2006

Glyoxylate Metabolism in Mycobacterium smegmatis

Lubomir Nikolaev Merkov

Glyoxylate Metabolism In *Mycobacterium smegmatis*

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Lubomir Nikolaev Merkov

June 2006

ABSTRACT

Much has been learned about *Mycobacterium tuberculosis*, the causative agent of tuberculosis, “the great white plague,” since the bacterium was isolated and initially characterized by Robert Koch over a century ago. Over the last decade, new genetic tools for manipulation of the bacterium have been developed, its genome has been sequenced, and the search for new vaccines and drug targets has greatly intensified. Yet, surprisingly little is known about which mycobacterial genes are truly important for the organism’s ability to persist in the tissues of its human hosts. The metabolic pathways used by the tubercle bacillus to establish and maintain a life-long infection have largely been ignored by researchers, yet they may represent promising new areas for therapeutic intervention. Recently, one enzyme of the glyoxylate shunt of *M. tuberculosis*, isocitrate lyase (ICL), was shown to be required for virulence in experimental infections of mice. The other enzyme of the glyoxylate shunt, malate synthase (MLS), may also be important for the intracellular survival of the tubercle bacillus; yet, no studies have been done to determine its *in vivo* role. We present here results of genetic studies of MLS in the saprophyte *Mycobacterium smegmatis*, and show that MLS, unlike ICL, is dispensable for growth on acetate or fatty acids. We also describe the d-glycerate pathway in *M. smegmatis*, which enables malate synthase-deficient bacteria to utilize acetate and fatty acids as sole carbon sources, and which allows *M. smegmatis* to grow on glyoxylate. The d-glycerate pathway, however, does not appear to exist in the pathogenic *Mycobacterium tuberculosis*.

ACKNOWLEDGMENTS

In early March of 1999, I visited The Rockefeller University to interview for admittance to its graduate program, and in the University's famous Faculty Club, I met John McKinney for the first time. John had just started his own laboratory, housed on the 5th floor of the Bronk building at the time, which he invited me to visit the next day. His lab at the time consisted of himself, Ernesto Munoz, a graduate student who had just joined the lab, and several cardboard boxes stacked on the benches and hiding the entrance to John's office. For about half an hour, John and I talked about tuberculosis, about his research, and about the future of his lab—and little did I know that day that exactly two years later, I would join his laboratory, now on the 9th floor of Bronk, for my doctoral research.

John has been a great mentor to me; patient, understanding, very helpful, always available (unless he was in South Africa or Germany or Switzerland), John gave me the freedom and opportunity to pursue many projects, and the chance to follow up on the malate synthase story in *M. smegmatis* through all the twists and turns that it would lead me. John, I couldn't ask for a better thesis advisor.

I also would like to thank Drs. Charlie Rice and Ralph Steinman for being members of my faculty committee and for providing invaluable advice during my dissertation work. Also, Dr. Kevin Reynolds kindly agreed to be a member of my committee, and to fly all the way from Portland, OR, for my title defense. I will always be extremely grateful to my committee for their help and support.

To my wife, Natalia: thank you for being my best friend and supporter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Tuberculosis: The Disease and The Bacillus

1.1. The duration of the mycobacterial infection

> A white woman had active pulmonary tuberculosis of 41 years' duration. She had bilateral bronchiectasis, cavitation in the right upper lobe, a collapsed left lung, pleural effusion, and an induced left pneumothorax of 36 years' duration. She had received no antituberculosis drugs for 28 years from the onset of her illness, and these drugs, when finally prescribed, were taken inadequately. Pleural effusion present for 28 years required aspiration.... The patient survived 41 years with advanced, active, pulmonary tuberculosis with cavitation, a functionless left lung, and pleural effusion, and she harbored a strain of *M. tuberculosis* resistant to all antituberculous drugs except streptomycin. [Edwards *et al.* 1970]

The report by Edwards and colleagues is noteworthy for several reasons. On the one hand, it highlights the resilience of the human body in the face of severe pathological destruction, as the patient survived for decades with one functional lung. On the other, however, it underlines the tenacity and recalcitrance of the tubercle bacillus. Despite the best surgical interventions of the time, the care by numerous physicians in six large military hospitals, and the administration of antibiotics, the patient was never cured; instead, the bacillus acquired resistance to all but one of the antibiotics and, by maintaining an active infection, it was undoubtedly transmitted to other individuals.

