooijen, A. M. C., Van Alphen, J. J. M, and Seehausen, O. (2008) Parasite-mediated sexual selection and species divergence in Lake Victoria cichlid fish. *Biol. J. Linn. Soc*. **94:** 53-60

Magrini, V., Creighton, C., White, D., Hartzell, P. L., and Youderian, P. (1998) The *aad*A gene of plasmid R100 confers resistance to spectinomycin and streptomycin in *Myxococcus xanthus*. *J. Bacteriol*. **180:** 6757-6760

McGowan, S. J., Gorham, H. C., and Hodgson, D. A. (1993) Light-induced carotenogenesis in *Myxococcus xanthus*: DNA sequence analysis of the CarR region. *Mol. Microbiol*. **10:** 713-735

Mignot, T., Shaevitz, J. W., Hartzell, P., and Zusman, D. R. (2007) Evidence that focal adhesion powers bacterial gliding motility. *Science*. **315:** 853-856

Mignot, T. (2007) The elusive engine in *Myxococcus xanthus* gliding motility. *Cell. Mol. Life Sci*. **64:** 2733-2745

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E. S., Knight, J. R., Irzyk, G. P., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Pringle, T., Schuster, S. C. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*. **456:** 387-390

Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., and Harastima, K. (1990) Elucidation of *the Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *E. coli*. *J. Bacteriol*. **172:** 6704-6712

Misawa, N., Satomi, Y., Kondo, K., Yokoyama, A., Kajiwara, S., Saito, T., Ohtani, T., and Miki, W. (1995) Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol*. **177:** 6575-6584

Mochimaru, M., Masukawa, H., and Takaichi, S. (2005) The cyanobacterium *Anabaena* sp. PCC 7120 has two distinct β-carotne ketolases: CrtO for echinenone and CrtW for ketomyxol synthesis. *FEBS*. **579:** 611-6114

Moraleda-Munoz, A., Perez, J., Fontes, M., Murillo, F.J., and Munoz-Dorado, J. (2005) Copper induction of carotenoid synthesis in the bacterium *Myxococcus xanthus*. *Mol. Microbiol*. **56:** 1159-1168

Moreno, A. J., Fontes, M., and Murillo, F. J. (2001) *ihfA* Gene of the bacterium *Myxococcus xanthus* and its role in activation of carotenoid genes by blue light. *J. Bacteriol*. **183:** 557-569

Morrison, T., Weis, J. J., and Wittwer, C. T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques*. **24:** 954-962

Mortensen, A.., Skibsted, L. H., Sampson, J., Rice-Evans, C., and Everett, S. A. (1997) Comparative mechanisms and rates of free radical scavenging by carotenoid antioxidants. *FEBS Letters*. **418:** 91-97

Navarro-Avilés, G., Jiménez, M. A., Pérez-Marin, M. C., González, C., Rico, M., Murillo, F. J., Elías-Arnanz, M., and Padmanabhan S. (2007) Structural basis for operator and antirepressor recognition by *Myxococcus xanthus* CarA repressor. *Mol Microbiol*. **63:** 980-994

Nicolas, F. J., Ruiz-Vazquez, R. M., and Murillo, F. J. (1994) A genetic link between light response and multicellular development in the bacterium *Myxococcus xanthus*. *Genes Dev*. **8:** 2375-2387

Nicolas, F. J., Cayuela, M. L., Martínez-Argudo, I., Ruiz-Vazquez, R. M., Murillo, F. J. (1996) High mobility group I(Y)-like DNA-binding domains on a bacterial transcription factor. *Proc. Natl. Acad. Sci. USA*. **93:** 6881-6885

Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., Komemushi, S., Miki, W., and Misawa, N. (2005) Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'-beta-hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanyhophylls. *Appl. Environ. Microbiol*. **71:** 4286-4296

