A thesis submitted for the degree of Doctor of Philosophy

# Functional and mathematical analysis of the glyoxylate shunt in *Streptomyces coelicolor*

Richard A. Reumerman

November 2015

Strathclyde Institute of Pharmacy and Biomedical Sciences,
University of Strathclyde, Glasgow

# Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.



Signed:

Date:

*Tempus fugit*

# Acknowledgements

Although writing this PhD thesis is a great personal achievement, nothing in it would have been possible if not for the continual support, help and advice of the people around me. It is fitting, then, to dedicate some time to acknowledge those who have contributed to this work in any way.

**T**hroughout my studies, my supervisor Dr. Paul Herron allowed me to pursue my interests, supported me when results remained elusive and never stopped believing in me. His suggestions and discussions helped shape my results into a coherent thesis.

**H**elpful support and advice were also provided by Dr. Paul Hoskisson and Dr. Nick Tucker in areas where my own supervisor was less knowledgeable, which is essential to multidisciplinary research. Many thanks also to the people of the microbiology group and the technical support staff at SIPBS. Jana Hiltner deserves a special mention here, as our collaboration was both pleasant and fruitful.

**A**mple funding was provided by SULSA (Scottish Universities Life Science Alliance), which not only paid for research and living in Glasgow but also for attending many interesting conferences which helped spark my enthusiasm for science. I would also like to thank the SGM (Society for General Microbiology) for enabling so many students to attend their excellent biannual meetings.

**N**obody provided me with more support, love and care than my girlfriend, Janna. We've been through ups and downs together and I'm glad we set out on this adventure. I would not have been where I am now without you.

**K**eeping everyone at home posted about the newest developments is not always easy, and my parents bore the brunt of this more than once. I would like to thank them from the bottom of my heart for all the support and patience they I received from them. The same goes for my friends and other family in the Netherlands, who wouldn't have expected me to move abroad even shortly before it suddenly happened.

**S**ince moving to Glasgow, I met many great and interesting people. They and many others contributed to my work by providing the necessary distractions. I would like to mention each of them individually but they are too many to list here. Therefore, to anyone I didn't mention here, thank you.

# Abstract

*Streptomyces coelicolor* is the model organism for the genus *Streptomyces*, which produces many bioactive secondary metabolites with clinical applications. Based on work done in *Escherichia coli*, the glyoxylate shunt was thought to be the main anapleurotic pathway in *S. coelicolor* during growth on fatty acids and therefore an important pathway in providing precursors for secondary metabolism. The *S. coelicolor* genome contains genes for a second anapleurotic pathway, the ethylmalonyl-CoA pathway. The relative importance of both to anapleurosis in streptomycete metabolism was unclear. The function of the glyoxylate shunt was investigated in this thesis using sequence analysis, genetic manipulation, transcriptomics and mathematical modelling.

Analysis of orthologues of *aceA*, *ccr* and genes encoding tricarboxylic acid (TCA) cycle genes revealed that all are subject to a similar level of purifying selection pressure. The operons of the glyoxylate shunt and the ethylmalonyl-CoA pathway share a 15 bp palindromic motif in their upstream sequences, which was also found upstream of other genes. This suggests an overlap in regulation and thus an overlap in function.

The sequence analysis is contradicted by results of experiments with an $aceA^{-}\,aceB1^{-}$ mutant, which did not display a phenotype during growth on Tween 40, a model carbon source for fatty acids. Results obtained by total RNA sequencing indicate that the ethylmalonyl-CoA pathway is the main anapleurotic pathway during growth of *S. coelicolor* on fatty acids whereas expression of the glyoxylate shunt is minimal. This apparent contradiction is resolved by hypothesising that the ethylmalonyl-CoA pathway is the main anapleurotic pathway, but that the glyoxylate shunt provides a backup when acyl-CoA thioesters are withdrawn from the ethylmalonyl-CoA pathway for secondary metabolite biosynthesis.

Enzymes of the isocitrate branchpoint were isolated following heterologous expression and analysed. The resulting kinetic parameters, as well as their specific activities measured during growth on Tween 40 and additional data from literature, were used to set up a mathematical model of the TCA cycle and the glyoxylate shunt. Simulations of this model predicted that, as growth proceeds from early to mid and late exponential phase, the relative concentrations of TCA cycle intermediates changed from promoting gluconeogenesis to accomodating secondary metabolism. Further model refinement is needed using data on the flux through the ethylmalonyl-CoA pathway as these were unavailable at the time of writing.

# Contents

# Acronyms and Abbreviations

**ACP** acyl carrier protein

**AMP** adenosine monophosphate

**AT** acyltransferase

**ATP** adenosine triphosphate

**CCR** crotonyl-CoA carboxylase / reductase

**CoA** coenzyme A

**CRP** cyclic AMP receptor protein

**CS** citrate synthase

**DCW** dry cell weight

**DH** dehydratase

**DMSO** dimethyl sulfoxate

**EGFP** enhanced green fluorescent protein

**ER** enoyl reductase

**FUM** fumarase

**HTH** helix-turn-helix

**IAA** isoamyl alcohol

**ICL** isocitrate lyase

**IDH** isocitrate dehydrogenase

**IKP** isocitrate dehydrogenase kinase/phosphatase

**ISP** ion sphere particle

**KR** $\beta$-ketoacyl-ACP reductase

**KS** $\beta$-ketoacyl-ACP synthase

**LDH** lactate dehydrogenase

**MCA** metabolic control analysis

**MDH** malate dehydrogenase

**MFA** metabolic flux analysis

**MNP** multiple nucleotide polymorphism

**MS** malate synthase

**NRP** nonribosomal peptide

**NRPS** nonribosomal peptide synthase

**OAA** oxaloacetate

**PAGE** polyacrylamide gel electrophoresis

**PCP** peptidyl carrier protein

**PDH** pyruvate dehydrogenase

**PEP** phosphoenolpyruvate

**PEPC** phosphoenolpyruvate carboxylase

**PKS** polyketide synthase

**PPDK** pyruvate phosphate dikinase

**PPP** pentose phosphate pathway

**RIN** RNA integrity number

**SCS** succinyl-CoA synthetase

**SDH** succinate dehydrogenase

**SDS** sodium dodecyl sulphate

**SNP** single nucleotide polymorphism

**TCA** tricarboxylic acid

**TE** thioesterase

# 1. Introduction

*Streptomyces* is a genus of Gram positive soil-dwelling bacteria that belongs to the family of sporulating Actinobacteria. On solid media, a germinating spore forms a loose network of branched and interconnected hyphae, called the vegetative mycelium (Claessen *et al.*, 2014). This mycelium grows into the surface of the medium, such as an agar plate or the organic remains of other organisms (Fig. 1.1, steps 1–3). Streptomycetes are capable of using these often insoluble remains as a carbon source because of their capability to break them down with hydrolytic exoenzymes (Chater *et al.* (2010), Fig. 1.1). This enables them to grow on such nutrients as chitin , which is one of the most abundant biopolymers found in insects and plants, respectively (Chater *et al.*, 2010).

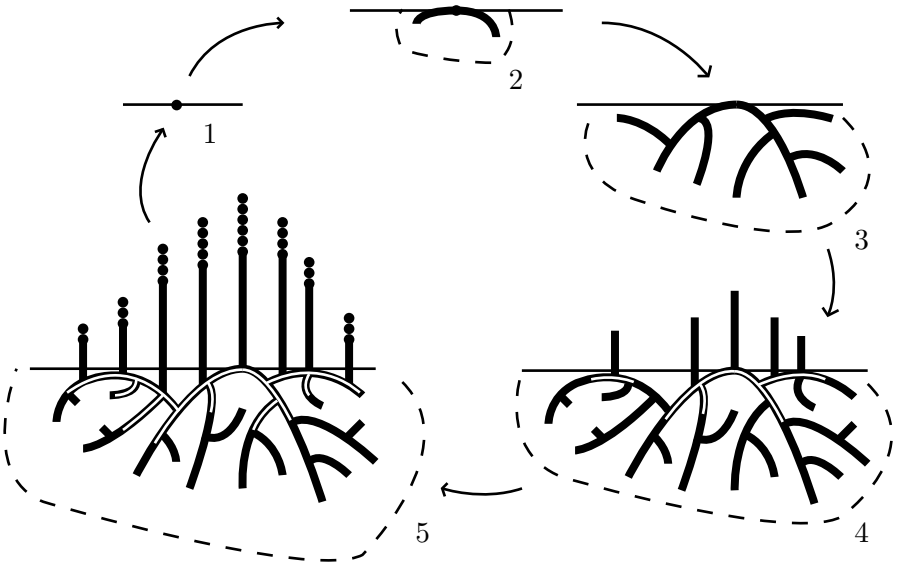



Figure 1.1.: Schematic representation of the *Streptomyces* life cycle. A single spore lands on a suitable surface (1) and germinates (2). The branching vegetative hyphae grow into the surface (3), secreting hydrolytic enzymes to break down complex substrates (dashed line). When nutrients become scarce, the aerial mycelium is formed from degrading vegetative mycelium (4) which eventually forms spore chains (5, Claessen *et al.* (2014)).

Over time, the nutrients start to deplete. *Streptomyces* is an immobile organism

and therefore unable to move to a more favourable position. Instead, it enters the developmental phase which starts with the degradation of the vegetative mycelium. The material derived from the degradation of the vegetative mycelium is used to form the aerial mycelium, which consists of hydrophobic hyphae growing through the moist surface into the air (Fig. 1.1, step 4). These hyphae eventually turn into chains of mono-nucleoid spores which, after maturation, are distributed over the wet environment because of their hydrophobicity and can form new colonies (Claessen *et al.* (2014), Fig. 1.1, steps 5 & 1).

During their development phase, streptomycetes produce many chemical compounds that are excreted into the environment. Many of these compounds can be used as antibiotics, fungicides, cytostatic compounds, modulators of the immune response and effectors of plant growth (Bérdy, 2005; Chater *et al.*, 2010; Watve *et al.*, 2001). The biosynthesis of these compounds is called secondary metabolism and it is this characteristic that makes the genus important for industrial applications.

## 1.1. Secondary metabolite biosynthesis

Three main types of secondary metabolite biosynthesis mechanisms can be distinguished. The first type of metabolites to be discussed here, polyketides, consist of a string of acyl-CoA derived building blocks that are connected by successive decarboxylative condensations. The many ketone bonds formed by these condensations are what gives polyketides their name (Hopwood & Sherman, 1990).

### 1.1.1. Type I polyketide synthases

The enzymes or enzyme complexes that produce polyketides' backbones, polyketide synthases (PKSs), can be subdivided into two main types. Type I PKSs are relatively large protein complexes consisting of distinct modules. Starting from the starting module, each module adds one extender unit to the polyketide and often carries out one or more reductions before passing the growing chain to the next module. Examples of Type I PKSs are those producing the antibiotic erythromycin (Cortes *et al.*, 1990), the immunosuppressant rapamycin (Schwecke *et al.*, 1995) and the antihelminthic avermectin (Ikeda *et al.*, 1999).

The type of extender unit added by a module is determined by the acyltransferase (AT) domain of the module, which usually selects for a malonyl-CoA, methylmalonyl-CoA or an ethylmalonyl-CoA substrate. The extender unit is condensed to the growing polyketide chain by the $\beta$-ketoacyl-ACP synthase (KS) domain, which also removes

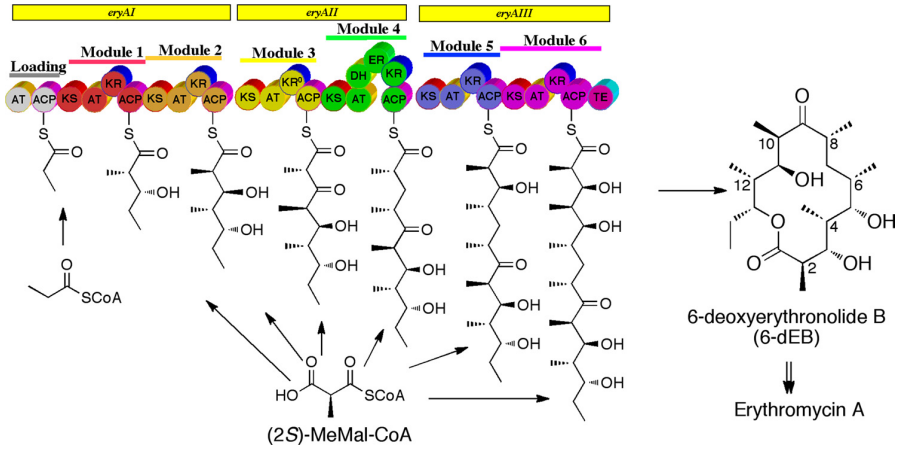



Figure 1.2.: Schematic representation of the PKS that produces the erythromycin backbone. The polyketide backbone grows by successive carboxylative condensation of methylmalonyl-CoA extender units, one at each module. This image was reproduced from Cane (2010).

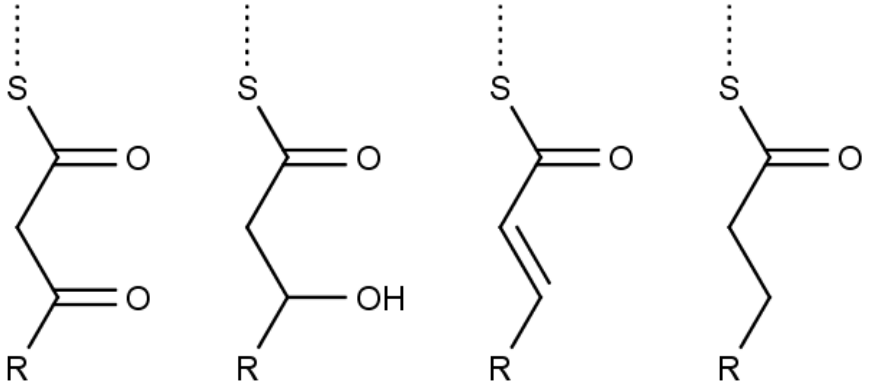
$CO_2$. At this point, the polyketide chain is attached to the acyl carrier protein (ACP) domain, which presents it to the reductive domains (if present) and consequently to the KS domain of the next module.

Three levels of reductions can be carried out on a polyketide's keto groups (Fig. 1.3), but these always occur on the extender unit attached by the previous module. A $\beta$-ketoacyl-ACP reductase (KR) domain reduces the $\beta$-keto group to a $\beta$-hydroxy group (Fig. 1.3b). A dehydratase (DH) domain can remove the hydroxy group, forming $H_2O$ and a double bond (Fig. 1.3c). An enoyl reductase (ER) domain can then reduce this double bond to a single bond (Fig. 1.3d). PKS modules carry out varying levels of reduction and attach different extender units to the polyketide, allowing a huge variety in structures for these compounds.

The last domain of a Type I PKS is a thioesterase (TE) domain, which hydrolyses the polyketide chain from the last ACP domain. Following hydrolysis, the polyketide is further derivatised by non-PKS enzymes to produce the final product.

### 1.1.2. Type II polyketide synthases

Type II PKSs are much smaller enzyme (non-modular) complexes, generally consisting of only three domains. Polyketide backbones are formed using decarboxylative condensations as with Type I PKSs, but these are all carried out by the same KS domain. The length of the polyketide backbone in these systems is determined by a chain length factor




(a) No reduction (b) Reduction by a KR domain (c) Reduction by a DH domain (d) Reduction by an ER domain

Figure 1.3.: Possible reductions carried out on the polyketide backbone by PKS modules. The reactions occur in the order shown and each reaction depends on completion of the previous reaction. Reductive reactions are always carried out on the extender unit added by the previous module.

domain rather than by the number of modules, as is the case with Type I polyketides.

### 1.1.3. Non-ribosomal peptide synthases

The third main class of secondary metabolite biosynthesis complexes consist of the nonribosomal peptide synthases (NRPSs). They display a modular structure like the Type I PKSs, but produce a peptide instead of a polyketide (Schwarzer *et al.*, 2003).

In NRPSs, extender amino acids are recognised by the adenylation-(A)-domain). The A-domain also activates the substrate by condensing it with adenosine monophosphate (AMP), which is derived from adenosine triphosphate (ATP). Following activation, the amino acid is transferred to the peptidyl carrier protein (PCP) domain, which binds it by means of a thioester bond. The PCP domain serves the same function as the ACP domain in a Type I PKS, transporting the substrate tethered to it between active sites of catalytic domains of the NRPS, such as the condensation-(C)-domain. A C-domain forms the peptide bond in nonribosomal peptides (NRPs). Other domains are found in NRPSs that each catalyse a different reaction (Schwarzer *et al.*, 2003).

The many different amino acids available for incorporation, both proteinogenic and non-proteinogenic (Marahiel *et al.*, 1997), combined with additional derivatisation after the polypeptide has been synthesised, allow for a huge variety in NRP size and structure

and therefore in bioactivity. Hybrid PKS-NRPS clusters, such as the one responsible for rapamycin biosynthesis (Schwecke *et al.*, 1995), incorporate both acyl-CoA thioesters and amino acids into their products.

### 1.1.4. $\beta$-lactam biosynthesis

$\beta$-lactams are a class of secondary metabolites that share a core structure. This core consists of a non-ribosomally synthesised tripeptide, (L-$\alpha$-aminoadipoyl)-L-cysteinyl-D-valine. In bacteria, L-$\alpha$-Aminoadipic acid is derived from lysine through a dedicated pathway. The cysteine and valine moieties are cyclised to form isopenicillin N after the tripeptide has been formed. Further derivatisations then lead to different end products (Hamed *et al.*, 2013).

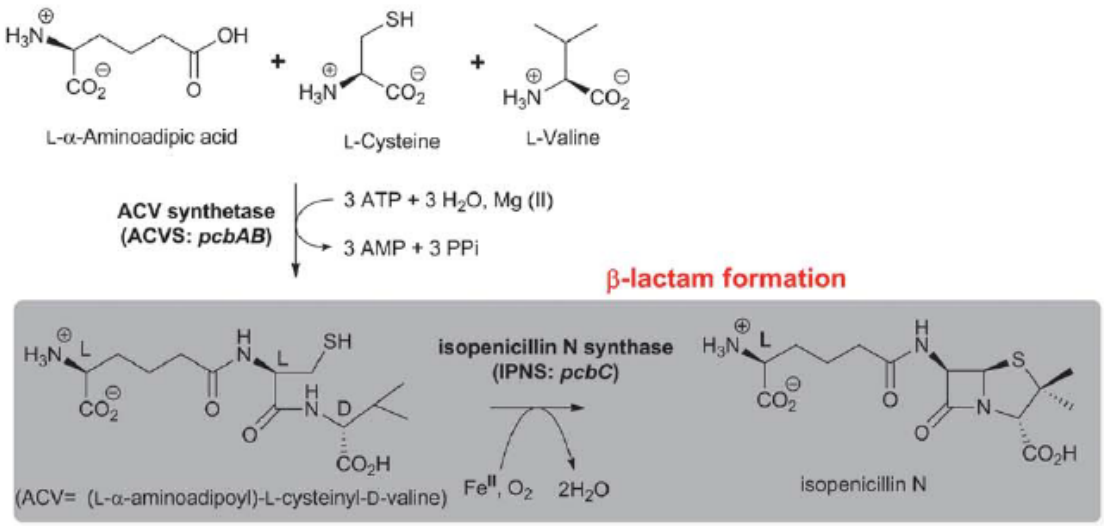



Figure 1.4.: Biosynthesis pathway leading to $\beta$-lactam production. Image adopted from Hamed *et al.* (2013).

$\beta$-lactams known so far are either antibiotics or $\beta$-lactamase inhibitors (Hamed *et al.*, 2013). This low variety in bioactivity is possibly due to the structural similarities between various compounds, all being derived from the same three amino acids.

## 1.2. Anapleurosis in *S. coelicolor*

Fatty acids are among the types of carbon sources that soil bacteria such as *S. coelicolor* are likely to encounter. Fatty acids make up a major part of the lipids of cell membranes and are also used to store energy by animals and plants. Fatty acids are often produced

as 16-carbon palmitic acid, which is the most common fatty acid in living organisms (Gunstone *et al.*, 2007). They are also often used as a carbon source in industrial media (Banchio & Gramajo, 1997). This makes studying their utilisation relevant to both fundamental and applied research.

Uptake of fatty acids is followed by their activation (attachment of a coenzyme A group) and stepwise degradation through $\beta$-oxidation. Each cycle of the $\beta$-oxidation pathway in prokaryotes yields one molecule of acetyl-CoA as well as energy by reduction of FAD and $NAD^+$ to $FADH_2$ and NADH, respectively (Fujita *et al.*, 2007). The acetyl-CoA molecules enter the tricarboxylic acid (TCA) cycle or provide carbon for biosynthesis reactions.

Since the environment does not always provide sufficient amino acids, nucleotides and carbohydrates for the biosynthesis of new biomass, bacteria possess metabolic pathways to synthesise these from available compounds. The precursors for these pathways are intermediates of primary carbon metabolic pathways: glycolysis, the pentose phosphate pathway (PPP) and the TCA cycle.

During growth on fermentable sugars, such as glucose or other carbohydrates, the flux of carbon through glycolysis and the PPP follows directly from carbon uptake and can therefore always occur (Wang *et al.*, 1958). The TCA cycle, on the other hand, relies on a pool of intermediates for its function; each acetyl moiety that enters the cycle needs the presence of a molecule of oxaloacetate to accept it. Oxaloacetate and 2-oxoglutarate are both precursors for aminoacid biosynthesis, while oxaloacetate, 2-oxoglutarate and succinate are precursors for various secondary metabolites in streptomycetes (Olano *et al.*, 2008). Succinate is also a precursor for heme biosynthesis. Withdrawal of TCA cycle intermediates for biosynthesis therefore occurs during all stages of the *Streptomyces* life cycle, but these intermediates must be replenished to prevent them from being depleted. The reactions that perform this function are called anapleurotic pathways.

An important anapleurotic reaction during growth on fermentable sugars is the carboxylation of phosphoenolpyruvate (PEP) to form oxaloacetate by the enzyme PEP carboxylase (Vorísek *et al.*, 1969). This reaction replenishes TCA cycle intermediates by producing them directly from PEP, which is an intermediate of glycolysis. However during growth on fatty acids, glycolysis operates in reverse direction (gluconeogenesis) in order to produce glycerol for triacylglycerol lipids (Olukoshi & Packter, 1994) and glucose for the synthesis of cell wall components and as a substrate for the PPP, which provides the cell with NADPH and pentose sugars for nucleotide biosynthesis. Gluconeogenesis withdraws oxaloacetate from the pool of TCA cycle intermediates as a precursor. No net production of oxaloacetate from pyruvate can take place in these conditions and

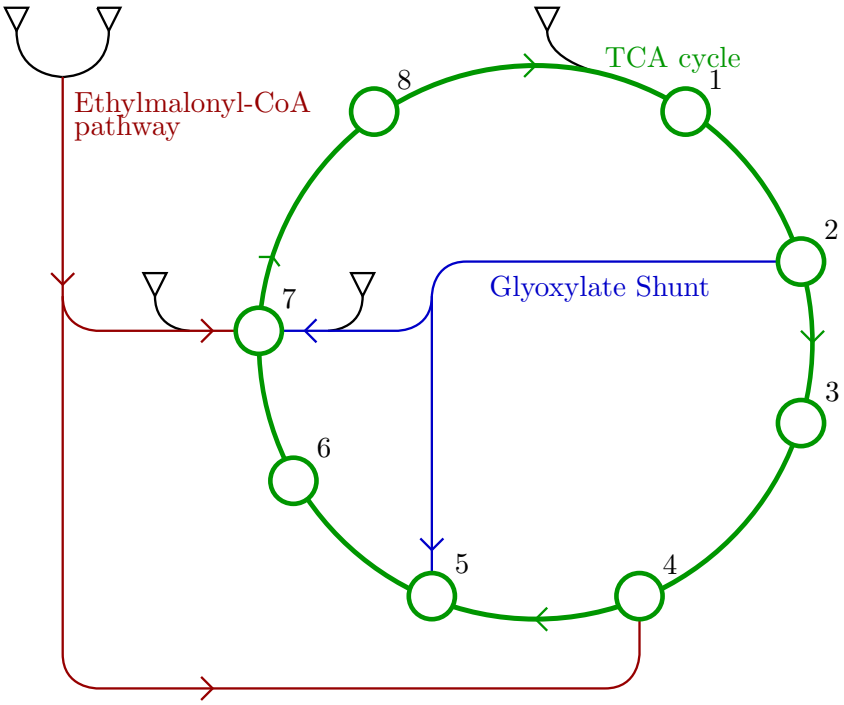



Figure 1.5.: Schematic visualisation of the TCA cycle (green), the glyoxylate shunt (blue) and ethylmalonyl-CoA pathway (red), and the flow of acetyl-CoA (black). Labelled intermediates are $\bigtriangledown$: acetyl-CoA, 1: citrate, 2: isocitrate, 3: $\alpha$-ketoglutarate, 4: succinyl-CoA, 5: succinate, 6: fumarate, 7: malate, 8: oxaloacetate.

anapleurosis must occur using acetyl-CoA molecules as precursors. *S. coelicolor* possesses the genetic potential for two such anapleurotic pathways, the glyoxylate shunt and the ethylmalonyl-CoA pathway (Fig. 1.5). More information on genes and pathways in *S. coelicolor* is found in the ScoCyc database (http://biocyc.org/SCO/organism-summary?object=SCO).

## 1.3. The glyoxylate shunt

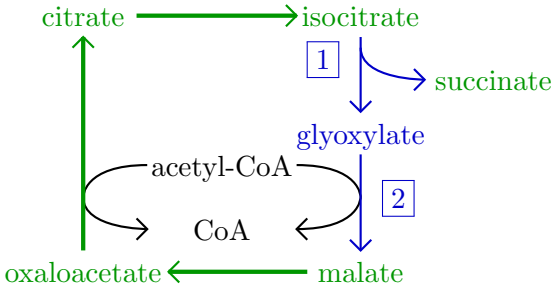



Figure 1.6.: Reactions of the glyoxylate shunt (blue) and three reactions of the TCA cycle (green). Glyoxylate shunt reactions are catalysed by the following enzymes: 1) isocitrate lyase (ICL), 2) malate synthase (MS).

The glyoxylate shunt is a metabolic pathway that is present in archaea, bacteria, protists, plants, fungi and nematodes (Dunn *et al.*, 2009). It is integrated with the TCA cycle, but bypasses its two decarboxylation reactions. The carbon that is preserved this way is used to produce a second TCA cycle intermediate, malate, by condensing it with the acetyl moiety of acetyl-CoA, making the glyoxylate shunt an anapleurotic pathway (Fig. 1.6).

The glyoxylate shunt branches off from the TCA cycle at the compound isocitrate. In the TCA cycle, isocitrate is converted to 2-oxoglutarate (also known as $\alpha$-ketoglutarate) and $CO_2$ by isocitrate dehydrogenase (IDH). As a substrate for the glyoxylate shunt, isocitrate is cleaved into succinate and glyoxylate by isocitrate lyase (ICL) instead. The next enzyme of the pathway, malate synthase (MS), uses glyoxylate, acetyl-coenzyme A (CoA) and $H_2O$ as substrates, producing CoA and malate (Fig. 1.6). Malate can be converted into oxaloacetate, completing the cycle. The succinate that was produced by ICL, can go through the TCA cycle reactions or be used as a substrate for biosynthetic processes (Fig. 1.5) (Dunn *et al.*, 2009).

The net result of the reactions of the glyoxylate shunt is the formation of a TCA cycle intermediate from two molecules of acetyl-CoA. Thus the glyoxylate shunt can be used

to replenish these intermediates when they are used for biosynthesis or gluconeogenesis.

## 1.4. The ethylmalonyl-CoA pathway

The ethylmalonyl-CoA pathway is a recently elucidated acetyl-CoA assimilation pathway (Erb *et al.*, 2009). The genes encoding enzymes of this pathway are present in the genomes of different bacteria, among which are members of the order of *Actinomycetales* (Erb *et al.*, 2009).

The pathway starts with the condensation of two acetyl-CoA molecules, and ends with the production of two TCA cycle intermediates (malate and succinyl-CoA, Fig. 1.7). It is an anapleurotic pathway similar to the glyoxylate shunt, but most of its intermediates are acyl-CoA thioesters and some of these, such as methylmalonyl-CoA and ethylmalonyl-CoA, can also serve as starter- or extender units in polyketide biosynthesis (see Section 1.1, Hertweck (2009); Wilson & Moore (2012)). An example is monensin A, produced by *Streptomyces cinnamonensis*, the biosynthesis of which uses both methylmalonyl-CoA and ethylmalonyl-CoA (Liu & Reynolds, 1999).

## 1.5. Relative impact of the ethylmalonyl-CoA pathway and the glyoxylate shunt on streptomycete metabolism

When comparing the energetic efficiencies of the glyoxylate shunt and the ethylmalonyl-CoA pathway, the former appears the require less energy; the formation of one mole of malate and succinyl-CoA each through the ethylmalonyl-CoA pathway costs 2 mol NADPH and 1 mol ATP in addition to the acetyl-CoA precursors (Erb *et al.*, 2009), while the glyoxylate shunt requires no additional cofactors. Moreover, the glyoxylate shunt consists of 2 enzymes, while the ethylmalonyl-CoA pathway involves the action of 12 enzymes, the production of which costs energy and amino acids.

A closer look however reveals that this comparison is insufficient. Equations (1.1) and (1.2) show simplified mass balances for the glyoxylate shunt and ethylmalonyl-CoA pathway, normalised to the production of 2 molecules of oxaloacetate (intermediate 8 in Fig. 1.5). Consumed cofactors (ADP, GDP, $NAD^+$, $NADP^+$ and FAD) were left out for clarity. The energy yield from reactions of the TCA cycle are included as well (succinyl-CoA synthetase (SCS) yields GTP, succinate dehydrogenase (SDH) yields $FADH_2$, SDH and malate dehydrogenase (MDH) yield NADH).

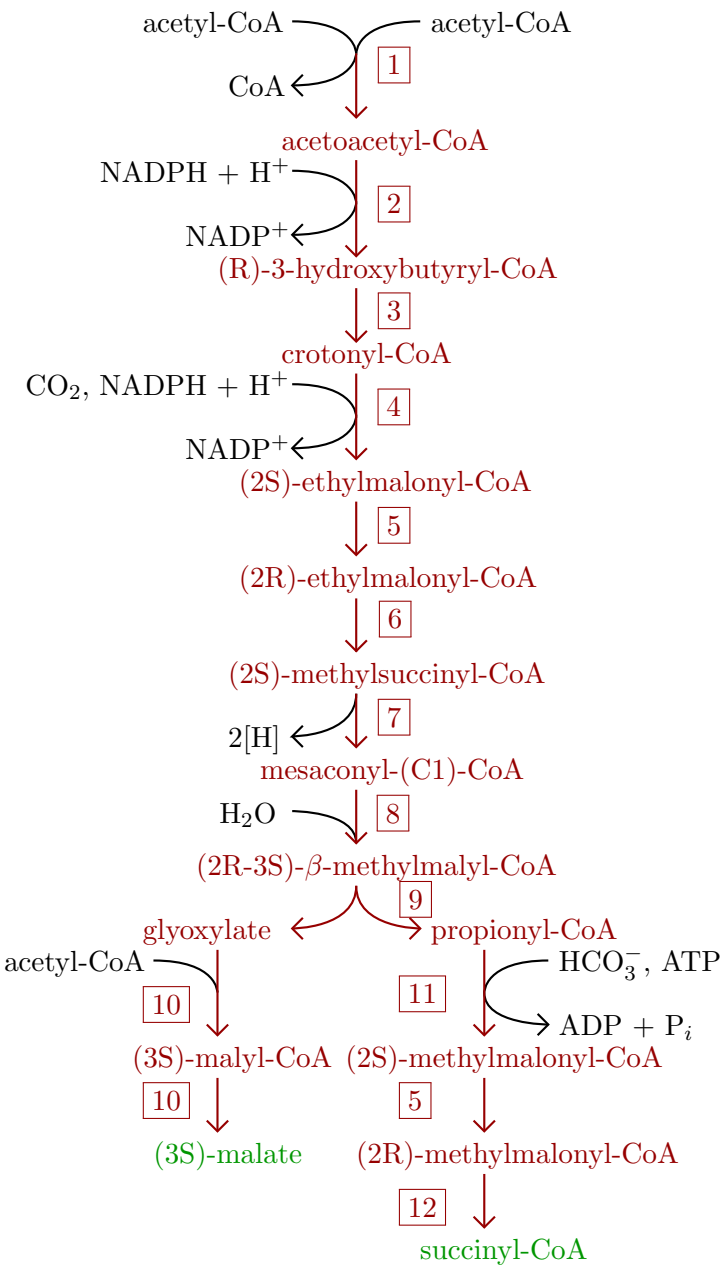



Figure 1.7.: Reactions of the ethylmalonyl-CoA pathway as given by Erb *et al.* (2009). Reactions are catalysed by the following enzymes: 1) $\beta$-ketothiolase (PhaA), 2) acetoacetyl-CoA reductase (PhaB), 4) crotonyl-CoA carboxylase / reductase (Ccr), 5) ethylmalonyl-CoA / methylmalonyl-CoA epimerase (Epi), 6) (2R)-ethylmalonyl-CoA mutase (Ecm), 7) (2S)-methylsuccinyl-CoA dehydrogenase (Mcd), 8) mesaconyl-CoA hydratase (Mch), 9) $\beta$-methylmalyl-CoA / (3S)-malyl-CoA lyase (Mcl-1), 10) (3S)-malyl-CoA thioesterase, 11) propionyl-CoA carboxylase (PccAB), 12) (2R)-methylmalonyl-CoA mutase (Mcm). Exact details of reaction 3 are uncertain (Erb *et al.*, 2009). The first two reactions of this pathway are shared with polyhydroxybutyrate biosynthesis (Erb *et al.*, 2009)

$$2\,\text{oxaloacetate} + 4\,\text{acetyl-CoA} \rightarrow 4\,\text{oxaloacetate} + 2\,\text{FADH}_2 + 4\,\text{NADH} \quad (1.1)$$

$$3\,\text{acetyl-CoA} + 2\,\text{CO}_2 + 2\,\text{NADPH} + 1\,\text{ATP} \rightarrow 2\,\text{oxaloacetate} + 1\,\text{GTP} + 2\,\text{FADH}_2 + 4\,\text{NADH} \quad (1.2)$$

$$1\,\text{oxaloacetate} + 1\,\text{acetyl-CoA} \rightarrow 1\,\text{NADPH} + 2\,\text{NADH} + 1\,\text{GTP} + 1\,\text{FADH}_2 + 1\,\text{oxaloacetate} \quad (1.3)$$

A comparison of Eq. (1.1) and Eq. (1.2) shows that the glyoxylate shunt indeed uses less energy than the ethylmalonyl-CoA pathway, with a difference of 2 NADPH. However, the latter requires only 3 acetyl-CoA molecules whereas the former requires 4. If this acetyl-CoA is oxidised using the TCA cycle, it has the energy yield shown in Eq. (1.3) (IDH yields NADPH). The difference in acetyl-CoA consumption between the two pathways is caused by the two carboxylation steps in the ethylmalonyl-CoA pathway (Section 1.4). Including the potential energy yield of this additional acetyl-CoA molecule in the comparison makes the ethylmalonyl-CoA pathway more energy efficient.

Additional evidence for the relative importance of the ethylmalonyl-CoA pathway compared to the glyoxylate shunt is provided by the observation that an $\text{ICL}^-$ strain, *Streptomyces collinus*, is capable of growth on acetate as the carbon source (Han & Reynolds, 1997). Disruption of the *ccr* gene in this strain however severely impaired growth on this carbon source. Expression of *S. coelicoloraceA* and *aceB* in a *ccr* mutant of *Streptomyces cinnamonensis*, which had a similar phenotype, did not restore growth on acetate. The higher efficiency of the ethylmalonyl-CoA pathway with respect to carbon utilisation, as well as the observation that certain strains lacking the glyoxylate shunt are still capable of growth on acetate, makes it likely to be the primary anapleurotic pathway in *S. coelicolor*, despite the higher requirement for enzyme production. The question of what the function of the glyoxylate shunt is, therefore remains open.

## 1.6. Regulation of the glyoxylate shunt

For a primary metabolic pathway, the steady state is the thermodynamically most efficient state. A metabolic pathway in steady state is characterized by a constant flux, meaning that the amount of substrate that is transformed into a product per unit of time is constant. When a parameter in the system changes, such as the concentration of

a substrate or the activity of an enzyme, all reaction rates and metabolites in the system will adapt to the changes until a new steady state is reached. Metabolism is generally assumed to be in steady state for modelling purposes (Singh & Ghosh, 2006; Tepper *et al.*, 2013; Visser & Heijnen, 2002; Wiechert & Noack, 2011).

Various factors influence the steady state, such as the concentrations of metabolites, activity levels of enzymes and their affinity for various substrates, products and inhibitors. These factors can be influenced by the organism by regulation at the transcriptional, translational or post-translational levels. The TCA cycle generates NADH, which is then used to generate ATP, by oxidation of acetyl-CoA but also provides precursors for biosynthesis when needed. Three different regulatory systems that balance these two functions have been described in *E. coli*, which are discussed below along with their relevance to *S. coelicolor* metabolism.

### 1.6.1. Phosphorylation and dephosphorylation of IDH

Posttranslational modification of an enzyme can influence its activity or substrate affinity. A relevant example is the phosphorylation and dephosphorylation of IDH by isocitrate dehydrogenase kinase/phosphatase (IKP) in *E. coli*, which is located in the same operon as the genes for ICL and MS. This bifunctional enzyme can both phosphorylate and dephosphorylate IDH at serine 113, which is part of the active site. Phosphorylation leads to complete inactivation of the enzyme, while dephosphorylation reactivates it (Cozzone, 1998).

With this system, *E. coli* can regulate fluxes at the branch point between the TCA cycle and the glyoxylate bypass. IDH has a much higher affinity for isocitrate ($K_M$ = 29 µM) than ICL ($K_M$ = 890 µM) (Ogawa *et al.*, 2007), hence flux through the TCA cycle under normal conditions is much higher than through the glyoxylate shunt. When a large fraction of IDH is phosphorylated, the level of active enzyme decreases and so does the rate of the reaction it catalyses. Due to this, isocitrate levels increase, promoting the isocitrate lyase reaction. Dephosphorylation leads to an increased rate of the reaction catalysed by IDH, lowering the isocitrate level and thus lowering the flux through the glyoxylate shunt (Cozzone, 1998).

Although this system is thought to help *E. coli* maintain homeostasis, the situation in *S. coelicolor* is different. One of the key differences lies in the enzyme IDH. In *E. coli*, IDH (GenBank ACB02329.1) has 416 amino acids and acts as a homodimer (Hurley *et al.*, 1989). *Streptomyces lividans* IDH (GenBank EU661252.1), which is 99 % similar to *S. coelicolor* IDH, is a monomer of 739 amino acids (Zhang *et al.*, 2009).

The crystal structure of the monomeric IDH of *Azotobacter vinelandii* has been solved

by X-ray christallography (Yasutake *et al.*, 2002). The structure showed a high similarity to the *E. coli* IDH dimer, even though the amino acid sequences are different. Serine 113, the phosphorylation target in *E. coli* IDH, turned out to be structurally conserved in *A. vinelandii* and *S. coelicolor* at position 132 (Yasutake *et al.*, 2002). This suggests these monomeric enzymes could be regulated in a similar way as the dimeric enzyme.

The conservation of this serine in *S. coelicolor*IDH leads to the expectation that it is possibly regulated similarly. However, *E. coli* IKP was unable to inhibit *S. coelicolor* IDH activity (Taylor, 1992) and no homologue of IKP could be found in its genome. A study analysing the phosphoproteome of *S. coelicolor* found a number of proteins phosphorylated at serine residues, but IDH was not among them (Parker *et al.*, 2010). The apparent lack of *S. coelicolor* IDH phosphorylation and the inability of *E. coli* IKP to effect it suggests that regulation of the flux through the glyoxylate shunt in *S. coelicolor* does not use this mechanism.

### 1.6.2. Transcriptional regulation of genes encoding glyoxylate shunt enzymes

Since the spread of microbial spores is driven by external forces and the interconnected hyphae of the vegetative mycelium are immobile, *S. coelicolor* needs to be capable of adapting to a wide range of environmental conditions. These conditions include the presence or absence of nutrients such as sugars, vitamins or amino acids. In order to adapt to these conditions, *S. coelicolor* needs to "reprogram" its metabolism to allow utilisation of the available compounds and synthesis of those that are absent.

The role of transcriptional regulation in prokaryotic metabolism is limited. The correlation between transcript abundance and the flux through the corresponding pathway was studied in *Bacillus subtilis* (Chubukov *et al.*, 2013). Enzymes that are required during specific circumstances, such as substrate importers or amino acid biosynthesis pathways, showed changes in the level of flux proportional to changes in transcript abundance. For enzymes of central metabolism however, changes in transcription could not explain the changes in metabolic flux. Chubukov *et al.* (2013) concluded that pathways which are switched either on or off, are regulated primarily at the level of transcription. Anapleurotic pathways such as the glyoxylate shunt only have a function in specific circumstances so strong transcriptional regulation is expected.

A likely transcriptional regulator is encoded by *Sco0981* (*aceR*), which is located next to *Sco0982* (*aceA*) and *Sco0983* (*aceB2*), the genes that encode ICL and MS, respectively (Fig. 1.8). AceR is a likely orthologue of RamB (49 % amino acid sequence similarity), a transcriptional repressor from *Corynebacterium glutamicum*. In this organism, RamB

represses the expression of *aceA* and *aceB* when grown on glucose (Gerstmeir *et al.*, 2004). The sequence similarity of AceR to RamB and its position relative to the glyoxylate shunt operon makes it a likely transcriptional regulator for these genes.

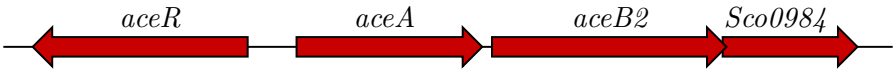



Figure 1.8.: Genes of the glyoxylate shunt, *aceR*, *aceA*, *aceB2*. *Sco0984* is not known to be involved in the pathway, but overlaps with *aceB2* by 4 nucleotides.

Another predicted regulator for the glyoxylate shunt is IclR, which is encoded by *Sco6246* (*iclR*). The *E. coli* orthologue (40% sequence similarity with *S. coelicolor* IclR), called IclR as well, has been studied extensively. It acts as a repressor for the *aceBAK* operon, which contains the genes for ICL and MS, as well as IKP (which is absent from *S. coelicolor*, see above and Taylor (1992)). It also binds to its own promoter. However, despite its similarity to *E. coli* IclR, *S. coelicolor* IclR was recently shown to act as a regulator of the allantoin catabolic pathway (Navone *et al.*, 2014). Although this pathway features glyoxylate as one of its intermediates, its function is unrelated to anapleurosis during growth on $C_2$ compounds.

A third possible regulator of glyoxylate shunt genes is cyclic AMP receptor protein (CRP), encoded by *Sco3571* (*crp*) (Derouaux *et al.*, 2004a). This gene has orthologues in both *E. coli* and *C. glutamicum*. In *E. coli*, CRP regulates the expression of genes involved in carbon metabolism and one of its three homologues in *C. glutamicum*, GlxR, was shown to repress the expression of *aceA*, *aceB* and genes involved in carbon catabolite repression, the system that represses utilization of certain carbon sources in favour of preferred ones. A *crp* knockout mutant of *S. coelicolor* had defects in spore germination and growth. Deletion of *glxR* in *C. glutamicum* also resulted in a severe growth defect, which could be restored by complementation with the *glxR* and *crp* genes from *C. glutamicum* and *S. coelicolor*, respectively (Park *et al.*, 2010).

The ability of *S. coelicolor crp* to complement *C. glutamicum glxR* suggests a similarity in function and mechanism. However, *S. coelicolor crp* could not complement a deletion of *crp* in *E. coli* while *C. glutamicum glxR* could. It was suggested that *S. coelicolor* CRP might need the presence of one or more additional proteins, since purified CRP failed to bind to CRP-like *cis*-acting elements (Derouaux *et al.*, 2004b). Consequently, the function of CRP in relation to the glyoxylate shunt genes remains unclear.

### 1.6.3. Allosteric regulation of glyoxylate shunt enzymes

Chubukov *et al.* (2013) concluded that for central metabolic pathways, differences in

transcription could not explain experimentally determined differences in metabolic flux. Surprisingly, variations in substrate levels for each reaction were also insufficient to explain flux changes, and a significant role for allosteric regulation was proposed.