Other details of the story make it even more interesting. The patient did survive for over four decades, but she was the exception rather than the rule, as studies of patients with such advance disease and cavitation revealed that 85 per cent died within a year (Barnes and Barnes, 1928). Furthermore, her husband of 30 years was with her at all times, served as her "personal doctor", and

administered pneumothorax fills at home, yet he never showed any evidence of the disease. It is clear, however, that the patient would have shed an enormous number of multi-drug resistant bacilli in her community, and while her husband never had an active infection, others may have contracted the disease from her.

This ability of the mycobacteria to persist in the face of the combined assault of the immune system and antibiotics is a hallmark of tuberculosis. It is estimated that 90% of infected individuals will not develop clinical disease; of the 10% who do develop disease, some of them will succumb within the first few years after infection, while the rest will control the infection initially, but develop tuberculosis later in life (Comstock, 1982). The persistence of the bacterium in the tissues of otherwise healthy humans has been well documented. Hernandez-Pando and colleagues (2000) used PCR to examine macroscopically normal lung tissue from individuals who had died from causes other than tuberculosis. In five of 13 people from Ethiopia and ten of 34 individuals from Mexico, DNA from *M. tuberculosis* was detected. The lungs of these humans presumably harbored a handful of live mycobacteria, which could eventually establish an active infection. A remarkable example of a reactivation of latent infection was reported in Denmark, where scientists obtained evidence of endogenous reactivation of *M. tuberculosis* after 33 years of latency (Lillebaek *et al*., 2002). The patient had presumably received the BCG vaccine prior to infection, did not show signs of disease after contracting the bacterium, yet presented with active tuberculosis 33 years later.

1.2. The global burden of the disease

The World Health Organization's website (www.who.int) provides statistics regarding the magnitude of the TB epidemic worldwide. Of the 2 billion people who are currently infected with the tubercle bacillus, 200 million are expected to develop active tuberculosis. 8.8 million new cases of TB occurred in 2003, with the overwhelming majority of cases concentrated in 22 countries. About half a million of these new cases were multi-drug resistant (MDR), with the highest rates in the former Soviet Union and China, where one in eight patients fails to respond to standard drug therapy. MDR-TB, however, has been detected in virtually all of the 109 countries surveyed by the WHO. TB has reached pandemic status in Africa, which accounts for a quarter of all TB cases, and in Southeast Asia, in which 6 countries (Bangladesh, China, India, Indonesia, Pakistan and Philippines) register half of all new cases (WHO, 2005).

TB, malaria and HIV/AIDS kill 6 million people each year, a third of whom die from TB. TB is the leading killer of HIV-infected individuals; a quarter of a million TB deaths, most of them in Africa, are associated with HIV/AIDS. The TB epidemic in the developing world mainly impacts on young adults, who are the economic drivers of society; in fact, TB is a leading cause of mortality among young women, especially in Africa. Due to the high rates of TB infection in Africa, the global incidence of the disease is growing at 1% per year. Areas of high HIV prevalence in Africa experience more than twice the rate of TB cases (400 per 100,000 people per year) compared to low HIV prevalence areas (WHO, 2005).

1.3. The physical nature of the bacillus

M. tuberculosis, and the related fast-growing saprophyte *M. smegmatis*, belong to the genus *Mycobacterium*. This genus comprises a number of species of acid-fast, rod-shaped bacteria, and is the only genus in the family *Mycobacteriaceae* within the order *Actinomycetales*. Like other related *Actinomycetales*, such as *Nocardia* and *Corynebacterium*, mycobacteria have high genomic DNA GC content and produce mycolic acids as major components of their cell wall. The tubercle bacilli measure between 1 and 4 μm in length and are 0.3 to 0.6 μm in diameter; they possess only a single cytoplasmic membrane, but they also have thick, waxy cell walls. Mycobacteria are obligate aerobes, despite possessing enzymes required for anaerobic fermentative growth (Cole *et al.*, 1998); however, they can grow at low O_2 concentrations and can even survive long periods of complete oxygen deprivation (Wayne and Hayes, 1996).