Ossa, F., Diodati, M. E., Caberoy, N. B., Giglio, K. M., Edwards, M., Singer, M., and Garza, A. G. (2007) The *Myxococcus xanthus* N1A4 protein is important for expression of stringent response-associated genes, ppGpp accumulation, and fruiting body development. *J. Bacteriol*. **189:** 8474-8483

Peñalver-Mellado, M., García-Heras, F., Padmanabhan, S., García-Moreno, D., Murillo, F. J., Elías-Arnanz, M. (2006) Recruitment of a novel zinc-bound transcriptional factor by a bacterial HMGA-type protein is required for regulating multiple processes in *Myxococcus xanthus. Mol. Microbiol*. **61:** 910-926

Pérez-Marin, M. C., Lopez-Rubio, J. J., Murillo, F. J., Elias-Arnanz, M., and Padmanabhan, S. (2004) The N-terminus of *M. xanthus* CarA repressor is an autonomously folding domain that mediates physical and functional interactions with both operator DNA and antirepressor protein. *J. Biol. Chem***. 279:** 33093-33103

Pérez-Marín, M. C., Padmanabhan, S., Polanco, M. C., Murillo, F. J., Elías-Arnanz, M. (2008) Vitamin B12 partners the CarH repressor to downregulate a photoinducible promoter in *Myxococcus xanthus. Mol. Microbiol*. **67**: 804-819

Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acid Res*. **29:** e45

Pham, V. D., Shebelut, C. W., Jose, I. R., Hodgson, D. A., Whitworth, D. E., and Singer, M. (2006) The response regulator PhoP4 is required for late developmental event in *Myxococcus xanthus*. *Microbiol*. **152:** 1609-1620

Pinel, N., Davidson, S. K., and Stahl, D. A. (2008) *Verminephrobacter eiseniae* gen. nov., sp. nov., a nephridal symbiont of the earthworm *Eisenia foetida* (Savigny). *Int. J. Syst. Evol. Microbiol*. **58:** 2147-2157

Rajagopalan, K. V., and Johnson, J. L. (1992) The pterin molybdenum cofactors. *J. Biol. Chem*. **267:** 10199-10202

Reichenbach, H., and Kleinig, H. (1971) the carotenoids of *Myxococcus fulvus* (Myxobacterales). *Arch. Mikrobiol*. **76:** 364-380

Reichenbach, H. (1999) the ecology of Myxobacteria. *Environ. Microbiol*. **1:** 15-21

Reichenbach, H. (2001) Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol***. 27:** 149-156

Rice, S. A., and Lampson, B. C. (1995) Phylogenetic of retron elements among the myxobacteria: evidence for vertical inheritance. *J. Bacteriol*. **177:** 37-45

Ruiz-Vazquez, R., Fontes, M., and Murillo, F. J. (1993) Clustering and co-ordinated activation of carotenoid genes in *Myxococcus xanthus* by blue light. *Mol. Microbiol*. **10:** 25-34

Sandre, S. L., Tammaru, T., Esperk, T., Julkunen-Titto, R., Mappes, J. (2007) Carotenoid-based colour polyphenism in a moth species: search for fitness correlates. *Entomol. Exp. Appl*. **124:** 269-277

Scanlan, D. J., Bloye, S. A., Mann, N. H., Hodgson, D. A., and Carr, N. G. (1990) Construction of *lacZ* promoter probe vectors for use in *Synechococcus*: application to the identification of CO_2-regulated promoters. *Gene*. **90:** 43-49

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H. P., and Zillig, W. (1995) *Picrophilus* gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J. Bacteriol*. **177:** 7050-7059