An example of this type of regulation of glyoxylate shunt enzymes was described in *E. coli*, where PEP is an allosteric cooperative inhibitor of IDH and an uncompetitive inhibitor of ICL (Ogawa *et al.*, 2007). Allosteric cooperative inhibition is of the competitive type. When (reversibly) bound to inhibitor, the enzyme cannot bind to the substrate and vice versa. Cooperative inhibition implies inhibition by two nonexclusive inhibitors. This is easily explained by the fact that IDH acts as a dimer in *E. coli* and thus has two distinct inhibitor binding sites, one on each monomer. The effect of binding of this inhibitor is that the apparent $K_{\mathrm{M}}$ is increased but the $\mathrm{V}^{\max}$ is unaltered. At a given substrate concentration, the reaction rate in presence of this inhibitor will therefore be a smaller fraction of $\mathrm{V}^{\max}$ than when the inhibitor is absent. The reaction schematic for cooperative competitive inhibition can be explained in the following way (X and I are both PEP in this case):

$$\begin{array}{ccccc} \mathrm{EX} & \rightleftharpoons & \mathrm{X} + & \mathrm{E} + \mathrm{S} \rightleftharpoons \mathrm{ES} \rightarrow \mathrm{E} + \mathrm{P} \\ + & & & + & \\ \mathrm{I} & & & \mathrm{I} & \\ \uparrow\downarrow & & & \uparrow\downarrow & \\ \mathrm{EXI} & \rightleftharpoons & \mathrm{X} + & \mathrm{EI} & \end{array}$$

When an inhibitor binds in an uncompetitive way (which is distinct from noncompetitive), it will reversibly bind to the enzyme-substrate complex to form an enzyme substrate-inhibitor complex. It only inhibits product formation and dissociation and does not bind to free enzyme. Effectively, it lowers $V^{\max}$ but because it decreases the amount of enzyme-substrate complex, it also decreases the apparent $K_{\mathrm{M}}$. This means that, at low substrate concentrations in presence of PEP, the reaction rate is increased compared to the rate that substrate concentration in the absence of PEP. The reaction schematic for uncompetetive inhibition can be explained in the following way:

$$\begin{array}{c} \mathrm{E + S \rightleftharpoons ES \rightarrow E + P} \\ + \\ \mathrm{I} \\ \uparrow\downarrow \\ \mathrm{ESI} \end{array}$$

The important difference between the two types of inhibition for this branchpoint is that in the case of allosteric cooperative inhibition, a higher substrate concentration relieves the inhibition because the equilibrium of E + S ⇌ ES is shifted towards ES. This is opposed to the uncompetitive type of inhibition, which is stronger at higher substrate concentrations because the higher level of ES will lead to a higher level of inhibition. At lower substrate concentration, however, inhibition is almost absent because the lower $V^{\max}$ is compensated for by the lowering of $K_{\mathrm{M}}$ (Segel, 1975). Therefore, at high PEP levels, either IDH or ICL will be inhibited, depending on the level of the substrate, isocitrate. At a low level of isocitrate, IDH is inhibited strongly and ICL is not inhibited. A high level of isocitrate results in strong inhibition of ICL and weak inhibition of IDH. At low PEP concentrations, no inhibition by PEP occurs.

## 1.7. Mathematical modelling of metabolic pathways

Engineering of metabolic pathways, with the aim of altering flux, requires a method to rationally determine which intervention is most likely to produce the desired result. In order to help understand the contribution of individual components to the flux through a metabolic pathway, a mathematical model of the pathway can be set up.

A mathematical model of a metabolic pathway consists of a list of intermediates and the reactions they are part of (the modelled system), and a set of equations that describe the rate of these reactions as a function of their participating metabolites (rate equations, Hadlich *et al.* (2009); Wiechert & Noack (2011)). After choosing a suitable set of starting concentrations for all intermediates, a simulation of the model can be carried out. During this simulation, the rate of each reaction is repeatedly calculated according to its rate equation, alternated by an adjustment of the concentration of each intermediate as a consequence of their production and consumption by these reactions for discrete time steps. An example of freely available software that implements this commonly used method is COPASI (Hoops *et al.*, 2006).

The net production or consumption of an intermediate affects the rates of all reactions that intermediate participates in, but changes in these rates in turn affect consumption and production of this intermediate. In a balanced model, the rates of producing and consuming reactions equate each other after simulation of a sufficient number of steps. When this is the case for each reaction in a modelled pathway, the system has reached a steady state and none of the reaction rates or intermediate concentrations change anymore. As mentioned in Section 1.6, the steady state is the thermodynamically most stable state of an open system. A model's steady state can be compared to flux distribution data, obtained by experimental approaches such as metabolic flux analysis (MFA) using $^{13}$C labelled compounds (Wiechert, 2001).

The set of reactions and intermediates described by a mathematical model is called the model system. It is always limited in scope and delimited by boundary conditions. Model boundaries are either intermediates or reactions whose concentrations or rates are under direct control of the modeller. The choice of model boundaries greatly influences the results of simulations. In the example system of Fig. 1.9, which shows the conversion of threonine to isoleucine, the logical system boundaries are threonine, pyruvate, NADPH, $NADP^+$, glutamate and 2-oxoglutarate. Their concentrations can be kept constant during simulation or varied to investigate their influence on the flux through this pathway.

Isoleucine is both the product of this biosynthesis pathway as well as an inhibitor of the pathway's first step (Yang *et al.*, 2005). As a result, choosing either isoleucine or its consumption in protein synthesis as a system boundary has an effect on the steady state flux of the pathway.

Different types of rate equation are available for use in a model. Two modelling frameworks are discussed below, along with the types of experimental data they require.

### 1.7.1. Models based on mechanistic rate equations

Mechanistic rate equations are based on classic enzyme kinetics and rely on full knowledge of each enzyme involved in the pathway. The simplest example of an equation of this type is the Michaelis-Menten equation for a unidirectional reaction with one substrate and one product. This equation is defined by Eq. (1.4), in which $v$ is the reaction rate, $V^{\max}$ the maximum reaction rate, $S$ the substrate concentration and $K_{\mathrm{S}}$ the Michaelis-Menten constant for the substrate. Equation (1.5) shows a similar equation, but for a reversible reaction. Here, $P$ is the reaction's product and $K_{\mathrm{P}}$ is the Michaelis-Menten constant for the product. $V_f^{\max}$ and $V_r^{\max}$ are the maximum reaction rates for the forward and reverse reactions, respectively. Other equations have been derived for enzymes

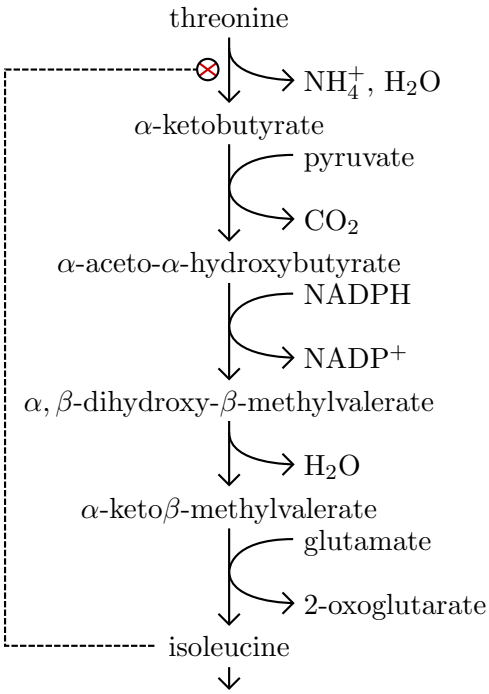



Figure 1.9.: Conversion of threonine to isoleucine in *E. coli*. The bottom arrow represents isoleucine consumption by protein synthesis

with more complex mechanisms (Segel, 1975).

$$v = V^{\max} \cdot \frac{S}{K_S + S} \tag{1.4}$$

$$v = \frac{V_f^{\max} \cdot \frac{S}{K_S} - V_r^{\max} \cdot \frac{P}{K_P}}{1 + \frac{A}{K_A} + \frac{P}{K_P}} \tag{1.5}$$

Setting up a mathematical model using mechanistic rate equations, such as those given in Eqs. (1.4) and (1.5), requires knowledge of the maximum reaction rates and Michaelis-Menten constants of all enzymes for each intermediate. These can be determined from assays on isolated enzymes or taken from literature. Although these equations are very precise across a broad range of intermediate concentrations, obtaining the model parameters is laborious when not much work has been done on the enzymes of interest or for longer pathways.

### 1.7.2. Approximative Kinetics

Approximative kinetic equations do not describe each enzyme in full detail but attempt to approximate its behaviour non-mechanistically. The reason for the development of this approach is that it simplifies algebraic treatment and reduces the number of required measurements. A number of approximative kinetic formats has been developed (Hadlich *et al.*, 2009), but only linlog kinetics (Visser & Heijnen, 2003) will be described here.

Linlog kinetics is an extension of the metabolic control analysis (MCA) framework (Visser & Heijnen, 2002). In MCA, control coefficients are calculated for each enzyme in a pathway. These control coefficients quantify the effect a change in the level of enzyme $j$ has on the flux through a pathway ($C_{ij}^{J0}$) or on the concentration of intermediate $i$ ($C_{ij}^{x0}$). In a similar way, elasticities ($\varepsilon_{ij}^{0}$) are a measure of the effect of the concentration of intermediate $j$ on each reaction rate $i$ separately. The mathematical definitions of these control coefficients and elasticities are given in Eqs. (1.6) to (1.8) (Visser & Heijnen, 2002).

$$C_{ij}^{J0} = \frac{e_j^0}{J_i^0}\frac{\mathrm{d}J_i}{\mathrm{d}e_j} \qquad \sum C_{ij}^{J0} = 1 \tag{1.6}$$

$$C_{ij}^{x0} = \frac{e_j^0}{x_i^0}\frac{\mathrm{d}x_i}{\mathrm{d}e_j} \qquad \sum C_{ij}^{x0} = 0 \tag{1.7}$$

$$\varepsilon_{ij}^{0} = \frac{x_j^0}{J_i^0}\frac{\partial v_i}{\partial x_j} \tag{1.8}$$

In linlog kinetics the rate equation of each reaction in a pathway consists of a linear sum of logarithmic concentration terms (Visser & Heijnen, 2003). Equation (1.9) is a linlog equation for the first reaction of the isoleucine biosynthesis pathway shown in Fig. 1.9, with terms for most of the intermediates omitted for clarity. This equation calculates the rate of the pathway's first reaction in a given state relative to a reference steady state value ($\frac{v_1}{J^0}$, with reaction rate $v_1$ and reference flux $J^0$). Each intermediate involved in the pathway is represented by the natural logarithm of its concentration divided by its concentration in the reference condition ($\ln \frac{\mathrm{thr}}{\mathrm{thr}^0}$ with $thr$ the concentration of threonine and $thr^0$ the concentration of threonine in the reference condition). This logarithm is multiplied by its elasticity value for that reaction ($\varepsilon_{1,\mathrm{thr}}^{0}$, Eq. (1.8)), as the elasticity quantifies the extent to which a metabolite influences the rate of a reaction. The sum of elasticity-corrected logarithms is multiplied by the level of enzyme activity divided by the level of enzyme activity in the reference condition ($\frac{e}{e^0}$).

$$\frac{v_1}{J^0} = \frac{e}{e^0}\left(1 + \varepsilon^0_{1,\text{thr}} \ln \frac{\text{thr}}{\text{thr}^0} + \varepsilon^0_{1,\alpha\text{-kb}} \ln \frac{\alpha\text{-kb}}{\alpha\text{-kb}^0} + ... + \varepsilon^0_{1,\text{ile}} \ln \frac{\text{ile}}{\text{ile}^0}\right) \tag{1.9}$$

This formulation has a few implications. The elasticity $\varepsilon^0_{ij}$ can have a positive or negative value, depending on whether it inhibits or promotes the reaction. In the example of Eq. (1.9), $\varepsilon^0_{1,\text{thr}}$ would be positive since threonine stimulates the reaction, while $\varepsilon^0_{1,\text{ile}}$ would be negative, since isoleucine inhibits this reaction. If a metabolite has no influence on the rate of a particular reaction, its elasticity value for that reaction $\varepsilon^0_{ij}$ equals 0 and any change in its concentration does not result in a change in reaction rate. If a metabolite does not change between the reference state and the new state the reaction rate is not influenced by it, since the ratio of its concentrations then equals 1, and $\ln 1 = 0$.

The elasticity values for the linlog rate equation can be calculated from steady state flux and corresponding metabolite concentration data. The linear character of the equation allows this to be done using multiple linear regression (Wu *et al.*, 2004).

### 1.7.3. Functional and mathematical analysis of the glyoxylate shunt in *S. coelicolor*

The work described in this thesis was carried out in order to achieve two objectives; to elucidate the function of the glyoxylate shunt in *S. coelicolor* and to build a mathematical model of the flux through this pathway, based on the hypothesis that it is an anapleurotic pathway that replenishes TCA cycle intermediates during growth on acetate and fatty acids. The mathematical model should allow *in silico* testing of hypotheses with regard to metabolic engineering and provide a rational basis for experimental design.

The additional presence of genes encoding enzymes of the ethylmalonyl-CoA pathway, which provides the organism with a second anapleurotic pathway, challenges this hypothesis. In order to investigate whether the glyoxylate shunt genes are potentially redundant, the numbers of polymorphisms in sequences of genes encoding central carbon metabolic enzymes in a number of streptomycete genomes were compared. In addition to this, sequences upstream of the glyoxylate shunt and ethylmalonyl-CoA pathway genes were searched for potential regulatory motifs.

The relevance of the glyoxylate shunt to fatty acid utilisation was investigated using both a genetic and a transcriptomic approach. Growth of a mutant *S. coelicolor* strain lacking the glyoxylate shunt genes was compared to that of the wild type in a defined minimal medium containing Tween 40 as the sole carbon source. Differential gene expression between *S. coelicolor* cultures grown on glucose or Tween 40 was measured using

total RNA sequencing.

In order to quantify flux through the TCA cycle and the glyoxylate shunt in *S. coelicolor*, a mathematical model was developed using mechanistic rate equations. The reaction mechanisms of the enzymes involved in these pathways have been studied in other prokaryotes and a mathematical model of this system was set up previously for *E. coli* (Singh & Ghosh, 2006). Additional kinetic parameters for key enzymes of these pathways in *S. coelicolor* were obtained by heterologous expression, purification and subsequent enzyme assays. These parameters were supplemented with data from the literature and used to adapt the *E. coli* model to describe *S. coelicolor* metabolism. The process was validated using experimental flux distribution data and simulations of the model were carried out to predict the effects of increasing extracellular fatty acid concentration on central carbon metabolism.

The work described here sheds new light on fatty acid utilisation by streptomycetes, which, in the humble opinion of the author, is an understudied subject. It proposes a function for the glyoxylate shunt which is potentially relevant to a genus that is widely relied on to produce bioactive secondary metabolites. This reliance can be expected to increase in the near future as antibiotic resistance is becoming increasingly problematic and gains more attention from society and politicians. Since fatty acids make up important components of industrial growth media as well as of the organic matter that forms the growth substrate of Streptomycetes in their natural habitat, increased knowledge of the glyoxylate shunt and of anapleurosis in streptomycete metabolism in general is important in both industrial settings as for more fundamental research.

# 2. Materials and methods

## 2.1. Bacterial strains

*E. coli* strains JM109, ET1256 / pUZ8002, BTH101, BL21 (DE3) and BW25113 / pIJ790 were used in this work Table 2.1. Strains were routinely grown at 37 °C, except for strains BTH101, BW25113 / pIJ790 and BL21 (DE3) expressing *S. coelicolor* proteins, which were grown at 30 °C. Liquid cultures were incubated in a rotary shaking incubator at 250 RPM.

*S. coelicolor* M145 and its derivatives were grown at 30 °C on solid SFM, R5 or Minimal Medium (Kieser *et al.*, 2000) at 30 °C. Liquid cultures were grown in YEME or HMM (Hobbs *et al.*, 1989) as 30 °C in a rotary shaking incubator at 250 RPM using conical flasks with baffles or springs.

| **Strain** | **Characteristics** | **Reference** |
|---|---|---|
| *E. coli* | | |
| DH5$\alpha$ | F$^-$ $\phi$80*lacZ*$\Delta$M15 $\Delta$(*lacZYA*$^-$ *argF U169 recA1*) *endA1 hsdR17* ($r_K^-$, $m_K^+$) *supE44* $\lambda^-$ *thi-1 gyrA96 relA1* | Sambrook *et al.* 1989 |
| JM109 | *endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB*$^+$$\Delta$(*lac-proAB*) *e14*$^-$ [F' *traD36 proAB*$^+$ *lacI*$^q$ *lacZ*$\Delta$M15] *hsdR17* ($r_K^-m_K^+$) | Sambrook *et al.* 1989 |
| ET12567 / pUZ8002 | *dam13*::Tn*9 dcm6 hsdM hsdR recF143 zjj201*::Tn*10 galK2 galT22 ara14 lacY1 xyl5 leuB6 thi1 tonA31 rpsL136 hisG4 tsx78 mtli glnV44* F$^-$ | Kieser *et al.* 2000; MacNeil *et al.* 1992 |
| BL21 (DE3) | F- *ompT gal dcm HsdS*$_B$($r_B^-m_B^-$) $\lambda$(DE3 [*lacI lacUV5-T7 ind1 sam7 nin5*]) | Studier & Moffatt 1986 |
| BW25113 / pIJ790 | *lacI*$^q$ *rrnB*$_{T14}$$\Delta$*lacZ*$_{WJ16}$ *hsdR514* $\Delta$*araBA*$^-$ D$_{AH33}$$\Delta$*rhaBAD*$_{LD78}$ | Datsenko & Wanner 2000 |
| *S. coelicolor* | | |
| M145 | SCP1$^-$ SCP2$^-$ | Kieser *et al.* 2000 |
| RAR101 | M145 $\Delta$0982–3::aac(3)IV·oriT | This work |
| RAR102 | RAR101 / pShunt | This work |
| RAR8660 | M145 / pIJ8660 | This work |
| RARccrO | M145 / pCCRO | This work |
| RARiclO | M145 / pICLO | This work |

Table 2.1.: Bacterial strains used in this thesis

## 2.2. Plasmids and cosmids

The plasmids and cosmids used in this study are listed and briefly described below.

| Name | Characteristics |
|---|---|
| 11D02 | Cosmid containing glyoxylate shunt genes (Bentley *et al.*, 2002) |
| pIJ773 | Plasmid containing redirect cassette conferring apramycin resistance |
| pIJ8660 | Plasmid containing $\phi$C31 integration system and promoter-less *egfp* |
| pRiz101 | *aceA* cloned into pUC19 |
| pRiz104 | *idh* cloned into pUC19 |
| pRiz110 | XbaI-BamHI fragment of 11D02 cloned into pIJ925 |
| pExpaceA | *aceA* subcloned from pRiz101 into pET26b$^+$ (Novagen) |
| pExpidh | *idh* subcloned from pRiz104 into pET26b$^+$ (Novagen) |
| pShunt | XbaI-EcoICRI fragment of pRiz110 cloned into pMS82 |
| pCcrO | Sequence upstream of *ccr* start codon cloned into pIJ8660 |
| pIclO | Sequence upstream of *aceA* start codon cloned into pIJ8660 |
| 11D02 $\Delta$0982–3 | Cosmid 11D02 with genes *aceA* and *aceB2* replaced by pIJ773 cassette |

Table 2.2.: Plasmids used in this thesis

## 2.3. Growth media for *E. coli* and *Streptomyces* and buffers

### 2.3.1. Solid media

LB agar contained (per litre): 10 g tryptone, 5 g NaCl, 5 g yeast extract, 15 g agar.

M63 medium was made by aseptically mixing (per litre medium): 200 ml 5x M63 salts solution, 800 ml sterilised $dH_2O$ containing 15 g agar, 3 ml 20 % (w/v) maltose solution, 400 µl 0.05 % (w/v) thiamine solution (vitamin B1), 200 µl $MgSO_4$ (1 M). 5x M63 salts solution contained (per litre): 10 g $(NH_4)_2SO_4$, 68 g $KH_2PO_4$ and 2.5 mg $FeSO_4 \cdot 7H_2O$, adjusted to pH = 7.0 with KOH.

SFM contained (per litre): 16 g soya flour, 20 g mannitol, 20 g agar. Tap water was used instead of $dH_2O$.

Solid minimal medium contained (per litre): 1 g $(NH_4)_2SO_4$, 0.5 g $K_2HPO_4$, 0.2 g $MgSO_4 \cdot 7\ H_2O$ and 10 g agar. Carbon source was added after autoclaving (10 g $l^{-1}$ glucose, 0.5 % (w/v) Na-Acetate or 1 % (w/v) Tween 40).

R5 medium contained (per litre): 103 g sucrose, 0.25 g $K_2SO_4$, 10.12 g $MgCl_2 \cdot 6H_2O$, 10 g glucose, 0.1 g casaminoacids, 2 ml trace element solution, 5 g yeast extract, 5.73 g TES buffer, 22 g Difco Bacto agar. At time of use the following was added (per 100 ml, all but NaOH autoclaved): 1 ml $KH_2PO_4$ (0.5 %), 0.4 ml $CaCl_2 \cdot 2H_2O$ (5 M), 1.5 ml L-proline (20 %), 0.7 ml NaOH (1 N).

### 2.3.2. Liquid media

LB contained (per litre): 10 g tryptone, 5 g NaCl, 5 g yeast extract.

2xYT contained (per litre): 16 g tryptone, 10 g yeast extract, 5 g NaCl.

SOB contained (per litre): 20 g tryptone, 5 g yeast extract, 0.5 g NaCl and 0.186 g KCl. Just before use, filter sterilized 10 ml $l^{-1}$ $MgCl_2$ (1 M) and 10 ml $l^{-1}$ $MgSO_4$ (1 M) were added. SOC consisted of SOB supplemented with 20 mM glucose.

YEME contained (per litre): 10 g glucose, 3 g yeast extract, 3 g malt extract, 5 g peptone, 340 g sucrose (optional).

Liquid minimal medium (HMM, (Hobbs *et al.*, 1989)) contained (per litre): 5 g NaCl, 5 g $Na_2SO_4$, 4.5 g $NaNO_3$, 1.2 g Tris, 1 g $MgSO_4$, 0.01 g $ZnSO_4$, 1 ml Trace element solution, 2 g $K_2HPO_4$. Trace element solution contained (per 100 ml): 204.9 mg $ZnSO_4 \cdot 7H_2O$, 102.2 mg $MnCL_2 \cdot 4H_2O$, 30.5 mg $H_3BO_3$, 41.2 mg $CuSO_4$, 20.7 mg $NaMoO_4$, 20.6 mg $CaCl_2 \cdot 2H_2O$, 870.2 mg $FeCl_3$. All was mixed in 800 ml $dH_2O$, the pH was set to 7.2 and the volume was increased to 900 ml. $K_2HPO_4$ (pH = 7.2) was autoclaved separately as a 20 g $l^{-1}$ solution and 100 ml $l^{-1}$ (final) was added just before use.

### 2.3.3. Protein isolation buffers

The resuspension buffer used for the isolation of ICL contained 50 mM $NaH_2PO_4$, 20 mM PIPES, 250 mM NaCl, 10 mM imidazole at pH 8.0. The elution buffer for ICL isolation was the same as the resuspension buffer but contained 500 mM imidazole instead.

The resuspension buffer used for the isolation of IDH contained 10 mM $K_2HPO_4$, 500 mM NaCl, 2 mM $MgCl_2$, 2 mM $\beta$-mercaptoethanol and 10 mM imidazole at a pH of 7.7. The elution buffer for IDH isolation was the same as the resuspension buffer but contained 500 mM imidazole instead.

## 2.4. DNA manipulation

Plasmids and cosmids were isolated from *E. coli* using alkaline lysis (Birnboim & Doly, 1979). Additional purifications were done by incubation with 30 µg $ml^{-1}$ RNaseA, ex-

traction with an equal volume of phenol-chloroform-isoamyl alcohol (IAA) (25:24:1), extraction with an equal volume pure chloroform (twice) followed by overnight precipitation in a final concentration of 0.8 M NaCl and 6.5 % PEG 6000 at 4 °C.

*S. coelicolor* genomic DNA was isolated by lysosyme treatment of mycelium grown in YEME, neutral lysis with sodium dodecyl sulphate (SDS) and extraction with phenol-chloroform-IAA and pure chloroform, followed by precipitation with isopropanol (Kieser *et al.*, 2000).

Restriction enzymes and $T_4$ ligase were obtained from Promega and NEB. Pfu (Stratagene) and MyTaq (Bioline) were used for PCR reactions. A list of primers can be found in Appendix A. Enzymes were used according to the manufacturer's instructions.

PCR reactions for the creation of gene knockout cosmids (ReDirect procedure) were done according to Gust *et al.* (2003). The template consisted of a gel-purified 1396 bp fragment of EcoRI and HindIII digested pIJ773. The long primers consisted of a 5' 39 bp sequence identical to the flanks of the targeted region and a 3' sequence identical to a priming sequence on the pIJ773 cassette. The reaction mixture contained 50 pmol of each primer, 50 ng template, 50 µM of each dNTP, 5 µl 10x concentrated reaction buffer, 2.5 µl dimethyl sulfoxate (DMSO), 2.5 U DNA polymerase and $dH_2O$ to a total volume of 50 µl. Cycle conditions were 94 °C for 2 min, 10 cycles of 94 °C for 45 s, 50 °C for 45 s and 72 °C for 90 s, followed by 15 cycles of 94 °C for 45 s, 55 °C for 45 s and 72 °C for 90 s, with a final extension step at 72 °C for 5 min.

## 2.5. *S. coelicolor* inoculum preparation

Inoculum for cultures of minimal medium were prepared in two different ways. Mycelium from a 16–24 h old culture grown in YEME was washed with and resuspended in a 10.3 M sucrose solution. The dry cell weight (DCW) content of this solution was determined as described in Section 2.6 and used to inoculate the medium to a desired DCW content.

Alternatively, spores from a 3 day old plate were harvested and pregerminated by incubating them in 2xYT at 30 °C in a shaking incubator at 220 RPM for 8 h. They were then harvested, washed and used to inoculate the medium.

## 2.6. Dry cell weight measurements

The DCW content of a culture was determined by taking a 5–10 ml sample, harvesting the biomass by centrifugation (4000 RPM for 10 min) and resuspending it in 5 ml $H_2O$. This suspension was applied to a pre-dried and pre-weighed filter paper under a pressure

gradient and subsequently washed with 20 ml $H_2O$. The loaded filter paper was then dried in a microwave oven (900 W, 10 min) and weighed. Dividing the mass difference by the sample volume resulted in the DCW in mg ml$^{-1}$.

## 2.7. Protein isolation from *E. coli* BL21 (DE3)

*E. coli* BL21 (DE3) was transformed with the appropriate expression plasmid and grown overnight in LB containing kanamycin. 1 L of LB with kanamycin was inoculated with 10 ml overnight culture and grown at 30 °C to an $OD_{600}$ of 0.4. IPTG was added to a fnial concentration of 1 mM and the culture was left to incubate for 2 h at 30 °C. Cells were harvested by centrifugation (5000 RPM for 10 min), resuspended in a resuspension buffer specific to the protein in question, containing protease inhibitors (Pierce Protease Inhibitor Mini Tablets (EDTA-free, Thermo Scientific), containing AEBSF, Aprotinin, Bestatin, E64, Leupeptin, Pepstatin A) and frozen at −80 °C for later use.

Cells were thawed gently and lysed using a sonic oscillator. Cell debris was removed by centrifugation (15 000 RPM for 30 min at 2 °C) and careful decanting. A Ni-sepharose column (GE Healthcare HisTrap FF crude, 1 ml) pre-equilibrated with resuspension buffer and mounted on an AKTA FPLC (GE Healthcare) was loaded with the supernatant and washed with resuspension buffer until absorption at 280 nm ($A_{280}$) of the flowthrough had returned to base levels. The protein was eluted with an imidazole gradient by applying a linearly increasing ratio of elution buffer to resuspension buffer to the column. Samples were taken when the $A_{280}$ showed a sharp increase. Imidazole was then removed from the protein preparation using a buffer exchange column (Merck Millipore Amicon Ultra-4 10K Centrifugal Filter Device) or overnight dialysis.

## 2.8. Activity assay for isocitrate lyase

Isocitrate lyase was assayed following the method of El-Mansi *et al.* (1987) by using the product glyoxylate as a substrate for lactate dehydrogenase, coupled with the oxidation of NADH in equimolar amounts. NADH absorbs at 340 nm whereas $NAD^+$ does not, allowing the reaction to be followed photospectrometrically.

The reaction mixture contained 50 mM MOPS/NaOH, pH 7.3, 5 mM $MgCl_2$, DL-isocitrate as needed, 0.2 mM NADH, pig heart lactate dehydrogenase (0.1 mg ml$^{-1}$). ICL was added last at a final concentration of 0.01 mg ml$^{-1}$. Reactions were carried out at 30 °C in a Shimadzu UV-2550 spectrophotometer with a CPS 240A cell positioner system.

Absorbance measurements were taken twice per second and the slope of the linear section in the curve was determined using the least squares method. The rate of production of glyoxylate was calculated using Eq. (2.2), derived from the Beer-Lambert law (Eq. (2.1)), in which $c$ is compound concentration (M), $A$ absorbance (AU), $\epsilon$ the molar absorbance coefficient ($\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$), $l$ the path length (cm), $r$ is the reaction rate ($\mathrm{mol\,min}^{-1}$), $V$ the reaction volume (l) and $t$ is time (min). For NADH, $\epsilon_{340nm}$ of $6220\,\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$ was used, with a path length $l$ of 1 cm. Volume $V$ was 1 ml.

$$c = \frac{A}{\epsilon \cdot l}\,(\mathrm{M}) \tag{2.1}$$

$$r = \frac{\Delta c \cdot V}{\Delta t}\,(\mathrm{mol\,min}^{-1})\,,\ \text{with}\ \Delta c = \frac{|A_2 - A_1|}{\epsilon \cdot l}\ \text{and}\ \Delta t = t_2 - t_1 \tag{2.2}$$

## 2.9. Activity assay for isocitrate dehydrogenase

Isocitrate dehydrogenase was assayed by photospectrometrically following the production of NADPH, which absorbs ultraviolet light at 340 nm with an absorption coeffecient $\epsilon_{340nm}$ of $6220\,\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$. The reaction mixture contained 50 mM MOPS/NaOH, pH 7.3, 0.5 mM $MnCl_2$ and $NADP^+$ and DL-isocitrate as needed. The reaction was started by addition of 0.1 µg IDH per ml reaction mixture and were carried out at 30 °C in a Shimadzu UV-2550 spectrophotometer with a CPS 240A cell positioner system. The Beer-Lambert law was used to calculate reaction rates using Eqs. (2.1) and (2.2).

## 2.10. Computational methods

Programming of algorithms for calculations and simulations were done using Python 2.7 (van Rossum & Drake, 2001), extended with the SciPy (Oliphant, 2007) and MatPlotLib (Hunter, 2007) packages.

## 2.11. Microscopic analysis

Samples for microscopic analysis were harvested by withdrawing 1 ml from a liquid culture. This was allowed to stand in a microcentrifuge tube for 1 min to let the mycelial pellets sink to the bottom. Most of the supernatant was then withdrawn and the pellets were resuspended in the remaining liquid. 20 µl of this suspension was then applied to a microscope slide and covered with a cover slip, which was then fixed with nail polish. The slides were frozen for later analysis.

Images were taken with a Nikon Eclipse TE2000-S microscope, fitted with a Hamamatsu C4742-95 camera at 20x magnification. Brightfield images were taken using phase contrast with a 1 ms exposure time. Fluorescence microscopy images were taken using a FITC filter with a 100 ms exposure time. Images were analysed using ImageJ (Abràmoff *et al.*, 2007).

## 2.12. Total RNA isolation from *S. coelicolor*

Total RNA was extracted using an adaptation of the Qiagen RNeasy Protect Bacteria Kit (Bucca *et al.*, 2009) (available from http://www.surrey.ac.uk/fhms/microarrays/Downloads/Protocols/index.htm). The method was adapted to allow isolation of RNA from smaller samples.

Biomass of 15 ml culture was harvested by centrifugation for 5 min at 4000 RPM and 4 °C in a pre-cooled centrifuge. The supernatant was discarded and the pellet was treated with 2 ml RNAprotect at room temperature for 5 min. The biomass was harvested as before and stored at −80 °C until further processing.

Frozen biomass was thawn on ice and resuspended in 1 ml TE buffer containing 15 mg ml$^{-1}$ lysozyme, followed by incubation at room temperature for 60 min whilst shaking. 1 ml RLT buffer (Promega RNeasy kit) and 10 µl $\beta$-mercaptoethanol were added and the samples were mixed using a vortex mixer.

The samples were extracted twice by addition of a volume of phenol-chloroform (1:1) equal to the sample volume, mixing using a vortex mixer, centrifugation at 4000 RPM for 10 min at 4 °C and transfer of the aqueous phase to a clean tube. This was repeated again, but with pure chloroform instead of a phenol-chloroform mixture. The sample volume was estimated to be 1.5 ml and 1.9 ml ethanol was added.

Promega RNeasy spin columns (one per sample) were mounted onto a collection tube and loaded with 700 µl of lysate. The columns were centrifuged at 10 000 RPM for 15 s at room temperature and the flow through was discarded. This procedure was repeated until the complete lysate was loaded onto the column.

The columns were loaded with 700 µl of Promega buffer RW1 and centrifuged at 10 000 RPM for 15 s at room temperature. They were then washed twice with 500 µl Promega buffer RPE with added ethanol as before, except that the last wash step involved a centrifugation step of 2 min. The tubes were placed in a clean 1.5 ml microcentrifuge tube and the samples were eluted with 30–50 µl RNAse free water by centrifugation at 10 000 RPM for 1 min.

## 2.13. Removal of genomic DNA from total RNA samples

Residual DNA was removed from the samples using DNaseI treatment reagents supplied with the Life Technologies (Ambion) RiboPure Bacteria kit. $1/9^{th}$ of the sample volume of DNase buffer was added, followed by 4 µl (2 U µl$^{-1}$) DNaseI. The samples were incubated at 37 °C, after which 20 % of the RNA sample volume of DNase inactivation reagent was added. The samples were mixed using a vortex mixer and incubated at room temperature for 2 min with occasional gentle mixing. The samples were centrifuged for 1 min at maximum speed and the supernatant containing the RNA was transfered to a clean microcentrifuge tube.

## 2.14. Concentration measurement and quality control of total RNA samples

Concentrations of RNA samples and sequencing libraries were measured with a Qubit 2.0 fluorometer, using a Qubit RNA HS Assay Kit (Life Technologies catalog number Q32852). In short, a working solution was prepared by dilution of Qubit RNA HS Reagent 1:200. A small volume of sample (1–20 µl) was mixed with working solution to a total volume of 200 µl in a thin-wall clear 0.5 ml PCR tube. The samples and standards were incubated at room temperature for 2 min, after which all tubes were measured. RNA concentrations in samples were calculated by the Qubit fluorometer.

Quality control of the isolated RNA was carried out on an Agilent BioAnalyzer (2100 Expert, using a Prokaryote Total RNA Pico chip, Agilent catalog number 5067-1513), following the manufacturer's instructions.

## 2.15. Removal of ribosomal RNA from total RNA samples

Ribosomal RNA was removed from total RNA samples using the Epicentre Ribo-Zero rRNA Removal Kit for Gram-Positive Bacteria (Epicentre catalog number MRZGP126) following the manufacturer's instructions. Magnetic beads were washed for each sample individually for this procedure.

## 2.16. Preparation of total RNA sequencing library

Transcriptome RNA fragmentation and barcoded library preparation were carried out using reagents supplied in the Ion Total RNA-Seq Kit v2 (Life Technologies catalog

number 4475936). Ion spheres were prepared using the Ion PGM Template OT2 200 Kit. The samples were sequenced in two sets of three on an Ion Torrent PGM using an Ion 316 Chip Kit v2 (Life Technologies catalog number 4483324).

# 3. Analysis of gene and upstream sequences of Streptomycete metabolic gene orthologues

The inability of ICL negative strains of *E. coli* to grow on acetate led to the conclusion that the glyoxylate shunt is essential for growth on this compound (Kornberg, 1966). However, the discovery that certain bacterial species, such as *Rhodobacter sphaeroides* or *Streptomyces cinnamonensis*, that lack a functional glyoxylate shunt, are still capable of growing on acetate or oil-based media shed doubt on this assertion (Akopiants *et al.*, 2006; Alber *et al.*, 2006). The recent discovery of (2S)-methylsuccinyl-CoA dehydrogenase in *R. sphaeroides* completed an alternative pathway to the glyoxylate shunt, the ethylmalonyl-CoA pathway (Erb *et al.*, 2009).

All complete streptomycete genomes contain at least one copy of *ccr*, a gene that encodes one of the enzymes of the ethylmalonyl-CoA pathway (crotonyl-CoA carboxylase / reductase). *Streptomyces avermitilis* and *Streptomyces scabies* contain two genes annotated as (putative) *ccr*, although their sequences vary significantly. The presence of a *ccr* gene in an organism was proposed to be indicative of the presence of the ethylmalonyl-CoA pathway (Erb *et al.*, 2007) and will be regarded as such in this thesis.

Not all of the sequenced streptomycete genomes contain homologues of the genes *aceR*, *aceA* and *aceB2*, meaning those strains (*S. scabies* and most notably *S. lividans*, which is closely related to *S. coelicolor*) do not contain an active glyoxylate shunt. Some other Actinobacteria, such as *Saccharopolyspora erythraea* and *Corynebacterium glutamicum*, have a gene encoding ICL but do not appear to have a *ccr* gene. The presence of two pathways that can seemingly fulfil similar functions in some *Streptomyces* strains and the absence of the glyoxylate shunt operon in others raises questions about the importance of the pathway to the metabolism of streptomycetes.

The ethylmalonyl-CoA pathway starts by the condensation of two acetyl-CoA molecules into acetoacetyl-CoA and ends, after twelve more reactions, with the production of a molecule of malate and a molecule of succinyl-CoA (Erb *et al.* (2009), Section 1.4). This provides a similar anapleurotic functionality as the glyoxylate shunt but optionally

provides the cell with various acyl-CoA intermediates, which can be used as starter or extender units for polyketide synthesis (Wilson & Moore, 2012). For example, the pathway was found to provide precursors for monensin A biosynthesis in *S. cinnamonensis* C730.1 (Liu & Reynolds, 1999), a strain in which no ICL activity could be detected (Akopiants *et al.*, 2006). It is possible that a common ancestor of the streptomycetes obtained the genes encoding the ethylmalonyl-CoA pathway, which allowed it to generate intermediates for the production of advantageous secondary metabolites. The glyoxylate shunt is potentially redundant in this environment since anapleurosis is already provided for, which is a possible reason for the absence of *aceA* from certain strains.

In this scenario, streptomycetes are not under evolutionary pressure to maintain the glyoxylate shunt genes, which as a result would accumulate mutations at a higher rate than related genes that are maintained. In order to test this hypothesis, TCA cycle, glyoxylate shunt and ethylmalonyl-CoA pathway genes from different streptomycetes were compared with respect to the single nucleotide polymorphisms (SNPs) in their sequences. A comparison of numbers of synonymous and nonsynonymous SNPs was made, as well as a more detailed analysis that calculates the purifying selection strength.

This is the effect of deleterious sequence variants being purged from a population over time due to the fitness cost that arises from them (Rocha *et al.*, 2006). In theoretical evolutionary biology it is assumed that the vast majority of nonsynonymous mutations in a protein coding sequence are deleterious to the protein's function, while synonymous mutations are assumed to be neutral in that respect. Therefore most nonsynonymous polymorphisms in protein coding sequences will in time disappear from a population due to purifying selection while synonymous polymorphisms will persist at a higher frequency (Hurst, 2002; Yang & Bielawski, 2000).

As a result of this effect, gene orthologues that contribute positively to an organism's fitness are expected to have a higher number of synonymous SNPs than nonsynonymous SNPs. The strength of the purifying selection pressure on a gene can be quantified by the value $\omega = \mathrm{d}N/\mathrm{d}S$, which is the ratio of nonsynonymous ($N$) and synonymous ($S$) SNPs, corrected for the inequality between the number of possible synonymous and nonsynonymous SNPs, and codon use (Hurst, 2002; Yang & Bielawski, 2000). The effect of this correction is that in the absence of selection pressure on a gene, a value of $\omega = 1$ is found. $\omega < 1$ indicates purifying selection while $\omega > 1$ indicates positive selection. The $\omega$ values calculated for a given set of gene orthologues are therefore a measure of the strength of purifying selection on that gene in those organisms (Hurst, 2002).

## 3.1. *aceA* orthologues do not contain a higher number of SNPs than related genes

Nucleotide sequences were obtained from the *Streptomyces* database (http://strepdb.streptomyces.org.uk/) or the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) and aligned using MEGA 6.0 (Tamura *et al.*, 2013). ClustalW for codons was selected as the algorithm using standard settings except for the codon table, which was set to "Bacterial Plastid". SNPs were extracted from the resulting alignment using a Python script. Annotation of SNPs was done with TRAMS (Reumerman *et al.*, 2013), using the *S. coelicolor* genes as reference, downloaded as a genbank file from the ncbi database.

Numbers of SNPs found are shown in Tables 3.1 to 3.3. They were divided into synonymous and nonsynonymous SNPs. multiple nucleotide polymorphisms (MNPs) occur when a codon contains multiple SNPs, which were annotated once according to their cumulative effect. Divergence was defined as the percentage of codons that had a nonsynonymous mutation in it.

| **Strain** | **SNPs** | **MNPs** | **Synonymous** | **Nonsynonymous** | **Divergence** (%) |
|---|---|---|---|---|---|
| *S. avermitilis* | 128 | 15 | 87 | 26 | 6.09 |
| *S. venezuelae* | 139 | 21 | 82 | 36 | 8.43 |
| *S. griseus* | 140 | 27 | 76 | 37 | 8.67 |
| *S. clavuligerus* | 184 | 26 | 109 | 49 | 11.48 |
| *C. glutamicum* | 374 | 122 | 128 | 124 | 29.04 |
| *M. tuberculosis* | 297 | 98 | 92 | 107 | 25.06 |
| *S. erythraea* | 279 | 95 | 87 | 99 | 23.19 |

Table 3.1.: Numbers of SNPs in *aceA* between various bacterial strains and *S. coelicolor* as a reference. SNPs are sorted into MNPs, synonymous and nonsynonymous SNPs. *S. coelicolor aceA* consists of 427 codons. Divergence was defined as the percentage of codons that had a nonsynonymous mutation in it.

When comparing the divergence between *aceA* and *ccr* genes in each strain, neither gene was clearly more divergent than the other. *aceA* contains more nonsynonymous SNPs than *ccr* in *S. venezuelae* and *S. griseus* but the reverse is true in *S. avermitilis* and *S. clavuligerus*.

Apart from the absolute number of SNPs, their locations are also important. A nonsynonymous SNP in the active site is likely to have more influence on enzymatic activity than a residue change on the surface. The enzyme activity of ICL involves displacement of an amino acid loop (shown in Fig. 3.1) upon binding of the substrate (Sharma *et al.*,

| **Strain** | **SNPs** | **MNPs** | **Synonymous** | **Nonsynonymous** | **Divergence** (%) |
|---|---|---|---|---|---|
| *S. lividans* | 2 | 0 | 2 | 0 | 0 |
| *S. avermitilis* (ccrA2) | 98 | 19 | 50 | 29 | 6.47 |
| *S. scabies* (1) | 107 | 18 | 62 | 27 | 6.03 |
| *S. venezuelae* | 124 | 21 | 67 | 36 | 8.04 |
| *S. griseus* | 129 | 17 | 82 | 30 | 7.03 |
| *S. clavuligerus* | 220 | 50 | 108 | 62 | 13.84 |
| *S. avermitilis* (ccrA1) | 458 | 190 | 63 | 204 | 45.54 |
| *S. scabies* (2) | 199 | 47 | 81 | 71 | 15.85 |

Table 3.2.: Numbers of SNPs in *ccr* between various strains and *S. coelicolor* as a reference. SNPs are sorted into MNPs, synonymous and nonsynonymous SNPs. *S. coelicolor ccr* consists of 448 codons. Divergence was defined as the percentage of codons that had a nonsynonymous mutation in it.

| **Strain** | **SNPs** | **MNPs** | **Synonymous** | **Nonsynonymous** | **Divergence** (%) |
|---|---|---|---|---|---|
| *S. lividans* | 3 | 0 | 2 | 1 | 0.14 |
| *S. scabies* | 208 | 33 | 97 | 78 | 10.54 |
| *S. avermitilis* | 254 | 52 | 116 | 86 | 11.62 |
| *S. venezuelae* | 321 | 82 | 105 | 134 | 18.11 |
| *S. clavuligerus* | 324 | 88 | 116 | 120 | 16.22 |
| *S. griseus* | 328 | 83 | 126 | 119 | 16.08 |
| *C. glutamicum* | 680 | 231 | 207 | 242 | 32.70 |

Table 3.3.: Numbers of SNPs in *idh* between various strains and *S. coelicolor* as a reference. SNPs are sorted into MNPs, synonymous and nonsynonymous SNPs. *S. coelicolor idh* consists of 740 codons. Divergence was defined as the percentage of codons that had a nonsynonymous mutation in it.

2000). Among these residues, only one was found to contain any nonsynonymous SNPs. This SNP was the same (residue 198) in all three strains that had it, which were the non-*Streptomyces* strains. The SNP caused a valine to be replaced by an isoleucine. Both are nonpolar aliphatic side chains that differ only in side chain length.

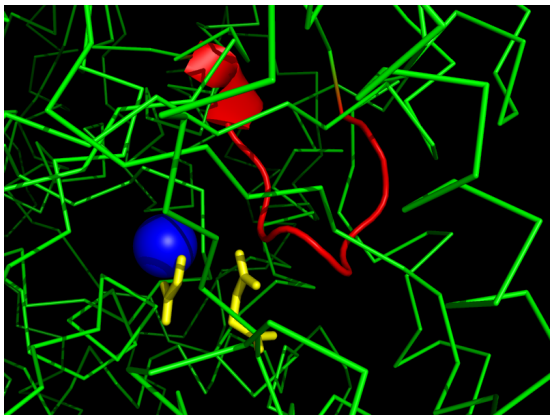


Figure 3.1.: The flexible loop of *M. tuberculosis* ICL in red with an $Mg^{2+}$ ion in blue and substrate in yellow. PDB ID: 1F8I (Sharma *et al.*, 2000)

Divergence values for both genes are slightly lower than that of *idh* in all strains. The observation that *aceA* orthologues do not show elevated numbers of SNPs compared to others and only one polymorphism is present in residues involved in the active site leads to the rejection of the hypothesis that isocitrate lyase is redundant. The gene is maintained instead and so must still serve an important function in streptomycete metabolism.