Unlike *M. leprae*, *M. tuberculosis* and *M. smegmatis* can be cultured in liquid or on solid media containing carbon and nitrogen sources as well as basic salts and trace elements (Wayne, 1994). *M. smegmatis* is more amenable to growth in minimal media than its pathogenic cousin, which makes it a useful model for metabolic studies. It also doubles every 3-4 hours in culture, and forms colonies in 3 to 4 days, while *M. tuberculosis* doubles every 20-24 hours and forms a colony on agar in 3-4 weeks. All of the experiments described in this thesis were carried out with *M. smegmatis*, largely because they were not feasible, or easily executable, with pathogenic *M. tuberculosis*.

The study of mycobacteria has been facilitated by the sequencing of several mycobacterial genomes over the last decade. They are high-GC (~ 65%) and lack pathogenicity islands typical of many bacterial pathogens. The genomes of the virulent laboratory strain H37Rv (Cole *et al.* 1998), the clinical isolate CDC1551 (Fleischmann *et al.*, 2002), the animal pathogen *M. bovis* (Garnier *et al.* 2003), as well as the *M. leprae* genome (Cole *et al.*, 2001), were all recently decoded. The Institute of Genomic Research (TIGR) has also sequenced and annotated the genomes of *M. aviam paratuberculosis* and *M. smegmatis* mc^2155. A perusal of the genome statistics reveals that *M. smegmatis* has duplicated a significant portion of its genetic material. However, this duplication does not include the glyoxylate shunt *glcB* gene (which encodes malate synthase) and the two isocitrate lyase genes (*icl1* and *icl2*).

Mycobacterial species have greatly duplicated the genes involved in fatty acid biosynthesis, which is perhaps not surprising given that mycobacteria possess an extremely complex lipid-rich cell wall (Brennan, 2003). Close to ten percent of the DNA of pathogenic mycobacteria is predicted to be involved in lipid metabolism. The *M. smegmatis* genome displays an equally extensive expansion. Focusing on the fatty acid degradation (*fad*) genes, The Institute for Genomic Research's (TIGR's) annotation suggests that there could be as many as 7 *fadA*, 9 *fadB*, around 40 *fadD*, and more than 40 *FadE* genes. Comparable duplication of the fatty acid biosynthesis (*fab*) genes exists in *M. smegmatis*.

1.4. General metabolic features of mycobacteria

The extensive duplication of lipid assimilation and biosynthesis genes appear be the most striking feature of mycobacterial metabolism. When grown in synthetic media, mycobacteria behave like most other bacterial species, and can use a range of substrates as carbon and energy sources (Wheeler and Ratledge, 1994). Mycobacteria possess all the enzymes for glycolysis, the pentose phosphate pathway, and the tricarboxylic acid and glyoxylate cycles, plus a number of oxygenases, oxidoreductases, and dehydrogenases. Mycobacteria can also utilize several anaerobic electron transport chains, such as the nitrate, fumarate, and possibly nitrite reductases (Cole *et al.*, 1998).

Little is known about the primary carbon sources utilized by mycobacteria *in vivo*, but it is thought that lipids might be more readily available for assimilation than carbohydrates (Wheeler and Ratledge, 1994). Mycobacteria may in fact use both lipids and carbohydrates in host tissues, and biosynthetic and degradative reactions could take place simultaneously. Studies by Segal and Bloch (1956) of virulent bacilli isolated from host tissues suggested that fatty acids might serve as the major source of carbon and energy *in vivo.* Recent studies in our laboratory (McKinney *et al.*, 2000; Muñoz-Elías and McKinney, 2005; Muñoz-Elías *et al.*, 2006) suggest that the ability to degrade and assimilate lipids for both energy and biosynthetic purposes is of crucial importance to the tubercle bacillus during the establishment of a persisten infection. A more complete overview of mycobacterial metabolism will be provided later.