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., Bartels, D., Bekel, T., Beyer, S., Bode, E., Bode, H. B., Bolten, C. J., Choudhuri, J. V., Doss, S., Elnakady, Y. A., Frank, B., Gaigalat, L., Goesmann, A., Groeger, C., Gross, F., Jelsbak, L., Jelsbak, L., Kalinowski, J., Kegler, C., Knauber, T., Konietzny, S., Kopp, M., Krause, L., Krug, D., Linke, B., Mahmud, T., Martinez-Arias, R., McHardy, A. C., Merai, M., Meyer, F., Mormann, S., Muñoz-Dorado, J., Perez, J., Pradella, S., Rachid, S., Raddatz, G., Rosenau, F., Rückert, C., Sasse, F., Scharfe, M., Schuster, S. C., Suen, G., Treuner-Lange, A., Velicer, G. J., Vorhölter, F. J., Weissman, K. J., Welch, R. D., Wenzel, S. C., Whitworth, D. E., Wilhelm, S., Wittmann, C., Blöcker,

H., Pühler, A., Müller, R. (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol*. **11:** 1281-1289

Soltis, P. S., and Soltis, D. E. (2003) Applying the bootstrap in phylogeny reconstruction. *Statistical Science*. **18:** 256-267

Stanley, N. R., Palmer, T., and Berks, B. C. (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. *J. Biol. Chem*. **275:** 11591-11596

Starruß, J., Bley, T., Søgaard-Andersen, L., and Deutsch, A. (2007) A new mechanism for collective migration in *Myxococcus xanthus*. *J. Stat. Phys*. **128:** 269-286

Steiger, S., Jackisch, Y., and Sandmann, G. (2005) Carotenoid biosynthesis in *Gloeobacter violaceus* PCC4721 involves a single *crtI*-type phytoene desaturase instead of typical cyanobacterial enzymes. *Arch. Microbiol*. **184:** 207-214

Studier, F. W., and Moffatt B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol*. **5:** 113-130

Sudo, S. Z., and Dworkin, M. (1969) Resistance of vegetative cells and microcysts of *Myxococcus xanthus*. *J. Bacteriol*. **98:** 883-887

Takaichi, S., and Mochimaru, M. (2007) Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. *Cell. Mol. Life Sci*. **64:** 2607-2619

Takano, H., Obitsu, S., Beppu, T., and Ueda, K. (2005) Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol*. **187:** 1825-1832

Tanvetyanon, T., and Bepler, G. (2008) Beta-carotene in multivitamins and the possible risk of lung cancer among smokers versus former smokers. *Cancer*. **113:** 150-157

Tao, L., and Cheng, O. (2004) Novel beta-carotene ketolases from non-photosynthetic bacteria for canthaxanthin synthesis. *Mol. Genet. Genomics*. **272:** 530-537

Tao, L., Rouvière, P. E., and Cheng, O. (2006) A carotenoid synthesis gene cluster from a non-marine *Brevundimonas* that synthesizes hydroxylated astaxanthin. *Gene*. **379:** 101-108

Truscott, T. G. (1990) The photophysics and photochemistry of the carotenoids. *Photochem. Photobiol. B. Biol*. **6:** 359-371

Tsuchiya, T., Takaichi, S., Misawa, N., Maoka, T., Miyashita, H., and Mimuro, M. (2005) The cyanobacterium *Gloeobacter violaceus* PCC 7421 uses bacterial-type phytoene desaturase in carotenoid biosynthesis. *FEBS*. **579:** 2125-2129

Tzeng, L., and Singer, M. (2005) DNA replication during sporulation in *Myxococcus xanthus* fruiting bodies. *Proc. Natl. Acad. Sci. USA*. **102:** 14428-14433

Umeno, D., Tobias, A. V., and Arnold, F. H. (2005) Diversifying carotenoid biosynthetic pathways by directed evolution. *Micobiol. Mol. Biol. Rev*. **69:** 51-78

Valko, M., Morris, H., and Cronin, M. T. D. (2005) Metals, toxicity and oxidative stress. *Curr. Med. Chem*. **12:** 1161-1208

Valko, M., Rhodes, C.J., Monool, J., Izakovic, M., and Mazur, M. (2006) Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem. Biol. Interact*. **160:** 1-40

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor-Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin,

D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. (2001) The sequence of the human genome. *Science*. **291:** 1304-1351