## 3.2. Purifying selection pressure on streptomycete genes encoding TCA-cycle enzymes, *aceA* and *ccr*

In order to compare the purifying selection strength on *aceA* with that on related genes, the $dN/dS$ ratios for pairs of genes from *S. coelicolor*, *Streptomyces scabies*, *Streptomyces avermitilis*, *Streptomyces venezuelae*, *Streptomyces clavuligerus* and *Streptomyces griseus* were calculated. The genes used in this study were drawn from the TCA cycle, glyoxylate shunt and ethylmalonyl-CoA pathway: *Sco2736* (*gltA*), *Sco5999* (*sacA*), *Sco7000* (*idh*), *Sco4808* (*sucC*), *Sco4809* (*sucD*), *Sco5044* (*fumB*), *Sco5042* (*fumC*) and *Sco4827* (*mdh*)), *Sco0982* (*aceA*) and *Sco6473* (*ccr*). The mean of the calculated ratios for *aceA* was then compared to those of the other genes to test whether the purifying selection on the glyoxylate shunt is relaxed, indicating redundancy of the metabolic pathway.

*Streptomyces lividans* was omitted from this analysis because the difference between its gene sequences and those of *S. coelicolor* were too few to result in a reliable ratio, d$N$ often equalling zero. It has also been noted that for very recently diverged genomes, d$N$/ d$S$ is more dependent on the time since divergence more than less closely related genomes (Rocha *et al.*, 2006). This constituted another reason to omit the *S. lividans* sequences, as calculating and compensating for time since divergence of *S. coelicolor* and *S. lividans* was beyond the scope of this study.

Gene sequences were obtained from the *Streptomyces* genome database (http://strepdb.streptomyces.org.uk/) and aligned using MEGA6 (Tamura *et al.*, 2013) as described in Section 3.1. The output was saved as a fasta file and consecutively converted to the "interleaved" format of the PAML package (Yang, 2007). The d$N$ and d$S$ values were calculated using the PAML package's program yn00. Ratios between the values were calculated using Microsoft Excel and statistical tests were performed using OriginPro 9. Normality tests used were OriginPro's Shapiro-Wilk, Lilliefors, Anderson-Darling and D'Agostino-K squared tests.

| **Gene** | **Mean d$N$/ d$S$** | **Std. Error of the Mean** |
|---|---|---|
| *gltA* | 0.02568 | 0.00256 |
| *sacA* | 0.02488 | 0.00114 |
| *idh* | 0.05256 | 0.00216 |
| *sucC* | 0.07545 | 0.00269 |
| *sucD* | 0.04087 | 0.00377 |
| *fumB* | 0.03088 | 0.00186 |
| *fumC* | 0.03628 | 0.00167 |
| *mdh* | 0.03405 | 0.00212 |
| *aceA* | 0.02466 | 0.003 |
| *ccr* | 0.02264 | 0.00249 |

Table 3.4.: d$N$/ d$S$ ratios of genes encoding enzymes of the TCA cycle, *aceA* and *ccr*. Ratios were calculated between orthologues in *Streptomyces* species and the mean and standard error of the mean were determined from them.

None of the calculated d$N$/ d$S$ ratios, shown in Table 3.4, are close to unity. This was expected, as the TCA cycle performs an important role in metabolism so species are under strong selective pressure to maintain the genes encoding its enzymes. The values for *aceA* and *ccr* are lower than those for most TCA cycle genes and are also very close to each other, which suggests that they are both under a similarly strong purifying selection pressure. The ratios were normally distributed, with *sucC* being the single

outlier (Fig. 3.2).

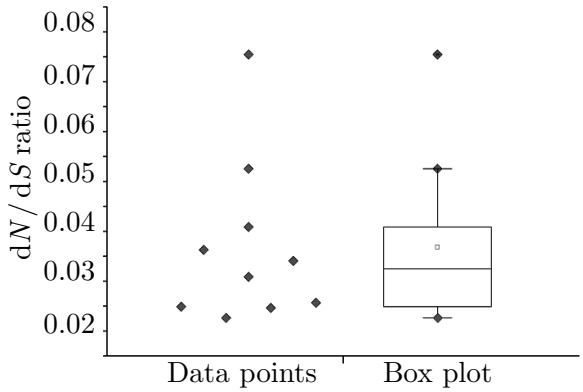



Figure 3.2.: Box plot of the $dN/dS$ ratios shown in Table 3.4. This plot reveals the value for *sucC* as an outlier in an otherwise normally distributed data set (calculated using Origin Pro)

### 3.2.1. The $dN/dS$ ratio calculation is sensitive to potential gene redundancy

In order to verify the validity of this method, $dN/dS$ ratios were calculated for *S. coelicolor* genes annotated as "pseudogenes". These genes are not expected to be subject to purifying selection because they no longer encode a functional protein. Pseudogenes can contain frameshifts due to insertions or deletions, which cause problems in this type of analysis. An insertion in one sequence could also be a deletion in another. To allow for this possibility, alignment was done using the regular ClustalW algorithm in MEGA6, using standard settings, rather than the ClustalW for codons algorithm which only allows gap sizes of multiples of three nucleotides.

Annotation of pseudogenes is often problematic and results are highly influenced by the method used (Karro *et al.*, 2007). Possibly because of this, not all features marked as a pseudogene are likely to be one and a few, such as *Sco0306*, *Sco0515* and *Sco1974*, indeed had a low $dN/dS$ ratio indicating purifying selection pressure on this sequence. Other features annotated as pseudogenes did not have any orthologues among Streptomycetes, which is a requirement for this analysis to be carried out. Finally, suitable values were obtained for genes *Sco0205*, *Sco0448* and *Sco4010*.

*Sco0205* is annotated as a "possible pyruvate formate-lyase ativating enzyme pseudogene". It only has orthologues in *S. venezuelae* and *S. scabies*, and has a mean $dN/dS$ ratio of 1.1809 with a standard deviation of 0.217. This value, although obtained from only a small number of strains clearly indicates neutral selection, since it is very close

to 1. The *S. coelicolor* gene contains two frameshift mutations (as identified from the alignment) and, with GTC as the first three bases, no start codon. These properties mark it as a pseudogene and the neutral selection, as indicated by the $dN/dS$ ratio, reflects that.

*Sco0448* is annotated as a "putative transmembrane efflux protein pseudogene" and only has one orthologue in *S. lividans*. This gene pair has a $dN/dS$ ratio of 0.268, but the single orthologue in a closely related strain makes this value statistically questionable. However, the value is clearly higher than the values found for the TCA cycle genes and selection pressure on this gene is not neutral as with *Sco0205*. This means that *Sco0448* is not a pseudogene, but purifying selection is somewhat relaxed compared to more important genes.

*Sco4010* is a "possible secreted protein". It has orthologues in all Streptomycetes for which a full genome sequence has been finished. It was proposed to be a pseudogene due to its homology to *Sco0762* (a protease inhibitor precursor) despite not being expressed under the same conditions (Kato *et al.*, 2005), although it not annotated as such in the genome database. The alignment of all orthologues results in many gaps, leaving a consensus sequence of only 276 nucleotides. A mean $dN/dS$ ratio of 1.413 with a standard deviation of 0.265 was obtained, which indicates neutral selection. This ratio, like that of *Sco0205*, is close to 1 which means *Sco4010* is likely to be a pseudogene.

The data presented here show that the $dN/dS$ ratios of a set of gene orthologues is a sensitive measure for a gene's importance, as it indicates neutral selection on two putative pseudogenes and a relaxed selection pressure on a third. The values for genes encoding TCA cycle enzymes and those for *aceA* and *ccr* clearly show that they are under a strong selection pressure and that there is no indication of reduced selection pressure on the glyoxylate shunt genes. Based on these results the hypothesis that the glyoxylate shunt is a redundant pathway was rejected.

## 3.3. Regulatory sequences upstream of *aceA* and *ccr*

After finding that the glyoxylate shunt genes appear to form a functional operon and given the overlap in proposed function between the glyoxylate shunt and the ethylmalonyl-CoA pathway, the possibility that expression of the two operons is at least partially under the control of a shared regulatory protein was investigated.

In order to explore this hypothesis, sequences located upstream of genes *aceA*, *aceB1* and *ccr* were compared using the MEME suite (Bailey *et al.*, 2009), a set of programs capable of discovering and analysing potential regulatory motifs. The sequences were

found to contain a shared palindrome, which could be the binding site of a transcriptional regulator.

A search for similar sites in the *S. coelicolor* genome sequence with FIMO 4.9.1 (Bailey *et al.*, 2009) revealed a list of genes that are possibly co-regulated. Some of these sites are conserved in other streptomycetes, although this conservation could not definitively be correlated with the presence of the glyoxylate shunt genes.

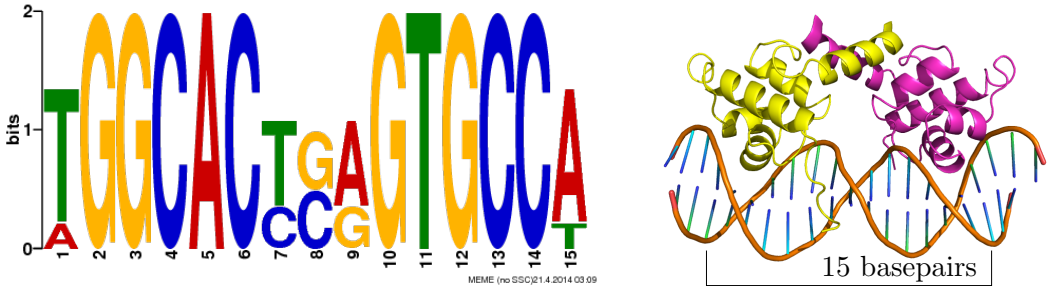
### 3.3.1. Upstream sequences of *aceA*, *aceB1* and *ccr* share a palindromic motif

In order to discover any potential shared regulatory motifs, the intergenic sequences upstream of genes *aceA*, *aceB1* and *ccr* in *S. coelicolor* were obtained from the *Streptomyces* database and submitted to MEME 4.9.1 (accessible via http://meme.nbcr.net/), using the following settings: Options Configuration: Advanced, Check Reverse Complement: yes, Check for Palindromes: yes, Expected Motif Distribution: Any Number of Repetitions. The genes *aceA* and *ccr* were chosen because they are the first genes in their apparent operons.

The analysis yielded a 15 bp palindromic motif, shown in Fig. 3.3a. Regulatory proteins containing helix-turn-helix (HTH) domains are known to bind palindromic sequences, often as a homodimer (Huffman & Brennan, 2002). An example of a palindromic sequence bound by a homodimer containing HTH domains is the phage lambda repressor, shown in Fig. 3.3b (Beamer & Pabo, 1992). Two $\alpha$-helices interact with the major groove of the DNA, astride a minor groove that makes up the origin of symmetry of the binding motif.

The motif found in this analysis shows a high degree of conservation, with only the three base pairs in the centre and the outermost two being varied. The presence of this motif suggests that transcriptional regulation of these genes overlaps at least partially.

The presence of this motif upstream of *aceB1* is particularly interesting, since it was recently shown that the product of this gene is involved in allantoin metabolism (Navone *et al.*, 2014). The proposed pathway contains glyoxylate as an intermediate, which can be used by either AceB1 to produce malate or by Gcl to produce tartronate semialdehyde. Both proteins were present in higher amounts in *S. coelicolor* when allantoin was used as nitrogen source than when casamino acids were included. Although this indicates that the function of AceB1 is different from the proposed function of the glyoxylate shunt, the presence of this shared motif upstream of *aceB1* may point towards a more complex regulatory system for this gene than one based solely on the presence of allantoin.




(a) Palindromic motif discovered in upstream sequences of *aceA*, *aceB1* and *ccr* in *S. coelicolor*.

(b) Crystal structure of the Lambda repressor in complex to the operator sequence. PDB ID: 1LMB.

Figure 3.3.: Palindromic motif found upstream of *aceA*, *aceB1* and *ccr* (Fig. 3.3a) and the phage Lambda repressor bound to such a palindromic motif (Fig. 3.3b)

### 3.3.2. Presence of the motif in the genomes of Streptomycetes

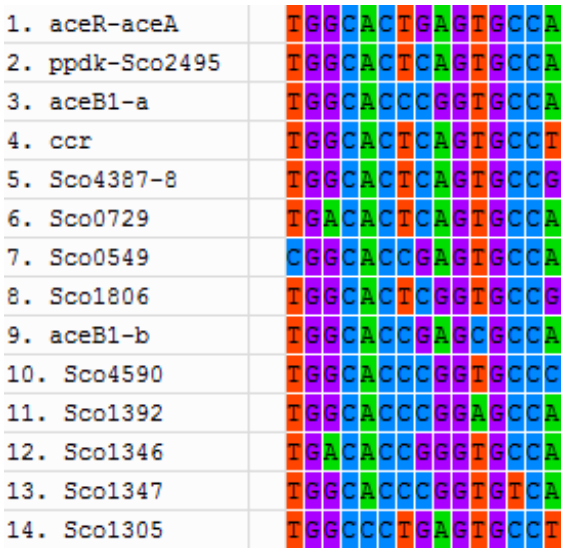
After discovery of this potential regulatory motif, the question of whether it was found in the upstream sequences of other *S. coelicolor* genes and to what extent it was conserved among other streptomycetes, especially with regards to presence or absence of the glyoxylate shunt genes, presented itself. To address this, the genome sequences of *S. coelicolor*, *S. avermitilis*, *S. clavuligerus*, *S. griseus*, *S. lividans*, *S. scabies* and *S. venezuelae* were searched for occurrences of the palindromic motif described in Section 3.3.1 using FIMO 4.9.1 (accessible via http://meme.nbcr.net/). The genome sequences were obtained from the ncbi database (accessible via http://www.ncbi.nlm.nih .gov/genome/) in fasta format. For *S. clavuligerus*, the large plasmid pSCL4 (Alvarez-Álvarez *et al.*, 2014) was searched as well since it contains the glyoxylate shunt operon.

The results of the FIMO search are shown in Table 3.5 and an alignment of the discovered motifs is shown in Fig. 3.4. Apart from the genes that were used to discover the motif (*aceA*, *aceB1* and *ccr*), a few hits stand out. *Sco2494* encodes pyruvate phosphate dikinase (PPDK), an enzyme that consumes pyruvate and produces phosphoenol pyruvate, thereby catalysing the first dedicated step of the gluconeogenic process. Since growth on acetate or fatty acids requires gluconeogenesis, potential overlap in regulation between this gene and the glyoxylate shunt and the ethylmalonyl-CoA pathway was expected. The motif was not detected upstream of other genes involved in gluconeogenesis however, meaning its importance in regulation of gluconeogenesis remains unclear.

The product of *Sco1346* (*fabG3*) has 43 % similarity to PhaB, the protein that was proposed to catalyse the second step in the ethylmalonyl-CoA pathway (Erb *et al.*, 2009). Presence of this potential regulatory motif provides additional support for this

| Strand | p-value | q-value | Downstream gene | Distance |
|---|---|---|---|---|
| + | 1.29e-09 | 0.00375 | *Sco0982*, *aceA*: isocitrate lyase | -240 |
| - | 1.29e-09 | 0.00375 | *Sco0981*, *aceR*: DNA-binding protein | -128 |
| + | 1.29e-09 | 0.00375 | *Sco2495*: hypothetical protein | -210 |
| - | 1.29e-09 | 0.00375 | *Sco2494*: pyruvate phosphate dikinase | -259 |
| - | 4.27e-09 | 0.00829 | *Sco6243*, *aceB1*: malate synthase | -23 |
| - | 6.85e-09 | 0.00997 | *Sco6473*, *ccr*: crotonyl-CoA carboxylase / reductase | -168 |
| + | 2.55e-08 | 0.0321 | *Sco4388*: citrate synthase 2 | -131 |
| - | 2.76e-08 | 0.0321 | *Sco4387*: pyridoxamine oxidase | -73 |
| + | 5.34e-08 | 0.054 | *Sco0729*: hypothetical protein | -101 |
| - | 8.32e-08 | 0.054 | *Sco0549*: acyl carrier protein | -550 |
| - | 8.32e-08 | 0.054 | *Sco1806*: ABC transporter ATP-binding protein | -1616 |
| - | 9.28e-08 | 0.054 | *Sco6243*, *aceB1*: malate synthase | -164 |
| - | 1.56e-07 | 0.067 | *Sco4590*: hypothetical protein | -470 |
| - | 1.57e-07 | 0.067 | *Sco1392*: hypothetical protein | -369 |
| - | 1.75e-07 | 0.0704 | *Sco1346*: 3-oxoacyl-ACP reductase | -101 |
| + | 1.97e-07 | 0.0766 | *Sco1347*: hypothetical protein | -151 |
| + | 2.17e-07 | 0.0788 | *Sco3705*: ABC transporter membrane subunit | -725 |

Table 3.5.: Occurrences of the palindromic motif in the *S. coelicolor* genome, with q-value (false discovery rate) $< 0.1$. The distance column shows the number of nucleotides between the motif and the start codon of the downstream gene. Instances which were not upstream of any gene were omitted, 15 in total.




(a) Multiple alignment (ClustalW) of FIMO hits listed in Table 3.5.


(b) Weblogo of ClustalW alignment of Fig. 3.4a.

Figure 3.4.: ClustalW alignment of motifs found in the FIMO search of the *S. coelicolor* genome (Fig. 3.4a). Figure 3.4b shows the consensus sequence.

connection.

Genome sequences of other streptomycetes were searched for the same motif in order to investigate if it is conserved. The results are shown in Table 3.6. Conservation differs widely between genes and strains. The motif is well conserved in the intergenic sequence between *Sco0981* and *Sco0982* for all strains that contain a glyoxylate shunt. The q-values are very similar for *Sco2494*, except in *S. clavuligerus* and *S. scabies*. The q-value is the expected false discovery rate when the occurrence and all occurrences with an equal or better score are accepted as positive. At a q-value threshold of 0.1, 10 % of hits is a false positive result.

Similar differences can be found for other genes in the table for different organisms. This variation could represent differences in life style or habitat. For example, it could be speculated that *S. scabies*, which is a plant pathogen that infects potatoes, might have a different requirement for gluconeogenesis and therefore has a different regulatory mechanism.

The sequence upstream of *Sco1346* was not conserved strongly. The product of this gene is possibly shared with other metabolic pathways, such as polyhydroxybutyrate biosynthesis (Alber *et al.*, 2006; Erb *et al.*, 2009). Diverging needs for other metabolic pathways of which this enzyme is part could be the reason conservation of this regulatory motif is weak.

## 3.4. Conclusions

The results from this chapter show that, despite overlap in proposed functions between the glyoxylate shunt and the ethylmalonyl-CoA pathway, gene sequences of both pathways are subject to selection pressure of similar strength to genes with a related function. It was concluded that, despite absence from certain streptomycete genomes, the glyoxylate shunt genes are not redundant.

Analysis of sequences upstream of *aceA*, *aceB1* and *ccr* revealed a highly conserved 15 bp palindromic motif which is a potential binding site for a regulatory protein. The motif was found upstream of several other genes, a few of which (*Sco2494* and *Sco1346*) are functionally related. Many sites were conserved among other streptomycete genomes as well, which provides further support for this motif as a regulatory sequence. Experimental evidence, such as from transcriptomic profiling by microarrays or RNA sequencing combined with isolation of the regulatory protein, is needed to show the importance of this motif.

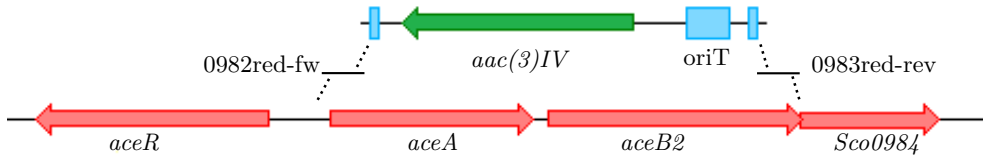
| **Downstream gene** | | **q-values** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| (*S. coelicolor*) | | (Sco) | (Sav) | (Sclav) | (Sgr) | (Sliv) | (Sscab) | (Sven) |
| *Sco0982*: | *isocitrate lyase* | 0.00375 | 0.0071 | 0.00414 | 0.0385 | - | - | 0.00639 |
| *Sco0981*: | *DNA-binding protein* | 0.00375 | 0.0071 | 0.00414 | 0.0385 | - | - | 0.00639 |
| *Sco2494*: | *pyruvate phosphate dikinase* | 0.00375 | 0.0071 | 0.242 | 0.00719 | 0.00735 | 0.27 | 0.056 |
| *Sco6243*: | *malate synthase* | 0.00829 | 0.0431 | 0.0431 | 0.276 | 0.0122 | 0.217 | 0.0871 |
| *Sco6473*: | *crotonyl-CoA carboxylase / reductase* | 0.00997 | 0.137 | 0.137 | 0.0309 | 0.013 | 0.0293 | 0.108 |
| *Sco4388*: | *citrate synthase 2* | 0.0321 | 0.0431 | 0.0431 | 0.239 | 0.0394 | 0.0315 | 0.279 |
| *Sco0729*: | conserved hypothetical protein | 0.054 | - | 0.0416 | 0.0385 | 0.0588 | - | 0.00363 |
| *Sco0549*: | *acyl carrier protein* | 0.054 | - | - | - | 0.0588 | - | - |
| *Sco6243*: | *malate synthase* | 0.054 | 0.137 | 0.137 | 0.276 | 0.0588 | 0.217 | 0.297 |
| *Sco1346*: | *3-oxoacyl-ACP reductase* | 0.0704 | 0.132 | 0.132 | 0.124 | 0.08 | 0.141 | 0.108 |

Table 3.6.: Conservation of the palindromic motif upstream of selected genes among Streptomycetes, expressed in q-values. Species names are abbreviated as follows. Sco: *S. coelicolor*, Sav: *S. avermitilis*, Sclav: *S. clavuligerus*, Sgr: *S. griseus*, Sliv: *S. lividans*, Sscab: *S. scabiei*, Sven: *S. venezuelae*.

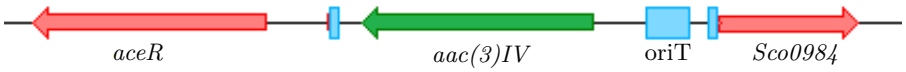
# 4. Glyoxylate shunt genes are not essential for growth on Tween 40 as carbon source

Earlier studies showed expression of glyoxylate shunt genes, specifically *aceA*, in *S. coelicolor* when grown on minimal medium using Tween 40 as sole carbon source (Chapman, 1994). In order to investigate whether the glyoxylate shunt is essential for growth in these conditions, a double knockout mutant of genes *aceA* and *aceB2* was made (RAR101) and growth in liquid and solid mimimal medium containing Tween 40 as sole carbon source was compared to that of parental strain M145 and complemented strain RAR102. These experiments showed that the glyoxylate shunt genes are not essential for growth on fatty acids, as their absence caused no impairment of growth under the tested conditions.

## 4.1. PCR targeted mutagenesis of glyoxylate shunt genes




(a) ReDirect primers consisted of a gene flanking sequence and a cassette priming sequence (homology indicated by dashed lines). A PCR product was inserted into cosmid 11D02 by homologous recombination.

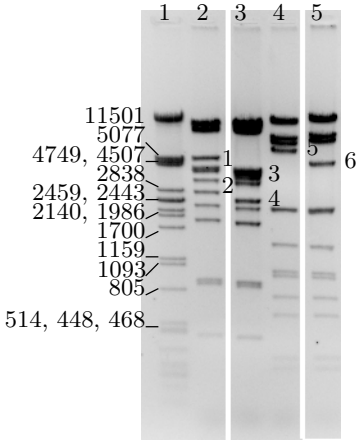



(b) The glyoxylate shunt genes were replaced with the pIJ773 cassette. Due to the primer design, the start- and stop codons were not removed.

Figure 4.1.: Two phases of gene deletion in *S. coelicolor*. Mutations were first performed on cosmid 11D02, which was then introduced into the target strain by conjugation. After a double cross over event the genes were removed from the *S. coelicolor* chromosome.

*4. Glyoxylate shunt genes are not essential for growth on Tween 40 as carbon source*

In order to investigate the role of the glyoxylate shunt in fatty acid utilisation, genes *aceA* and *aceB2* were knocked out jointly using PCR-targeted mutagenesis. A cassette containing the *aac(3)IV* gene conferring apramycin resistance and an *oriT* sequence enabling conjugal transfer, flanked by FLP-recognition sites, was fitted with sequences flanking the target genes using PCR (see Section 2.4, Fig. 4.1a and (Gust *et al.*, 2003)). This cassette was used to transform BW25113/pIJ790 (made competent while expressing λRED) carrying cosmid Sc11D02 (Redenbach *et al.*, 1996), which contains the targeted genes. λRED-mediated homologous recombination between the cosmid and the linear cassette resulted in a mutated cosmid that had *aceA* and *aceB2* replaced by the knockout cassette (resulting in Sc11D02Δ*aceA–aceB2*). The mutant cosmid was verified by restriction analysis (Fig. 4.2) and introduced into *S. coelicolor* M145 by conjugation from ET12567/pUZ8002. An apramycin$^r$ & kanamycin$^s$ phenotype in *S. coelicolor* indicated a double crossover event and successful gene replacement (see Fig. 4.1b), resulting in strain RAR101.




(a) Gel electrophoresis of digested cosmids.

| PstI | | NcoI | |
|---|---|---|---|
| Sc11D02 (2) | Δ*aceA–aceB2* (3) | Sc11D02 (4) | Δ*aceA–aceB2* (5) |
| 11887 | 11887 | 15106 | 15106 |
| 9616 | 9616 | 8319 | 8319 |
| [1]**5100** | 4188 | 7500 | 7500 |
| 4188 | 3981 | [5]**6329** | [6]**4730** |
| [2]**3981** | [3]**3728** | 2339 | 2339 |
| 3393 | 3393 | 2315 | 2315 |
| 2793 | [4]**2566** | 1486 | 1486 |
| 2357 | 2357 | 1074 | 1074 |
| 1894 | 1894 | 1025 | 1035 |
| 921 | 921 | 777 | 777 |
| 884 | 881 | 616 | 616 |
| 441 | 441 | 314 | 314 |
| | | 255 | 255 |

(b) Expected band sizes (bp) for digested cosmids.

Figure 4.2.: Restriction analysis of cosmid Sc11D02 and Sc11D02Δ*aceA–aceB2*. Figure 4.2a: Agarose gel electrophoresis of digestions. Samples were run on the same gel but in different lanes. Lane 1: λ/PstI, Lane 2: Sc11D02/PstI, Lane 3: 11D02Δ*aceA–aceB2*/PstI, Lane 4: Sc11D02/NcoI, Lane 5: 11D02Δ*aceA–aceB2*/NcoI. Figure 4.2b: Expected band sizes. Diagnostic bands in the table and on the gel are labelled with corresponding numbers. Numbers in headers indicate corresponding lanes on the gel in Fig. 4.2a

## 4.2. Complementation of gene knockout

The knocked out genes were complemented by reintroducing the wild type sequence into the genome using plasmid pMS82, which integrates into the $\phi$BT1 integration site of *Sco4848*. It also carries a hygromycin resistance gene and an origin of transfer, which enables conjugation into *S. coelicolor* (Gregory *et al.*, 2003).

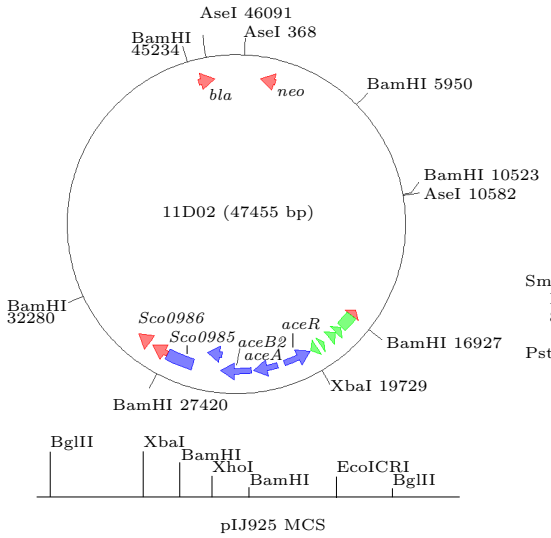
To construct the complementation vector, cosmid Sc11D02 was digested with XbaI, BamHI and AseI. The desired fragment containing genes *aceA* and *aceB2* was flanked by XbaI and BamHI sites as a result and could be ligated into pIJ925, also cut with XbaI and BamHI. The extra digestion of Sc11D02 with AseI limited the occurence of false positive colonies by cutting inside the *bla* gene, which is a selection marker on both pIJ925 and Sc11D02 (Fig. 4.3a). Ligation of the two resulted in plasmid pRiz110 (Fig. 4.3b).

After confirmation of the construct by restriction analysis (Fig. 4.4), pRiz110 was digested with XbaI and EcoICRI and pMS82 was digested with SpeI and EcoRV (Fig. 4.3c). This resulted in blunt ends on one side and compatible staggered ends on the other. The fragment and vector were isolated from an agarose gel and ligated together, resulting in plasmid pShunt (Fig. 4.3d), which was verified by restriction analysis (Fig. 4.5). Conjugation of pShunt into strain RAR101 resulted in strain RAR102.

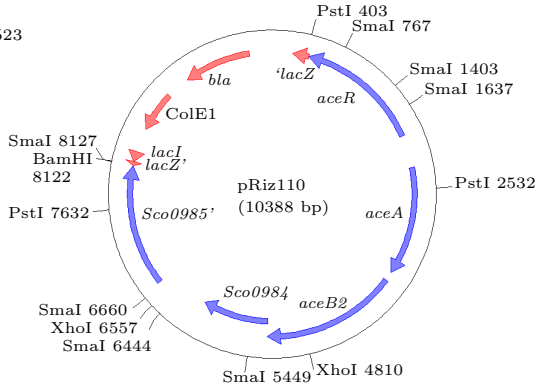
## 4.3. Verification of correct insertion of knockout and complementation constructs

Gene deletion and complementation were verified by PCR. Primers were designed such that one set annealed when the wild type sequence was present ($L_1$ and $R_1$ in Fig. 4.6a) and another set annealed when the genes were replaced by the deletion cassette ($L_2$ and $R_2$ in Fig. 4.6b). Primer sets were used on gDNA of M145, RAR101 and RAR102 with previously verified cosmids as control. The PCR mixtures contained 5 µl 5x MyTaq buffer, 1 µl of each primer (20 pmol µl$^{-1}$ stock), 1 µl DMSO, 1 U Taq polymerase, 1 µl of template and 15 µl $dH_2O$. Cycling conditions were 95 °C for 5 min, then 30 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, followed by a final incubation at 72 °C for 7 min.

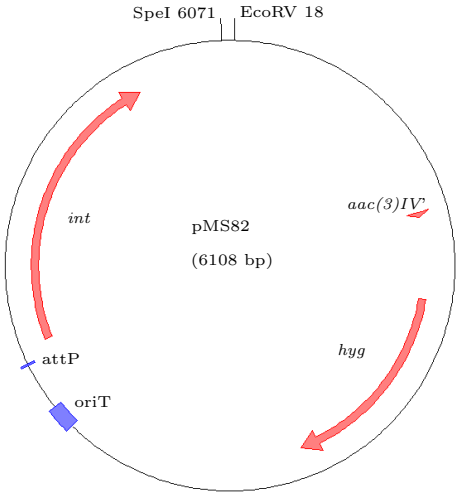
Gel photos for this experiment are shown in Fig. 4.7. The primer pairs for $L_1$ and $R_1$, $L_2$ and $R_2$ are specific for the wild type and mutant situation, respectively, as shown by the gel photos for the cosmid reactions (Sc11D02 and 11D02Δ*aceA–aceB2*). Genes *aceA* and *aceB2* were deleted from the M145 chromosome and replaced by the resistance



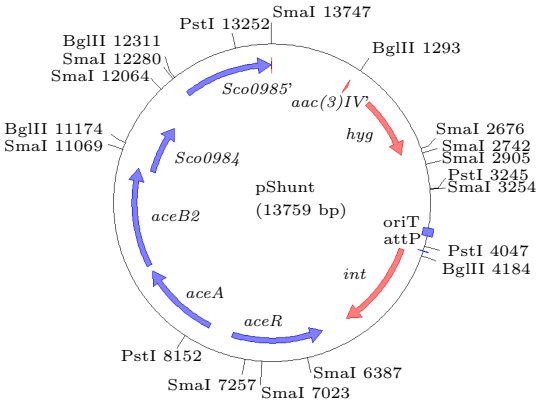

(a) Cosmid Sc11D02 (top) and the MCS of pIJ925 (bottom).




(b) pRiz110; the fragment containing glyoxylate shunt genes from Sc11D02 cloned into pIJ925.

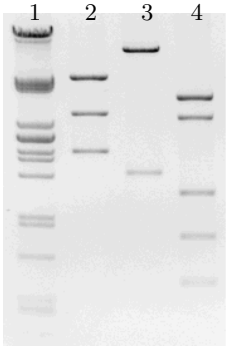



(c) Integrating vector pMS82 (Gregory *et al.*, 2003)




(d) Complementation construct pShunt.

Figure 4.3.: Plasmids involved in construction of complementation vector pShunt. Indicated restriction sites were used in cloning or checking of correctness.

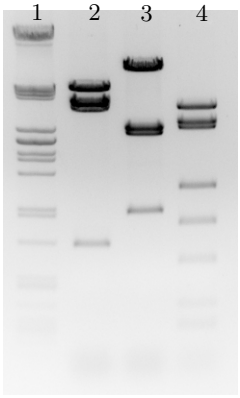



(a) Gel electrophoresis of digested pRiz110.

| **PstI** (2) | **XhoI** (3) | **SmaI** (4) |
|---|---|---|
| 5100 | 8641 | 3812 |
| 3159 | 1747 | 3028 |
| 2129 | | 1467 |
| | | 995 |
| | | 636 |
| | | 234 |
| | | 216 |

(b) Expected band sizes (bp) for digested pRiz110.

Figure 4.4.: Restriction analysis of plasmid pRiz110. Figure 4.4a: Agarose gel electrophoresis of digested plasmid. Ladder in lane 1 is $\lambda$/PstI. Figure 4.4b: Expected band sizes for digestions of pRiz110. Numbers in the headers indicate corresponding lanes on the gel in Fig. 4.4a.



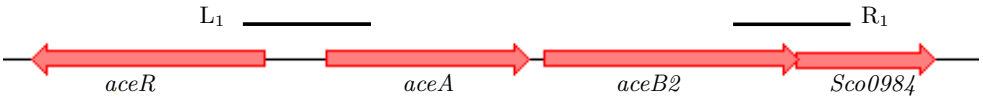

(a) Gel electrophoresis of digested pShunt.

| **PstI** (2) | **BglII** (3) | **SmaI** (4) |
|---|---|---|
| 5100 | 6990 | 3812 |
| 4105 | 2891 | 3122 |
| 3752 | 2741 | 2688 |
| 802 | 1137 | 1467 |
| | | 995 |
| | | 636 |
| | | 349 |
| | | 234 |
| | | 216 |
| | | 163 |
| | | 66 |

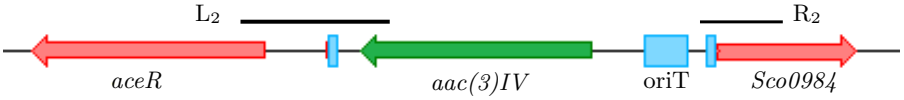
(b) Expected band sizes (bp) for digested pShunt.

Figure 4.5.: Restriction analysis of plasmid pShunt. Figure 4.5a: Agarose gel electrophoresis of digested plasmid. Ladder in lane 1 is $\lambda$/PstI. Figure 4.5b: Expected band sizes for digestions of pShunt. Numbers in the headers indicate corresponding lanes on the gel in Fig. 4.5a.

cassette, as shown in the gel for RAR101 gDNA by absence of $L_1$ and $R_1$ products and presence of $L_2$ and $R_2$ products. The presence of all four PCR products in RAR102 shows that both the wild type and mutant sequences are present, indicating that the genes were successfully replaced but not at the original location.




(a) Glyoxylate shunt genes with the approximate location of PCR products $L_1$ and $R_1$.




(b) Glyoxylate shunt genes replaced by deletion cassette, with the approximate location of PCR products $L_2$ and $R_2$.

Figure 4.6.: Approximate locations of PCR products used to verify successful deletion of glyoxylate shunt genes.

## 4.4. Genes *aceA* and *aceB2* are not essential for growth on Tween 40

Strains M145, RAR101 and RAR102 were grown in minimal medium (described in Section 2.3.2), with a Tween 40 concentration of $10\,g\,l^{-1}$. The inoculum was prepared by overnight growth in YEME (Section 2.5) and inoculated to a final DCW of $0.05\,mg\,ml^{-1}$. Single cultures of each strain were grown simultaneously, but due to variation in sampling times no statistical analysis could be done. Semi-logarithmic plots of DCW data are shown in Fig. 4.8.

The growth rates were calculated separately from the data shown in Figs. 4.8a and 4.8b. Both the specific growth rates and the biomass yields are highly similar in these cultures, showing that deletion of the genes does not inhibit growth of *S. coelicolor* in these conditions.

Strains M145, RAR101 and RAR102 were also grown on minimal medium agar (as decribed in Section 2.3) containing Tween 40 as carbon source. Photos were taken at regular intervals but no difference between M145 and RAR101 (Fig. 4.9) or RAR101 and RAR102 (not shown) was observed.

Neither the agar plates, the graphs shown in Fig. 4.8 or their associated $\mu$-values show a clear effect of the loss of genes *aceA* and *aceB2* on the growth rate or biomass yield of *S. coelicolor* on Tween 40. Despite clear evidence that the glyoxylate shunt genes

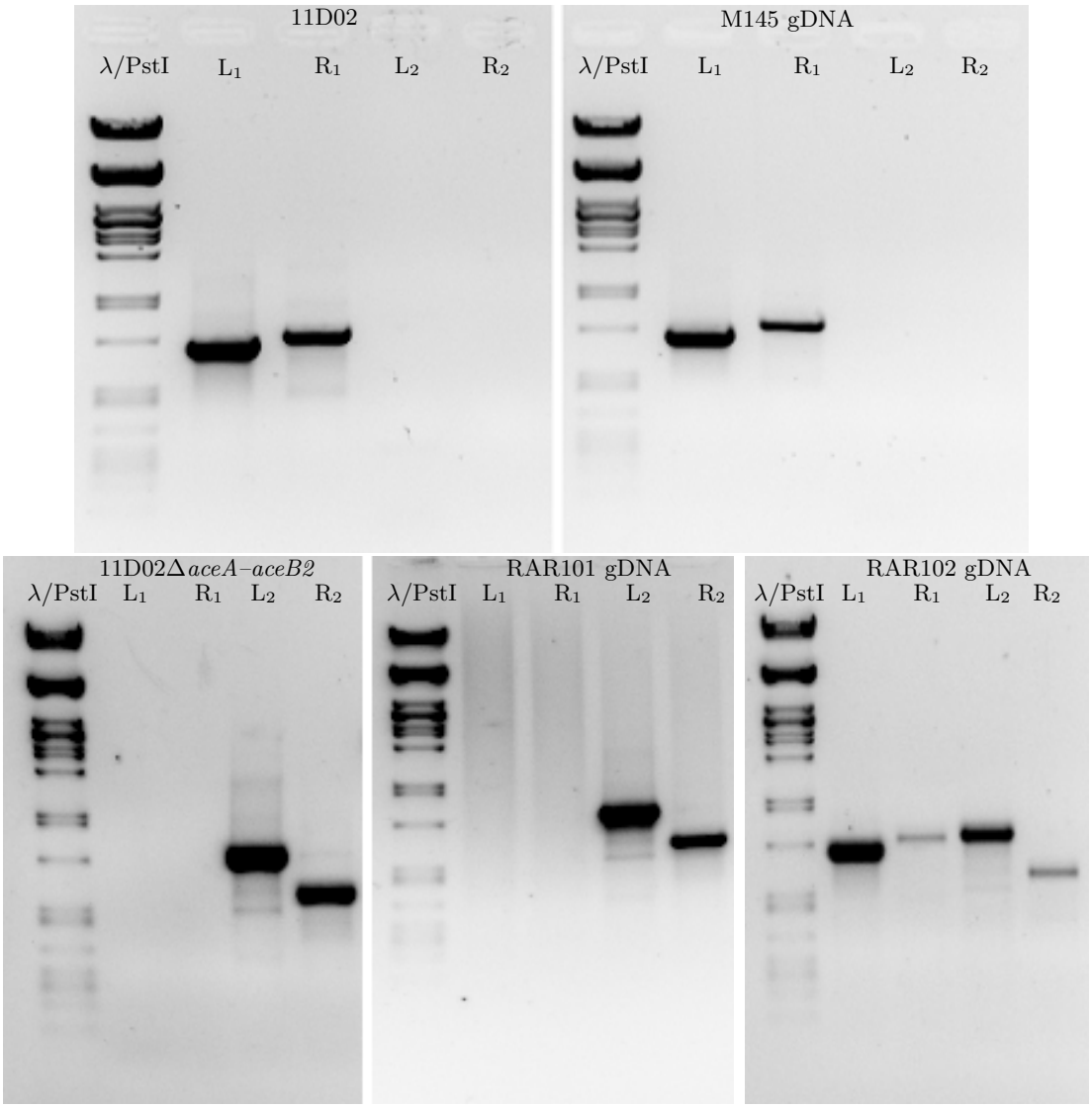
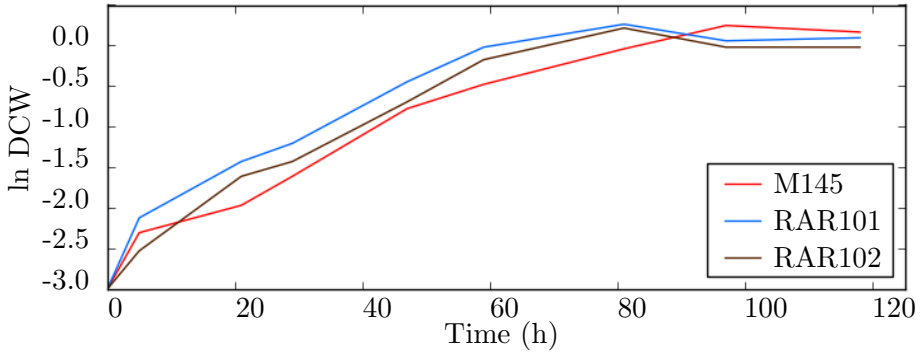
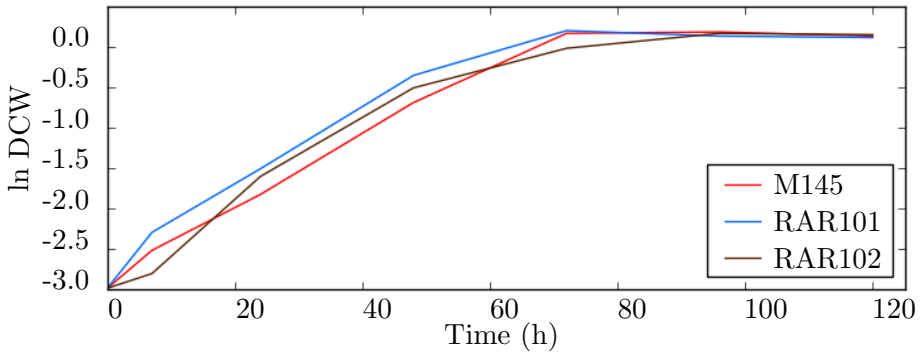



Figure 4.7.: Gene deletion and complementation were checked by PCR and visualised by agarose gel electrophoresis. Products for $L_1$ and $R_1$ primer sets were only found when the wild type sequence of Fig. 4.6a was present while products for $L_2$ and $R_2$ primer sets were found where the glyoxylate shunt genes were replaced with the knockout cassette, as shown in Fig. 4.6b. Cosmids Sc11D02 and 11D02Δ*aceA–aceB2*, previously checked by restriction analysis, were used as controls. Complemented mutant strain RAR102 was positive for all primer sets.

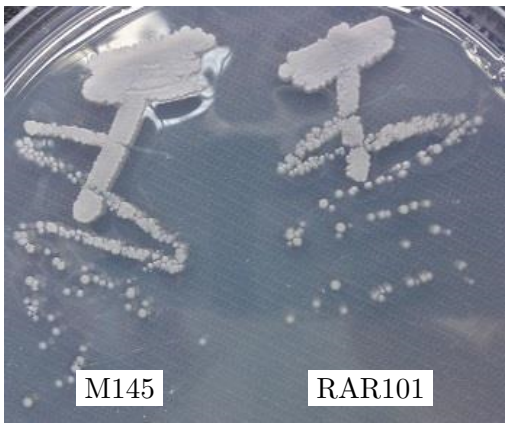



(a) Specific growth rates during the logarithmic phase (5–59 h) were $\mu_{M145} = 0.036\,h^{-1}$, $\mu_{RAR101} = 0.039\,h^{-1}$ and $\mu_{RAR102} = 0.042\,h^{-1}$.

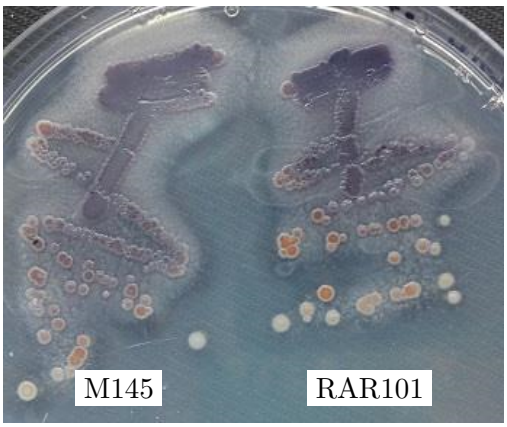



(b) Specific growth rates during the logarithmic phase (7–48 h) were $\mu_{M145} = 0.045\,h^{-1}$, $\mu_{RAR101} = 0.048\,h^{-1}$ and $\mu_{RAR102} = 0.056\,h^{-1}$.

Figure 4.8.: Growth of *S. coelicolor* M145, RAR101 and RAR102 in liquid minimal medium containing Tween 40 as sole carbon source, plotted semi-logarithmically.




(a) 72 h




(b) 120 h

Figure 4.9.: *S. coelicolor* M145 and RAR101 grown on minimal medium agar containing Tween 40 as carbon source, photographed after 72 h (4.9a) and 120 h (4.9b) of growth. No difference between the two strains was observed.

are expressed during growth on these conditions (Chapter 6, Chapman (1994)), which suggests involvement in fatty acid utilisation, loss of the pathway is not detrimental to the organism's fitness in these conditions.

A possible explanation for this contradiction is that the absence of the glyoxylate shunt can be compensated for by the ethylmalonyl-CoA pathway, which can perform the same net anapleurotic reactions as the glyoxylate shunt but follows a more complex route (Section 1.4).

The ethylmalonyl-CoA pathway is also known to produce precursors for certain secondary metabolites (Liu & Reynolds, 1999; Wilson & Moore, 2012). Withdrawal of intermediates for biosynthesis could mean that the anapleurotic function can no longer be fulfilled, in which case the glyoxylate shunt becomes necessary. Should the strains be grown in a phosphate- or nitrogen limited medium, conditions which are known to induce secondary metabolism (Hobbs *et al.*, 1990), a phenotype might be apparent.

## 4.5. Conclusions

The experiments in this chapter have shown that the glyoxylate shunt, which was shown previously to be activated during fatty acid utilisation, is not essential for growth on this carbon source. Flux through the ethylmalonyl-CoA pathway may perform anapleurosis under these conditions. The glyoxylate shunt may be important in conditions where intermediates from the ethylmalonyl-CoA pathway are withdrawn for biosynthesis of secondary metabolites.

# 5. Enzymes of the isocitrate branchpoint

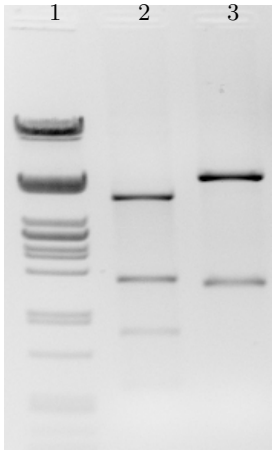
Flux distribution between the TCA cycle and the glyoxylate shunt is determined by the kinetic parameters of enzymes IDH and ICL, their relative activity levels and the concentration of isocitrate. The two enzymes have been well studied in other organisms such as *E. coli*, and IDH has been studied in *S. lividans* (Zhang *et al.*, 2009), which is closely related to *S. coelicolor*, and in *Streptomyces diastaticus* (Zhang *et al.*, 2013). In this chapter, ICLs and IDHs from different organisms were compared and kinetic parameters for *S. coelicolor* ICL and IDH were determined for use in the mathematical model.

## 5.1. Cloning of *S. coelicolor aceA* and *idh*

For the isolation of ICL and IDH, the genes encoding the proteins (*aceA*, *idh*) were amplified by PCR using Pfu DNA polymerase and primers 0982exp-fw / 0982exp-rev and 7000-fw / 7000-rev (Appendix A). The PCR products were cloned into pUC19 digested with SmaI and their identity was confirmed by sequencing (not shown). The resulting plasmids, pRiz101 and pRiz104, were digested with NdeI and XhoI and the inserts were subcloned into pET26b$^+$, also digested with NdeI and XhoI. This produced plasmids pExpaceA and pExpidh, which consequently encoded ICL and IDH, respectively, under the control of the T7 promoter and adds a C-terminal 6-histidine tag. They were checked by restriction digest (Figs. 5.1 and 5.2).

## 5.2. Overexpression and isolation of *S. coelicolor* ICL

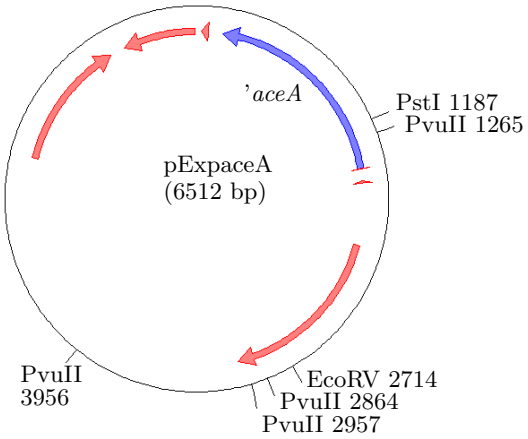
Plasmid pExpaceA was introduced into *E. coli* strain BL21 (DE3). Overexpression and enzyme isolation was performed following the method described in Section 2.7, using buffers described in Section 2.3.3. Overexpression was performed in multiple 1 l cultures simultaneously and was checked by comparison of pre- and post induction samples by SDS polyacrylamide gel electrophoresis (PAGE) (not shown). The protein was purified using the C-terminal 6-histidine tag, which allowed it to bind to a Ni-sepharose col-




(a) Plasmid pExpaceA digested and visualised by agarose gel electrophoresis.

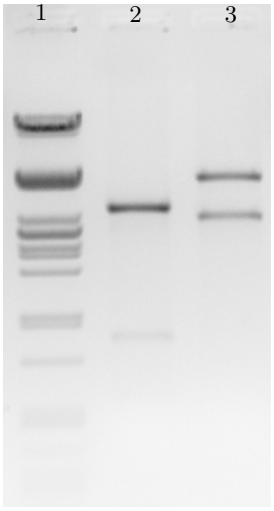
| **PvuII** (2) | **PstI + EcoRV** (3) |
|---|---|
| 3821 | 4987 |
| 1599 | 1525 |
| 999 | |
| 93 | |

(b) Expected band sizes (bp) for digested pExpaceA.




(c) Plasmid map for pExpaceA.

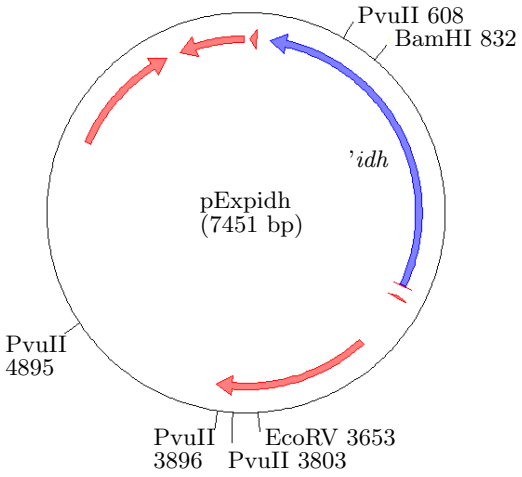
Figure 5.1.: Restriction analysis of plasmid pExpaceA. Figure 5.1a: Agarose gel electrophoresis of digestions. Lane 1: $\lambda$/PstI, Lane 2: pExpaceA/PvuII, Lane 3: pExpaceA/PstI + EcoRV. Figure 5.1b: Expected band sizes. Numbers in headers indicate corresponding lanes on the gel in Fig. 5.1a




(a) Plasmid pExpidh digested and visualised by agarose gel electrophoresis.

| **PvuII** (2) | **BamHI + EcoRV** (3) |
|---|---|
| 3195 | 4628 |
| 3164 | 2823 |
| 999 | |
| 93 | |

(b) Expected band sizes (bp) for digested pExpidh.




(c) Plasmid map for pExpidh

Figure 5.2.: Restriction analysis of plasmid pExpidh. Figure 5.2a: Agarose gel electrophoresis of digestions. Lane 1: $\lambda$/PstI, Lane 2: pExpidh/PvuII, Lane 3: pExpidh/BamHI + EcoRV. Figure 5.2b: Expected band sizes. Numbers in headers indicate corresponding lanes on the gel in Fig. 5.2a

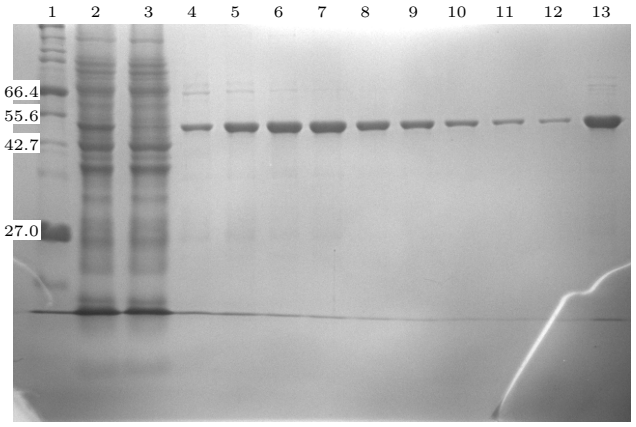



Figure 5.3.: SDS-PAGE of samples taken during isocitrate lyase isolation. Lane 1: broad range marker with selected band sizes indicated in kDa, lane 2: cell lysate (supernatant), lane 3: column flow-through, lane 4–12: collected fractions, lane 13: pooled fractions after buffer exchange.

umn. The 6-histidine tag was not removed post isolation. Samples were taken before loading the column, of column flow-through, elution fractions that had significant $A_{280}$ absorption and of combined fractions after buffer exchange.

Samples were analysed by SDS PAGE as shown in Fig. 5.3. A band clearly stands out in the soluble fraction of the cell lysate (lane 2) that has mostly disappeared after passing through the column (lane 3). Elution fractions (lanes 4–12) show a peak in protein concentration. Hardly any contaminating proteins are present in the fractions and none can be observed in the final solution.

Protein concentrations were determined using absorption at 280 nm with a calculated absorption coefficient $\epsilon_{280nm}$ of $57.8 \times 10^3\,\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$ and molar weight of 46.5 kDa. The accuracy of this method was verified in a preliminary experiment by comparing concentrations calculated from $A_{280}$ measurement and a Bradford assay using BSA as a standard (Bradford, 1976). The concentration was found to be $2.09\,\mathrm{mg\,ml}^{-1}$. The protein was diluted to a concentration of $1.0\,\mathrm{mg\,ml}^{-1}$ using storage buffer containing 80 % glycerol, stored at −20 °C overnight and assayed the following day.

## 5.3. Assay of isolated ICL and calculation of kinetic parameters

ICL is known to follow a steady-state ordered uni-bi kinetic mechanism from studies in *E. coli* (Ko & McFadden, 1990). Since this enzyme has a highly similar amino acid sequence to *S. coelicolor* ICL, similarity in kinetic mechanism was assumed. The reaction

schematic is shown in Eq. (5.1), and the rate for this type of reaction is described by Eq. (5.3) (with $I$ for isocitrate, $S$ for succinate and $G$ for glyoxylate) (Segel, 1975).

None of the reactants can be detected photospectroscopically. The assay thus involved a coupled reaction, where glyoxylate is used as a substrate for lactate dehydrogenase (LDH) in equimolar ratio with NADH, which absorbs at 340 nm (Eq. (5.2)).

$$\text{Isocitrate} \rightleftharpoons \text{Glyoxylate} + \text{Succinate} \tag{5.1}$$

$$\text{Glyoxylate} + \text{NADH} \rightleftharpoons \text{Glycolate} + \text{NAD}^+ \tag{5.2}$$

Although complicated, the rate equation of Eq. (5.3) can easily be simplified by assuming concentrations of glyoxylate and succinate to be 0 or close to it. Setting $[S]$ and $[G]$ to 0 in Eq. (5.3) allows the terms of the formula containing these concentrations to be removed (Eq. (5.4)), since $X \cdot 0 = 0$ for any X. For example, $V_{max_f}\left([I] - \frac{[S][G]}{K_{eq}}\right)$ becomes $V_{max_f}\left([I] - \frac{0}{K_{eq}}\right)$, or simply $V_{max}[I]$. Removing all terms containing $[S]$ or $[G]$ reduces Eq. (5.4) to Eq. (5.5), which is the classic Michaelis-Menten equation (Eq. (1.4)). Since no product is added to the assay mixture, this assumption is valid at the start of the reaction.

$$v = \frac{V_f^{\max}\left([I] - \dfrac{[S][G]}{K_{\text{eq}}}\right)}{K_{\text{I}} + [I] + \dfrac{V_f^{\max} K_{\text{G}}[S]}{V_r^{\max} K_{\text{eq}}} + \dfrac{V_f^{\max} K_{\text{S}}[G]}{V_r^{\max} K_{\text{eq}}} + \dfrac{[I][S]}{K_{\text{ip}}} + \dfrac{V_f^{\max}[S][G]}{V_r^{\max} K_{\text{eq}}}} \tag{5.3}$$

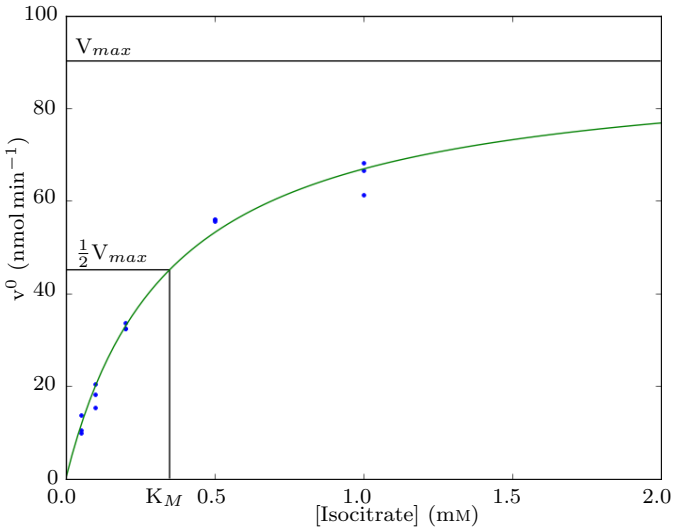
$$v = \frac{V_f^{\max}\left([I] - \cancelto{0}{\dfrac{[S][G]}{K_{\text{eq}}}}\right)}{K_{\text{I}} + [I] + \cancelto{0}{\dfrac{V_f^{\max} K_{\text{G}}[S]}{V_r^{\max} K_{\text{eq}}}} + \cancelto{0}{\dfrac{V_f^{\max} K_{\text{S}}[G]}{V_r^{\max} K_{\text{eq}}}} + \cancelto{0}{\dfrac{[I][S]}{K_{\text{ip}}}} + \cancelto{0}{\dfrac{V_f^{\max}[S][G]}{V_r^{\max} K_{\text{eq}}}}} \tag{5.4}$$

$$v = \frac{V_f^{\max}[I]}{K_{\text{I}} + [I]} \tag{5.5}$$

Activity of 0.01 mg enzyme was assayed in triplicate at isocitrate concentrations of 0.05–1.0 mM as described in Section 2.8. Substrate concentrations and corresponding reaction rates, shown in Fig. 5.4a, were used to calculate kinetic parameters $K_M$ and $V^{max}$ using a weighted non-linear regression method (Wilkinson (1961), described in

| [Substr.] mM | Rate AU min$^{-1}$ | Rate nmol min$^{-1}$ |
|---|---|---|
| 0.05 | 0.062 | 9.892 |
| | 0.085 | 13.653 |
| | 0.065 | 10.508 |
| 0.10 | 0.095 | 15.285 |
| | 0.113 | 18.127 |
| | 0.128 | 20.504 |
| 0.20 | 0.209 | 33.620 |
| | 0.201 | 32.353 |
| | 0.202 | 32.493 |
| 0.50 | 0.348 | 55.939 |
| | 0.346 | 55.627 |
| | 0.348 | 55.920 |
| 1.00 | 0.424 | 68.197 |
| | 0.414 | 66.602 |
| | 0.381 | 61.181 |

(a) Enzymatic rates at different substrate concentrations, determined in triplicate.




(b) Plot of measured reaction rates (blue) and theoretical rates (green). Black lines represent $V_{max}$, $\frac{1}{2}V_{max}$ and $K_M$ as indicated.

Figure 5.4.: Results of ICL assays. Reaction rates in AU min$^{-1}$ and nmol min$^{-1}$ are given in Fig. 5.4a and were plotted in a graph, shown in Fig. 5.4b.

Appendix B). This resulted in a value for $K_M$ of 0.349 mM with a standard error of 0.035 and a $V^{max}$ of 90.2 nmol min$^{-1}$ with a standard error of 3.8. Experimental rates are plotted in Fig. 5.4b (blue dots) together with theoretical rates (green line) based on the calculated $K_M$ and $V^{max}$, which are also outlined (black).

From the value of $V^{max}$ and the amount of enzyme in the reaction, the $k_{cat}$ was calculated following Eq. (5.6). This parameter is normally expressed in $\mu mol_{substrate}\, \mu mol^{-1}_{enzyme}\, s^{-1}$, or $s^{-1}$ in short. From the data given above, a $k_{cat}$ of 6.975 $s^{-1}$ was calculated.

$$k_{cat} = \frac{V^{max}}{[E]}\ (\mu mol_{substrate}\, \mu mol^{-1}_{enzyme}\, s^{-1}) \tag{5.6}$$

## 5.4. Comparison between *S. coelicolor* ICL and those of other organisms

Kinetic parameters obtained for *S. coelicolor* ICL were compared to values for the enzyme from other organisms. $K_M$ and specific activity values were retrieved from BRENDA (http://www.brenda-enzymes.org/) (Schomburg *et al.*, 2002), an online database for enzyme-related information. Values are shown in Table 5.1.

| **Source** | **$K_M$** (mM) | **$k_{cat}$** ($s^{-1}$) | **Reference** |
|---|---|---|---|
| *S. coelicolor* | 0.35 | 6.98 | This work |
| *Streptomyces clavuligerus* | | 1.74 | Soh *et al.* 2001 |
| *C. glutamicum* ATCC 13032 | 0.28 | 24.00 | Reinscheid *et al.* 1994 |
| *M. tuberculosis* (Icl) | 0.145 | 1.08 | Höner Zu Bentrup *et al.* 1999 |
| *E. coli* ML308 | 0.063 | 28.13 | MacKintosh & Nimmo 1988 |
| *E. coli* 23559 | 0.604 | | Walsh & Koshland 1984 |
| *E. coli* K-12 W3110 | 0.89 | | Ogawa *et al.* 2007 |
| *Rhodopseudomonas* sp. 7 | 0.136 | 9.02 | Tahama *et al.* 1990 |
| *M. tuberculosis* (AceA) | 1.3 | 0.34 | Höner Zu Bentrup *et al.* 1999 |
| *Pinus pinea* | 0.033 | 11.11 | Pinzauti *et al.* 1986 |
| *Fomitopsis palustris* | 1.6 | 30.3 | Munir *et al.* 2002 |
| *Yarrowia lipolytica* | 0.6 | 6.17 | Hönes *et al.* 1991 |
| *Papilia machaon* | 1.4 | 6.23 | Popov *et al.* 2005 |
| *Rattus norvegicus* | 0.07 | | Popov *et al.* 1996 |
| *Acinetobacter calcoaceticus* | 0.04 | 3.09 | Hoyt *et al.* 1991 |

Table 5.1.: $K_M$ and $k_{cat}$ values for isocitrate lyases from various organisms.

When comparing the obtained ICL kinetic parameters with those of other organisms, considerable variation is observed (see Table 5.1). Between *E. coli* strains, determined $K_M$ values differ by over 10 fold. MacKintosh & Nimmo (1988) point to a high concentration of chloride ions (supplied as KCl) in the assay mixture in the work of Walsh & Koshland (1984) and find a similar $K_M$ at the higher KCl concentration. The third study (Ogawa *et al.*, 2007) finds a similarly high value but does not use KCL in their assay. They do use 100 mM potassium phosphate buffer in their reaction, which suggests that not chloride but potassium ions may be responsible for the discrepancy. Neither the assay mixture used in Reinscheid *et al.* (1994) or in this work uses a high concentration of either potassium or chloride ions and find $K_M$ values in between the ones found in *E. coli*.

ICL activity has also been detected in several eukaryotic species, among which plants, animals and fungi, although genes encoding the enzyme have not always been identified (Popov *et al.*, 2005). Here, values vary widely as well.

The large spread in kinetic parameter values reported in literature and the many possible causes for them make it difficult to make comparisons. Assay conditions such as salt concentration can influence the $K_M$ value, as proposed by MacKintosh & Nimmo (1988). Other factors, such as enzyme purity and storage conditions and duration will have an equally large influence on the specific activity. ICL isolated from *S. coelicolor*

showed a marked decrease in activity after storage at −20 °C for longer than 24 h (not shown) and it is possible that this effect accounts for at least some of the differences between values.

Comparing *S. coelicolor* ICL with those of its closest relatives for which data were found in literature, the kinetic parameter values are within the same order of magnitude as the values for *C. glutamicum* and *M. tuberculosis*.

## 5.5. Overexpression and isolation of *S. coelicolor* IDH

Plasmid pExpidh was introduced into *E. coli* strain BL21 (DE3). Overexpression and enzyme isolation was performed following the method described in Section 2.7, using buffers described in Section 2.3.3. Overexpression was performed in multiple 1 l cultures simultaneously and was checked by comparison of pre- and post induction samples by SDS PAGE (not shown). The protein was purified using the C-terminal 6-histidine tag, which allowed it to bind to a Ni-sepharose column. The 6-histidine tag was not removed post isolation. Samples were taken before loading the column, of column flow-through and of elution fractions that had significant $A_{280}$ absorption.

Samples were analysed by SDS PAGE as shown in Fig. 5.5. A band clearly stands out in the soluble fraction of the cell lysate (lane 4) that was not present in the pre-induction sample (lane 2) and which has mostly disappeared after passing through the column (lane 5), indicating that the protein was bound to the column. Elution fractions (lanes 6–13) show a peak in protein concentration. A contamination can be seen at 55.6 kDa in lanes 7–9, but the bands are of such low relative intensity that this is unlikely to be significant. Elution fractions containing the protein were combined, concentrated using a buffer exchange column and diluted in 20 mM Tris·HCl (pH = 8) and 250 mM NaCl three times.

The protein concentration was determined using absorption at 280 nm with a calculated absorption coefficient $\epsilon_{280nm}$ of $69.3 \times 10^3\,\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$ and molar weight of 79.5 kDa. The accuracy of this method was verified in a preliminary experiment by comparing concentrations calculated from $A_{280}$ measurement and a Bradford assay using BSA as a standard (Bradford, 1976). The protein was diluted to a concentration of $0.8\,\mathrm{mg\,ml}^{-1}$ using storage buffer containing 20 mM Tris·HCl (pH = 8), 250 mM NaCl and glycerol (final concentration 40 %). DTT was added to a concentration of 1 mM, the protein was stored at −20 °C and assayed the following day.

$$\text{Isocitrate} + \mathrm{NADP^+} \rightleftharpoons \text{2-ketoglutarate} + \mathrm{CO_2} + \mathrm{NADPH} + \mathrm{H^+} \tag{5.7}$$

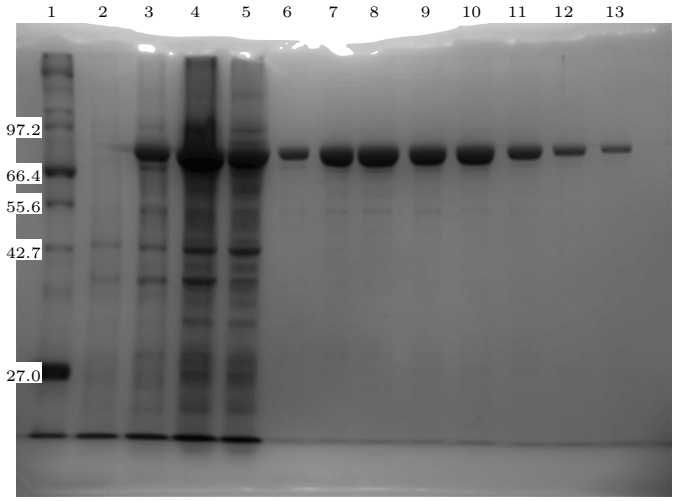


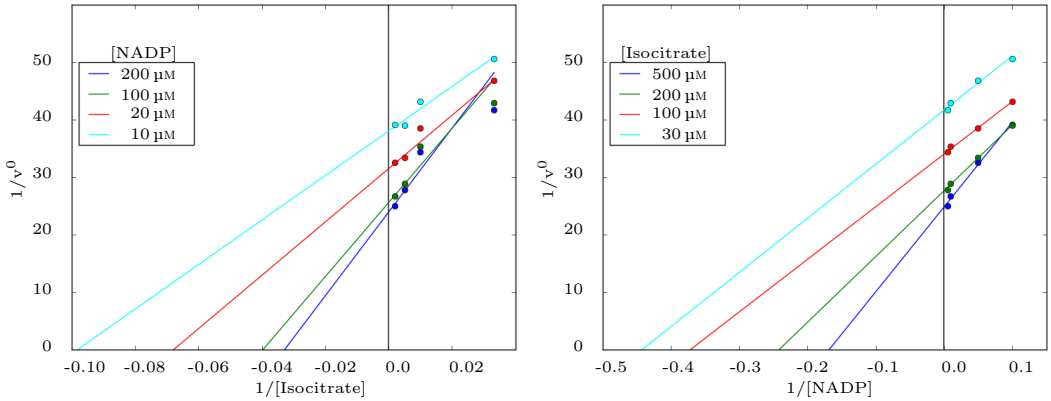

Figure 5.5.: SDS-PAGE of samples taken during isocitrate dehydrogenase isolation. Lane 1: broad range marker with selected band sizes indicated in kDa, lane 2: pre-induction cells, lane 3: harvested cells, lane 4: cell lysate (supernatant), lane 5: column flow-through, lane 6–13: collected fractions (1–8).

IDH was assayed as described in Section 2.9. IDH activity (Eq. (5.7)) can be detected photospetroscopically, as the produced NADPH absorbs at 340 nm whereas $NADP^+$ does not. Because data from literature suggested low $K_M$ values, isocitrate concentrations of 30–500 µM and $NADP^+$ concentrations of 10–200 µM were used initially. The resulting reaction rates are shown in Fig. 5.7.

Depending on the reaction mechanism and the relative values of the kinetic parameters, regression lines through data points in a reciprocal plots intersects in either quadrant of the plot at $\frac{1}{[S]} \leq 0$ (Segel, 1975). To accurately calculate the kinetic parameters, a range of reaction rates is ideally determined at substrate concentrations below, at and above the $K_M$ value (Segel, 1975). The data plotted in Fig. 5.7 (closed circles) did not match that criterion, as experimental reaction rates were all near $V^{max}$ and therefore accurate parameter estimation was not possible. Possibly because of this, regression lines in reciprocal plots did not intersect near one point at $\frac{1}{[S]} < 0$ (Fig. 5.6).

In an attempt to obtain an approximate value for the $K_M$ parameters, the measured reaction rates were fitted into the standard Michaelis-Menten equation, shown in Eq. (1.4), using Origin Pro 9.0. The resulting values were used to plot theoretical lines shown in Fig. 5.7 (lines). It must be stressed that these obtained values do not represent the actual $K_M$ values, because the data were fitted to an incorrect rate equation. Using the highest $V^{max}$ value found (395.5 nmol min$^{-1}$), $k_{cat}$ was calculated to be 654.55 s$^{-1}$

Based on the obtained values, ranging from 2.2–6 µM for NADP and 10–24 µM for




(a) Double reciprocal plot of initial reaction rates versus isocitrate concentration

(b) Double reciprocal plot of initial reaction rates versus NADP concentration

Figure 5.6.: Reciprocal plots of the reaction rates vs substrate concentrations

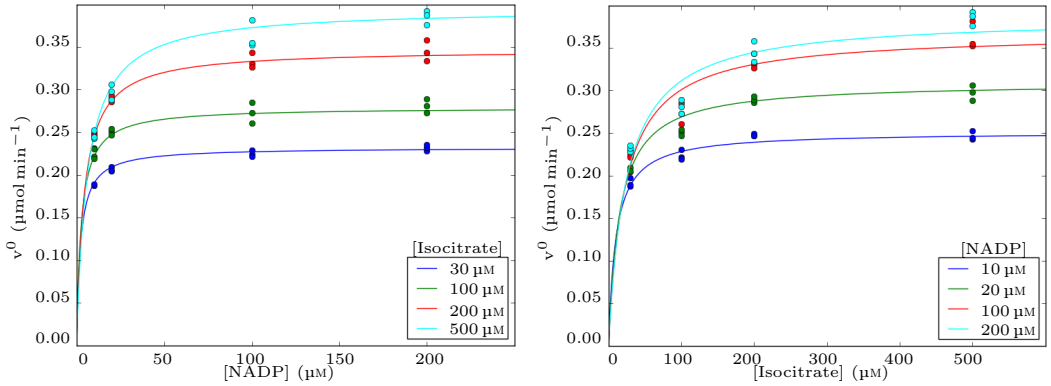
isocitrate, a new assay was carried out using isocitrate concentrations of 5–50 µM and $NADP^+$ concentrations of 4–32 µM. At these low substrate concentrations, the reaction ran to completion within seconds resulting in unreliable reaction rates. The enzyme was isolated freshly, but dialysed overnight against a buffer containing 20 mM Tris·HCl (pH = 8) and 100 mM NaCl. A final enzyme concentration of 0.1 µg $ml^{-1}$ was used in the reaction mixture the following day. The resulting reaction rates are shown in Fig. 5.9.

In this data set, as with the one described previously, the lines do not intersect at or near a single point and some intersect at $\frac{1}{[S]} > 0$ (Fig. 5.8).

In part, this problem is likely to be caused by noise in the data set. A possible source of noise is the low enzyme concentration, resulting in very low reaction rates. At low reaction rates, absolute differences in rates at different substrate concentrations are small, so slight errors in pipetting, calculation of the slope and noise in the measured absorbance values have a relatively large effect. Due to the large number of required data points and the requirement of collecting them on a single day, measurements for each data point were done only once. These errors were therefore not smoothed out, as would be the case when triplicate measurements would have been taken. Because of this, the kinetic parameters could not be determined from these data using reciprocal plots.

As with the previous data set, the data were divided into sets of the same $NADP^+$ or isocitrate concentrations and fit to the standard Michaelis-Menten equation. The results of this are shown in Fig. 5.9.

The estimated $K_M$ values are between 0.75–1.23 µM for $NADP^+$ and 1.76–2.98 µM for isocitrate. Using a $V^{max}$ of 1.90 nmol $min^{-1}$, $k_{cat}$ was calculated to be 25.18 $s^{-1}$.




(a) Enzymatic rates versus NADP concentration at various isocitrate concentrations

(b) Enzymatic rates versus isocitrate concentration at various NADP concentrations

| **[Isocitrate]** µM | $\mathbf{V}_{max}$ nmol min$^{-1}$ | $\mathbf{K}_{M,NADP}$ µM |
|---|---|---|
| 30 | 231.5 | 2.2 |
| 100 | 279.7 | 2.6 |
| 200 | 347.3 | 4.2 |
| 500 | 395.5 | 6.0 |

(c) Calculated values for kinetic parameters for curves shown in Fig. 5.7a

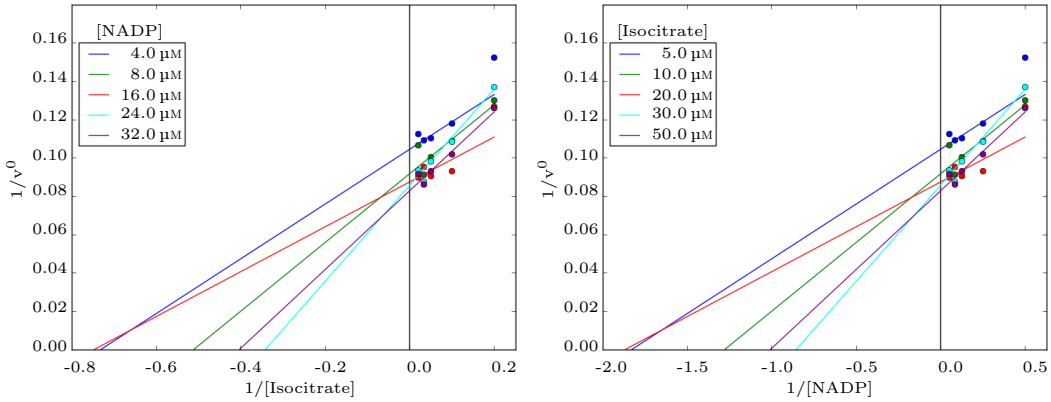
| **[Isocitrate]** µM | $\mathbf{V}_{max}$ nmol min$^{-1}$ | $\mathbf{K}_{M,isocitrate}$ µM |
|---|---|---|
| 10 | 250.8 | 10 |
| 20 | 308.7 | 15 |
| 100 | 366.6 | 22 |
| 200 | 385.9 | 24 |

(d) Calculated values for kinetic parameters for curves shown in Fig. 5.7b

Figure 5.7.: Enzymatic reaction rates of IDH at various substrate concentrations. Figure 5.7a shows experimental rates and fitted curves sorted by isocitrate concentrations and Fig. 5.7b shows experimental rates and fitted curves sorted by NADP concentrations. Reactions were performed at 30 °C in a 1 ml volume containing 0.8 µg enzyme. Estimated kinetic parameters are for the curves are shown in Figs. 5.7c and 5.7d.

For $NADP^+$, the $K_M$ values are consistent with the ones calculated earlier as they lie within the same order of magnitude. The isocitrate values showed a higher deviation than the ones calculated before. Assuming these values give an accurate indication of the magnitude of the $K_M$ values, the substrate concentrations used were still too high for accurate kinetic parameter calculation. However, because of the issues with noise due to the low reaction rates, experiments with lower substrate concentrations were not conducted.

These results indicate that the $K_M$ value for $NADP^+$ is in the range of 1–4 µM. The $K_M$ value for isocitrate is less clear, but likely to be between 2 and 20 µM.




(a) Double reciprocal plot of initial reaction rates versus isocitrate concentration

(b) Double reciprocal plot of initial reaction rates versus NADP concentration

Figure 5.8.: Reciprocal plots of the reaction rates vs substrate concentrations.

## 5.6. Discussion

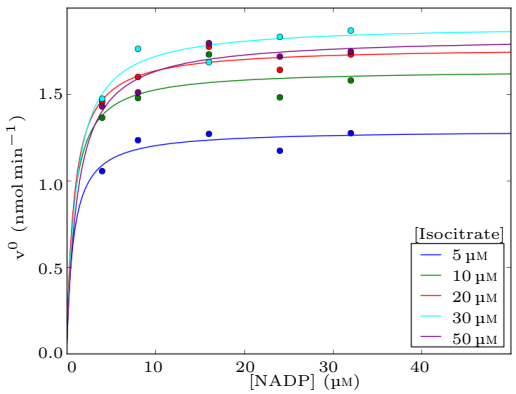
Two enzymes utilising isocitrate in *S. coelicolor*, ICL and IDH were successfully expressed and isolated. Enzyme assays of ICL showed a $K_M$ for isocitrate of 0.35 mM and a specific activity of 9.0 U mg$^{-1}$. These parameters are similar to those of other ICLs.

Analysis of IDH was hampered by technical difficulties. Due to the need to test a range of concentrations of two substrates, it was necessary to collect many data points. Activity of the enzyme was also found to decrease upon prolonged storage, even at −20 °C, therefore all assays had to be performed on the same day as purification which limited the number of replicates that could be done.

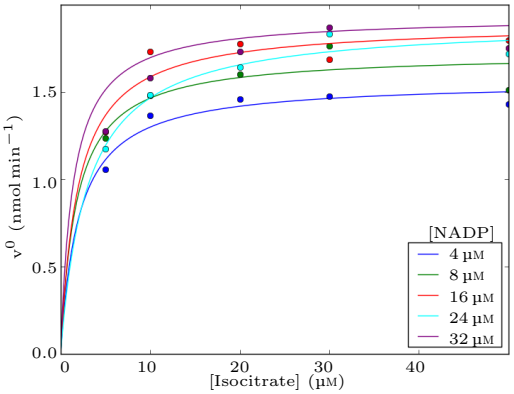
The estimated values, a $K_M$ of 1–4 µM for $NADP^+$ and a $K_M$ or 2–20 µM for isocitrate, are in the same order of magnitude as the values calculated for other IDHs. Zhang *et al.* (2009) report a $K_M$ of 2.78 µM for $NADP^+$ in *S. lividans*, which differs from *S. coelicolor* IDH by one amino acid, indicating that they are likely to be correct.

The $K_M$ for isocitrate of IDH is approximately 100-fold lower than the $K_M$ for isocitrate of ICL. Since IDH thus has a much higher affinity for this shared substrate than ICL, the flux distribution between the glyoxylate shunt and the TCA cycle in the mathematical model simulations can be expected to be in favour of the latter.

Specific activity values for IDH, as with ICL, vary wildly. Zhang *et al.* (2009) report a value of 20.8 U $mg^{-1}_{enzyme}$ for *S. lividans* ICL while Coze *et al.* (2013) report a specific activity of 132.76 ±4.3 µmol min$^{-1}$ $mg^{-1}_{enzyme}$ and strict $NADP^+$ dependence. In this study, specific activity values differed by a factor 26 between enzyme preparations. Values for ICL also varied significantly (Table 5.1) even between related species. This




(a) Enzymatic rates versus NADP concentration at various isocitrate concentrations




(b) Enzymatic rates versus isocitrate concentration at various NADP concentrations

| **[Isocitrate]** µM | $\mathbf{V}_{max}$ nmol min$^{-1}$ | $\mathbf{K}_{M,NADP}$ µM |
|---|---|---|
| 5 | 1.29 | 0.75 |
| 10 | 1.64 | 0.77 |
| 20 | 1.77 | 0.84 |
| 30 | 1.90 | 1.07 |
| 50 | 1.83 | 1.23 |

(c) Calculated values for kinetic parameters for curves shown in Fig. 5.9a

| **[NADP]** µM | $\mathbf{V}_{max}$ nmol min$^{-1}$ | $\mathbf{K}_{M,Isocitrate}$ µM |
|---|---|---|
| 4 | 1.56 | 2.03 |
| 8 | 1.72 | 1.76 |
| 16 | 1.89 | 1.93 |
| 24 | 1.90 | 2.98 |
| 32 | 1.93 | 2.39 |

(d) Calculated values for kinetic parameters for curves shown in Fig. 5.9b

Figure 5.9.: Enzymatic reaction rates of IDH at various substrate concentrations. Figure 5.9a shows experimental rates and fitted curves sorted by isocitrate concentrations and Fig. 5.9b shows experimental rates and fitted curves sorted by NADP concentrations. Reactions were performed at 30 °C in a 1 ml volume containing 0.1 µg enzyme. Estimated kinetic parameters are for the curves are shown in Figs. 5.9c and 5.9d.

indicates that the values for this parameter are heavily influenced by factors unrelated to the enzyme, such as the isolation method, and cannot be effectively compared. For modeling purposes, specific activity values measured from raw cell extracts will provide the necessary data.

# 6. Investigation of spatial and temporal distribution of *aceA* and *ccr* expression

Unlike unicellular prokaryotic organisms, members of the genus *Streptomyces* follow a developmental cycle of growth of branched substrate mycelium, aerial mycelium and finally sporulation. To orchestrate this development, streptomycetes employ elaborate regulatory networks (McCormick & Flärdh, 2012) which are likely to influence activity of metabolic pathways over time. Gene expression is also expected to differ between inner and outer regions of a colony (on solid media) or a pellet (in liquid media) due to gradients in oxygen and nutrient concentrations (Nieminen *et al.*, 2013).

When modeling metabolic pathways, this variable morphology and life cycle present difficulties. In *E. coli* and other unicellular prokaryotes, the system under consideration can be assumed to be homogeneous (Zhao & Shimizu, 2003) and measurements of enzyme activity or transcript levels are sufficient for model parameterisation. In *S. coelicolor*, in order to use these types of data, some effort must be made to rule out the factors outlined above.

This chapter aims to address the problem of potential spatial and temporal distribution in the expression of genes representative of the glyoxylate shunt and the ethylmalonyl-CoA pathway. The promoters for genes *aceA* and *ccr*, representative of the two pathways, were used to control *egfp* expression in *S. coelicolor*. *egfp* encodes enhanced green fluorescent protein (EGFP), the expression of which can be followed with microscopy.

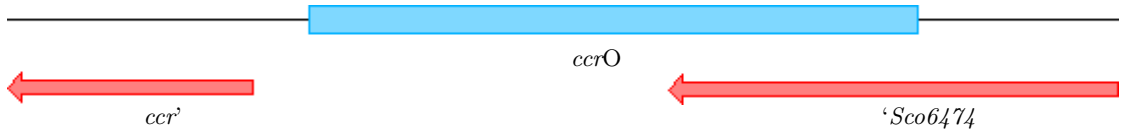
Strains were grown in liquid HMM containing glucose or Tween 40 as carbon source and viewed using a microscope after 24 and 48 h. ICL activity was found when *S. coelicolor* was grown on Tween 40 (Chapman, 1994), while no expression was found during growth on glucose (Chapman, 1994; Gubbens *et al.*, 2012). Since the ethylmalonyl-CoA pathway was proposed to enable acetate utilisation in $icl^-$ strains, a similar expression profile was expected for the genes encoding its enzymes.

Cultures containing both glucose and Tween 40 were grown simultaneously to test whether expression of the genes is subject to carbon catabolite repression. Glucose repression is known to operate at the transcriptional level and is proposed to be medi-

ated by glucose kinase in a growth-phase dependent manner (van Wezel *et al.*, 2007). Therefore if either the glyoxylate shunt or the ethylmalonyl-CoA pathway are subject to glucose repression, a difference in *egfp* expression measured by fluorescence intensity in the respective strain should be visible between cultures grown on Tween 40 and cultures grown on glucose and Tween 40.

## 6.1. Operators were cloned into plasmid pIJ8660 and introduced into strain M145

Two genes were chosen as representatives of the pathways of interest. For the glyoxylate shunt, *aceA* was chosen because its product ICL catalyses the first reaction in the pathway. The ethylmalonyl-CoA pathway is represented by *ccr* because its presence in an organism was proposed to be indicative of presence of the ethylmalonyl-CoA pathway (Erb *et al.*, 2007). Both genes are also the first genes in apparent operons.



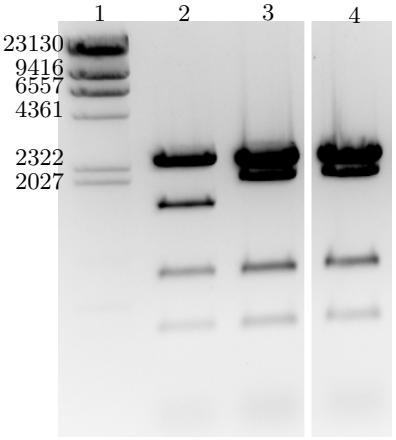

(a) Fragment amplified for construction of pCcrO


(b) Fragment amplified for construction of pIclO


(c) Schematic of operator sequence cloned upstream of promotorless *egfp*. Location of forward (Fw) and reverse (Rv) primers for amplification as well as the EZL2 primer for construct confirmatin are indicated. RBS is the ribosome binding site present on the plasmid

Figure 6.1.: Fragments were amplified for plasmids pCcrO and pIclO both in their original genomic context (Figs. 6.1a and 6.1b) and cloned into plasmid pIJ8660 (Fig. 6.1c).

EGFP is a reporter used extensively in *S. coelicolor*, often as a translational fusion to investigate protein localisation (Bishop *et al.*, 2004; Sun *et al.*, 1999; Willemse & van Wezel, 2009). In this study, transcriptional fusions were made using plasmid pIJ8660 (Sun *et al.*, 1999). This plasmid contains a promotorless *egfp* as a reporter gene




(a) Gel electrophoresis of digested plasmids

| **pIJ8660** (2) | **pccrO** (3) | **piclO** (4) |
|---|---|---|
| 2750 | 2750 | 2750 |
| 2630 | 2630 | 2630 |
| 1520 | 2020 | 2020 |
| 800 | 800 | 800 |

(b) Expected band sizes

Figure 6.2.: Restriction analysis of pIJ8660 and operater probing constructs, digested with PstI. Figure 6.2a: Agarose gel electrophoresis of digestions. Samples were run on the same gel but in different lanes. Lane 1: $\lambda$/HindIII, Lane 2: pIJ8660/PstI, Lane 3: pccrO/PstI, Lane 4: piclO/PstI. Figure 6.2b: Expected band sizes for digestions of plasmids with PstI. Bands of 2750 bp and 2630 bp are seen as a single band. An unexpected band was visible at $\pm$400 bp in all samples.