Vos, M., and Velicer, G. J. (2006) Genetic population structure of the soil bacterium *Myxococcus xanthus* at the centimetre scale. *Appl. Environ. Microbiol*. **72:** 3615-3625

Watanabe, S., Kobayashi, T., Saito, M., Sato, M., Nimura-Matsune, K., Chibazakura, T., Taketani, S., Nakamoto, H., and Yoshikawa, H. (2006) Studies on the role of HtpG in the tetrapyrrole biosynthesis pathway of the cyanobacterium *Synechococcus elongatus* PCC 7492. *Biochem. Biophys. Res. Commun*. **352:** 36-41

Whitworth, D. E., and Hodgson, D. A. (2001) Light-induced carotenogenesis in *Myxococcus xanthus*: Evidence that CarS acts as an anti-repressor of CarA. *Mol. Microbiol***. 42:** 809-819

Whitworth, D. E., Bryan, S. J., Berry, A. E., McGowan, S. J., and Hodgson, D. A. (2004) Genetic dissection of the light-inducible *carQRS* promoter region of *Myxococcus xanthus*. *J. Bacteriol*. **186:** 7836-7846

Wilding, E. I, Brown, J.R., Bryant, A.P., Chalker, A.F., Holmes, D.J., Ingraham, K.A., Iordanescu, S., so, C.Y., Rosenberg, M., and Gwynn, M.N. (2000) Identification, evolution, and essentiality of the mevalonate pathway for isopentyl diphosphate biosynthesis in gram-positive cocci. *J. Bacteriol*. **182:** 4319-4327

Williamson, D (1993) The unstable haemoglobins. *Blood Rev*. **7:** 146-163

Wireman, J. W., and Dworkin, M. (1975) Morphogenesis and developmental interactions in Myxobacteria. *Sceince*. **189:** 516-523

Wolgemuth, C., Hoiczyk, E., Kaiser, D., and Oster, G. (2002) How Myxobacteria glide. *Curr. Biol*. **12:** 369-377

Woodall, A. A., Lee, S. W., Weesie, R. J., Jackson, M. J., and Britton, G. (1997) Oxidation of carotenoids by free radicals: relationship between structure and reactivity. *Biochim. Biophys. Acta*. **1336:** 33-42.

Wu, Z-H., Jiang, D-M., Li, P., and Li, Y-Z. (2005) Exploring the diversity of Myxobacteria in a soil niche by myxobacteria-specific primers and probes. *Environ. Microbiol*. **7:** 1602-1610

Ye, R. W., Stead, K. J., Yao, H., and He, H. (2006) Mutational and functional analysis of the beta-carotene ketolase involved in the production of canthaxanthin and astaxanthin. *Appl. Environ. Microbiol*. **72:** 5829-5837

Zafriri, D., Rosenberg, E., and Mirelman, D. (1981) Mode of action of *Myxococcus xanthus* antibiotic TA. *Antimicrob. Agents. Chemother*. **19:** 349-351

Zhao, J-Y., Zhang, L., Shen, M-J., Xia, Z-J., Cheng, Q-X., Sun, X., Zhao, G-P., Li, Y-Z., and Qin, Z-J. (2008) Discovery of the autonomously replicating plamid pMF1 from *Myxococcus fulvus* and development of a gene cloning system in *Myxococcus xanthus. Appl. Environ. Microbiol*. **74:** 1980-1987

Books

Throughout the investigation one book was constantly used for reference:

Molecular cloning – A Laboratory Manual, 3rd edition, Sambrook and Russell, Spring Harbor Laboratory Press (2001).

Computer programs

(The majority of the specialised computer software utilised during this investigation is detailed in the main body of text)

When constructing plasmid maps the program BVTech Plasmid was used, version 2.0. This is a BV Tech Inc program.

When predicting transmembrane regions from protein sequences the program TMpred was used from a Swiss website (ch.EMBnet.org)

The ACD/Chemsketch program (V 1.0) was utilised for drawing carotenoid substrates and precursors throughout the thesis.