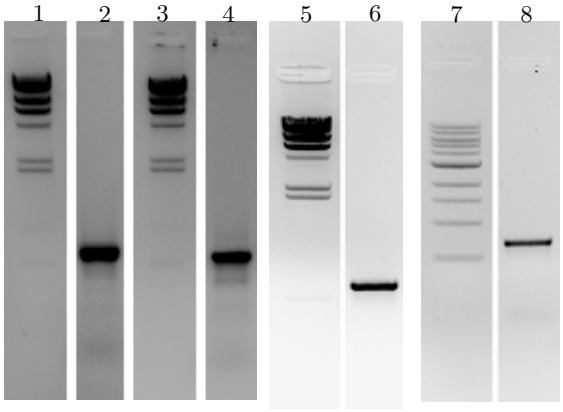



Figure 6.3.: Gel electrophoresis of PCR products with EZL2 and iclO or ccrO primers as appropriate. Presence of a PCR product indicates operater-controlled *egfp*. Lane 1,3,5: $\lambda$/HindIII, Lane 7: NEB 1 kb ladder, Lane 2: piclO, Lane 4: pccrO, Lane 6: RARiclO, Lane 8: RARccrO.

and integrates into the *S. coelicolor* genome using $\phi$C31 integrase after conjugation.

500 bp sequences upstream of the start codons were amplified by PCR and ligated into EcoRV-digested pIJ8660 (Fig. 6.1). Successful ligation resulted in plasmids pCcrO and pIclO. The presence and correct orientation of the insert was checked by restriction analysis (Fig. 6.2) and PCR using the forward primer of the operator fragment and the EZL2 primer located inside *egfp* (Bishop *et al.*, 2004), where presence of a product indicated correct insertion of the target sequence. After introduction of the plasmids into *S. coelicolor* by conjugation, insertion into the genome was checked using the same method (Fig. 6.3). The resulting strains were called RARccrO and RARiclO, which contain pCcrO and pIclO, respectively.

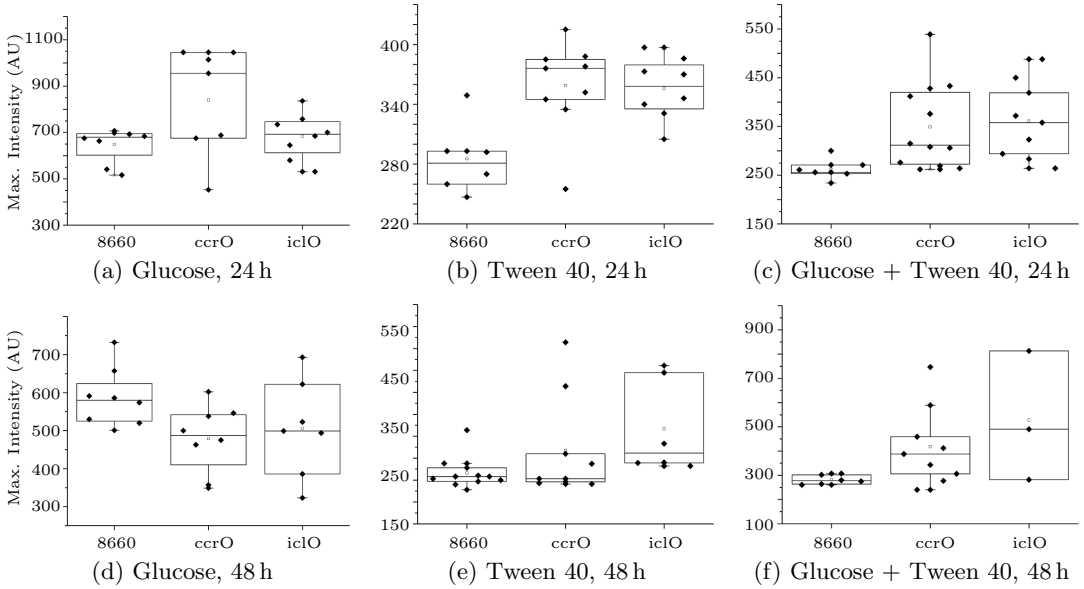
Studies using fluorescent proteins are often hampered by a high background signal, caused by autofluorescence of the biomass. To compensate for this effect, an *S. coelicolor* M145 strain containing pIJ8660 was made (RAR8660) and used as a negative control. Fluorescence in samples taken from this strain were assumed to be due to autofluorescence and its intensity was set as a threshold for other samples grown in the same conditions.

## 6.2. Analysis of strains using fluorescence microscopy

Seed cultures of strains RAR8660, RARccrO and RARiclO (Table 2.1) were grown in YEME for 24 h and the mycelium was harvested and washed with 0.3 M sucrose. After resuspension in 0.3 M sucrose, 50 ml liquid HMM containing either glucose (1 %), Tween 40 (1 %) or glucose and Tween 40 combined (1 % each) was inoculated to give a DCW content of 0.05 mg $ml^{-1}$ and incubated at 30 °C in a shaking incubator. Samples were taken after 24 and 48 h and images were taken with an exposure time of 1 ms for brightfield images and 100 ms for fluorescence images.

Maximum intensity values for each pellet were determined using ImageJ (Abràmoff *et al.*, 2007) and plotted in a box-and-whiskers plot (Fig. 6.4) to inspect the distribution within samples. Save for a few outliers in Figs. 6.4b, 6.4c and 6.4e, values obtained for strain RAR8660 within culture conditions show relatively little deviation from the median. This indicates that intensity of autofluorescence does not change over time, varying perhaps only slightly with pellet size.

Between culture conditions, intensity of autofluorescence varied more widely. The median values for RAR8660 grown on glucose for 24 and 48 h are 679 and 580, respectively. For cultures containing Tween 40 these values are much lower: 281 and 258 when the strain is grown on Tween 40 and 256 and 277 when Tween 40 and glucose were both




(a) Glucose, 24 h (b) Tween 40, 24 h (c) Glucose + Tween 40, 24 h

(d) Glucose, 48 h (e) Tween 40, 48 h (f) Glucose + Tween 40, 48 h

Figure 6.4.: Box plots of signal intensity of pellets for all samples. Intensity values are unitless numbers assigned to each pixel by the microscope camera, with 0 for no signal and the maximum number for saturated pixels. Data points above or below upper or lower whisker, respectively, were considered outliers.

present.

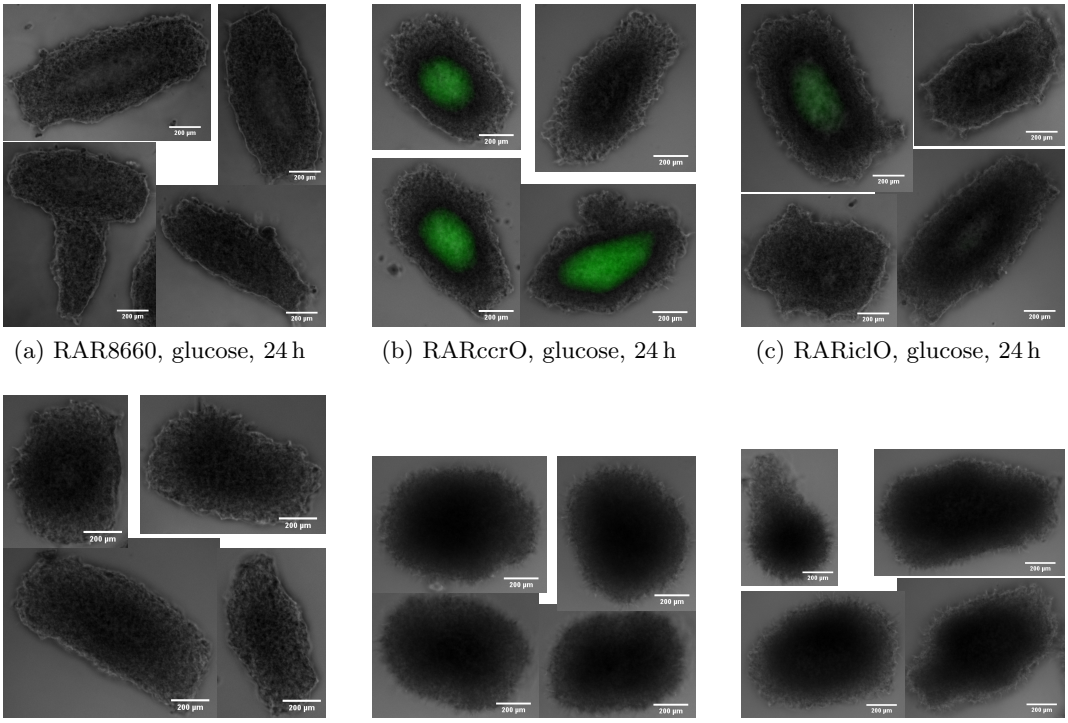
To correct for this autofluorescence, the highest signal intensity values obtained for strain RAR8660 that were not judged outliers were substracted from intensity values in RARccrO and RARiclO within each set of samples. The fluorescent images, in green pseudocolour, were overlayed over the brightfield images to ease interpretation. The results are shown in Figs. 6.5 to 6.7.

Although fluorescence intensities between pellets within samples showed considerable variation in both RARccrO and RARiclO (as seen in Fig. 6.4), certain trends were observed.

### 6.2.1. Cultures grown in the presence of Tween 40 show reduced autofluorescence

Autofluorescence, measured in a strain containing a promoterless *egfp* gene, varied with the carbon source used in the culture. Intensity values for cultures grown in presence of Tween 40 (both with and without glucose) were lower than those for cultures grown on glucose alone.

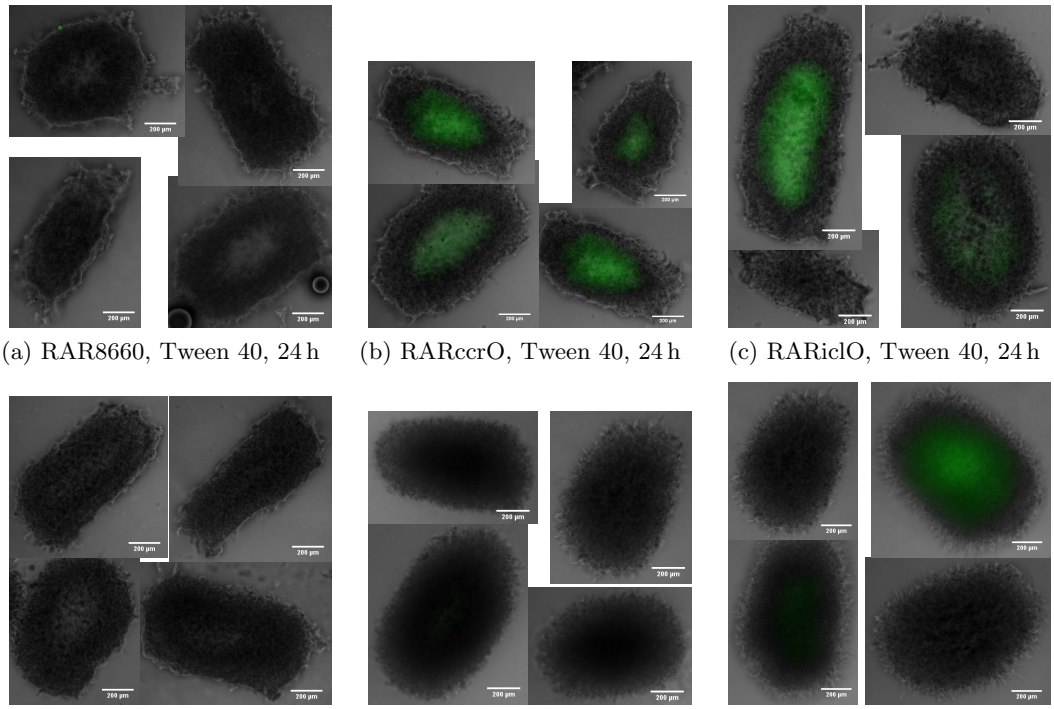
Although by itself not sufficient to support any conclusions, this observation indicates



(a) RAR8660, glucose, 24 h (b) RARccrO, glucose, 24 h (c) RARiclO, glucose, 24 h

(d) RAR8660, glucose, 48 h (e) RARccrO, glucose, 48 h (f) RARiclO, glucose, 48 h

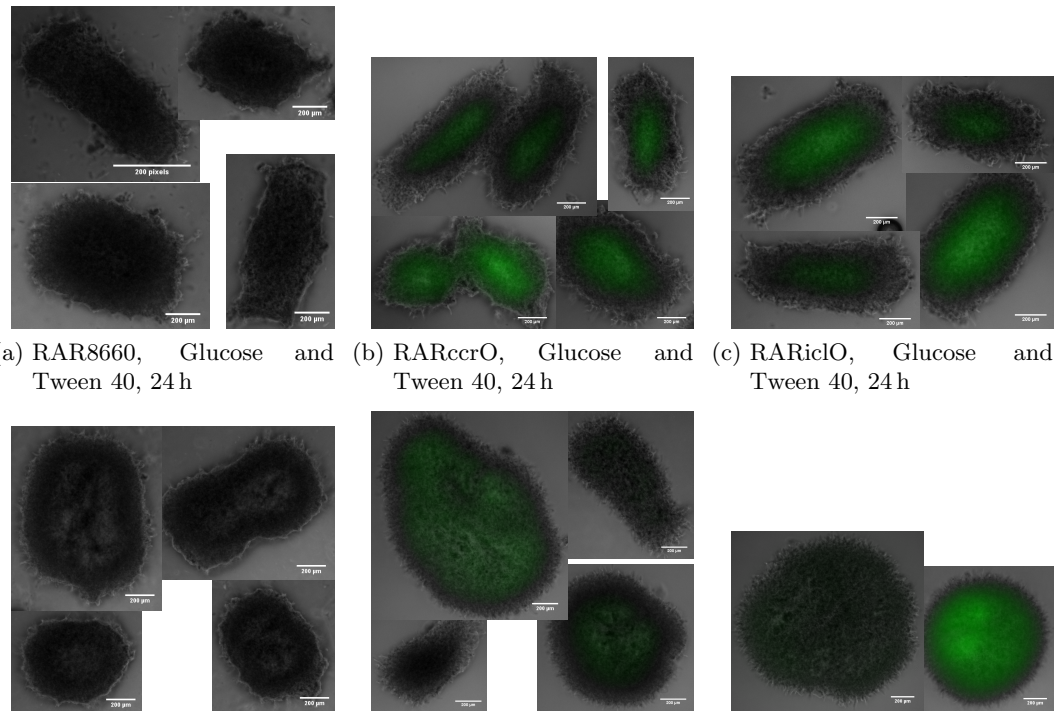
Figure 6.5.: Selected microscopic images from cultures grown in HMM with glucose, sampled after 24 or 48 h. White horizontal bars are 200 µm.



(a) RAR8660, Tween 40, 24 h (b) RARccrO, Tween 40, 24 h (c) RARiclO, Tween 40, 24 h

(d) RAR8660, Tween 40, 48 h (e) RARccrO, Tween 40, 48 h (f) RARiclO, Tween 40, 48 h

Figure 6.6.: Selected microscopic images from cultures grown in HMM with Tween 40, sampled after 24 or 48 h. White horizontal bars are 200 µm.



(a) RAR8660, Glucose and Tween 40, 24 h (b) RARccrO, Glucose and Tween 40, 24 h (c) RARiclO, Glucose and Tween 40, 24 h

(d) RAR8660, Glucose and Tween 40, 48 h (e) RARccrO, Glucose and Tween 40, 48 h (f) RARiclO, Glucose and Tween 40, 48 h

Figure 6.7.: Selected microscopic images from cultures grown in HMM with glucose and Tween 40, sampled after 24 or 48 h. White horizontal bars are 200 µm.

that the metabolic state of *S. coelicolor* mycelium grown on both carbon sources resembled that of *S. coelicolor* grown on Tween 40 more than the state of *S. coelicolor* grown on glucose. The precise causes of autofluorescence in *S. coelicolor* are not known (Willemse & van Wezel, 2009), but these observations indicate that carbon metabolism plays an important role in it. It could be hypothesised that proteins or metabolites with fluorescent properties are present in different levels, that a fluorophore is quenched by a metabolite that is more abundant when *S. coelicolor* is grown on fatty acids, or a combination of these. Localisation studies using EGFP fusions could benefit from this observation, as adding Tween 40 to the culture media could reduce the signal-to-noise ratio without requiring the use of a specially developed strain (Willemse & van Wezel, 2009).

### 6.2.2. Gene expression shows no clear spatial distribution

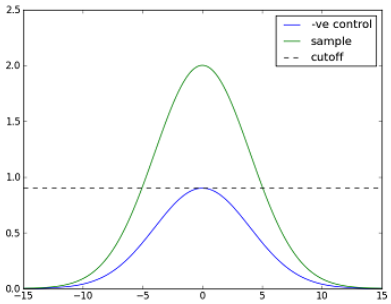
Whenever pellets showed fluorescence relative to the negative control, it was primarily located in the centre of the pellet. Assuming that the pellets' shapes resemble oblong spheres with mycelium being densest in the centre, a gaussian-like distribution of fluorescence intensity along the pellet radius was expected (Fig. 6.8a, where the blue line resembles the negative control and the green line resembles a pellet expressing *egfp*).

When intensity values obtained from microscope images were plotted on a surface, it indeed resembled this shape (Fig. 6.8c). After correction for autofluorescence (Fig. 6.8d), a more or less conical distribution remains. As shown in Fig. 6.8a, this result corresponds with a model where *egfp* expression is homogenous throughout the mycelium. This observation was made in both strains under all culture conditions tested.

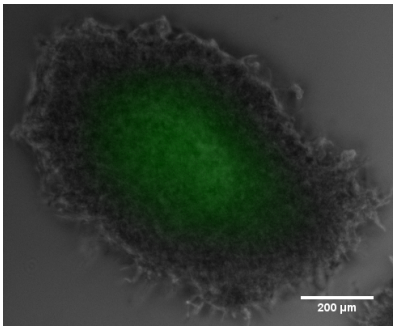
### 6.2.3. *aceA* is expressed in presence of Tween 40 and not repressed in the presence of glucose

*S. coelicolor* does not express *aceA* when grown on glucose, as shown by a lack of fluorescence in Figs. 6.5c and 6.5f. This finding is in line with earlier studies, that found no ICL activity (Chapman, 1994) or trace of the enzyme itself (Gubbens *et al.*, 2012) in glucose-grown cultures. However, when grown in presence of Tween 40, expression was induced (Figs. 6.6c and 6.6f).

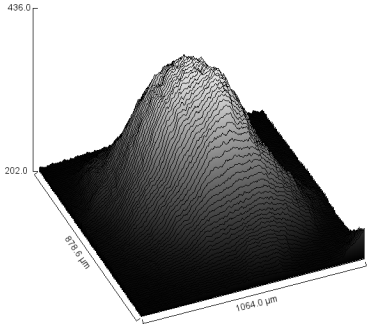
Presence of glucose was insufficient to repress expression in presence of Tween 40, as shown in Fig. 6.7. Thus, assuming glucose is utilised alongside the fatty acids, the glyoxylate shunt genes are not regulated by carbon catabolite repression. This means that expression of this pathway is likely to be repressed by default, but activated when



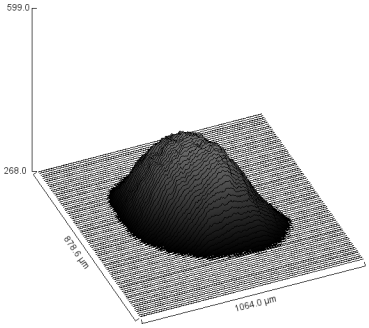

(a) Expected distributions of intensity, assuming homogenous expression




(b) Example overlay of Brightfield and FITC images (RARccrO)




(c) Plot of fluorescence intensity values as obtained from the microscope




(d) Plot of intensity values after compensating for autofluorescence

Figure 6.8.: Example of how fluorescent signal intensity is distributed along the pellet radius. Figure 6.8a shows a two-dimensional plot of two gaussian distributions, representing an image of a negative control and one of a sample image. After the peak intensity of the negative control ("cutoff") has been substracted from the sample image, a conical intensity distribution remains. Figures 6.8c and 6.8d are raw and compensated signal intensities for Fig. 6.8b, showing that the model presented in Fig. 6.8a is a correct approximation and the gene is expressed homogenously throughout the pellet.

fatty acids are present in the environment.

Ubiquity in the environment and a relatively high energy content make fatty acids attractive carbon sources, perhaps too valuable not to utilise in a competitive environment. This is corroborated by the discovery that enzymes catalysing $\beta$-oxidation are constitutively expressed, even without any fatty acids in the growth medium (Banchio & Gramajo, 1997). Expression of *egfp* in the RARiclO strains when grown in the presence of Tween 40 indicates that the glyoxylate shunt plays an important role in environments that are rich in fatty acids, even when other carbon sources are available.

## 6.3. Conclusions

In this chapter, the spatial distribution of *aceA* and *ccr* expression was investigated using fluorescence microscopy. Sequences upstream of both genes were cloned into pIJ8660 and successfully conjugated into *S. coelicolor* M145. The cloned sequences proved capable of expressing *egfp* in response to culturing conditions.

During the analysis of the results flaws in experimental design became apparent. The method of inoculum preparation resulted in pellets with a high level of maturity that displayed significant autofluorescence in their centres. Samples were only taken after 24 h and 48 h, when the mycelium has reached stationary phase. Due to the high level of autofluorescence, as well as the long time between inoculation and sampling, the obtained results are unlikely to represent gene expression during the growth phase.

Analysis of microscopic images has shown that in both strains *egfp* was expressed as a result of growth on fatty acids, independent of the presence of glucose in the medium. No evidence for differential expression between inner- and outer regions of the pellet was found. This means that, for modelling and transcriptomic purposes, expression levels of these pathways can reasonably be assumed to be homogenous throughout the mycelium.

Although no spatial distribution in *egfp* expression was found, a more detailed study might find subtle differences. One such possible improvement is a combination of fluorescence microscopy with biomass measurements, as done in Section 7.1. Strains should be grown in a 400 ml culture and sampled at different time points, starting at only a few hours. Setting up the experiment this way would give a more accurate picture of temporal distribution of expression.

Data analysis can be improved by the use of a threshold gradient over the pellet radius. In this study, the peak intensity of the brightest pellet of the negative control samples was set as a baseline for the whole of each sample image (Fig. 6.8a). By expressing signal intensity as a function of the normalised distance to the geometric centre, aut-

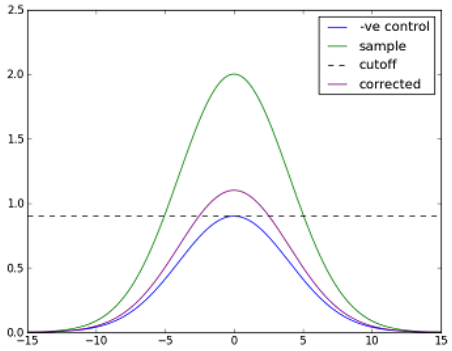



Figure 6.9.: Using the peak intensity of the negative control (blue line) as a cutoff value, only fluorescence in the centre of the sample pellet (green line) can be seen. If the cutoff value is calculated depending on the distance to the pellet centre, a more accurate picture of *egfp* expression can be drawn.

ofluorescence can be corrected for in a way that takes position in the pellet into account (as illustrated in Fig. 6.9). Development of this technique was beyond the scope of this study, however, but is recommended for anyone who wishes to investigate spatial gene expression patterns on a pellet-wide basis where the protein is not expected to localise to a specific location in the hyphae.

# 7. The isocitrate branch point *in vivo*

Mathematical modelling of a metabolic pathway using rate equations requires knowledge of the kinetic parameters of each enzyme involved in the pathway. The kinetic parameters for ICL and IDH were determined before (Chapter 5) as they constitute the branch point between the glyoxylate shunt and the TCA cycle. However, these values only provide information on the activity of pure enzyme, not of activity levels *in vivo*. Moreover, knowledge on the function of a pathway and the conditions under which its genes are expressed are required in order to correctly interpret modelling results. Previously it was assumed that the glyoxylate shunt was the main anapleurotic pathway during growth on Tween 40, but results of Chapter 4 shed doubt on that hypothesis.

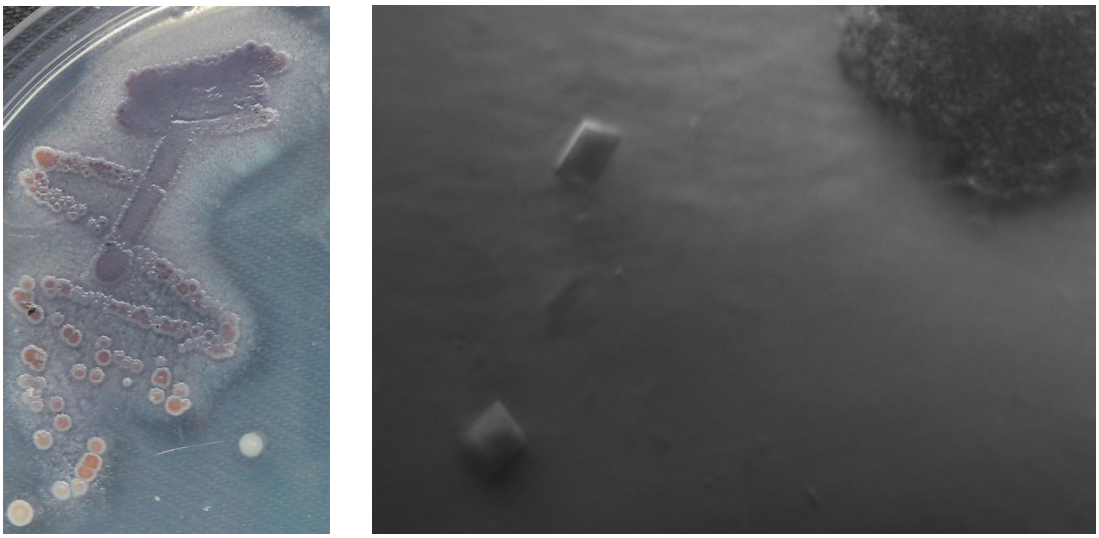
In this chapter, activity of the glyoxylate shunt and in particular the isocitrate branch point is investigated *in vivo*. Cultures of *S. coelicolor* M145 were grown in a minimal medium containing either glucose or Tween 40 as sole carbon source. Samples were taken during the mid-exponential growth phase and tested for ICL and IDH activity. Total RNA was also extracted from samples taken at this time point and, following rRNA depletion, was sequenced in order to determine which genes were differentially expressed between the two conditions. These experiments were performed in collaboration with Jana Hiltner, MSc.

## 7.1. Growth of *S. coelicolor* on Tween 40

Tween 40 is a brand name for the surfactant Polyoxyethylene (20) sorbitan monopalmitate, an ester of polyethyleneglycosylated sorbitan and the fatty acid palmitate. The utilized carbon source, derived from Tween 40, was the palmitate group, as Tween molecules are a substrate for extracellular ester bond breaking enzymes (esterases) which release the fatty acid group from the PEG-ylated Sorbitan (Plou *et al.*, 1998; Pratt *et al.*, 2000; Sakai *et al.*, 2002). Utilisation of Tween molecules by microorganisms this way was shown in the Gram-positive bacterium *Microthris parvicella*. This organism had a similar biomass yield when grown on Tween 80 or oleic acid, either with or without the presence of polyoxyethylene and sorbitan (Slijkhuis, 1983). To our knowledge no such

study has been done in *S. coelicolor*, but since PEG cannot be utilised as a carbon source by *S. coelicolor* (Hobbs *et al.*, 1989), fatty acid uptake and oxidation is the only route through which Tween 40 can be utilised. The surfactant was chosen over palmitate for reasons of solubility and because use of palmitate as carbon source in liquid cultures leads to clumping of the mycelium, preventing biomass-based determination of growth rate of *S. coelicolor* (Banchio & Gramajo, 1997).

Hydrolysis of Tween 40 could be observed after prolonged growth on solid media, when esterase action took place beyond the reach of the mycelium and magnesium palmitate formed a halo of insoluble crystals (Fig. 7.1a). In liquid media containing Tween 40, crystals also formed after prolonged incubation (Fig. 7.1b). Precipitation of fatty acid molecules in the presence of divalent ions, either in liquid or solid cultures, is the basis of standard esterase assays (Pratt *et al.*, 2000). The precipitation in Fig. 7.1 shows that esterase activity indeed takes place in these cultures. Interestingly, when *S. coelicolor* was grown on agar media containing Tween 40 as carbon source, no surface subsidence could be observed, indicating that there was no or hardly any agarase (encoded by *Sco3471*, *dagA*) activity under these conditions.



(a) Palmitic acid precipitation in a solid medium.

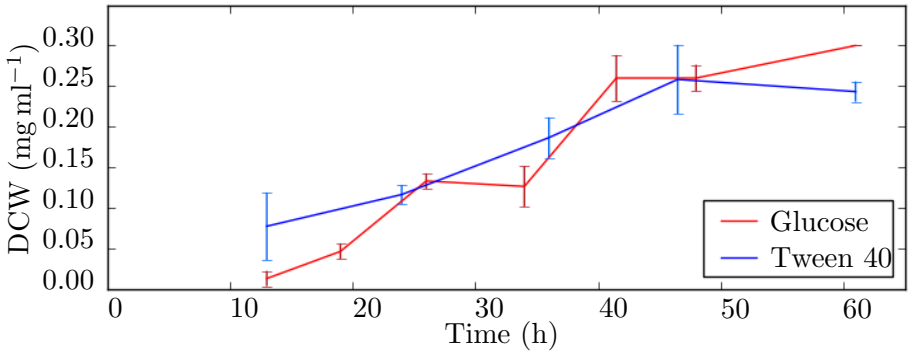
(b) Palmitic acid crystals in a liquid medium.

Figure 7.1.: Palmitic acid crystallises when Tween 40 is hydrolysed. On solid media the crystals formed a turbid halo around dense mycelium (Fig. 7.1a), in liquid media the crystals could be observed as turbidity or microscopically (Fig. 7.1b). Crystallisation of fatty acids from Tween molecules is the basis of esterase assays (Pratt *et al.*, 2000).

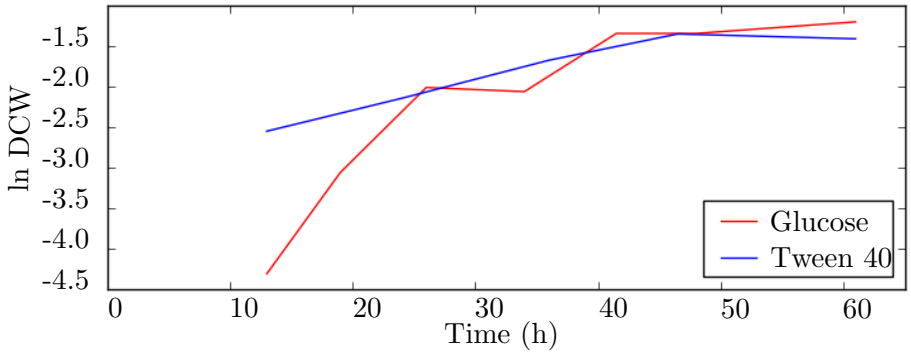
Utilisation of Tween 40 as a carbon source was compared to that of glucose during growth in a minimal liquid medium (Hobbs *et al.* 1989; Section 2.3.2). Spores of *S. coelicolor* M145 were pregerminated in 2xYT as described in Section 2.5. Spores from six confluent plates were used to inoculate six 2 l shakeflasks containing 500 ml HMM medium, three of which contained glucose and three of which contained Tween 40 as the sole carbon sources. Concentrations of glucose and Tween were such that both media contained an equal C-molar concentration of utilisable carbon. The C-molar concentration was defined as the molar concentration of a compound multiplied by the number of utilisable carbon atoms (6 for glucose and 16 for Tween 40). $5\,\mathrm{g\,l^{-1}}$ (55.5 mM) glucose was chosen as a reference concentration, which corresponds to 166.5 mM utilisable carbon. The molar weight of Tween 40 is $1277\,\mathrm{g\,mol^{-1}}$ (Sigma-Aldrich) and only the 16 carbon atoms from the palmitate moiety were considered utilisable, which results in a concentration of Tween 40 of $13\,\mathrm{g\,l^{-1}}$. Growth was monitored by DCW measurements, as described in Section 2.6.

The biomass content of the three biological replicates were plotted against time (shown in Fig. 7.2a). In a semi-logarithmic plot of the biomass content (Fig. 7.2b) the datapoints of the logarithmic growth phase form a straight line, the slope of which equals the specific growth rate. The plot shown in Fig. 7.2b reveals that growth on glucose occurred at a higher specific growth rate ($\mu_{\mathrm{Glc}} = 0.18\,\mathrm{h^{-1}}$ vs $\mu_{\mathrm{Tw}} = 0.04\,\mathrm{h^{-1}}$), but the logarithmic growth phase ended 24 h after inoculation. Growth on Tween 40 was slower but logarithmic growth continued until 48 h after inoculation and a higher biomass concentration was reached at the start of the transition phase.

One factor which possibly limits growth in Tween 40 cultures is the need for esterase activity, which releases palmitate molecules from Tween 40. During growth on glucose, all sugar is dissolved and directly available for uptake, enabling rapid consummation and consequently rapid growth. During growth of *S. clavuligerus* on vegetable oils the concentrations of unsaturated fatty acids, released by esterase-mediated breakdown of triglycerides, was three to four orders of magnitude lower than the concentration of oil in the medium throughout growth (Efthimiou, 2008). In cultures containing Tween 40 as sole carbon source, fatty acid molecules are released from the surfactant by esterases the same way they are released from triglycerides. The C-molar concentrations of carbon added to both cultures described here were equal, but the need for release of palmitate molecules from the surfactant in the Tween 40 cultures means the concentration of carbon available for uptake was likely to have been lower than in the glucose cultures, possibly leading to carbon-limited growth and thus causing the observed difference in growth rate.




(a) Growth of *S. coelicolor* in minimal medium containing either glucose or Tween as sole carbon source, expressed in DCW. Error bars show the standard deviation calculated from measurements of three biological replicates.




(b) DCW data from Fig. 7.2a expressed in a semi-logarithmic plot, showing a logarithmic growth phase between 13–24 h for growth on glucose and 13–48 h for growth on Tween 40. Specific growth rates for the logarithmic phases were $\mu_{\mathrm{Glc}} = 0.18\,\mathrm{h}^{-1}$ and $\mu_{\mathrm{Tw}} = 0.04\,\mathrm{h}^{-1}$ for growth on Tween 40.

Figure 7.2.: Growth of *S. coelicolor* M145 in liquid minimal medium containing either glucose or Tween 40 as sole carbon source.

## 7.2. Specific activity of IDH and ICL in raw cell extracts

Samples were taken at mid-log phase of cultures grown in triplicate in minimal medium containing either glucose (sampled at 19 h) or Tween 40 (sampled at 36 h) as sole carbon source, the mycelium was harvested by centrifugation for 10 min, resuspended in a buffer containing 100 mM $K_2HPO_4$ (pH 7.6), 100 mM NaCl, 20 mM imidazole and 10 % glycerol and frozen at −80 °C. For the assays, the mycelium was thawed and sonicated four times (10 s on, 20 s off on ice), then centrifuged at 2 °C, 15 000 RPM for 30 min. Soluble protein concentration was measured using a Bradford assay using BSA as a standard (Bradford, 1976).

Enzyme assays for IDH were done following Section 2.9, using an isocitrate concentration of 2.5 mM, an NADP concentration of 0.4 mM and 25 µl cell extract per ml reaction mixture. Enzyme assays for ICL were done following Section 2.8, using an isocitrate concentration of 5 mM and 100 µl cell extract per ml reaction mixture. The results are shown in Table 7.1. The specific activity measurements were tested for normal distribution using Shapiro-Wilk, Lilliefors, Anderson-Darling and D'Agostino-K squared tests and statistical significance was tested for using a two sample t-test, all in Origin Pro 9.0.

| **Sample** | **S.A.** (IDH) $\mathrm{U\,mg^{-1}_{protein}}$ | **S.A.** (ICL) $\mathrm{U\,mg^{-1}_{protein}}$ |
|---|---|---|
| Glucose A | $0.876 \pm 0.039$ | $0.014 \pm 0.005$ |
| Glucose B | $1.041 \pm 0.024$ | $0.020 \pm 0.010$ |
| Glucose C | $1.005 \pm 0.036$ | $0.020 \pm 0.007$ |
| Tween A | $1.137 \pm 0.092$ | $0.032 \pm 0.009$ |
| Tween B | $1.174 \pm 0.264$ | $0.031 \pm 0.018$ |
| Tween C | $1.123 \pm 0.136$ | $0.026 \pm 0.004$ |

Table 7.1.: Specific Activity values for IDH and ICL. Values are the mean and standard deviation of triplicate measurements.

The specific activity values of the triplicates of all glucose and Tween 40 samples were normally distributed and were significantly different between the two carbon sources. The values displayed in Table 7.1 show that specific activity of both enzymes are higher in the Tween 40-grown cultures than in the glucose-grown cultures, differing by a factor 1.18 for IDH and 1.67 for ICL. Since IDH has a much higher activity value than ICL (a factor 54.1 for growth on glucose and a factor 38.2 for growth on Tween 40), as well as a lower $K_M$ for isocitrate (Sections 5.2 and 5.5), the reaction rate of IDH can be expected to be much higher than that of ICL, and because of that the flux through the glyoxylate

shunt will be low compared to that through the TCA cycle.

## 7.3. Transcriptomic analysis of *S. coelicolor* M145 grown in minimal medium containing glucose or Tween 40

In order to investigate the total transcriptome of *S. coelicolor* during growth on glucose or Tween 40, an RNA sequencing experiment was set up. Samples were taken from the cultures described in Section 7.1 at mid-log phase (19 h for glucose, 36 h for Tween 40) and total RNA was isolated using an adaptation of the Promega RNeasy Protect Bacteria Kit (Bucca *et al.*, 2009) (available from http://www.surrey.ac.uk/fhms/microarrays/Downloads/Protocols/index.htm).

Total protein extracts from samples taken simultaneously were analysed using SDS PAGE, as shown in Fig. 7.3. The protein content in samples of cultures grown on the same carbon source, which were normalised with respect to total protein concentration, appeared very similar while clear differences could be observed between the two groups. This shows that the replicates of growth on the two carbon sources are consistent with respect to protein content and therefore likely with respect to gene expression as well.

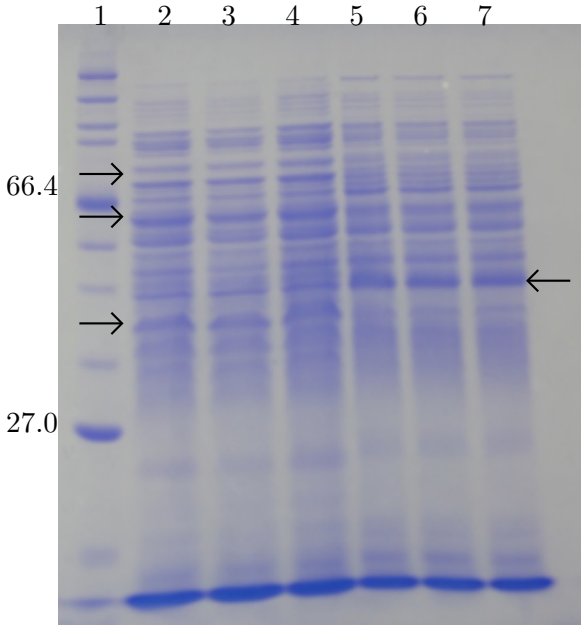
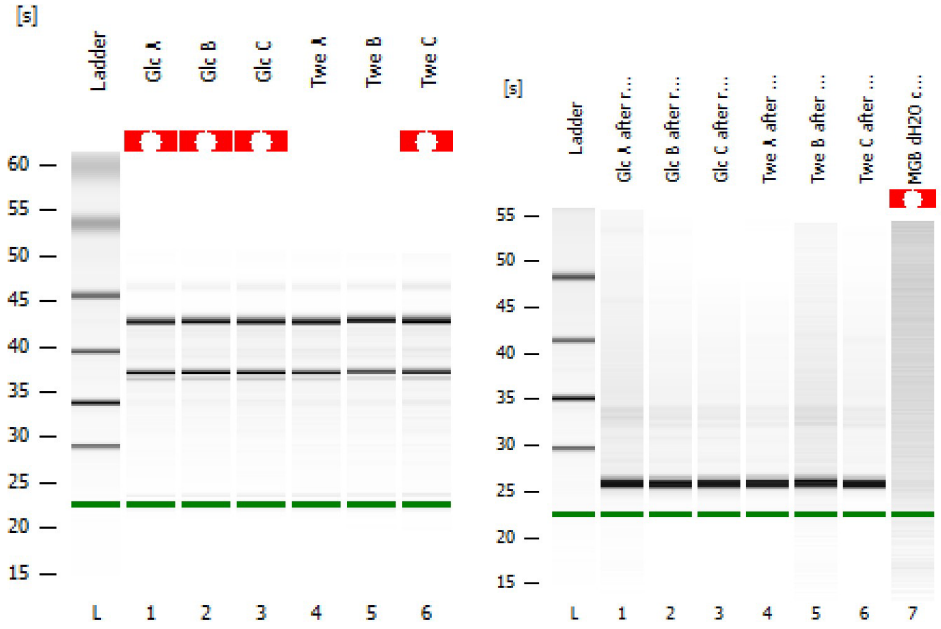



Figure 7.3.: SDS PAGE of total protein samples taken from *S. coelicolor* M145 grown on glucose (lanes 2–4) and Tween 40 (lanes 5–7). Arrows indicate clear differences between the glucose and Tween 40 cultures. Lane 1 contains a broad range protein ladder (NEB)




(a) Bioanalyzer result for total RNA samples.

(b) Bioanalyzer result for total RNA samples after rRNA depletion.

Figure 7.4.: Bioanalyzer traces of RNA samples before and after ribosomal RNA depletion, visualised analogously to an agarose gel.

Quality control of the isolated RNA was done on an Agilent BioAnalyzer (2100 Expert, using a Prokaryote Total RNA Pico chip) following the manufacturer's instructions. The result is shown graphically in Fig. 7.4a. RNA quality is expressed in a RNA integrity number (RIN), which is calculated from the ratio of 16S and 23S ribosomal RNA peak intensities. A RIN score ranges between 0 and 10, with 7.6 being the minimum acceptable quality. The RIN scores for some of the samples could not be calculated because the marker could not be recognised (indicated by the red icon at the top), but because the electrophoresis traces (Fig. 7.5) were very similar between the samples and RIN scores were very high for all samples isolated thus far (8.5–9.5), it was decided to accept all samples.

Ribosomal RNA was removed from the samples using the Epicentre Ribo-Zero rRNA Removal Kit for Gram-Positive Bacteria following the manufacturer's instructions. RNase III digestion, cDNA synthesis, library preparation (including sample barcoding) and sequencing was carried out using Life Technologies Ion Torrent PGM kits, also according to the manufacturer's instructions. An internal standard (ERCC spike-in mixture) was added after the ribosomal depletion step. Samples were pooled according to carbon source of the original culture. Pools of glucose and Tween 40 samples were sequenced

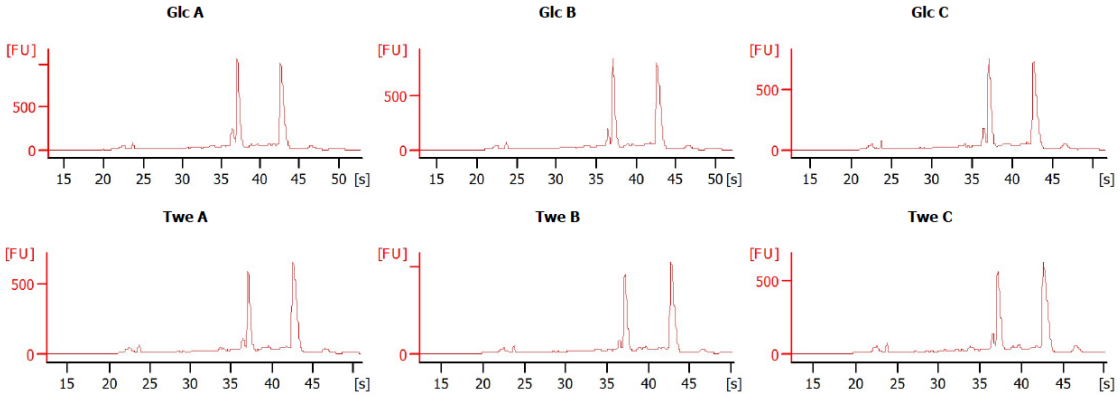



Figure 7.5.: Traces of all samples shown in Fig. 7.4. "FU" are fluorescence units.

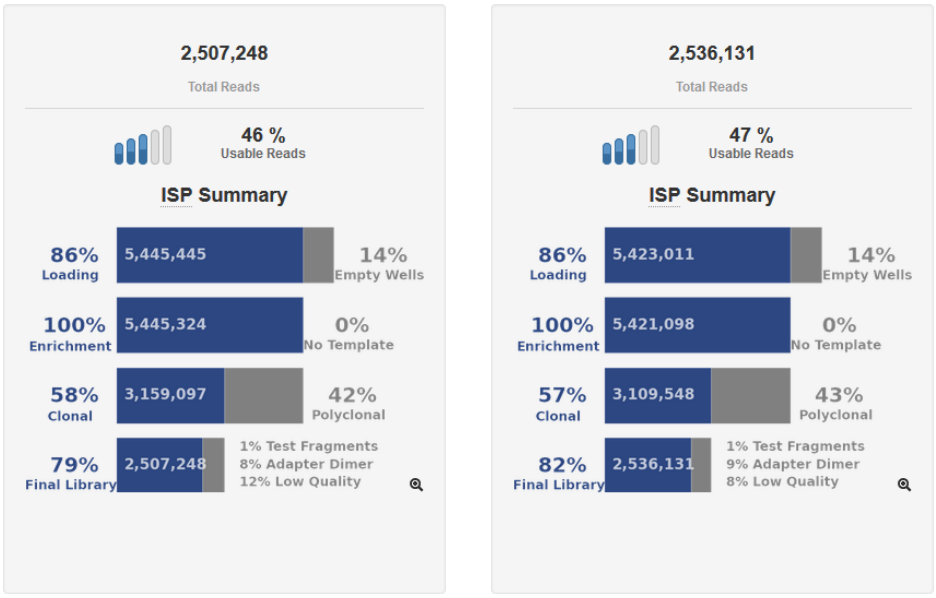
using a Life Technologies Ion 316 v2 chip following manufacturer's instructions.

A summary of the library and sequencing quality is shown in Fig. 7.6. The size of the set of usable data was limited, mainly by the fraction of polyclonal ion sphere particles (ISPs) and the capacity of the 316 chip (5.3 million reads according to the data in Fig. 7.6). The run resulted in an average of $\pm 0.83 \times 10^6$ reads per sample.

Data analysis was done using CLC bio software version 7.5, and differential expression of genes was determined using the EGDE algorithm (Robinson & Smyth, 2008; Robinson *et al.*, 2010). In total, expression of 20 % of the genes was detected. Analysis of the ERCC transcripts revealed that 22 and 55 of the 92 in the mix were not detected in the glucose and Tween 40 samples, respectively. The transcripts that were not detected were those with a low concentration, which is an indicator of sequencing depth being insufficient to detect genes with low expression levels. The following results are presented as differences in gene expression in the Tween 40 samples versus the glucose samples. At the time of writing, data analysis for this experiment was still ongoing, but preliminary data relevant to this thesis are discussed below.

### 7.3.1. The ethylmalonyl-CoA pathway is the primary anapleurotic pathway under these conditions

*S. coelicolor* has the genetic potential for two known anapleurotic pathways, the glyoxylate shunt and the ethylmalonyl-CoA pathway. Of these two, only expression of genes involved in the latter was detected in this experiment. No expression of glyoxylate shunt genes was found. Moreover, the genes encoding enzymes of the ethylmalonyl-CoA pathway were significantly upregulated in the cultures grown on Tween 40 compared to those grown on glucose (Table 7.2).




(a) (b)

Figure 7.6.: Breakdown of the results obtained with the Ion Torrent PGM 316v2 chips for glucose (Fig. 7.6a) and Tween (Fig. 7.6b) samples.

| **Gene** | **fold change** | **FDR-corrected p-value** |
|---|---|---|
| *Sco1346* (*fabG3* / *phaB*) | 11.25 | $1.28 \times 10^{-20}$ |
| *Sco6469* (*mcd*) | 5.769 | $3.42 \times 10^{-24}$ |
| *Sco6470* (*mch*) | 9.338 | $1.96 \times 10^{-34}$ |
| *Sco6471* (*mcl-1*) | 6.085 | $2.94 \times 10^{-22}$ |
| *Sco6472* (*ecm*) | 17.224 | $6.64 \times 10^{-48}$ |
| *Sco6473* (*ccr*) | 8.241 | $1.57 \times 10^{-25}$ |
| *Sco5398* (*epi*) | 2.22 | $2.95 \times 10^{-5}$ |
| *Sco5399* (*thiL* / *phaA*) | 2.695 | $1.41 \times 10^{-8}$ |
| *Sco4926* (*pccB*) | 3.48 | $2.76 \times 10^{-8}$ |

Table 7.2.: Differential expression of ethylmalonyl-CoA pathway genes between Tween 40 and glucose grown cultures. Gene identities were taken from Erb *et al.* (2009). *Sco1346* and *Sco5399*, listed in Table 7.2, were identified as *phaA* and *phaB* by a BLAST search of the *S. coelicolor* genome using the respective *Rhodobacter sphaeroides* genes (Erb *et al.*, 2009) as query sequences, although they were annotated differently in the genome database.

The lack of expression of the glyoxylate shunt genes as measured in this experiment is contradictory to the ICL activity measurement (Section 7.2). However, the specific activity value for ICL was very low (Table 7.1), indicating that expression of *aceA* was probably very low as well. Since the sequencing depth in this experiment was insufficient to detect the low level transcripts of the ERCC spike-in mix, it must have been insufficient to detect transcripts of genes with low expression levels as well. Considering the low specific activity value for ICL, the glyoxylate shunt genes were expressed but their expression level was probably below the detection limit.

Since the expression of the ethylmalonyl-CoA pathway genes was significantly upregulated in the Tween 40 samples compared to the glucose ones and expression of the glyoxylate shunt genes was below the detection limit of this experiment, it is clear that the ethylmalonyl-CoA pathway is the main anapleurotic pathway in *S. coelicolor* under these conditions.

### 7.3.2. Expression of an extracellular esterase and $\beta$-oxidation genes

Apart from the glyoxylate shunt and ethylmalonyl-CoA pathway, differential expression of a few other genes was noteworthy. One group which was differentially expressed was involved in utilisation of Tween 40 as a carbon source. *Sco5420* is annotated to encode a cholesterol esterase, based on similarity with a cholesterol esterase from *Streptomyces lavendulae* (Nishimura & Sugiyama, 1994). This is a secreted enzyme and due to its esterase activity is likely to be the enzyme that hydrolyses the ester bond between the palmitate and head groups of Tween 40 in the culture medium. It was upregulated almost 37 times during growth on Tween 40.

A number of genes involved in $\beta$-oxidation showed higher expression during growth on Tween 40 as well. Since this pathway is responsible for the oxidation of fatty acids to acetyl-CoA molecules this was to be expected. Certain genes, such as *Sco2773* and *Sco6730*, are not directly involved in long-chain fatty acid oxidation but may be co-expressed as part of an operon.

$\beta$-oxidation yields one mole of $FADH_2$ and one mole of NADH per mole of acetyl-CoA released from fatty acids. The generation of ATP this allows reduces the requirement for generation of NADH and $FADH_2$ through the TCA cycle. The gene encoding citrate synthase (CS) (*Sco5832*) was downregulated during growth on Tween 40 (Table 7.3), which is consistent with this hypothesis. The downregulation of *Sco5832* expression could reflect a mechanism to reduce flux of acetyl-CoA and oxaloacetate into the TCA cycle in favour of anapleurotic and gluconeogenetic pathways, respectively.

| Gene | Gene product | fold change | FDR-corrected p-value |
|---|---|---|---|
| *Sco2131* | Long chain fatty acid CoA ligase | 4.066 | $8.34 \times 10^{-8}$ |
| *Sco2773* | Acyl-CoA thioesterase II | 15.999 | $1.17 \times 10^{-18}$ |
| *Sco2774* | Acyl-CoA dehydrogenase | 7.418 | $5.59 \times 10^{-32}$ |
| *Sco3079* | Acetyl-CoA acetyltransferase | 4.745 | $4.95 \times 10^{-15}$ |
| *Sco5420* | Cholesterol esterase | 36.914 | $8.74 \times 10^{-93}$ |
| *Sco6026* | Fatty acid oxidation complex $\alpha$-subunit | 12.96 | $8.85 \times 10^{-103}$ |
| *Sco6027* | Acetyl-CoA acetyltransferase | 7.768 | $1.08 \times 10^{-34}$ |
| *Sco6730* | Racemase | 3.834 | $9.05 \times 10^{-3}$ |
| *Sco6731* | Acetyl-CoA acetyltransferase | 2.534 | $2.0 \times 10^{-2}$ |
| *Sco6732* | Fatty acid oxidative multifunctional enzyme | 3.791 | $1.21 \times 10^{-4}$ |
| *Sco5832* | Citrate Synthase | −5.57 | $1.66 \times 10^{-3}$ |

Table 7.3.: Differential expression of genes involved in fatty acid utilisation between Tween 40 and glucose grown cultures.

### 7.3.3. Expression of genes sharing an upstream palindromic sequence

*Sco6473* and *Sco1346*, which were discussed above (Section 7.3.1), were also found in the search of the *S. coelicolor* genome for the 15 bp palindromic motif in Section 3.3.1. Since this motif is a potential regulatory sequence, a search for the genes listed in Table 3.5 was done among the genes that were significantly differentially regulated between the glucose and Tween 40 samples. Collective upregulation combined with the shared upstream palindromic sequence provides strong evidence for shared regulation of these genes. The genes that were both present in Table 3.5 and upregulated in the Tween 40 cultures in this experiment are listed in Table 7.4.

Besides *Sco6473* and *Sco1346* of the ethylmalonyl-CoA pathway, two other genes were found to be upregulated while also containing the motif of Section 3.3.1 in their upstream sequences. *Sco2494* encodes a pyruvate phosphate dikinase, a gluconeogenic enzyme that generates PEP from pyruvate. It was shown earlier that, when *S. coelicolor* is grown on glucose, expression of *Sco2494* is repressed 8–35-fold in a *glkA*-independent manner (Gubbens *et al.*, 2012). This means that repression could simply be relieved during growth on Tween 40 versus growth on glucose, as the fold changes in expression between this experiment and that of Gubbens *et al.* (2012) are similar to each other.

From this it follows that regulation at the palindromic motif could be responsible for repression of *Sco2494*. However, in the work of Gubbens *et al.* (2012), none of the other

genes listed in Table 7.4 showed any significant change in expression in the absence of glucose. Culturing conditions in Gubbens *et al.* (2012) were also different from the ones described here, since 0.5 % (w/v) casamino acids were added. This suggests that regulation of this gene and the others listed in Table 7.4, assuming the palindromic motif is indeed relevant, entails more than just glucose repression.

The last gene listed in Table 7.4 is *Sco0729*, which encodes a hypothetical protein. To our knowledge this gene has not been described before in the literature and no clear domains were annotated in the genome database, hence no inferences can be made about the relevance of the expression of this gene. However, considering the potential shared regulatory system between this gene and others in Table 7.4 it is a lead for future investigation of fatty acid utilisation in *S. coelicolor*.

| **Gene** | **fold change** | **FDR-corrected p-value** |
|---|---|---|
| *Sco0729* (unknown) | 5.895 | $9.21 \times 10^{-20}$ |
| *Sco1346* (*fabG3* / *phaB*) | 11.25 | $1.28 \times 10^{-20}$ |
| *Sco2494* (*ppdk*) | 30.768 | $4.34 \times 10^{-66}$ |
| *Sco6473* (*ccr*) | 8.241 | $1.57 \times 10^{-25}$ |

Table 7.4.: Genes identified in Section 3.3.1 that were differentially expressed between Tween 40 and glucose grown cultures.

### 7.3.4. Glycolysis and the pentose phosphate pathway

During growth on glucose, catabolism occurs through the Embden-Meyerhof and pentose phosphate pathways (Borodina *et al.*, 2008; Cochrane *et al.*, 1953; Wang *et al.*, 1958). During growth on Tween 40 however, the need for gluconeogenesis reverses the direction of the flux through this pathway. Since most reactions are reversible, no changes in the expression level of the genes that encode the corresponding enzymes was expected between the two growth conditions. A few genes did show differential expression, however (Table 7.5, top).

Two genes in Table 7.5 that encode proteins that are specific for glycolysis are *Sco5578* (*glcP*) and *Sco1214* (*pfkA3*). *Sco5578* encodes a transporter that combines glucose import with phosphorylation. *Sc01214* encodes a 6-phosphofructokinase, which catalyses the irreversible phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate. Since neither of these functions are required in the absence of glucose, downregulation of these genes during growth on Tween 40 is unsurprising. The step is bypassed by fructose

1,6-biphosphatase during gluconeogenesis, which is encoded by *Sco5047* (*glpX*) but no expression of this gene was detected in this experiment.

Expression of the glucose transporter *glcP* is regulated in a *glkA*-dependent manner (Gubbens *et al.*, 2012). However, *S. coelicolor* has two homologues for *glcP*: *Sco5578* and *Sco7153*. The sequences of the two differ by two basepairs and the aminoacid sequences of their products are identical. Despite the high similarity of the two genes, only deletion of *Sco5578* affects uptake of glucose in *S. coelicolor* and only minimal expression of *Sco7153* was detected using RT-PCR (van Wezel *et al.*, 2005). Expression of *Sco7153* was not detected in this experiment, which is consistent with the earlier findings.

The three genes encoding the enzyme glyceraldehyde-3-phosphate dehydrogenase (*Sco1947*, *Sco7040* and *Sco7511*) show opposing trends, two being underexpressed and one being overexpressed during growth on Tween 40. Opposing expression profiles were also seen by Gubbens *et al.* (2012) and it was proposed that the product of *Sco7040* antagonises the other two. These expression data support this hypothesis.

Many genes of the PPP were downregulated as well (Table 7.5, below the line). It is possible that *S. coelicolor* has less need of these enzymes during growth on Tween 40 because glucose, the pathway's substrate, is less abundant. The relatively low growth rate in these conditions possibly also results in a lower demand for nucleotides, the production of which relies on pentose sugars from this pathway for carbon (Ogata & Sadako, 1990; Yamada *et al.*, 1990).

| **Gene** | **fold change** | **FDR-corrected p-value** |
|---|---|---|
| *Sco1214* (*pfkA3*) | −3.084 | $8.57 \times 10^{-15}$ |
| *Sco1947* (*gap1*) | −2.774 | $2.31 \times 10^{-14}$ |
| *Sco5578* (*glcP1*) | −25.784 | $4.70 \times 10^{-14}$ |
| *Sco7040* (*gap2*) | 10.577 | $2.98 \times 10^{-38}$ |
| *Sco7511* (*gap2*) | −19.276 | $5.93 \times 10^{-10}$ |
| *Sco1936* (*tal2*) | −1.772 | $1.37 \times 10^{-3}$ |
| *Sco3877* (*6pgd*) | −1.91 | $6.58 \times 10^{-4}$ |
| *Sco6658* (*6pgd*) | −6.585 | $9.17 \times 10^{-13}$ |
| *Sco6660* (unknown) | −8.131 | $4.35 \times 10^{-5}$ |
| *Sco6661* (*zwf*) | −2.968 | $5.29 \times 10^{-8}$ |
| *Sco6662* (*tal1*) | −2.593 | $6.62 \times 10^{-4}$ |

Table 7.5.: Genes involved in glycolysis (top) and the pentose phosphate pathway (bottom) that were differentially expressed between Tween 40 and glucose grown cultures.

### 7.3.5. Differential expression of biosynthesis clusters

When the *S. coelicolor* genome sequence was published, more than 20 clusters for the biosynthesis of known or predicted secondary metabolite were identified (Bentley *et al.*, 2002). The genes of four of them were differentially expressed in this experiment and are listed in Table 7.6.

Isorenieratene is a carotenoid, a family of compounds that assist with light harvesting in photosynthetic organisms and serve as photoprotectants in organisms frequently exposed to sunlight (Olson & Krinsky, 1995) . Its biosynthesis cluster was described in *S. griseus*, but application of an inducible promoter was necessary to induce expression (Krügel *et al.*, 1999). Carotenoid production in *S. coelicolor* is also induced as a response to blue light (Takano *et al.*, 2005) and expression of its biosynthetic genes is repressed by the two-component regulatory system DraR-K. Increased expression of the isorenieratene biosynthesis operon in this experiment suggests that induction was mediated by growth on fatty acids. *S. coelicolor* is not a photosynthetic organism, so isorenieratene is probably synthesised to protect against photodamage. No explanation for the increase in expression during growth on Tween 40 can be given however without further experiments.

Removed: unclear photo of culture supernatant.

The second cluster to be upregulated is a hopanoid biosynthesis cluster. Hopanoids were proposed to protect against water loss through the plasma membrane in the aerial mycelium (Poralla *et al.*, 2000). The hopanoid that was previously detected in *S. coelicolor* is aminotrihydroxybacteriohopane, but it was only found during differentiation on solid media (Poralla *et al.*, 2000). Regulation of biosynthesis was also found to be decoupled from antibiotic biosynthesis, since it did not depend on the presence of intact BldB (Poralla *et al.*, 2000). Hopanoids induce higher packaging of lipids and thus protect against desiccation of the aerial mycelium (Poralla *et al.*, 2000). In the liquid cultures in this experiment, the production of the hopanoid may be a response to the presence of the Tween 40, which as a surfactant potentially disrupts the membrane. Hopanoids could "stabilise" the membrane and thus reduce diffusion of water across the membrane out of the cell, which would otherwise be considerable due to the high concentration of salts in the growth medium.

During growth on Tween 40, the actinorhodin biosynthesis cluster was downregulated compared to growth on glucose. The precursor molecule for actinorhodin is malonyl-CoA, which can be produced from acetyl-CoA by acetyl-CoA carboxylase. It is possible that this cluster is downregulated during growth on fatty acids to prevent actinorhodin biosynthesis from starving the organism of carbon.

The final biosynthesis cluster that is downregulated is one encoding a NRPS. The function of this cluster is unknown, therefore the implications of differential expression of this cluster remain open to speculation.

| Gene | Gene product | Fold change | FDR-corrected p-value |
|---|---|---|---|
| Isorenieratene biosynthesis | | | |
| *Sco0185* | Geranylgeranyl pyrophosphate synthase | 8.825 | $2.52 \times 10^{-9}$ |
| *Sco0186* | Phytoene dehydrogenase | 5.551 | $4.17 \times 10^{-7}$ |
| *Sco0187* | Phytoene dehydrogenase | 4.294 | $5.07 \times 10^{-3}$ |
| *Sco0188* | methylesterase | 3.618 | $1.1 \times 10^{-2}$ |
| *Sco0191* | lycopene cyclase | 3.129 | $4.1 \times 10^{-2}$ |
| Aminotrihydroxybacteriohopane biosynthesis | | | |
| *Sco6768* | 1-deoxy-D-xylulose-5-phosphate synthase | 2.17 | $2.2 \times 10^{-2}$ |
| *Sco6764* | Squalene-hopene cyclase | 1.973 | $4.8 \times 10^{-2}$ |
| *Sco6766* | Hypothetical protein | 1.9 | $2.0 \times 10^{-2}$ |
| NRPS biosynthesis | | | |
| *Sco6429* | Hypothetical protein | −5.075 | $9.30 \times 10^{-19}$ |
| *Sco6430* | Hypothetical protein | −2.946 | $1.06 \times 10^{-9}$ |
| *Sco6431* | Peptide synthase | −3.279 | $1.72 \times 10^{-8}$ |
| *Sco6432* | Peptide synthase | −2.738 | $1.78 \times 10^{-4}$ |
| *Sco6433* | Hypothetical protein | −2.101 | $9.51 \times 10^{-3}$ |
| *Sco6434* | Oxidoreductase | −2.152 | $1.9 \times 10^{-2}$ |
| *Sco6436* | tRNA synthetase | −2.592 | $1.69 \times 10^{-3}$ |
| *Sco6437* | Hypothetical protein | −3.221 | $5.24 \times 10^{-3}$ |
| *Sco6438* | Diaminopimelate decarboxylase | −2.041 | $8.70 \times 10^{-3}$ |
| Actinorhodin biosynthesis | | | |
| *Sco5071* | Hydroxylacyl-CoA dehydrogenase | −4.529 | $1.12 \times 10^{-5}$ |
| *Sco5072* | Hydroxylacyl-CoA dehydrogenase | −2.995 | $3.01 \times 10^{-3}$ |
| *Sco5073* | Oxidoreductase | −4.975 | $7.48 \times 10^{-5}$ |
| *Sco5074* | Dehydratase | −4.069 | $4.01 \times 10^{-3}$ |
| *Sco5075* | Oxidoreductase | −2.477 | $3.9 \times 10^{-2}$ |
| *Sco5076* | Hypothetical protein | −2.814 | $5.56 \times 10^{-6}$ |

Table 7.6 – continued on next page

Table 7.6 – continued from previous page

| Gene | Gene product | Fold change | FDR-corrected p-value |
|---|---|---|---|
| *Sco5077* | Hypothetical protein | −2.623 | $3.79 \times 10^{-7}$ |
| *Sco5078* | Hypothetical protein | −3.803 | $5.48 \times 10^{-3}$ |
| *Sco5079* | Hypothetical protein | −5.067 | $6.55 \times 10^{-6}$ |
| *Sco5080* | Hydrolase | −4.329 | $2.39 \times 10^{-4}$ |
| *Sco5081* | Hypothetical protein | −4.623 | $7.01 \times 10^{-14}$ |
| *Sco5082* | Transcriptional regulator | −2.846 | $2.15 \times 10^{-5}$ |
| *Sco5083* | Actinorhodin transporter | −2.666 | $4.62 \times 10^{-3}$ |
| *Sco5084* | Hypothetical protein | −1.939 | $3.7 \times 10^{-2}$ |
| *Sco5086* | Ketoacyl reductase | −4.485 | $5.25 \times 10^{-4}$ |
| *Sco5087* | Actinorhodin polyketide beta-ketoacyl synthase subunit $\alpha$ | −4.556 | $6.55 \times 10^{-5}$ |
| *Sco5088* | Actinorhodin polyketide beta-ketoacyl synthase subunit $\beta$ | −5.883 | $1.92 \times 10^{-10}$ |
| *Sco5089* | Actinorhodin polyketide synthase ACP | −11.493 | $1.47 \times 10^{-13}$ |
| *Sco5090* | Actinorhodin polyketide synthase bifunctional cyclase / dehydratase | −7.725 | $4.72 \times 10^{-11}$ |
| *Sco5091* | Cyclase | −7.458 | $1.30 \times 10^{-14}$ |
| *Sco5092* | Actinorhodin polyketide dimerase | −6.913 | $1.96 \times 10^{-9}$ |

Table 7.6.: Differential expression of secondary metabolism clusters between Tween 40 and glucose grown cultures. Secondary metabolite identities were taken from (Bentley *et al.*, 2002).

## 7.4. Conclusions

The experiments described in this chapter show that growth on Tween 40 is markedly different from growth on glucose. Tween is a suitable growth substrate for *S. coelicolor*, with a biomass yield similar to that of growth on glucose (Fig. 7.2). The growth rate was much lower however, potentially due to the need for extracellular esterase activity which releases fatty acids into the environment.

The most obvious differences in gene expression between growth of either of the two carbon sources were in carbon source utilisation. During growth on Tween 40, genes involved in fatty acid utilisation, anapleurosis and gluconeogenesis were upregulated while genes involved in glucose uptake, glycolysis and the pentose phosphate pathway were upregulated during growth on glucose.

Not all genes involved in these pathways were found to be differentially regulated between growth on glucose or Tween 40, possibly because they did not meet the statistical threshold. For example, *Sco5261* encodes a malic enzyme (Rodriguez *et al.*, 2012), an enzyme that catalyses the conversion of malate to pyruvate and is therefore a gluconeogenic enzyme. *Sco5261* was upregulated 1.444-fold during growth on Tween 40 but had an FDR-corrected p-value of 0.05 and was not considered statistically significant.

Of the two anapleurotic pathways *S. coelicolor* is known to possess, the ethylmalonyl-CoA pathway was found to be the most highly expressed. Expression of the glyoxylate shunt genes was not detected using RNA sequencing, but measurements of ICL activity showed that there is at least low level expression. Based on these experiments however, it was concluded that the ethylmalonyl-CoA pathway is the main anapleurotic pathway during mid-exponential growth on fatty acids.

Earlier *in silico* analysis revealed a 15 bp palindromic motif (Section 3.3.1) upstream of a number of genes in *S. coelicolor* (Section 3.3.2). Four of these genes were found to be upregulated during growth on Tween 40. Since the function of three of these genes is related to utilisation of gluconeogenic substrates it is reasonable to assume that the fourth (*Sco0729*), for which no functional annotation is available yet, has a similar function. Future reseach could reveal if the palindromic motif is indeed relevant to transcriptional regulation of these genes and what the function of *Sco0729* is during growth of *S. coelicolor* on fatty acids.

Different secondary metabolites were produced during growth on Tween 40 than during growth on glucose. Biosynthesis of actinorhodin and a non-ribosomal peptide was downregulated. Instead a carotenoid and a lipid were produced, potentially as a response to potential damage by generation of radicals and destabilisation of the membrane, respectively. A yellow compound, potentially the carotenoid, could be extracted from Tween 40-grown mycelium using methanol. Earlier experiments revealed that growth on a combination of glucose and Tween 40 resulted in mycelium with an orange-yellow colour and a yellow culture supernatant. Further experiments are needed to investigate if this yellow colour indeed comes from production of isorenieratene, whether it indeed protects the organism against oxidative damage from radicals and how its biosynthesis can be optimised in the potential interest of industry.

# 8. Modelling of the glyoxylate shunt

Prediction of steady state fluxes through metabolic networks under different circumstances requires numerical simulation of the system under consideration. Starting with known concentrations for all intermediates, initial rates are calculated for all reactions in the system. Then, for discrete time steps, new values for concentrations and reaction rates are alternately calculated. Given small enough time steps, a seemingly continuous trace of concentrations and reaction rates is generated.

The huge number of calculations necessary for any meaningful simulation is ideally done by a computer. A program was developed in Python (van Rossum & Drake, 2001) that is capable of simulating a system of metabolic pathways using kinetic rate equations. The system is defined in a text file that contains information on enzymes (reaction mechanism, kinetic parameters, substrates and products) and reactants (name, starting concentration) as well as instructions concerning total simulation time and number of steps per second. System boundaries can be defined by keeping certain concentrations or reaction rates constant and system perturbations can be simulated by a defined permanent or temporary increase in intermediate concentrations. The program reads and interprets this file, carries out the necessary calculations and produces graphs for rates and concentrations versus time.

## 8.1. Creation of a TCA cycle and glyoxylate shunt model

In order to test the ability of this program to correctly simulate metabolic fluxes, a published metabolic model (Singh & Ghosh, 2006) in *E. coli* was adapted for use in *S. coelicolor*. Although in the first instance errors within the original model were corrected using published data for *E. coli*. The model describes the TCA cycle and the glyoxylate shunt under growth on acetate and glucose. Rate equations for uni- and bireactant reactions were used where appropriate (Equations (8.1) and (8.2), (Segel, 1975), in which $v$ is the net reaction rate, $V_f$ and $V_r$ are the maximum forward and reverse reaction rates, $A$ and $B$ the reaction substrates, $P$ and $Q$ the reaction products and $K_{A/B/P/Q}$ the Michaelis-Menten constants for the reaction substrates and products).

The concentrations of boundary intermediates (oxaloacetate and acetyl-CoA) were kept constant throughout the simulation. A biosynthesis reaction using 2-ketoglutarate was artificially set at a fraction of the IDH rate, the fraction being calculated from experimental data (Zhao & Shimizu, 2003). A graphic representation of the modelled system is shown in Fig. 8.1.

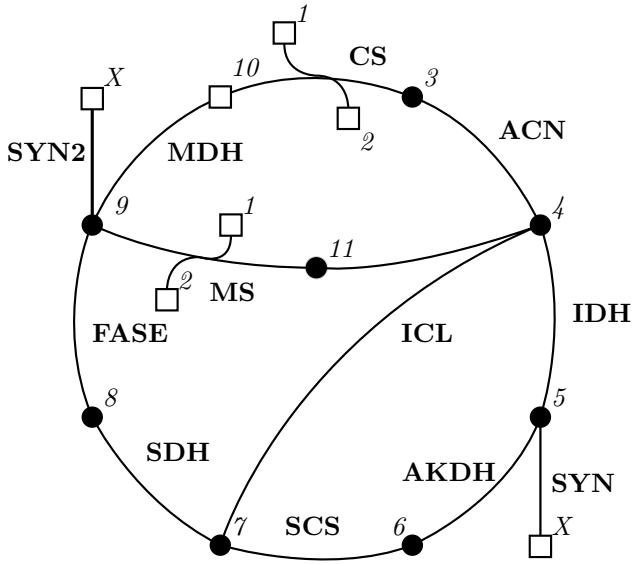



Figure 8.1.: Schematic representation of the modelled system. Reactions are denoted by the name of the enzyme that catalyses them (bold font) and intermediates are represented by numbered circles and boxes, where a box (□) represents a system boundary and a circle (•) represents a free intermediate. Numbered intermediates are 1: Acetyl-CoA, 2: CoA, 3: Citrate, 4: Isocitrate, 5: 2-Ketoglutarate, 6: Succinyl-CoA, 7: Succinate, 8: Fumarate, 9: Malate, 10: Oxaloacetate, 11: Glyoxylate and X: Biomass. Biosynthesis reaction SYN2 was not present in the original model by Singh & Ghosh (2006).

To adapt this model, the same rate equations and kinetic parameters were used, with one exception. The published model does not state the values used for inhibition constants ($K_i$) for glyoxylate and succinate in the reaction catalysed by ICL nor an assumption on estimating equilibrium constants; although they do state the reaction follows an ordered Uni Bi mechanism (Eq. (8.3), (Segel, 1975), with parameters as in Eqs. (8.1) and (8.2)). In our simulation, the missing inhibitor values were taken from literature (Ko & McFadden, 1990).

$$v = \frac{V_f \cdot \frac{A}{K_A} - V_r \cdot \frac{P}{K_P}}{1 + \frac{A}{K_A} + \frac{P}{K_P}} \tag{8.1}$$

$$v = \frac{V_f \cdot \frac{A}{K_A}\frac{B}{K_B} - V_r \cdot \frac{P}{K_P}\frac{Q}{K_Q}}{\left(1 + \frac{A}{K_A} + \frac{P}{K_P}\right)\left(1 + \frac{B}{K_B} + \frac{Q}{K_Q}\right)} \tag{8.2}$$

$$v = \frac{V_f \cdot \left(A - \frac{V_r K_A P Q}{V_f K_{iq} K_P}\right)}{K_A + A + \frac{K_Q K_A P}{K_{iq} K_P} + \frac{Q K_A}{K_{iq}} + \frac{AP}{K_{ip}} + \frac{K_A P Q}{K_{iq} K_p}} \tag{8.3}$$

### 8.1.1. Comparison of the models and correction of mistakes

| Reaction | Glucose | | Acetate | | |
|---|---|---|---|---|---|
| | Reference mM min$^{-1}$ | This study mM min$^{-1}$ | Reference mM min$^{-1}$ | This study mM min$^{-1}$ | Corrected mM min$^{-1}$ |
| CS | 4.187 | 4.187 | 8.006 | 8.006 | 8.012 |
| ACN | 4.187 | 7.187 | 8.006 | 8.006 | 8.012 |
| IDH | 4.179 | 4.177 | 6.125 | 6.144 | 5.855 |
| AKDH | 3.394 | 3.392 | 5.916 | 5.935 | 5.655 |
| SCS | 3.394 | 3.392 | 5.916 | 5.935 | 5.655 |
| SDH | 3.401 | 3.402 | 7.798 | 7.796 | 7.812 |
| FASE | 3.401 | 3.402 | 7.798 | 7.796 | 7.812 |
| MDH | 3.409 | 3.412 | 9.679 | 9.658 | 9.333 |
| ICL | 0.008 | 0.010 | 1.882 | 1.862 | 2.157 |
| MS | 0.008 | 0.010 | 1.882 | 1.862 | 2.157 |
| SYN | 0.786 | 0.785 | 0.209 | 0.210 | 0.200 |
| SYN2 | n/a | n/a | n/a | n/a | 0.636 |

Table 8.1.: Comparison of steady state fluxes in the reference model (Singh & Ghosh, 2006) and their reproduction in both the glucose and acetate situation. The "Corrected" column shows values obtained after correction of errors in the description of ICL and MS

A comparison of steady state flux values for different situations is shown in Table 8.1. Predicted values for growth on glucose agree with the experimental values for all reactions. Growth on acetate however reveals a few differences, starting from different relative fluxes of IDH and ICL. This difference is small and likely to be the result of a difference in values for $K_{ip}$ and $K_{iq}$, as mentioned in Section 8.1.

However, while attempting to recreate the published model, a number of errors became

apparent. The column "Corrected" in Table 8.1 contains steady state fluxes of a model where these errors have been adressed.

One such error was in the reaction mechanism of ICL. Although an ordered Uni Bi mechanism was used, the order in which the two products are released was confused in the implementation. This confused identity of $P$ and $Q$ in Eq. (8.3) results in an altered reaction rate. Indeed, introducing this error in the reproduction of the model resulted in steady state fluxes that resembled the reference fluxes from the published model more closely (Table 8.1, column "This study") than the ones where only this specific error was corrected (not shown).

A second error was made in the description of the glyoxylate shunt. It was necessary to estimate various parameters that were unavailable in literature. One such estimated parameter is the $V_f$ value for MS, which was assumed to be equal to the $V_f$ value of ICL. The authors state this was done because the two genes encoding them are located in the same operon (Singh & Ghosh (2006), supplementary table 1). Although this may be true, it is not possible to conclude that the maximum reaction rate is the same. The $V_f$ value is a product of enzyme concentration and specific activity and there is no reason to assume that the latter is equal to that of another enzyme, even assuming similar enzyme concentrations as a result of co-transcription. To address this, the ratio of specific activities of ICL and MS measured in crude extracts of *E. coli* ML308 cultures grown on acetate was calculated (MS:ICL = 1.88, (El-Mansi *et al.*, 1987)) and applied to the $V_f$ value of MS.

A third error became apparent only after fluxes in the model were compared to experimental values obtained using $^{13}$C tracer experiments and MFA (Zhao & Shimizu, 2003), as was done with the published model (Singh & Ghosh, 2006). Fluxes in this study were expressed as relative values, normalised to substrate uptake. To allow comparison, it was necessary to convert fluxes from the model, resulting in Table 8.2 (showing only values for growth on acetate). Considerable differences between experimental and simulated fluxes were observed for IDH, glyoxylate shunt reactions and MDH. The rate of MDH is relatively inconsequential to this model because its product is the boundary metabolite oxaloacetate (OAA). However, malate is used as a substrate by two malic enzymes: $NAD^+$ dependent (MDH) and $NADP^+$ dependent malic enzyme, with the latter decarboxylating malate to produce pyruvate and NADPH (Rodriguez *et al.*, 2012). A new biosynthetic reaction that was absent in the published model, "SYN2" was introduced to resemble this. In a similar fashion as with the other biosynthesis reaction (SYN), the rate of SYN2 was set to 0.0814 x fumarase (FUM).

After correcting these three errors, new (corrected) steady state fluxes were gener-

| Reaction | Experimental | Reference | This study | Corrected |
|---|---|---|---|---|
| CS | 73.4 | 73.4 | 73.4 | 73.5 |
| ACN | 73.4 | 73.4 | 73.4 | 73.5 |
| IDH | 52.8 | 56.1 | 56.3 | 53.6 |
| AKDH | 51.0 | 54.2 | 54.4 | 51.8 |
| SCS | 51.0 | 54.2 | 54.4 | 51.8 |
| SDH | 71.6 | 71.5 | 71.5 | 71.6 |
| FASE | 71.6 | 71.5 | 71.5 | 71.6 |
| MDH | 86.3 | 88.7 | 88.5 | 85.5 |
| ICL | 20.6 | 17.2 | 17.0 | 19.7 |
| MS | 20.6 | 17.2 | 17.0 | 19.7 |
| SYN | 1.8 | 1.9 | 1.9 | 1.8 |
| SYN2 | 5.8 | n/a | n/a | 5.8 |

Table 8.2.: Comparison of fluxes from experimental data (Zhao & Shimizu, 2003) and the reference model (Singh & Ghosh, 2006), as well as those from the reproduced model and the corrected model, for *E. coli* grown on acetate. Rates were expressed in relative units, with substrate uptake equal to 100.

ated. The corrected values show a larger deviation from the reference fluxes than the uncorrected values (Table 8.1). When all rates are compared to the experimental data however, the results of the corrected model are the most accurate of the three sets (Table 8.2).

Besides reaction rates, concentrations of intermediates were also recorded over time. The concentrations of all intermediates at steady state are shown in Table 8.3 next to their initial values.

Contrary to the steady state reaction rates, there is little similarity between initial compound concentrations and steady state values from the model. This was predictable, as the rate of, for example, CS at the initial concentrations using the given rate equation and concentrations is $0.608\,\mathrm{mM\,min^{-1}}$, which is only 7.5 % of the predicted rate at steady state. The initial rate of MDH is $904.07\,\mathrm{mM\,min^{-1}}$, nearly a hundred times the rate at steady state, while FUM has a negative initial reaction rate of $-0.305\,\mathrm{mM\,min^{-1}}$.

A simulation was carried out with the concentrations of all but the boundary metabolites (marked with * in Table 8.3) set to 0 mM. Simulations starting with these conditions resulted in the exact same steady state fluxes and intermediate concentrations as those done before (not shown). The only difference between simulations was that less time was needed to reach a steady state when initial intermediate concentrations were 0.

| Compound name | Initial (mM) | Reproduction (mM) | Corrected (mM) |
|---|---|---|---|
| Ac-CoA* | 0.5000 | 0.5000 | 0.5000 |
| CoA* | 0.0001 | 0.0001 | 0.0001 |
| Citrate | 9.0000 | 0.0232 | 0.0227 |
| Isocitrate | 0.1500 | 0.1118 | 0.0662 |
| 2-ketoglutarate | 0.2000 | 0.0120 | 0.0114 |
| Succinyl-CoA | 0.0200 | 0.0394 | 0.0343 |
| Succinate | 6.0000 | 0.0162 | 0.0163 |
| Fumarate | 0.1500 | 0.0125 | 0.0124 |
| Malate | 5.0000 | 0.0297 | 0.0191 |
| Oxaloacetate* | 0.0014 | 0.0014 | 0.0014 |
| Glyoxylate | 0.0200 | 0.1456 | 0.0872 |

Table 8.3.: Concentrations of various intermediates initially and after reaching steady state for the reproduced published model and for the corrected model. Compounds labeled with * are boundary compounds, the concentration of which is kept constant throughout the simulation. Steady state concentrations for the published model were not given (Singh & Ghosh, 2006)

### 8.1.2. Overview of the developed model

A model describing the TCA cycle and glyoxylate shunt in *E. coli* was successfully adapted. Several errors in the design of the original model were found and corrected, resulting in an improved prediction of flux distributions under growth on acetate.

Simulations with initial intermediate concentrations set to 0 showed that this system's steady state values depend solely on the values of the system's boundary conditions (concentrations of acetyl-CoA and oxaloacetate) and not on any others.

It is important to note that the various predicted fluxes were compared to experimental flux distribution data resulting from MFA, not with absolute fluxes. The flux distributions were calculated from labelling patterns in proteinogenic aminoacids, which only allows calculation of absolute flux values when compared to for example substrate uptake rates. This was not done in either of the papers and the substrate uptake rates were not given (Singh & Ghosh, 2006; Zhao & Shimizu, 2003).

The predicted absolute reaction rates can therefore not be compared to experimental absolute reaction rates. It then follows that, despite appearances, this model may not predict these rates accurately at all but only informs on their relative values.

The metabolite concentrations in the model and experimental data differ significantly, both in absolute terms and relative to each other. Whether the model predictions

or the experimental data are the most accurate cannot be determined from just these observations. The correspondence between relative flux values and those calculated from MFA favour the model's predictions, but only as relative values.

## 8.2. Adaptation of the model to *S. coelicolor*

The model described in Section 8.1 was adapted to describe the TCA cycle and glyoxylate shunt in *S. coelicolor*. Kinetic parameters for IDH, ICL and MS were replaced with the ones calculated for *S. coelicolor* in Chapter 5 or from the literature and different boundary conditions were used; oxaloacetate was changed from a system boundary to a free metabolite and an estimation for the uptake of fatty acids from a medium containing Tween 40 as carbon source was made. This resulted in a realistic value for the intracellular acetyl-CoA utilisation, as both its production and consumption were under the influence of the model. Amino acid biosynthesis and gluconeogenesis were defined using 2-ketoglutarate and oxaloacetate as substrates, respectively. In absence of kinetic information on their enzymes, these reactions (labelled SYN and SYN2) were set to a cumulative rate equal to that of ICL. This admittedly artificial constraint allowed the system to reach a steady state.

### 8.2.1. Model parameters

Several parameters and dimensions of the *S. coelicolor* mycelium were estimated in order to allow use of enzyme kinetic data in the mathematical model. These numbers are summarised in Table 8.4.

Intracellular MS activity was estimated based on a few assumptions. Kinetic data were available only for MS1 (*Sco6243*) in *S. coelicolor* (Loke *et al.*, 2002) but not for MS2 (*Sco0982*). The specific activity for MS1 was found to be $26.2\,\mu\text{mol}\,\text{min}^{-1}\,\text{mg}_{\text{MS}}^{-1}$ (Loke *et al.*, 2002), with a $K_M$ of $3.5\,\mu\text{M}$ for glyoxylate and $0.035\,\mu\text{M}$ for acetyl-CoA. It was assumed that kinetic parameters for MS2 were similar to those for MS1 and that ICL and MS2 were present in equimolar concentrations, based on their location in a single operon. The specific activity in raw cell extract for MS was calculated using the ratio of specific activity values between ICL (Eq. (8.4)) and MS (Eq. (8.5)) and their molar weights. Based on the calculations and the assumptions mentioned above we estimate the specific activity of MS in raw cell extracts to be 2.20 times the specific activity of ICL in raw cell extracts (Table 8.4 and Eq. (8.6)): $0.04\,\text{U}\,\text{mg}_{\text{protein}}^{-1}$ for glucose grown cells and $0.066\,\text{U}\,\text{mg}_{\text{protein}}^{-1}$ for Tween 40 grown cells. Expression of only *aceB2* was assumed,

| Parameter | Value | Reference |
|---|---|---|
| Cell diameter | 0.5–1 µm | Nieminen *et al.* (2013) |
| Cell length | 1.9 µm | Nieminen *et al.* (2013) |
| Cell volume | $5.97\,\mu m^3$ | Nieminen *et al.* (2013) |
| Cell dry weight | 2021.93 fg | Loferer-Krössbacher *et al.* (1998) |
| Protein content Glc / Tw | $0.42 / 0.43\,g_{protein}\,g_{DCW}^{-1}$ | Shahab *et al.* (1996) |
| IDH S.A. Glc / Tw | $0.97 / 1.14\,U\,mg_{protein}^{-1}$ | This study |
| IDH $K_M$ (Isocitrate / NADP) | 10 / 2 µM | This study |
| ICL S.A. Glc / Tw | $0.18 / 0.30\,U\,mg_{protein}^{-1}$ | This study |
| ICL $K_M$ (Isocitrate) | 0.35 mM | This study |
| MS S.A. Glc / Tw | $0.40 / 0.66\,U\,mg_{protein}^{-1}$ | Estimated from Loke *et al.* (2002) |
| MS $K_M$ (Glyoxylate / Acetyl-CoA) | 3.49 / 0.035 µM | Loke *et al.* (2002) |

Table 8.4.: Parameter values used in the model. Values for cell dimensions and dry weight are approximations. The protein content for glucose and tween grown cultures are based on their growth rate, as calculated in Fig. 7.2. An enzymatic unit (U) was defined as the conversion of 1 µmol of substrate $min^{-1}$.

since *aceB1* was recently shown to be involved in allantoin metabolism (Navone *et al.*, 2014).

$$\text{ICL} : \frac{9.0}{46}\left(\frac{\mu mol\,min^{-1}\,mg_{ICL}^{-1}}{mg\,mmol^{-1}}\right) = 0.20\,\mu mol\,min^{-1}\,mmol_{ICL}^{-1} \tag{8.4}$$

$$\text{MS} : \frac{26.2}{59}\left(\frac{\mu mol\,min^{-1}\,mg_{MS}^{-1}}{mg\,mmol^{-1}}\right) = 0.44\,\mu mol\,min^{-1}\,mmol_{MS}^{-1} \tag{8.5}$$

$$\text{S.A.(MS)} = \frac{0.44}{0.19} = 2.2 \cdot \text{S.A.(ICL)} \tag{8.6}$$

From the specific activity values given in Table 8.4, $V^{max}$ values could be calculated for IDH, ICL and MS. Given the values for the DCW and protein content (the latter of which is dependent on growth rate (Shahab *et al.*, 1996)), a "cell" contains $849.21 \times 10^{-12} / 864.43 \times 10^{-12}\,mg_{protein}$ during growth on glucose or tween, respectively. $V^{max}$ values were obtained by multiplying the specific activity values shown in Table 8.4 by the calculated amount of protein and dividing the result by the cell volume ($5.97\,\mu m^3$, or $5.97 \times 10^{-12}$ ml). The calculation for ICL in the glucose culture is shown in Eqs. (8.7) to (8.9) as an example and the values for all enzymes in both

conditions are given in Table 8.5.

$$\begin{aligned}
\text{S.A.}_{\text{ICL}}(\text{Glc}) &: 0.018\,\mu\text{mol}\,\text{min}^{-1}\,\text{mg}^{-1}_{\text{protein}} \\
\text{One cell contains} &: 849.21 \times 10^{-12}\,\text{mg}_{\text{protein}} \\
0.018 \cdot 849.21 \times 10^{-12} &= 15.286 \times 10^{-12}\,\mu\text{mol}\,\text{min}^{-1}\text{cell}^{-1} \qquad (8.7) \\
\text{Volume of a cell} &: 5.97 \times 10^{-12}\,\text{ml} \\
\frac{15.286 \times 10^{-12}}{5.97 \times 10^{-12}} &= 2.56\,\mu\text{mol}\,\text{min}^{-1}\,\text{ml}^{-1} \qquad (8.8) \\
1\,\mu\text{mol}\,\text{ml}^{-1} &= 1\,\text{mM} \\
V^{max}_{\text{ICL}}(\text{Glc}) &= 2.56\,\text{mM}\,\text{min}^{-1} \qquad (8.9)
\end{aligned}$$

| **Enzyme** | $\mathbf{V}^{max}_{Glucose}$ ($\text{mM}\,\text{min}^{-1}$) | $\mathbf{V}^{max}_{Tween\ 40}$ ($\text{mM}\,\text{min}^{-1}$) |
|---|---|---|
| IDH | 137.98 | 165.07 |
| ICL | 2.56 | 4.34 |
| MS | 5.69 | 9.56 |

Table 8.5.: $V^{max}$ values for IDH, ICL and MS, calculated according to Eqs. (8.7) to (8.9).

For growth of *S. coelicolor* on Tween 40, acetyl-CoA availability was estimated from published kinetic data of fatty acid uptake and fatty acid concentrations in olive oil-grown *S. clavuligerus*. Fatty acid uptake was found to consist of passive diffusion for a concentration of up to 100 µM, the rate of which was linearly proportional to the fatty acid concentration. At concentrations of 100 µM and higher active transport occurred as well, which followed Michaelis-Menten kinetics with a $K_M$ value of 97.8 µM and a $V^{max}$ value of $19.3\,\text{nmol}\,\text{min}^{-1}\,\text{mg}^{-1}_{\text{protein}}$ (Banchio & Gramajo, 1997).

$$v^0 = \begin{cases} 1.59 \cdot C_{F.A.} & \text{for } C_{F.A.} < 0.1\,\text{mM} \\ 1.59 \cdot C_{F.A.} + V^{max}\frac{(C_{F.A.}-0.1)}{K_M+(C_{F.A.}-0.1)} & \text{for } C_{F.A.} \geq 0.1\,\text{mM} \end{cases} \qquad (8.10)$$

The concentration of fatty acid molecules increases over time during growth of *S. clavuligerus* on olive oil, ranging from $20\,\text{mg}\,\text{l}^{-1}$ during early-log phase to $60\,\text{mg}\,\text{l}^{-1}$ in late-log phase (Efthimiou *et al.*, 2008). For growth of *S. coelicolor* M145 on Tween 40 it was assumed that the palmitic acid concentration, released from the surfactant due to esterase activity, would be in a similar range. Using a molar mass for palmitic acid of $256.4\,\text{g}\,\text{mol}^{-1}$, concentrations of extracellular free fatty acids between 78 µM ($20\,\text{mg}\,\text{l}^{-1}$) and 156 µM

($40\,\mathrm{mg\,l^{-1}}$) were used for the simulations. Equation (8.10) was introduced into the model to calculate fatty acid uptake, which was consecutively multiplied by 8 to give the rate of acetyl-CoA production by $\beta$-oxidation (palmitic acid has 16 carbon atoms, which results in 8 acetyl-CoA molecules following oxidation). Fatty acid uptake and oxidation to acetyl-CoA were combined in a reaction term labeled "$\beta$-Ox" in the model (Fig. 8.2) where appropriate.

### 8.2.2. Simulation using oxaloacetate as a free metabolite

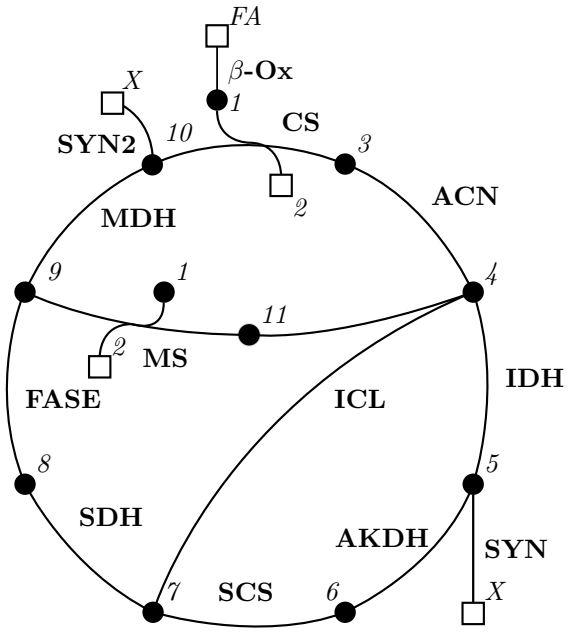



Figure 8.2.: Schematic representation of the modelled system during growth on fatty acids. Reactions are denoted by the name of the enzyme that catalyses them (bold font) and intermediates are represented by numbered circles and boxes, where a box (□) represents a system boundary and a circle (•) represents a free intermediate. Numbered intermediates are 1: Acetyl-CoA, 2: CoA, 3: Citrate, 4: Isocitrate, 5: 2-Ketoglutarate, 6: Succinyl-CoA, 7: Succinate, 8: Fumarate, 9: Malate, 10: Oxaloacetate, 11: Glyoxylate and X: Biomass. *FA* denotes extracellular fatty acids.

In the model defined in Section 8.1, oxaloacetate was defined as a system boundary and kept at a constant concentration during simulations, as in the published model (Singh & Ghosh, 2006). Oxaloacetate is the TCA cycle intermediate that is withdrawn in gluconeogenesis during growth on fatty acids, as well as a substrate of CS. Because of this, its concentration is an important parameter in the flux through both gluconeo-

genesis and the TCA cycle. It was decided to make the concentration dependent on its production from MDH and consumption by CS and gluconeogenesis.

The concentration of oxaloacetate was no longer kept constant and a biosynthetic reaction was introduced for gluconeogenesis that had oxaloacetate as substrate and biomass as product. This reaction (SYN2 in Fig. 8.2) replaced the biosynthesis reaction using malate as substrate introduced into the earlier model (SYN2 in Fig. 8.1).

New restrictions were introduced into the system by the change of oxaloacetate from a system boundary to a free metabolite. With oxaloacetate at a fixed concentration (Fig. 8.1) the TCA cycle is reduced to the equivalent of a linear pathway; any excess production by MDH or consumption by CS, which would normally result in accumulation or depletion of intermediates, is absorbed by this system boundary.

The new system (Fig. 8.2) cannot absorb accumulation or depletion of intermediates this way. Therefore, in order to allow the system to reach a meaningful steady state, the production of intermediates must equal their consumption. Since production of TCA cycle intermediates in this system only occurs through the glyoxylate shunt, it follows that the cumulative rates of all intermediate consuming reactions (SYN and SYN2) should equal the rate through the glyoxylate shunt (Eq. (8.11)). Since no kinetic information or activity levels of these reactions were known, the rates of SYN and SYN2 were artificially set to 0.3· ICL and 0.7· ICL, respectively. These relative rates of SYN1 and SYN2 are approximately the same as they were in the recreated model of Section 8.1 (Table 8.2).

$$\mathrm{SYN} + \mathrm{SYN2} = \mathrm{ICL/MS} \tag{8.11}$$

Simulations were carried out with enzyme activity values determined in the Tween 40 mycelium inserted into the *E. coli* model for growth on acetate (Fig. 8.1). Three starting concentrations of oxaloacetate were used and the steady state fluxes were compared to those found in Section 8.1.1. The concentration of acetyl-CoA (metabolite 1 in Fig. 8.2) was kept constant (0.5 mM) as in the original model. The results are shown in Table 8.6.

All three simulations shown in Table 8.6 reached a similar steady state. The oxaloacetate concentration at the steady state, regardless of the starting concentration, was very similar in the three simulations: 1.349 µM when the starting concentration was 0.7 µM or 1.4 µM, and 1.351 µM at a starting concentration of 2.8 µM. Concentrations of other intermediates were also similar across simulations when the starting concentrations of oxaloacetate or other intermediates (such as isocitrate or glyoxylate) were varied (not shown).

| | **Reaction rates** ($\mathrm{mM\,min^{-1}}$) | | | |
|---|---|---|---|---|
| **Reaction** | **OAA fixed** | **OAA** 0.7 | **OAA** 1.4 | **OAA** 2.8 |
| CS | 8.0196 | 7.7384 | 7.7436 | 7.7539 |
| ACN | 8.0196 | 7.7384 | 7.7436 | 7.7539 |
| IDH | 8.0121 | 7.7311 | 7.7363 | 7.7466 |
| AKDH | 7.7389 | 7.7290 | 7.7342 | 7.7444 |
| SCS | 7.7389 | 7.7290 | 7.7342 | 7.7444 |
| SDH | 7.7464 | 7.7362 | 7.7414 | 7.7517 |
| FASE | 7.7464 | 7.7362 | 7.7414 | 7.7517 |
| MDH | 7.1234 | 7.7434 | 7.7487 | 7.7589 |
| ICL | 0.0075 | 0.0072 | 0.0072 | 0.0073 |
| MS | 0.0075 | 0.0072 | 0.0072 | 0.0073 |
| SYN | 0.2732 | 0.0022 | 0.0022 | 0.0022 |
| SYN2 | 0.6306 | 0.0051 | 0.0051 | 0.0051 |

Table 8.6.: Simulations of the model shown in Fig. 8.2 with a fixed acetyl-CoA concentration (0.5 mM).

These simulations show that removing oxaloacetate as a system boundary resulted in a stable system that reached a steady state over a range of starting concentrations. A side effect of this removal was that, unlike the system discussed in Section 8.1, the starting concentrations of intermediates influenced the values in the steady state, but only to a minor extent (0.1 % difference in the final concentration of oxaloacetate and ±5 % difference in the steady state rate of CS). The reason for this is probably the equality of the flux through the glyoxylate shunt, and SYN and SYN2, which prevents any net production or consumption of TCA cycle intermediates. Indeed, the sums of concentrations of all intermediates at the start and end of the three simulations were nearly equal, indicating that no net consumption or production of TCA cycle intermediates has taken place, and the differences between total final concentrations between simulations was equal to the difference in starting concentration of oxaloacetate (not shown).

### 8.2.3. Simulation using $\beta$-oxidation as a source of acetyl-CoA

The model described in Section 8.1 was based on growth of *E. coli* on acetate, for which no uptake rates were known. An estimation for the uptake rate of fatty acids and the resulting acetyl-CoA production from concurrent $\beta$-oxidation was described in Section 8.2.1. These were introduced into the model described in Section 8.2.2, replacing the fixed acetyl-CoA concentration.

Simulations were carried out using the model described in Section 8.2.2, with an

oxaloacetate starting concentration of 1.4 µM and a range of extracellular fatty acid concentrations (0.078–0.136 mM). The resulting reaction rates are shown in Table 8.7

| **Reaction** | **Steady state reaction rates (mM min$^{-1}$) at different extracellular FA concentrations** | | | | |
|---|---|---|---|---|---|
| | 0.078 mM | 0.10 mM | 0.11 mM | 0.12 mM | 0.136 mM |
| $\beta$-Ox | 0.9922 | 1.2720 | 3.4727 | 5.3213 | 7.7439 |
| CS | 0.9914 | 1.2710 | 3.4698 | 5.3167 | 7.7367 |
| ACN | 0.9914 | 1.2710 | 3.4698 | 5.3167 | 7.7367 |
| IDH | 0.9906 | 1.2700 | 3.4669 | 5.3120 | 7.7294 |
| AKDH | 0.9904 | 1.2697 | 3.4660 | 5.3106 | 7.7273 |
| SCS | 0.9904 | 1.2697 | 3.4660 | 5.3106 | 7.7273 |
| SDH | 0.9911 | 1.2707 | 3.4689 | 5.3153 | 7.7345 |
| FASE | 0.9911 | 1.2707 | 3.4689 | 5.3153 | 7.7345 |
| MDH | 0.9919 | 1.2717 | 3.4718 | 5.3199 | 7.7417 |
| ICL | 0.0008 | 0.0010 | 0.0029 | 0.0046 | 0.0072 |
| MS | 0.0008 | 0.0010 | 0.0029 | 0.0046 | 0.0072 |
| SYN | 0.0002 | 0.0003 | 0.0009 | 0.0014 | 0.0022 |
| SYN2 | 0.0005 | 0.0007 | 0.0020 | 0.0032 | 0.0051 |

Table 8.7.: Reaction rates of simulations of the model shown in Fig. 8.2 at a range of extracellular fatty acid concentrations. Acetyl-CoA production from $\beta$-oxidation was taken as 8 times the fatty acid uptake rate.

The simulations that include $\beta$-oxidation show that the system has a maximum capacity for acetyl-CoA uptake of $\pm 7.744$ mM min$^{-1}$. At extracellular fatty acid concentrations above 0.136 mM, with a subsequent acetyl-CoA production rate of 7.75 mM min$^{-1}$ or higher, no steady state was reached as acetyl-CoA accumulated indefinitely (not shown).

While the reaction rates generally increased proportional to the acetyl-CoA production rates, the concentrations of the TCA cycle intermediates followed a different pattern. The cumulative concentration of all intermediates (not including acetyl-CoA) was the same between simulations, but the relative concentrations varied widely. Figure 8.3a shows the steady state concentrations of all intermediates for simulations with various acetyl-CoA production rates. The concentration of most intermediates increased steadily with the reaction rates with a few notable exceptions, which are shown in Fig. 8.3b.

The steady state concentration of acetyl-CoA (Ac-CoA in Fig. 8.3) remained very low in all simulations except at an acetyl-CoA production rate of 7.744 mM min$^{-1}$, where it suddenly increased. This is probably caused by the low concentration of oxaloacetate, which was far below the $K_M$ value of CS for that intermediate ($K_{OAA} = 70$ µM). The

low reaction rate led to a low consumption of acetyl-CoA, which then accumulated. The higher acetyl-CoA concentration then increased the rate of CS until it equated the rate of acetyl-CoA production again.

Malate and oxaloacetate (Mal and OAA in Fig. 8.3) concentrations were high when the rate of acetyl-CoA production was low, but dropped at higher production rates. This is possibly the result of an increase of SYN and SYN2 rates, which were coupled to the rate of ICL, as well as an increase in CS rate, which also consumes OAA.

The concentration of succinyl-CoA (SCoA in Fig. 8.3) increased exponentially at higher acetyl-CoA production rates, most likely because the rate of production of both 2-ketoglutarate and succinate were increased. The first induces succinyl-CoA production while the latter inhibits succinyl-CoA consumption in this pathway. Increased production and inhibited consumption led to net accumulation to a level where it inhibited its own production and increased its own consumption.

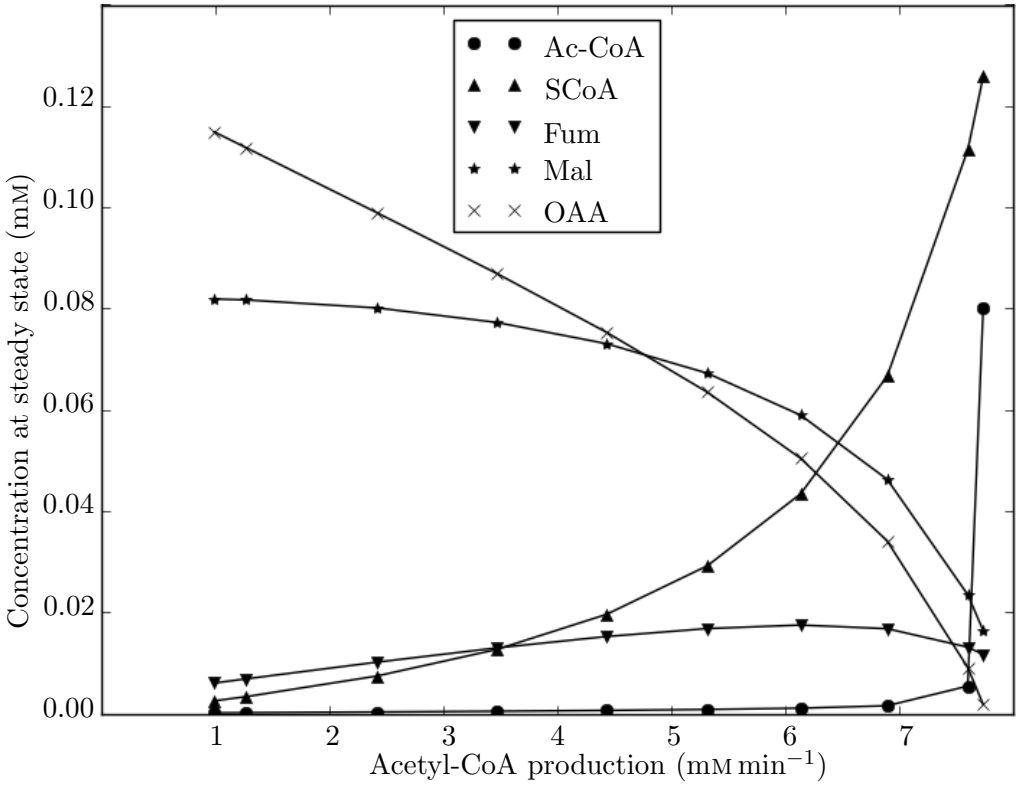
As mentioned in Section 8.2.1, the extracellular fatty acid concentration increases during growth of *Streptomyces* on fatty acid esters (Efthimiou *et al.*, 2008). As a consequence, the release of acetyl-CoA into central carbon metabolism from $\beta$-oxidation will steadily increase during the exponential growth phase. As shown in the steady state values of these simulations (Fig. 8.3) the predicted availability of substrates for gluconeogenesis (malate and oxaloacetate) decreased at higher acetyl-CoA availability. The availability of succinyl-CoA however, along with acetyl-CoA itself a precursor of starter and extender units for polyketide biosynthesis (Coze *et al.* (2013); Haydock *et al.* (2004), Section 1.4), increased. The conclusion is that, as the system progressed from the conditions of early to those of mid and late exponential growth, the increased availability of acetyl-CoA automatically reconfigured the TCA cycle from providing precursors for gluconeogenesis to providing precursors for secondary metabolism. Whether this shift actually occurs *in-vivo* and also coincides with the onset of polyketide biosynthesis requires experimental verification.

### 8.2.4. Perturbation experiments

To explore whether the model was sensitive to changes in its boundary condition, two simulations were carried out where the extracellular fatty acid concentration changed in time, starting from a steady state (Figs. 8.4 and 8.5). In the first, the fatty acid concentration started at 0.11 mM but was temporarily increased by a gaussian peak (Fig. 8.4) centered at 15 s, with a maximum of 0.136 mM. In the second, the concentration of fatty acid was permanently increased from 0.11 mM to 0.136 mM (Fig. 8.5). The concentrations of 9 metabolites were recorded, normalised to the initial (steady state)

| **Compound** | **Steady state concentrations (µM) at different acetyl-CoA production rates (mM min$^{-1}$)** | | | | |
|---|---|---|---|---|---|
| | 0.9922 | 1.2720 | 3.4727 | 5.3213 | 7.7439 |
| Free FA | 78 | 100 | 110 | 120 | 136 |
| Ac-CoA | 0.1077 | 0.1397 | 0.4293 | 0.7787 | 80.1380 |
| Cit | 2.6828 | 3.4410 | 9.4274 | 14.4890 | 21.1677 |
| Icit | 0.0632 | 0.0816 | 0.2352 | 0.3784 | 0.5966 |
| 2-KG | 1.7643 | 2.2751 | 6.5277 | 10.5358 | 17.6772 |
| SCoA | 2.4860 | 3.3031 | 12.6305 | 29.1576 | 125.7738 |
| Suc | 1.2074 | 1.5769 | 5.0409 | 8.9553 | 15.9849 |
| Fum | 6.0020 | 6.8253 | 12.9401 | 16.7629 | 11.6475 |
| Mal | 81.8590 | 81.6860 | 77.2357 | 67.2027 | 16.3404 |
| OAA | 114.8763 | 111.7534 | 86.9068 | 63.4622 | 1.7579 |
| Gly | 0.0090 | 0.0075 | 0.0056 | 0.0061 | 0.0039 |

(a) Steady state intermediate concentrations after simulations of the model shown in Fig. 8.2 with acetyl-CoA production from $\beta$-oxidation. Free fatty acid concentrations are extracellular




(b) Concentrations of selected TCA cycle intermediates at steady state (from Fig. 8.3a), plotted versus the rate of acetyl-CoA production.

Figure 8.3.: The effect of differing acetyl-CoA production rates on the concentration of TCA cycle intermediates at steady state.

concentrations and plotted against time. Due to the normalisation, all traces start at 1.0 and change in time according to their fold increase or decrease.

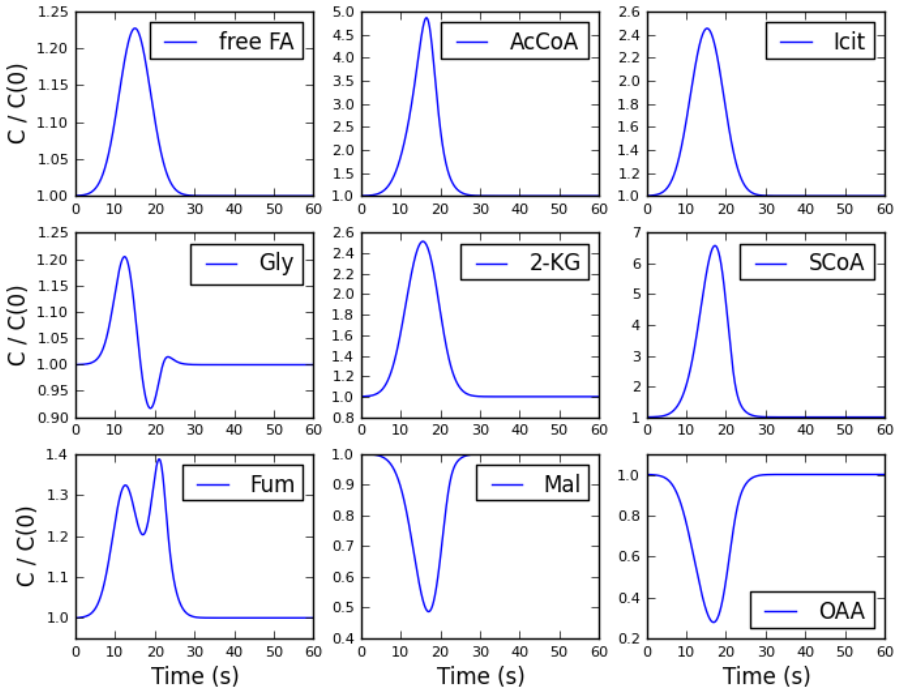



Figure 8.4.: Response of the model to a temporary (Gaussian) increase in extracellular fatty acid concentration. The graphs show concentrations of intermediates normalised to their starting concentrations

Figure 8.4 shows the effects of a temporary increase in fatty acid concentration. Most intermediates showed a temporary increase in concentration as well, except for malate and oxaloacetate which decreased temporarily. The valleys in the traces of the latter two coincided exactly with the peak in acetyl-CoA concentration (which occurred slightly later than the peak in extracellular fatty acid concentration), with an equally abrupt return to their steady state levels. The most likely reason for the decrease of these intermediates is an increased consumption of oxaloacetate by CS at higher acetyl-CoA concentrations, as stated above.

The concentration of glyoxylate increased, dropped below its initial value, increased again and then returned to its steady state level. The valley in the glyoxylate trace almost coincided with the peak in acetyl-CoA trace and the valley in the malate trace, which means that the rate of MS was probably temporarily increased as a result of a higher acetyl-CoA supply and lower product inhibition from malate.

The fumarate concentration follows a more complex pattern. An initial increase was

caused by an increase in succinate supply due to increased fluxes through the TCA cycle and the glyoxylate shunt. The concentration then decreased when the concentration of malate was at its lowest, during which time fumarase experienced a relatively low product inhibition. The concentration of fumarate then increased again when the concentration of malate returned to its original level. The concentration of fumarate then returned to its steady state level, possibly because the supply of succinate had also returned to its pre-perturbation level.

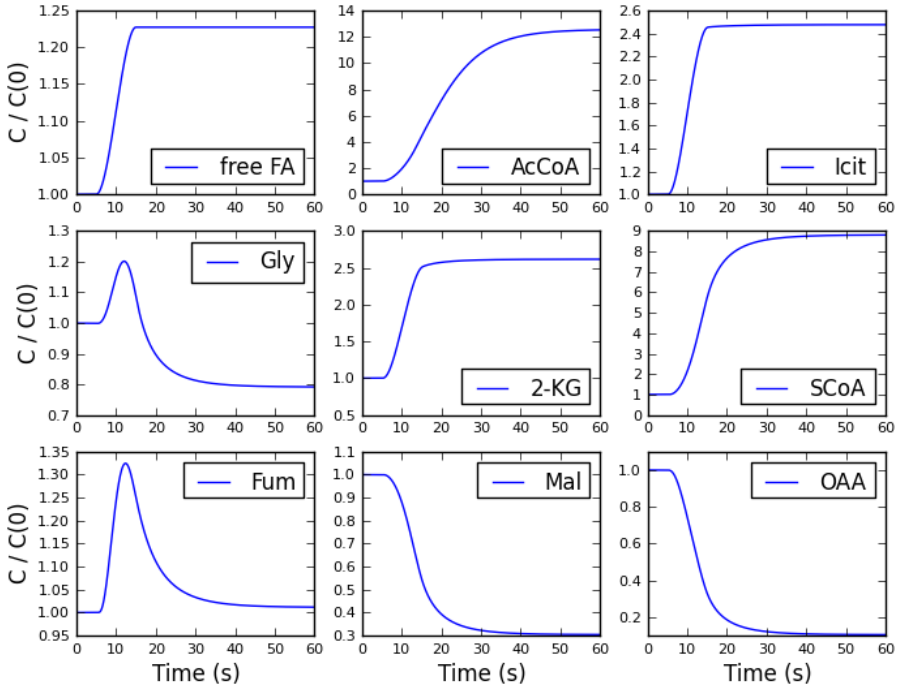



Figure 8.5.: Response of the model to a permanent increase in extracellular fatty acid concentration. The graphs show concentrations of intermediates normalised to their starting concentrations

Figure 8.5 shows the same model, but with a permanent increase in extracellular fatty acid concentration instead of a temporary one. The concentration of acetyl-CoA increased as with the previous perturbation, but kept increasing since the fatty acid concentration did not return to its original value.

The concentration of glyoxylate increased initially, probably due to a higher ICL rate, but then settled at a lower value. The concentrations of malate and oxaloacetate also settled at lower values, and again these can be explained by an increase of consumption made possible by the higher availability of acetyl-CoA.

Fumarate initially increased in concentration but then returned to a slightly higher

level than the starting concentration. The increase is likely to be the result of an increase in production of its precursor, succinate, via both the TCA cycle and the glyoxylate shunt. The following decrease in its concentration is probably caused by an increased consumption by fumarase, since a lower concentration of malate gives lower product inhibition of that enzyme.

### 8.2.5. Validation of the *S. coelicolor* model by simulation using published glucose uptake values and metabolic fluxes

Singh & Ghosh (2006) validated their model by comparing predicted fluxes with experimental flux distribution data published by Zhao & Shimizu (2003). The model described in Section 8.2.3 describes metabolism during growth on fatty acids. To our knowledge no flux distribution data for *S. coelicolor* grown in these conditions are available, therefore no such comparison can be made. In order to test whether the results of Section 8.2.3 are plausible, the model for growth on glucose that was described in Section 8.1 was adapted to describe the TCA cycle of *S. coelicolor* the same way the model for acetate utilisation was adapted to describe *S. coelicolor* growth on fatty acids. The results of this model's simulation were compared to published flux distribution data (Coze *et al.*, 2013).

In order to adapt the model, a number of adaptations were made that mimicked those described in Sections 8.2.2 and 8.2.3. Oxaloacetate was converted from a boundary condition to a free metabolite. A biosynthesis reaction was added which consumed oxaloacetate to produce biomass (SYN2, replacing SYN2 from the model of Section 8.1), as well as an anapleurotic reaction producing oxaloacetate from phosphoenol pyruvate (labelled phosphoenolpyruvate carboxylase (PEPC)). Acetyl-CoA production from pyruvate by the pyruvate dehydrogenase (PDH) complex was calculated from flux data (Coze *et al.*, 2013) using Table 8.4. Since no flux through the glyoxylate shunt was detected, these reactions were omitted from the model (Coze *et al.*, 2013). A schematic representation of the model system is shown in Fig. 8.6 and an overview of the values for fluxes through PEPC, PDH, SYN and SYN2 is given in Table 8.8.

The model shown in Fig. 8.6 was simulated until it reached a steady state. The predicted reaction rates are shown in Table 8.9, alongside the experimental fluxes reported in Coze *et al.* (2013). The reaction rates in both columns are nearly identical, showing that the *E. coli* model described in Section 8.1, following adaptation according to the process described in Section 8.2.3, is capable of simulating *S. coelicolor* metabolism realistically. Since no intracellular metabolite concentrations were measured in the experimental work, these predictions could not be validated. Samples for metabolomic

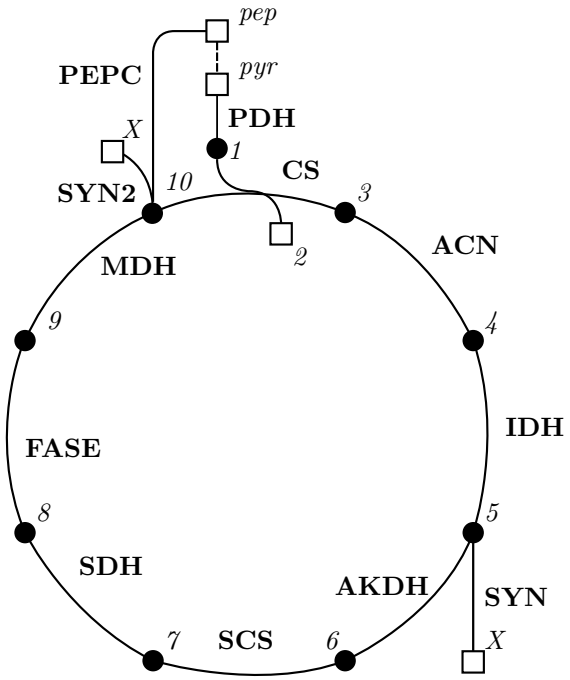



Figure 8.6.: Schematic representation of the modelled system during growth on glucose. Reactions are denoted by the name of the enzyme that catalyses them (bold font) and intermediates are represented by numbered circles and boxes, where a box (□) represents a system boundary and a circle (•) represents a free intermediate. Numbered intermediates are 1: Acetyl-CoA, 2: CoA, 3: Citrate, 4: Isocitrate, 5: 2-Ketoglutarate, 6: Succinyl-CoA, 7: Succinate, 8: Fumarate, 9: Malate, 10: Oxaloacetate and X: Biomass. *pep* and *pyr* denote phosphoenol pyruvate and pyruvate, respectively.

| **Reaction** | **Exp. rate** $\mu\text{mol}\,g_{DCW}^{-1}\,h^{-1}$ | **Rate** $\text{mM}\,\text{min}^{-1}$ |
|---|---|---|
| PEPC | 138.8 | 0.337 |
| PDH | 676.8 | 1.612 |
| SYN | 29.7 | $0.0438 \cdot \text{IDH}$ |
| SYN2 | 109.1 | $0.1686 \cdot \text{MDH}$ |

Table 8.8.: Experimental (Coze *et al.*, 2013) and converted rates for PEPC, PDH, SYN and SYN2. In the model, the fluxes of SYN and SYN2 were scaled to the rates of IDH and MDH, respectively, as was done in Section 8.1.

| **Reaction** | **Rates** ($\mathrm{mM\,min^{-1}}$) | |
|---|---|---|
| | **Experimental** | **Predicted** |
| PDH | 1.642 | 1.642 |
| PEPC | 0.337 | 0.337 |
| CS | 1.642 | 1.642 |
| ACN | 1.642 | 1.642 |
| IDH | 1.642 | 1.642 |
| AKDH | 1.571 | 1.570 |
| SCS | 1.571 | 1.570 |
| SCH | 1.571 | 1.570 |
| FASE | 1.571 | 1.570 |
| MDH | 1.571 | 1.570 |
| SYN | 0.072 | 0.072 |
| SYN2 | 0.265 | 0.265 |

Table 8.9.: Reference (Coze *et al.*, 2013) and predicted fluxes through the TCA cycle during growth on glucose, calculated by simulating the model shown in Fig. 8.6.

analysis were taken during the experiments described in Chapter 7, but at the moment of writing the analysis of these data is still ongoing.

## 8.3. Discussion

In this chapter a mathematical model for the TCA cycle and glyoxylate shunt was created using published *E. coli* data. Predicted flux distributions corresponded to experimental flux distribution data. Since no information on carbon uptake was available for the experimental data and the acetyl-CoA concentration in the model was fixed, rather than dependent on its production at a known rate, no conclusions could be drawn concerning the validity of the absolute flux values.

The mathematical model was adapted to describe the TCA cycle and glyoxylate shunt in *S. coelicolor*. Oxaloacetate was changed from a system boundary to a free intermediate and the fixed acetyl-CoA concentration was replaced with a production term, based on oxidation of palmitic acid. Realistic values for extracellular fatty acid concentrations and an expression combining passive and active uptake of this carbon source based the model on relevant growth conditions.

The model is sensitive to changes in the concentration of extracellular fatty acids, as shown by simulations run with differing starting conditions as well as by perturbation of steady state conditions. The method of transforming the *E. coli* acetate model to

an *S. coelicolor* fatty acids model was validated by applying the same transformation to the glucose model and comparing the predicted fluxes to published experimental data. The predicted and experimental fluxes corresponded, meaning the predictions of the *S. coelicolor* model for growth on fatty acids are credible with respect to fluxes. Predicted concentrations of individual metabolites could not be validated this way.

The cumulative flux of biosynthesic reactions SYN and SYN2 was equated to the flux through the glyoxylate shunt in the model for growth on fatty acids, allowing the system to reach a steady state. No *in-vivo* flux distribution data for growth in these conditions were available, making this artificial constraint necessary.

During simulations carried out at increasing extracellular fatty acid concentrations, the configuration of the TCA cycle changed from supplying precursors for growth through gluconeogenesis, enabled by a high concentrations of malate and oxaloacetate, to supplying precursors for polyketide biosynthesis by an increased concentration of acetyl-CoA and succinyl-CoA.

Between the simulations using the lowest and highest acetyl-CoA production rates, the concentrations of malate and oxaloacetate decreased by 80 % and 98.5 % respectively, while the concentrations of succinyl-CoA and acetyl-CoA increased 50 and 740-fold. Whether this reconfiguration occurs *in-vivo* and coincides with the onset of secondary metabolism remains unclear, as the model could not be validated with respect to metabolite concentrations. However, since realistic extracellular fatty acid concentrations were used in the simulations, the general trends revealed in Fig. 8.3 are possibly significant. Quantification of intracellular metabolite pools over a time course spanning the various growth phases, such as performed by Wentzel *et al.* (2012), is necessary to investigate this phenomenon and could provide powerful evidence for the accuracy of this model.

A ratio between the steady state reaction rates of IDH and ICL during simulation at various extracellular fatty acid concentrations was calculated. This ratio decreased linearly with an increasing acetyl-CoA production rate: 1238 for $0.99\,\mathrm{mM\,min^{-1}}$ to 1074 for $7.74\,\mathrm{mM\,min^{-1}}$ as a higher concentration of acetyl-CoA allowed a higher flux through the glyoxylate shunt. This ratio did not decrease further at acetyl-CoA production rates higher than $7.74\,\mathrm{mM\,min^{-1}}$. The values of this ratio show that IDH is always at least three orders of magnitude more active than ICL under a wide range of conditions.

This relatively low activity of ICL is not surprising, since the activity of IDH under these growth conditions is 38-fold higher (Table 8.5) and its $K_M$ value for isocitrate is 35-fold lower. It follows that the glyoxylate shunt is unlikely to have much impact on this part of metabolism, especially since RNA sequencing data show that the genes of the ethylmalonyl-CoA pathway, which provides similar anapleurotic functionality as

the glyoxylate shunt, are much more highly expressed (Section 7.3.1) than those of the glyoxylate shunt.

# 9. General discussion

Members of the genus *Streptomyces* produce many clinically relevant secondary metabolites. Biosynthesis of these compounds utilises precursors from primary metabolic pathways, such as the TCA cycle. Withdrawing of TCA cycle intermediates results in their depletion, but they are replenished through anapleurotic reactions. This is especially important during growth on fatty acids, during which gluconeogenesis imposes an extra burden on the level of TCA cycle intermediates. *S. coelicolor* possesses the genetic potential for two anapleurotic pathways capable of replenishing these during growth on fatty acids; the glyoxylate shunt and the ethylmalonyl-CoA pathway (Chapter 1). In this thesis, the function of the glyoxylate shunt in streptomycete metabolism was investigated using sequence analysis, genetic manipulation, transcriptomics and mathematical modelling.

## 9.1. Sequence analysis

Of the two anapleurotic pathways, only the ethylmalonyl-CoA pathway is present in all streptomycete genomes. The glyoxylate shunt is absent from certain strains, but is highly conserved in those that possess it. The two genes that are considered indicative of the presence of either pathway, *ccr* and *aceA*, are subject to purifying selection pressure (Section 3.2). This leads to the conclusion that both pathways are functional.

Analysis of sequences upstream of *aceA*, *ccr* and *aceB1* revealed a 15 bp palindromic motif. A search of the *S. coelicolor* genome for this motif resulted in a list of potentially coregulated genes. Some of these sites were found to be conserved in other streptomycete genomes (Section 3.3).

## 9.2. Mutagenesis

The function of the glyoxylate shunt as an anapleurotic pathway during growth of *S. coelicolor* on fatty acids was investigated using a $aceA^{-}aceB2^{-}$ mutant. The genes were deleted from the genome and the mutation was complemented by reintroducing the

genes. Tween 40 was used as a growth substrate because its solubility is better than that of fatty acids. However no effect of the gene deletion on growth in this medium was detected. It was concluded that the glyoxylate shunt is not essential for growth on fatty acids (Chapter 4).

## 9.3. Enzymes of the isocitrate branchpoint

When the glyoxylate shunt genes are expressed, the TCA cycle intermediate isocitrate forms a branchpoint between the enzymes IDH and ICL. The relative fluxes through the two enzymes will depend mostly on their kinetic parameters, which are essential parameters for the mathematical model. In order to determine these parameters the genes *aceA* and *idh* were cloned, expressed heterogously and purified. Enzyme assays showed that the $K_M$ of IDH for isocitrate is much lower than that of ICL as well as a higher specific activity. It can therefore be expected that the flux through the TCA cycle is higher than the flux through the glyoxylate shunt (Chapter 5).

## 9.4. Spatial distribution

When modelling metabolic pathways, homogeneity of the cytoplasm is usually assumed. Although this assumption is generally valid for single prokaryotic cells, *S. coelicolor*'s mycelial growth presents a more complex situation. In liquid cultures, the mycelium forms small clumps. These clumps experience a radial availability gradient of oxygen, substrates and other solutes (Nieminen *et al.*, 2013). In these conditions homogeneity cannot be assumed. In order to address this, distribution of *ccr* and *aceA* expression was investigated using fluorescence microscopy. No evidence for a spatial distribution was found for either gene (Chapter 6), although higher resolution analysis data could reveal as yet hidden patterns. Since the development of such a method was beyond the scope of this investigation it was not pursued further. More accurate correlation of the data with growth phases is also desirable.

## 9.5. Gene expression and enzyme activity *in-vivo*

A comparison was made between gene expression of *S. coelicolor* grown in minimal medium containing either glucose or Tween 40 using total RNA sequencing. No expression of glyoxylate shunt genes was detected while genes of the ethylmalonyl-CoA pathway were clearly upregulated in the Tween 40 cultures compared to the glucose

cultures. Enzyme assays on cell extracts revealed low ICL activity compared to the IDH activity (Chapter 7).

The results described here appear to directly contradict those of the microscopy experiments described in Chapter 6 (Section 9.4). Due to the careful timing and relatively high precision of the RNA sequencing experiment, these results are more reliable. It is possible however that, due to the differences in inoculum preparation and culture size, growth phases of the mycelium between the two experiments differ greatly. It is expected that, were the microscopy experiments to be repeated using the same culturing conditions used in the RNA sequencing experiments, results would correlate more accurately.

Other genes involved in gluconeogenesis were also upregulated while those involved in glucose metabolism were downregulated. Biosynthesis of a carotenoid (isorenieratene) and a lipid (aminotrihydroxybacteriohopane) were upregulated, presumably to protect the mycelium against the deleterious effects of fatty acid utilisation and the surfactant in the medium, respectively.

Aminotrihydroxybacteriohopane prevents dessication of aerial mycelium by influencing membrane packing (Poralla *et al.*, 2000) and probably serves a similar function during growth in the medium used here. The carotenoid isorenieratene is produced as a response to illumination with blue light to combat photooxidative stress (Takano *et al.*, 2005). Oxidative stress possibly also occurs as a result of fatty acid utilisation (Speijer *et al.*, 2014). The expression of the isorenieratene biosynthesis pathway may therefore be a response to oxidative stress in general and not just when it is photoinduced.

## 9.6. Mathematical modelling of the TCA cycle and the glyoxylate shunt

A mathematical model was created that describes the TCA cycle and the glyoxylate shunt. It was first created and verified using *E. coli* data (Singh & Ghosh (2006), Section 8.1) and then adapted to *S. coelicolor*. Expressions were introduced for the production of acetyl-CoA for simulation of growth on both glucose and Tween 40 (Sections 8.2.3 and 8.2.5). In the former, an anapleurotic reaction for the production of oxaloacetate from PEP was introduced as well.

Since no experimental data on metabolic fluxes of *S. coelicolor* grown on Tween 40 were available, the model predictions could not be verified. However, the conversion of the *E. coli* model to a *S. coelicolor* model was verified using flux distribution data for growth on glucose, for which accurate data were available (Coze *et al.*, 2013).

Simulations of the *S. coelicolor* model for Tween 40 showed that the TCA cycle has a maximum capacity for acetyl-CoA uptake. When the acetyl-CoA production from $\beta$-oxidation approaches this limit as a consequence of a higher extracellular fatty acid concentration during growth, the relative concentrations of TCA cycle intermediates were predicted to change to limit gluconeogenesis and increase biosynthesis of polyketides.

Whether these predictions are biologically relevant is unknown, since the model does not take into account the flux through the ethylmalonyl-CoA pathway, potential overflow metabolism and does not contain experimental estimates for the rate of gluconeogenesis and biosynthesis of amino acids. However, if these data become available in the future, extra constraints can easily be introduced into the developed model to increase its accuracy.

## 9.7. Relative impact of the ethylmalonyl-CoA pathway and the glyoxylate shunt on streptomycete metabolism

Growth on $C_2$ substrates or fatty acids requires gluconeogenesis, and hence anapleurosis. Based on work in *E. coli*, the glyoxylate shunt was thought to be the main anapleurotic pathway in *S. coelicolor* in these conditions (Kornberg, 1966). However, the experiments described in this thesis show that this is not the case, as deletion of the genes encoding the glyoxylate shunt enzymes has no effect on growth on fatty acids. Moreover, transcriptomic analysis of mid-exponential growth of *S. coelicolor* failed to detect expression of these genes under those conditions. In contrast, expression of genes encoding enzymes of the ethylmalonyl-CoA pathway, which can provide anapleurosis and precursors for secondary metabolism, was detected and at significantly higher levels during growth on Tween 40 than during growth on glucose. This is consistent with earlier experiments with *Streptomyces collinus*, where no ICL activity was found in oil-grown cultures (Akopiants *et al.*, 2006). A *ccr*$^-$ mutant of *S. collinus* was unable to grow on this medium (Akopiants *et al.*, 2006).

ICL activity was detected in *S. coelicolor* using enzyme assays on raw cell extracts, but activity was very low and mathematical modeling indicated that the flux through the glyoxylate shunt was minimal compared to that through the TCA cycle. The glyoxylate shunt was concluded not to impact *S. coelicolor* growth on Tween 40. The ethylmalonyl-CoA pathway is therefore likely to be the main anapleurotic pathway during growth on these conditions, as it is in other bacterial strains (Akopiants *et al.*, 2006; Alber *et al.*, 2006; Peyraud *et al.*, 2009). The absence of the glyoxylate shunt genes from the closely

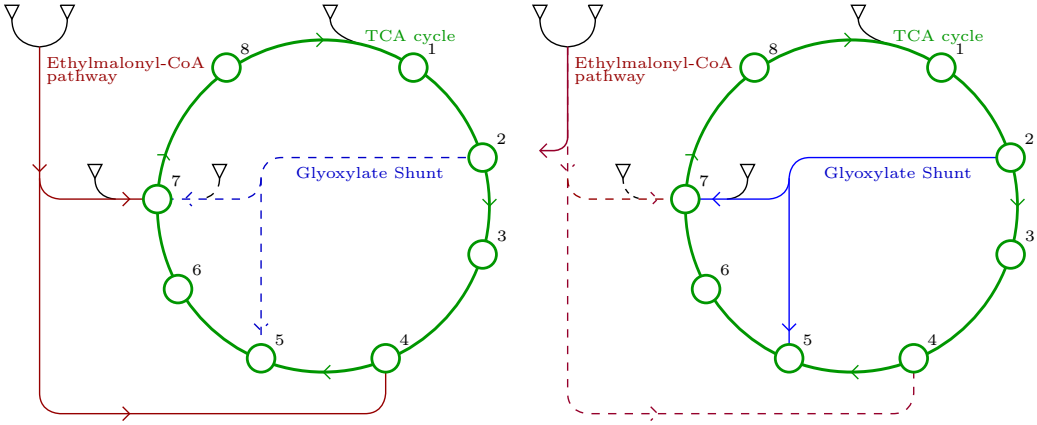
related *S. lividans* genome also supports this conclusion (Cruz-Morales *et al.*, 2013; Lewis *et al.*, 2010).

These results could be interpreted to mean that the pathway is redundant in *S. coelicolor*, especially since all streptomycete genomes contain orthologues of the ethylmalonyl-CoA pathway genes, but sometimes lack glyoxylate shunt gene orthologues. However, analysis of the sequences of *aceA*, *ccr* and genes that encode TCA cycle enzymes or subunits thereof shows that *aceA* undergoes purifying selection pressure to a similar extent as other central carbon metabolic genes. Therefore natural selection on *aceA* is at least as strong as on genes encoding other enzymes of central carbon metabolism. It also encodes a fully functional ICL enzyme, which displays kinetic parameters similar to ICLs from other organisms. This means the pathway still has a function in streptomycete metabolism, but it is different from the one it was originally hypothesised to be. In contrast, *Sco4010* was hypothesised to be a pseudogene by Kato *et al.* (2005) and indeed had a mean $dN/dS$ ratio close to 1.

Analysis of the intergenic sequences upstream of *aceA* and *ccr* also revealed a shared 15 bp palindromic motif, one which was also found upstream of other genes that are involved in anapleurosis and gluconeogenesis. It is therefore likely that the function of the glyoxylate shunt is at least related to providing anapleurosis during growth on $C_2$ substrates.

Except for glyoxylate, intermediates of the ethylmalonyl-CoA pathway are all acyl-CoA thioesters (Erb *et al.*, 2009). Many acyl-CoA thioesters are precursors for polyketides, either as starter- or extender units (Hertweck, 2009; Wilson & Moore, 2012). Before the ethylmalonyl-CoA pathway was elucidated, Liu & Reynolds (1999) described how crotonyl-CoA carboxylase / reductase (CCR) provides precursors for monensin A production in *S. cinnamonensis*. It is possible that when a *Streptomyces* strain grows (mainly) on fatty acids, the ethylmalonyl-CoA pathway is the primary anapleurotic pathway (Fig. 9.1a). At the onset of secondary metabolism however, intermediates of the ethylmalonyl-CoA pathway can be withdrawn as precursors for biosynthesis of polyketides (depending on the polyketide). A backup anapleurotic pathway would then be needed to provide the organism with new TCA cycle intermediates and allow biosynthesis of cell constituents to proceed. The glyoxylate shunt could be this backup pathway (Fig. 9.1b). ICL acitivity in *S. clavuligerus* was found in acetate-grown cultures by Soh *et al.* (2001) but only during cessation of growth, which supports this hypothesis.

As shown in the introduction (Section 1.5), the ethylmalonyl-CoA pathway is more carbon- and energy efficient than the glyoxylate shunt but its operation requires the production of many more proteins, which is costly in terms of both energy- and carbon




(a) Anapleurosis using the ethylmalonyl-CoA pathway during exponential growth on fatty acids

(b) Withdrawal of ethylmalonyl-CoA pathway intermediates necessitates use of the glyoxylate shunt for anapleurosis

Figure 9.1.: Schematic visualisation of the TCA cycle (green), the glyoxylate shunt (blue) and ethylmalonyl-CoA pathway (red), and the flow of acetyl-CoA (black) during exponential growth (Fig. 9.1a) and secondary metabolism (Fig. 9.1b). Labelled intermediates are $\bigtriangledown$: Acetyl-CoA, 1: Citrate, 2: Isocitrate, 3: $\alpha$-ketoglutarate, 4: Succinyl-CoA, 5: Succinate, 6: Fumarate, 7: Malate, 8: Oxaloacetate.

costs. This represents a trade-off with respect to efficiency since organisms have limited protein synthesis capacity (Nyström, 2004).

Trade-offs between different phenotypes are commonplace in biological systems, e.g. between expressing genes involved in growth and survival (Nyström, 2004). Maximum fitness in a given environment is often an optimum balance between two phenotypes specialised for single tasks (archetypes) (Shoval *et al.*, 2012). Streptomycetes may face this optimisation problem when dividing carbon and energy between the production of cell constituents (anapleurosis and gluconeogenesis) and secondary metabolites. When biosynthesis of an antibiotic relies heavily on the ethylmalonyl-CoA pathway for its precursor supply, expression of the glyoxylate shunt genes may be the answer to this optimisation problem, as it allows continuing production of cell constituents where secondary metabolism would otherwise prevent that.

Different streptomycetes produce different secondary metabolites and the production of these have different precursor requirements. The synthesis of one molecule of Monensin A by *S. cinnamonensis* requires one molecule of ethylmalonyl-CoA. *S. cinnamonensis* does not possess the glyoxylate shunt genes, but the ethylmalonyl-CoA requirement for Monensin A production may not be high enough to necessitate an alternative anapleu-

rotic pathway. A difference in need for a backup pathway due to a difference in secondary metabolism may explain why the glyoxylate shunt genes are absent from some strains while they are subject to strong purifying selection pressure in others.

## 9.8. Future work

Various experiments could be carried out to test the hypotheses outlined in Section 9.7. The importance of the ethylmalonyl-CoA pathway as the main anapleurotic pathway in *S. coelicolor* could easily be tested by producing a *ccr*$^-$ strain and growing it in a minimal medium containing Tween 40 as sole carbon source. This mutation is likely to result in significant growth defects under these conditions. The glyoxylate shunt could possibly compensate for this mutation, ICL activity would then be much higher than that measured in Chapter 7. This is as yet unclear, since an *S. cinnamonensis ccr*$^-$ mutant that expressed *S. coelicolor* glyoxylate shunt genes grew on oil, but not on acetate as the sole carbon source whereas the wild type strain grew on both carbon sources (Akopiants *et al.*, 2006).

The function of the glyoxylate shunt as a backup system could be tested by introducing an inducible pathway or biosynthesis cluster that withdraws ethylmalonyl-CoA pathway intermediates into *S. coelicolor* M145 and the *aceA*$^-$*aceB2*$^-$ strain RAR101. Growing these strains in minimal medium containing Tween 40 as sole carbon source should also result in phenotypic differences, either in the growth characteristics or in the secondary metabolite yield.

# Appendix A.

# PCR primers

This appendix contains a table of PCR primers used for this thesis. In the primer names, *fw* denotes a forward primer, *rev* denotes a reverse primer.

| **Primer name** | **Sequence** |
|---|---|
| 0982exp-fw | 5'-ttttcatatggcagaagcgaggacg |
| 0982exp-rev | 5'-ttttctcgaggtggaactgctcctcctccg |
| 7000exp-fw | 5'-ttttcatatgactgactcgaccatcatct |
| 7000exp-rev | 5'-ttttctcgaggccgagggacgccagcg |
| 0982red-fw | 5'-tggcaatcggtacagaccgaggagacggtg-acagtcatgattccggggatccgtcgacc |
| 0983red-rev | 5'-caccgccgaccacgccgacgcgacggatgc-ccccgctcatgtaggctggagctgcttc |
| ccrO-fw | 5'-gagctggaccacgccttc |
| ccrO-rev | 5'-gttacgtcggtgctgctgag |
| iclO-fw | 5'-gctccatctggttcaggtagc |
| iclO-rev | 5'-gactgtcaccgtctcctcgg |
| EZL2 (Bishop *et al.*, 2004) | 5'-tccagctcgaccaggatg |

Table A.1.: PCR primers used in this report. Restriction sites used in cloning are underlined.

# Appendix B.

# Non-linear regression for estimation of parameters in enzyme kinetics

In order to calculate the kinetic parameters for the forward reaction of ICL, a weighted and non-linear regression method was used (Wilkinson, 1961), as outlined in algorithm B.1. This algorithm provides $K_{\mathrm{M}}$ and $V^{\max}$, as well as the relevant standard errors. Provisional estimates for $K_{\mathrm{M}}$ and $V^{\max}$ were calculated with Eqs. (B.1) and (B.2), and refined by repeated iterations of a second set of calculations Eqs. (B.5) and (B.6). In the equations, $s$ is the set of experimental substrate concentrations, $v$ the set of corresponding reaction rates and $n$ is the number of experimental data points.

As the refinement is repeated, $b_1$ and $b_2$ (Eqs. (B.3) and (B.4)) approach 1 and 0, respectively. A threshold value $T$ was set to which $b_1$ was compared. When $|b_1| - T \leq 1$, the calculated $K_{\mathrm{M}}$ and $V^{\max}$ were accepted.

## B.1. Implementation in Python

Although the calculations outlined above can be done by hand, they tend to be tedious. A script was written in the Python programming language to alleviate this. The code is given below:

$$
\begin{aligned}
x &= v^2 \\
y &= \frac{v^2}{s} \\
\alpha &= \sum vx \\
\beta &= \sum x^2 \\
\gamma &= \sum vy \\
\delta &= \sum xy \\
\epsilon &= \sum y^2 \\
\Delta &= \alpha\epsilon - \gamma\delta \\
K_{\mathrm{M}}^0 &= \frac{\beta\gamma - \alpha\delta}{\Delta} \qquad \text{(B.1)} \\
V^{\max,0} &= \frac{\beta\epsilon - \delta^2}{\Delta} \qquad \text{(B.2)}
\end{aligned}
$$

$$
\begin{aligned}
f &= \frac{V^{\max,0} \cdot s}{s + K_{\mathrm{M}}^0} \\
f' &= -\frac{V^{\max,0} \cdot s}{(s + K_{\mathrm{M}}^0)^2} \\
\alpha &= \sum f^2 \\
\beta &= \sum f'^2 \\
\gamma &= \sum ff' \\
\delta &= \sum vf \\
\epsilon &= \sum vf' \\
\Delta &= \alpha\beta - \gamma^2 \\
b_1 &= \frac{\beta\delta - \gamma\epsilon}{\Delta} \qquad \text{(B.3)} \\
b_2 &= \frac{\alpha\epsilon - \gamma\delta}{\Delta} \qquad \text{(B.4)} \\
K_{\mathrm{M}} &= K_{\mathrm{M}}^0 + \frac{b_2}{b_1} \qquad \text{(B.5)} \\
V^{\max} &= b_1 V^{\max,0} \qquad \text{(B.6)} \\
S^2 &= \frac{\sum v^2 - b_1\delta - b_2\epsilon}{n-2} \\
S.E.(K_{\mathrm{M}}) &= \frac{S}{b_1}\sqrt{\frac{\alpha}{\Delta}} \qquad \text{(B.7)} \\
S.E.(V^{\max}) &= V^0 S\sqrt{\frac{\beta}{\Delta}} \qquad \text{(B.8)}
\end{aligned}
$$

Algorithm B.1: Calculations for provisional estimates (left) and the consecutive refinement of these estimates (right). Upon iteration of the refinement, the $K_{\mathrm{M}}$ and $V^{\max}$ resulting from a previous refinement are used as $K_{\mathrm{M}}^0$ and $V^{\max,0}$ in the next iteration.

```
import numpy as np

Concs = np.array( [...] ) # Instert list of experimental concentrations
Rates = np.array( [...] ) # Instert list of determined reaction rates

def InitEstimates( s, v ):
    x = np.square( v )
    y = np.square( v ) / s
    alpha = np.sum( v * x )
    beta = np.sum( np.square( x ) )
    gamma = np.sum( v * y )
    delta = np.sum( x * y )
    epsilon = np.sum( np.square( y ) )
    Delta = ( alpha * epsilon ) - ( gamma * delta )
    Km = ( ( beta * gamma ) - ( alpha * delta ) ) / Delta
    Vmax = ( ( beta * epsilon ) - delta**2 ) / Delta
    return Km, Vmax

def KinCalc( s, v, Km0, Vmax0 ):
    # Calculates kinetic parameters according to method of G.N. Wilkinson
    Km0 = np.array( Km0 )
    Vmax0 = np.array( Vmax0 )
    divisor = s + Km0
    f = ( Vmax0 * s ) / ( s + Km0 )
    fprim = ( -Vmax0 * s ) / ( ( s + Km0 )**2 )
    alpha = np.sum( f**2 )
    beta = np.sum( fprim**2 )
    gamma = np.sum( f * fprim )
    delta = np.sum( v * f )
    epsilon = np.sum( v * fprim )
    Delta = ( alpha * beta ) - gamma**2
    b1 = ( ( beta * delta ) - ( gamma * epsilon ) ) / Delta
    b2 = ( ( alpha * epsilon ) - ( gamma * delta ) ) / Delta
    Km = np.array( Km0 + ( b2 / b1 ) )
    Vmax = np.array( b1 * Vmax0 )
```

```
    if ( np.abs( b1 ) - Threshold > 1 ):
        return KinCalc( s, v, Km, Vmax )
    else:
        Ssq = ( ( np.sum( np.square( v ) ) - ( b1 * delta ) -
( b2 * epsilon ) ) / ( len( s ) - 2 ) )
        SEKm = ( np.sqrt( Ssq ) / b1 ) * np.sqrt( alpha / Delta )
        SEVmax = ( Vmax0 * np.sqrt( Ssq ) ) * np.sqrt( beta / Delta )
        return Km, Vmax, SEKm, SEVmax

Threshold = 0.0001
Km0, Vmax0 = InitEstimates( Concs, Rates )
print 'initial Km + Vmax:', Km0, Vmax0
Km, Vmax, SEKm, SEVmax = KinCalc( Concs, Rates, Km0, Vmax0 )
print 'done (Km, Vmax, SEKm, SEVmax):', Km, Vmax, SEKm, SEVmax
```

# Appendix C.

# SNP extraction from ClustalW alignments

Nucleotide sequences were aligned in Mega 6.0 (Tamura *et al.*, 2013) using the ClustalW algorithm for codons. The result was saved in FASTA format. SNPs were then extracted by the Python script below and saved in a format accessible by TRAMS (Reumerman *et al.*, 2013). Text following a hash sign (#) are comments and have no influence on the working of the script.

```
from Bio import SeqIO
handle = open("aceA_nt_alignment_codons.fasta", "rU")
records = list(SeqIO.parse(handle, "fasta"))
handle.close()

Strains = []# List of strains
Lib = {}# Library of sequences
SNPs = {}# Library

Ref_strain = records[0].id
Ref_seq = records[0].seq
Start = 0
for i,base in enumerate(Ref_seq):
    if base != '-':
        Start = i
        break

Num = len(Ref_seq)
```

```
for record in records[1:]:
    Strains.append(record.id)
    Lib[record.id] = record.seq
    SNPs[record.id] = ['']*len(record.seq)

Output = '\t'.join(['Pos', 'Ref'] + Strains)

Inserts = 0# Keeps track of number of inserted bases
for i in range(Start,Num):
    Rb = Ref_seq[i]# Rb for Reference base
    if Rb == '-':
        Inserts += 1
        continue

    for strain in Strains:
        if Lib[strain][i] not in [Rb, '-']:
            if Lib[strain][i].lower() in ['a','t','g','c']:
                SNPs[strain][i] = Lib[strain][i]

    Line = [SNPs[strain][i] for strain in Strains]
    if Line == ['']*len(Strains):
        continue
    else:
        # Gene starts at 194 in gbk file
        Output = '\n'.join([Output, '\t'.join([`i+194-Start-Inserts`, Rb] + Line)])

Outfile = open("snps_codons.txt", 'w')
Outfile.write(Output)
Outfile.write('\n')
Outfile.close()
print 'finished'
```

# Bibliography


Abràmoff, Michael D, Hospitals, Iowa, Magalhães, Paulo J, & Abràmoff, Michael. 2007. Image Processing with ImageJ. *Biophotonics International*, **11**, 36–42.

Akopiants, Konstantin, Florova, Galina, Li, Chaoxuan, & Reynolds, Kevin A. 2006. Multiple pathways for acetate assimilation in *Streptomyces cinnamonensis*. *Journal of Industrial Microbiology & Biotechnology*, **33**, 141–50.

Alber, Birgit E, Spanheimer, Regina, Ebenau-Jehle, Christa, & Fuchs, Georg. 2006. Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. *Molecular Microbiology*, **61**, 297–309.

Alvarez-Álvarez, R, Rodríguez-García, A, Martínez-Burgo, Y, Robles-Reglero, V, Santamarta, I, Pérez-Redondo, R, Martín, J F, & Liras, P. 2014. A 1.8-Mb-reduced *Streptomyces clavuligerus* genome: relevance for secondary metabolism and differentiation. *Applied Microbiology and Biotechnology*, **98**, 2183–95.

Bailey, Timothy L, Boden, Mikael, Buske, Fabian a, Frith, Martin, Grant, Charles E, Clementi, Luca, Ren, Jingyuan, Li, Wilfred W, & Noble, William S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37**, W202–8.

Banchio, C, & Gramajo, Hugo C. 1997. Medium- and long-chain fatty acid uptake and utilization by *Streptomyces coelicolor* A3(2): first characterization of a gram-positive bacterial system. *Microbiology (Reading, England)*, **143**, 2439–47.

Beamer, L J, & Pabo, C O. 1992. Refined 1.8 A crystal structure of the lambda repressor-operator complex. *Journal of Molecular Biology*, **227**, 177–96.

Bentley, S D, Chater, K F, Cerdeño Tárraga, A-M, Challis, G L, Thomson, N R, James, K D, Harris, D E, Quail, M A, Kieser, H, Harper, D, Bateman, A, Brown, S, Chandra, G, Chen, C W, Collins, M, Cronin, A, Fraser, A, Goble, A, Hidalgo, J, Hornsby, T, Howarth, S, Huang, C-H, Kieser, T, Larke, L, Murphy, L, Oliver, K, O'Neil, S, Rabbinowitsch, E, Rajandream, M-A, Rutherford, K, Rutter, S, Seeger, K, Saunders, D, Sharp, S, Squares, R, Squares, S, Taylor, K, Warren, T, Wietzorrek, A, Woodward,

J, Barrell, B G, Parkhill, J, & Hopwood, D A. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141–7.

Bérdy, János. 2005. Bioactive Microbial Metabolites. *Journal of Antibiotics*, **58**, 1–26.

Birnboim, H.C., & Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, **7**, 1515–1523.

Bishop, Amy, Fielding, Sue, Dyson, Paul, & Herron, Paul. 2004. Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production. *Genome Research*, **14**, 893–900.

Borodina, Irina, Siebring, Jeroen, Zhang, Jie, Smith, Colin P, van Keulen, Geertje, Dijkhuizen, Lubbert, & Nielsen, Jens. 2008. Antibiotic overproduction in *Streptomyces coelicolor* A3 2 mediated by phosphofructokinase deletion. *The Journal of Biological Chemistry*, **283**, 25186–99.

Bradford, M M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, **72**, 248–54.

Bucca, Giselda, Laing, Emma, Mersinias, Vassilis, Allenby, Nicholas, Hurd, Douglas, Holdstock, Jolyon, Brenner, Volker, Harrison, Marcus, & Smith, Colin P. 2009. Development and application of versatile high density microarrays for genome-wide analysis of *Streptomyces coelicolor*: characterization of the HspR regulon. *Genome Biology*, **10**, R5.

Cane, David E. 2010. Programming of erythromycin biosynthesis by a modular polyketide synthase. *Journal of Biological Chemistry*, **285**, 27517–27523.

Chapman, Paul G. 1994. *Studies of the glyoxylate bypass of Streptomyces coelicolor A3(2) : purification of isocitrate lyase and cloning of the icl gene.* PhD Thesis, University of Glasgow.

Chater, Keith F, Biró, Sandor, Lee, Kye Joon, Palmer, Tracy, & Schrempf, Hildgund. 2010. The complex extracellular biology of *Streptomyces*. *FEMS Microbiology Reviews*, **34**, 171–98.

Chubukov, Victor, Uhr, Markus, Le Chat, Ludovic, Kleijn, Roelco J, Jules, Matthieu, Link, Hannes, Aymerich, Stephane, Stelling, Jörg, & Sauer, Uwe. 2013. Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Molecular Systems Biology*, **9**, 1–13.

Claessen, Dennis, Rozen, Daniel E, Kuipers, Oscar P, Sø gaard Andersen, Lotte, & van Wezel, Gilles P. 2014. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nature Reviews. Microbiology*, **12**, 115–24.

Cochrane, Vincent W., Peck Jr., Harry D., & Harrison, Anne. 1953. The metabolism of species of *Streptomyces*. VII. The hexosemonophosphate shunt and associated reactions. *Journal of bacteriology*, **66**, 17–23.

Cortes, J, Haydock, S F, Roberts, G A, Bevitt, D J, & Leadlay, P F. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature*, **348**, 176–178.

Coze, Fabien, Gilard, Françoise, Tcherkez, Guillaume, Virolle, Marie-Joëlle, & Guyonvarch, Armel. 2013. Carbon-Flux Distribution within *Streptomyces coelicolor* Metabolism: A Comparison between the Actinorhodin-Producing Strain M145 and Its Non-Producing Derivative M1146. *PloS One*, **8**, e84151.

Cozzone, Alain J. 1998. Regulation of acetate metabolism by protein phosphorylation in enteric bacteria. *Annual Review of Microbiology*, **52**, 127–64.

Cruz-Morales, Pablo, Vijgenboom, Erik, Iruegas-Bocardo, Fernanda, Girard, Geneviève, Yáñez Guerra, Luis Alfonso, Ramos-Aboites, Hilda E, Pernodet, Jean-Luc, Anné, Jozef, van Wezel, Gilles P, & Barona-Gómez, Francisco. 2013. The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biology and Evolution*, **5**, 1165–75.

Datsenko, K a, & Wanner, B L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6640–5.

Derouaux, a., Halici, S., Nothaft, H., Neutelings, T., Moutzourelis, G., Dusart, J., Titgemeyer, F., & Rigali, S. 2004a. Deletion of a Cyclic AMP Receptor Protein Homologue Diminishes Germination and Affects Morphological Development of *Streptomyces coelicolor*. *Journal of Bacteriology*, **186**, 1893–1897.

Derouaux, Adeline, Dehareng, Dominique, Lecocq, Elke, Halici, Serkan, Nothaft, Harald, Giannotta, Fabrizio, Moutzourelis, Georgios, Dusart, Jean, Devreese, Bart, Titgemeyer, Fritz, Van Beeumen, Jozef, & Rigali, Sébastien. 2004b. Crp of *Streptomyces coelicolor* is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP. *Biochemical and Biophysical Research Communications*, **325**, 983–90.

Dunn, M F, Ramírez-Trujillo, J A, & Hernández-Lucas, I. 2009. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology (Reading, England)*, **155**, 3166–75.

Efthimiou, G, Thumser, a E, & Avignone-Rossa, C a. 2008. A novel finding that *Streptomyces clavuligerus* can produce the antibiotic clavulanic acid using olive oil as a sole carbon source. *Journal of Applied Microbiology*, **105**, 2058–64.

Efthimiou, Giorgos. 2008. *The role of olive oil utilisation in clavulanic acid biosynthesis by Streptomyces clavuligerus*. PhD Thesis, University of Surrey.

El-Mansi, E M, MacKintosh, C, Duncan, K, Holms, W H, & Nimmo, H G. 1987. Molecular cloning and over-expression of the glyoxylate bypass operon from *Escherichia coli* ML308. *The Biochemical Journal*, **242**, 661–5.

Erb, Tobias J, Berg, Ivan a, Brecht, Volker, Müller, Michael, Fuchs, Georg, & Alber, Birgit E. 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10631–6.

Erb, Tobias J, Fuchs, Georg, & Alber, Birgit E. 2009. (2S)-Methylsuccinyl-CoA dehydrogenase closes the ethylmalonyl-CoA pathway for acetyl-CoA assimilation. *Molecular Microbiology*, **73**, 992–1008.

Fujita, Yasutaro, Matsuoka, Hiroshi, & Hirooka, Kazutake. 2007. Regulation of fatty acid metabolism in bacteria. *Molecular Microbiology*, **66**, 829–39.

Gerstmeir, Robert, Cramer, Annette, Dangel, Petra, Schaffer, Steffen, & Eikmanns, Bernhard J. 2004. RamB , a Novel Transcriptional Regulator of Genes Involved in Acetate Metabolism of *Corynebacterium glutamicum*. *Journal of Bacteriology*, **186**, 2798–2809.

Gregory, Matthew A, Till, Rob, & Smith, Margaret C M. 2003. Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *Journal of Bacteriology*, **185**, 5320–3.

Gubbens, Jacob, Janus, Marleen, Florea, Bogdan I, Overkleeft, Herman S, & van Wezel, Gilles P. 2012. Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics. *Molecular Microbiology*, **86**, 1490–507.

Gunstone, Frank D., Harwood, John L., & Dijkstra, Albert J. 2007. *The Lipid Handbook*. 3 edn. Boca Raton (FL, USA): Taylor & Francis Group.

Gust, Bertolt, Challis, Greg L, Fowler, Kay, Kieser, Tobias, & Chater, Keith F. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1541–6.

Hadlich, Frieder, Noack, Stephan, & Wiechert, Wolfgang. 2009. Translating biochemical network models between different kinetic formats. *Metabolic Engineering*, **11**, 87–100.

Hamed, Refaat B., Gomez-Castellanos, J. Ruben, Henry, Luc, Ducho, Christian, McDonough, Michael a., & Schofield, Christopher J. 2013. The enzymes of $\beta$-lactam biosynthesis. *Nat. Prod. Rep.*, **30**, 21–107.

Han, L, & Reynolds, K a. 1997. A novel alternate anaplerotic pathway to the glyoxylate cycle in streptomycetes. *Journal of Bacteriology*, **179**, 5157–64.

Haydock, Stephen F, Mironenko, Tatiana, Ghoorahoo, Haroun I, & Leadlay, Peter F. 2004. The putative elaiophylin biosynthetic gene cluster in *Streptomyces sp.* DSM4137 is adjacent to genes encoding adenosylcobalamin-dependent methylmalonyl CoA mutase and to genes for synthesis of cobalamin. *Journal of Biotechnology*, **113**, 55–68.

Hertweck, Christian. 2009. The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International ed. in English)*, **48**, 4688–716.

Hobbs, G., Frazer, C. M., Gardner, D. C. J., Flett, F., & Oliver, S. G. 1990. Pigmented antibiotic production by *Streptomyces coelicolor* A3(2): kinetics and the influence of nutrients. *Microbiology*, **136**, 2291–2296.

Hobbs, Glyn, Frazer, Catherine M, Gardner, David C J, Cullum, John A, & Oliver, Stephen G. 1989. Dispersed growth of *Streptomyces* in liquid culture. *Applied Microbiology and Biotechnology*, **31**, 272–277.

Höner Zu Bentrup, K, Miczak, Andras, Swenson, Dana L, & Russell, David G. 1999. Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *Journal of Bacteriology*, **181**, 7161–7.

Hönes, I., Simon, M., & Weber, H. 1991. Characterization of isocitrate lyase from the yeast *Yarrowia lipolytica*. *Journal of Basic Microbiology*, **31**, 251–258.

Hoops, Stefan, Sahle, Sven, Gauges, Ralph, Lee, Christine, Pahle, Jürgen, Simus, Natalia, Singhal, Mudita, Xu, Liang, Mendes, Pedro, & Kummer, Ursula. 2006. COPASI–a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, **22**, 3067–74.

Hopwood, David a, & Sherman, David H. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annual review of genetics*, **24**, 37–66.

Hoyt, J C, Johnson, K E, & Reeves, H C. 1991. Purification and characterization of *Acinetobacter calcoaceticus* isocitrate lyase. *Journal of Bacteriology*, **173**, 6844–8.

Huffman, Joy L, & Brennan, Richard G. 2002. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology*, **12**, 98–106.

Hunter, John D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, **9**, 90–95.

Hurley, J H, Thorsness, P E, Ramalingam, V, Helmers, N H, Koshland, D E, & Stroud, R M. 1989. Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 8635–9.

Hurst, Laurence D. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics : TIG*, **18**, 486.

Ikeda, H, Nonomiya, T, Usami, M, Ohta, T, & Omura, S. 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 9509–9514.

Karro, John E, Yan, Yangpan, Zheng, Deyou, Zhang, Zhaolei, Carriero, Nicholas, Cayting, Philip, Harrrison, Paul, & Gerstein, Mark. 2007. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Research*, **35**, D55–60.

Kato, Jun-ya, Hirano, Setsu, Ohnishi, Yasuo, & Horinouchi, Sueharu. 2005. The *Streptomyces* subtilisin inhibitor (SSI) gene in *Streptomyces coelicolor* A3(2). *Bioscience, Biotechnology, and Biochemistry*, **69**, 1624–9.

Kieser, Tobias, Bibb, Mervyn J, Buttner, Mark J., Chater, Keith F, & Hopwood, David A. 2000. *Practical Streptomyces Genetics*. 1 edn. Norwich: The John Innes Foundation.

Ko, Young-hee, & McFadden, Bruce A. 1990. The inhibition of isocitrate lyase from *Escherichia coli* by glyoxylate. *Current Microbiology*, **21**, 313–315.

Kornberg, H L. 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. *The Biochemical Journal*, **99**, 1–11.

Krügel, H, Krubasik, P, Weber, K, Saluz, H P, & Sandmann, G. 1999. Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *Biochimica et Biophysica Acta*, **1439**, 57–64.

Lewis, Richard a, Laing, Emma, Allenby, Nicholas, Bucca, Giselda, Brenner, Volker, Harrison, Marcus, Kierzek, Andrzej M, & Smith, Colin P. 2010. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *BMC Genomics*, **11**, 682.

Liu, H, & Reynolds, K a. 1999. Role of crotonyl coenzyme A reductase in determining the ratio of polyketides monensin A and monensin B produced by *Streptomyces cinnamonensis*. *Journal of Bacteriology*, **181**, 6806–13.

Loferer-Krössbacher, M, Klima, J, & Psenner, R. 1998. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology*, **64**, 688–94.

Loke, P, Goh, L L, Seng Soh, B, Yeow, P, & Sim, T S. 2002. Purification and characterization of recombinant malate synthase enzymes from *Streptomyces coelicolor* A3(2) and *S. clavuligerus* NRRL3585. *Journal of Industrial Microbiology & Biotechnology*, **28**, 239–43.

MacKintosh, C, & Nimmo, H G. 1988. Purification and regulatory properties of isocitrate lyase from *Escherichia coli* ML308. *The Biochemical Journal*, **250**, 25–31.

MacNeil, D J, Gewain, K M, Ruby, C L, Dezeny, G, Gibbons, P H, & MacNeil, T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, **111**, 61–8.

Marahiel, Mohamed a, Stachelhaus, Torsten, & Mootz, Henning D. 1997. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews*, **97**, 2651–2674.

McCormick, Joseph R, & Flärdh, Klas. 2012. Signals and regulators that govern *Streptomyces* development. *FEMS Microbiology Reviews*, **36**, 206–31.

Munir, Erman, Hattori, Takefumi, & Shimada, Mikio. 2002. Purification and characterization of isocitrate lyase from the wood-destroying basidiomycete *Fomitopsis palustris* grown on glucose. *Archives of Biochemistry and Biophysics*, **399**, 225–31.

Navone, Laura, Casati, Paula, Licona-Cassani, Cuauhtémoc, Marcellin, Esteban, Nielsen, Lars K, Rodriguez, Eduardo, & Gramajo, Hugo. 2014. Allantoin catabolism influences the production of antibiotics in *Streptomyces coelicolor*. *Applied Microbiology and Biotechnology*, **98**, 351–60.

Nieminen, Leena, Webb, Steven, Smith, Margaret C M, & Hoskisson, Paul A. 2013. A flexible mathematical model platform for studying branching networks: experimentally validated using the model actinomycete, *Streptomyces coelicolor*. *PloS One*, **8**, e54316.

Nishimura, M, & Sugiyama, M. 1994. Cloning and sequence analysis of a *Streptomyces* cholesterol esterase gene. *Applied Microbiology and Biotechnology*, **41**, 419–24.

Nyström, Thomas. 2004. Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Molecular Microbiology*, **54**, 855–62.

Ogata, Seiya, & Sadako, Yamada. 1990. Characterization of Adenine-Thiamine Auxotrophic Mutant of *Streptomyces azureus* ATCC 14921. *Actinomycetologica*, **4**, 7–10.

Ogawa, Tadashi, Murakami, Keiko, Mori, Hirotada, Ishii, Nobuyoshi, Tomita, Masaru, & Yoshin, Masataka. 2007. Role of phosphoenolpyruvate in the NADP-isocitrate dehydrogenase and isocitrate lyase reaction in *Escherichia coli*. *Journal of Bacteriology*, **189**, 1176–8.

Olano, Carlos, Lombó, Felipe, Méndez, Carmen, & Salas, José a. 2008. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metabolic Engineering*, **10**, 281–92.

Oliphant, Travis E. 2007. Python for Scientific Computing. *Computing in Science & Engineering*, **9**, 10–20.

Olson, James Allen, & Krinsky, Norman I. 1995. Introduction: The colorful, fascinating world of the carotenoids: important physiologic modulators. *FASEB Journal*, **9**, 1547–1550.

Olukoshi, E. R., & Packter, N. M. 1994. Importance of stored triacylglycerols in *Streptomyces*: Possible carbon source for antibiotics. *Microbiology*, **140**, 931–943.

Park, Sun-Yang, Moon, Min-Woo, Subhadra, Bindu, & Lee, Jung-Kee. 2010. Functional characterization of the glxR deletion mutant of *Corynebacterium glutamicum* ATCC 13032: involvement of GlxR in acetate metabolism and carbon catabolite repression. *FEMS Microbiology Letters*, **304**, 107–15.

Parker, Jennifer L, Jones, Alexandra M E, Serazetdinova, Liliya, Saalbach, Gerhard, Bibb, Mervyn J, & Naldrett, Mike J. 2010. Analysis of the phosphoproteome of the multicellular bacterium *Streptomyces coelicolor* A3(2) by protein/peptide fractionation, phosphopeptide enrichment and high-accuracy mass spectrometry. *Proteomics*, **10**, 2486–2497.

Peyraud, Rémi, Kiefer, Patrick, Christen, Philipp, Massou, Stephane, Portais, Jean-Charles, & Vorholt, Julia a. 2009. Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 4846–51.

Pinzauti, G, Giachetti, E, Camici, G, Manao, G, Cappugi, G, & Vanni, P. 1986. An isocitrate lyase of higher plants: analysis and comparison of some molecular properties. *Archives of Biochemistry and Biophysics*, **244**, 85–93.

Plou, Francisco J, Ferrer, Manuel, Nuero, Oscar M, Calvo, Maria V, Alcalde, Miguel, Reyes, Fuensanta, & Ballesteros, Antonio. 1998. Analysis of Tween 80 as an esterase / lipase substrate for lipolytic activity assay. *Biotechnology Techniques*, **12**, 183–186.

Popov, V N, Igamberdiev, A U, & Volvenkin, S V. 1996. [Purification and properties of isocitrate lyase and malate synthase from fasting rat liver]. *Biokhimiia (Moscow, Russia)*, **61**, 1898–903.

Popov, V. N., Moskalev, E. a., Shevchenko, M. U., & Eprintsev, a. T. 2005. Comparative Analysis of Glyoxylate Cycle Key Enzyme Isocitrate Lyase from Organisms of Different Systematic Groups. *Journal of Evolutionary Biochemistry and Physiology*, **41**, 631–639.

Poralla, K, Muth, G, & Härtner, T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters*, **189**, 93–5.

Pratt, Joel, Cooley, J Danny, Purdy, Charles W, & Straus, David C. 2000. Lipase activity from strains of *Pasteurella multocida*. *Current Microbiology*, **40**, 306–9.

Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., & Hopwood, D. a. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Molecular Microbiology*, **21**, 77–96.

Reinscheid, Dieter J, Eikmanns, Bernhard J, & Sahm, Hermann. 1994. Characterization of the isocitrate lyase gene from *Corynebacterium glutamicum* and biochemical analysis of the enzyme. *Journal of Bacteriology*, **176**, 3474–83.

Reumerman, Richard A., Tucker, Nicholas P., Herron, Paul R., Hoskisson, Paul a., & Sangal, Vartul. 2013. Tool for rapid annotation of microbial SNPs (TRAMS): a simple program for rapid annotation of genomic variation in prokaryotes. *Antonie van Leeuwenhoek*, **104**, 431–434.

Robinson, Mark D, & Smyth, Gordon K. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics (Oxford, England)*, **9**, 321–32.

Robinson, Mark D, McCarthy, Davis J, & Smyth, Gordon K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, **26**, 139–40.

Rocha, Eduardo P C, Smith, John Maynard, Hurst, Laurence D, Holden, Matthew T G, Cooper, Jessica E, Smith, Noel H, & Feil, Edward J. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology*, **239**, 226–35.

Rodriguez, Eduardo, Navone, Laura, Casati, Paula, & Gramajo, Hugo. 2012. Impact of malic enzymes on antibiotic and triacylglycerol production in *Streptomyces coelicolor*. *Applied and Environmental Microbiology*, **78**, 4571–9.

Sakai, Yasuo, Hayatsu, Masahito, & Hayano, Koichi. 2002. Use of tween 20 as a substrate for assay of lipase activity in soils. *Soil Science and Plant Nutrition*, **48**, 729–734.

Sambrook, J, Fritsch, E.F., & Maniatis, T. 1989. *Molecular Cloning : a laboratory manual E/F*. 2 edn. New York: Cold Spring Harbor Laboratory Press.

Schomburg, Ida, Chang, Antje, & Schomburg, Dietmar. 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, **30**, 47–9.

Schwarzer, Dirk, Finking, Robert, & Marahiel, Mohamed a. 2003. Nonribosomal peptides: from genes to products. *Natural product reports*, **20**, 275–287.

Schwecke, T, Aparicio, J F, Molnár, I, König, a, Khaw, L E, Haydock, S F, Oliynyk, M, Caffrey, P, Cortés, J, & Lester, J B. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 7839–7843.

Segel, Irwin H. 1975. *Enzyme Kinetics*. 1 edn. New York: John Wiley & Sons Inc.

Shahab, N, Flett, F, Oliver, S G, & Butler, P R. 1996. Growth rate control of protein and nucleic acid content in *Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r. *Microbiology (Reading, England)*, **142**, 1927–35.

Sharma, V, Sharma, S, Hoener zu Bentrup, K, McKinney, J.D., Russell, D.G., Jacobs Jr., W.R., & Sacchettini, J.C. 2000. Crystal structure of isocitrate lyase:Nitropropionate:glyoxylate complex from *Mycobacterium tuberculosis*. *Nat.Struct.Biol*, **7**, 663–668.

Shoval, O, Sheftel, H, Shinar, G, Hart, Y, Ramote, O, Mayo, a, Dekel, E, Kavanagh, K, & Alon, U. 2012. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science (New York, N.Y.)*, **336**, 1157–60.

Singh, Vivek Kumar, & Ghosh, Indira. 2006. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets. *Theoretical Biology & Medical Modelling*, **3**, 27.

Slijkhuis, Herman. 1983. Microthrix parvicella, a filamentous bacterium isolated from activated sludge: cultivation in a chemically defined medium. *Applied and Environmental Microbiology*, **46**, 832–9.

Soh, B.S., Paxton, L, & Sim, T-S. 2001. Cloning, heterologous expression and purification of an isocitrate lyase from *Streptomyces clavuligerus* NRRL 3585. *Biochimica et Biophysica Acta - Gene Structure and Expression*, **1522**, 112–117.

Speijer, D, Manjeri, G R, & Szklarczyk, R. 2014. How to deal with oxygen radicals stemming from mitochondrial fatty acid oxidation. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **369**, 20130446.

Studier, F W, & Moffatt, B A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, **189**, 113–130.

Sun, J, Kelemen, G H, Fernández-Abalos, J M, & Bibb, M J. 1999. Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology (Reading, England)*, **145**, 2221–7.

Tahama, Hirotsugu, Shinoyama, Hirofumi, Ando, Akikazu, & Fujii, Takaaki. 1990. Purification and Characterization of Isocitrate Lyase from Rhodopseudomonas sp . No . 7. *Agricultural and Biological Chemistry*, **54**, 3177–3183.

Takano, Hideaki, Obitsu, Saemi, Beppu, Teruhiko, & Ueda, Kenji. 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. *Journal of Bacteriology*, **187**, 1825–32.

Tamura, Koichiro, Stecher, Glen, Peterson, Daniel, Filipski, Alan, & Kumar, Sudhir. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–9.

Taylor, Richard D. 1992. *Purification and characterisation of the Isocitrate Dehydrogenase from Streptomyces coelicolor and cloning of its gene.* PhD Thesis, University of Glasgow.

Tepper, Naama, Noor, Elad, Amador-Noguez, Daniel, Haraldsdóttir, Hulda S, Milo, Ron, Rabinowitz, Josh, Liebermeister, Wolfram, & Shlomi, Tomer. 2013. Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PloS One*, **8**, e75370.

van Rossum, G., & Drake, F.L. 2001. *Python Reference Manual.* Virginia, USA: Python Software Foundation.

van Wezel, Gilles P, Mahr, Kerstin, König, Miriam, Traag, Bjørn a, Pimentel-Schmitt, Elisângela F, Willimek, Andreas, & Titgemeyer, Fritz. 2005. GlcP constitutes the major glucose uptake system of *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, **55**, 624–36.

van Wezel, Gilles P, König, Miriam, Mahr, Kerstin, Nothaft, Harald, Thomae, Andreas W, Bibb, Mervyn, & Titgemeyer, Fritz. 2007. A new piece of an old jigsaw: glucose kinase is activated posttranslationally in a glucose transport-dependent manner in

streptomyces coelicolor A3(2). *Journal of Molecular Microbiology and Biotechnology*, **12**, 67–74.

Visser, Diana, & Heijnen, Joseph J. 2002. The mathematics of metabolic control analysis revisited. *Metabolic Engineering*, **4**, 114–23.

Visser, Diana, & Heijnen, Joseph J. 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering*, **5**, 164–76.

Vorísek, J., Powell, A.J., & Vank, Z. 1969. Regulation of biosynthesis of secondary metabolites. IV. Purification and properties of phosphoenolpyruvate carboxylase in *Streptomyces* aureofaciens. *Folia Microbiologica*, **14**, 398–405.

Walsh, K, & Koshland, D E. 1984. Determination of flux through the branch point of two metabolic cycles. The tricarboxylic acid cycle and the glyoxylate shunt. *The Journal of Biological Chemistry*, **259**, 9646–54.

Wang, C. H., Bialy, J. J., Klungsoyr, S., & Gilmour, C. M. 1958. Studies on the biochemistry of streptomyces. III. Glucose catabolism in *Streptomyces griseus*. *Journal of bacteriology*, **75**, 31–37.

Watve, M G, Tickoo, R, Jog, M M, & Bhole, B D. 2001. How many antibiotics are produced by the genus *Streptomyces*? *Archives of Microbiology*, **176**, 386–90.

Wentzel, Alexander, Sletta, Havard, Consortium, Stream, Ellingsen, Trond E, & Bruheim, Per. 2012. Intracellular Metabolite Pool Changes in Response to Nutrient Depletion Induced Metabolic Switching in *Streptomyces coelicolor*. *Metabolites*, **2**, 178–94.

Wiechert, W. 2001. 13C metabolic flux analysis. *Metabolic Engineering*, **3**, 195–206.

Wiechert, Wolfgang, & Noack, Stephan. 2011. Mechanistic pathway modeling for industrial biotechnology: challenging but worthwhile. *Current Opinion in Biotechnology*, **22**, 604–10.

Wilkinson, G N. 1961. Statistical estimations in enzyme kinetics. *The Biochemical Journal*, **80**, 324–32.

Willemse, Joost, & van Wezel, Gilles P. 2009. Imaging of *Streptomyces coelicolor* A3(2) with reduced autofluorescence reveals a novel stage of FtsZ localization. *PloS One*, **4**, e4242.

Wilson, Micheal C, & Moore, Bradley S. 2012. Beyond ethylmalonyl-CoA: The functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Natural Product Reports*, **29**, 72.

Wu, Liang, Wang, Weiming, van Winden, Wouter A, van Gulik, Walter M, & Heijnen, Joseph J. 2004. A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *European Journal of Biochemistry / FEBS*, **271**, 3348–59.

Yamada, S., Hayashida, S., & Ogata, S. 1990. Purine auxotrophic mutants with altered spore color in *Streptomyces azureus* ATCC 14921. *Applied and Environmental Microbiology*, **56**, 575–577.

Yang, Chin-Rang, Shapiro, Bruce E, Hung, She-Pin, Mjolsness, Eric D, & Hatfield, G Wesley. 2005. A mathematical model for the branched chain amino acid biosynthetic pathways of *Escherichia coli* K12. *The Journal of Biological Chemistry*, **280**, 11224–32.

Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang, Z, & Bielawski, Jp. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**, 496–503.

Yasutake, Yoshiaki, Watanabe, Seiya, Yao, Min, Takada, Yasuhiro, Fukunaga, Noriyuki, & Tanaka, Isao. 2002. Structure of the monomeric isocitrate dehydrogenase: evidence of a protein monomerization by a domain duplication. *Structure (London, England : 1993)*, **10**, 1637–48.

Zhang, Bei-Bei, Wang, Peng, Wang, Ao, Wang, Wen-Cai, Tang, Wang-Gang, & Zhu, Guo-Ping. 2013. Expression and characterization of a novel isocitrate dehydrogenase from *Streptomyces diastaticus* No. 7 strain M1033. *Molecular Biology Reports*, **40**, 1615–23.

Zhang, Beibei, Wang, Baojuan, Wang, Peng, Cao, Zhengyu, Huang, Enqi, Hao, Jiasheng, Dean, Antony M, & Zhu, Guoping. 2009. Enzymatic characterization of a monomeric isocitrate dehydrogenase from *Streptomyces lividans* TK54. *Biochimie*, **91**, 1405–10.

Zhao, Jiao, & Shimizu, Kazuyuki. 2003. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of Biotechnology*, **101**, 101–17.