1.5. Mycobacterial features examined in this thesis

Mycobacterial metabolism, molecular genetics, pathophysiology and persistence, the immune response of the infected host, and the interaction between the bacillus and its host at the cellular level have been reviewed extensively (Wheeler and Ratledge, 1994; Dannerberg and Rook, 1994; McKinney *et al*. 1998; Wayne and Sohaskey, 2001; Russell, 2001; Clark-Curtiss and Haydell, 2003; Boshoff and Barry, 2005; Muñoz-Elías and McKinney, 2006). This thesis will focus on the central metabolism of *M. tuberculosis* and its saprophytic cousin *M. smegmatis.* Chapter 2 describes our current understanding of the metabolism of bacteria grown *in vitro*, the main organisms studied being *Escherichia coli* and some relatives of the *Mycobacteriaseae* family, such as *Corynebacterium* and *Streptomyces*; the role of the glyoxylate shunt enzymes for growth *in vitro* and for the establishment of persistent infection *in vivo* will be emphasized. Chapter 3 describes the results of studies on the role of the gyoxylate shunt enzyme malate synthase in the utilization of carbohydrates and fatty acids as carbon sources, and the role of this pathway in mycobacterial anaplerosis during growth on two-carbon molecules (acetate and glyoxylate) and compounds catabolized to acetyl-CoA, such as fatty acids. The experiments were done in *M. smegmatis*, but the results have important implications for understanding the metabolism of *M. tuberculosis*.

CHAPTER 2

Bacterial Metabolism *In Vitro*

2.1. General aspects of *in vitro* metabolism

Our knowledge of the *in vitro* metabolism of eubacteria has largely been derived from studies of enteric organisms, particularly *Escherichia coli*. The ease of biochemical and genetic manipulation of *E. coli* has enabled scientists to assign unambiguous functions to many of the genes involved in central metabolism. Thus, in many cases the function of a newly sequenced gene can be adduced from its homology to functionally chyaracterized *E. coli* genes. This chapter will focus on aspects of *E. coli* metabolism and will summarize our current knowledge of mycobacterial metabolic pathways, highlighting important differences.

Many species of eubacteria will grow readily on solid or in liquid media containing basal salts and essential elements in trace amounts, as long as carbon and nitrogen sources are provided. As mentioned above, most mycobacteria can be cultured *in vitro* in synthetic media, *M. leprae* being a notable exception. Some organisms with mutations in biosynthetic pathways may require media supplementation with compounds like thiamine or lipoic acid, vitamins or amino acids, or precursors for the biosynthesis of coenzymes such as NAD^+ and $NADH^+$, or flavin-containing molecules such as FMN and FAD. Both the pathogenic tubercle bacillus and its non-pathogenic relative *M. smegmatis* grow without requiring such supplements, as can be expected given the completeness of their genomes for genes involved in the biosynthesis of essential precursors.

2.2. Carbohydrate metabolism

The most common type of sugar used to culture eubacteria is glucose, a six-carbon substrate. Glucose can be oxidized to pyruvate via the Embden-Meyerhof pathway (EMP, a.k.a. glycolysis) or the Entner-Doudoroff pathway (EDP), or it can be converted to pentose sugars via the pentose phosphate pathway (PPP) (Fraenkel, 1996). Acetyl-CoA generated from pyruvate is further oxidized to carbon dioxide and water via the tricarboxylic acid (TCA) cycle. Pentose sugars are precursors for nucleotides and nucleic acids. The EMP, EDP, and PPP pathways of eubacteria and eukaryotes are depicted in Figure 2.1.

The first step of glycolysis is the activation of glucose to glucose 6-phosphate, which can be done by glucokinase (*glk*). Enterics use the phosphoenolpyruvate (PEP) phosphotransferase system (PTS) as the main route of glucose activation and transport into the cell. *M. tuberculosis* lacks PTS and glucokinase homologs, and most likely uses polyphosphate glucokinase (PPGK) for glucose activation. Glucose-6-phosphate is converted to fructose-6-phosphate by phosphoglucose isomerase (*pgi*); *E. coli pgi* mutants can grow on glucose by using the PPP (Fraenkel and Horecker, 1965). The conversion of fructose-6-phosphate to fructose-1,6-bisphosphate is catalyzed by phosphofructokinase (*pfk*); two copies of the enzyme are present in *E. coli* and in *M. tuberculosis*, where functional redundancy seems to exist (Sassetti *et al.*, 2003). The hexose fructose-1-6-bisphosphate is finally split into the trioses glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate, by fructose-1,6-bisphosphate

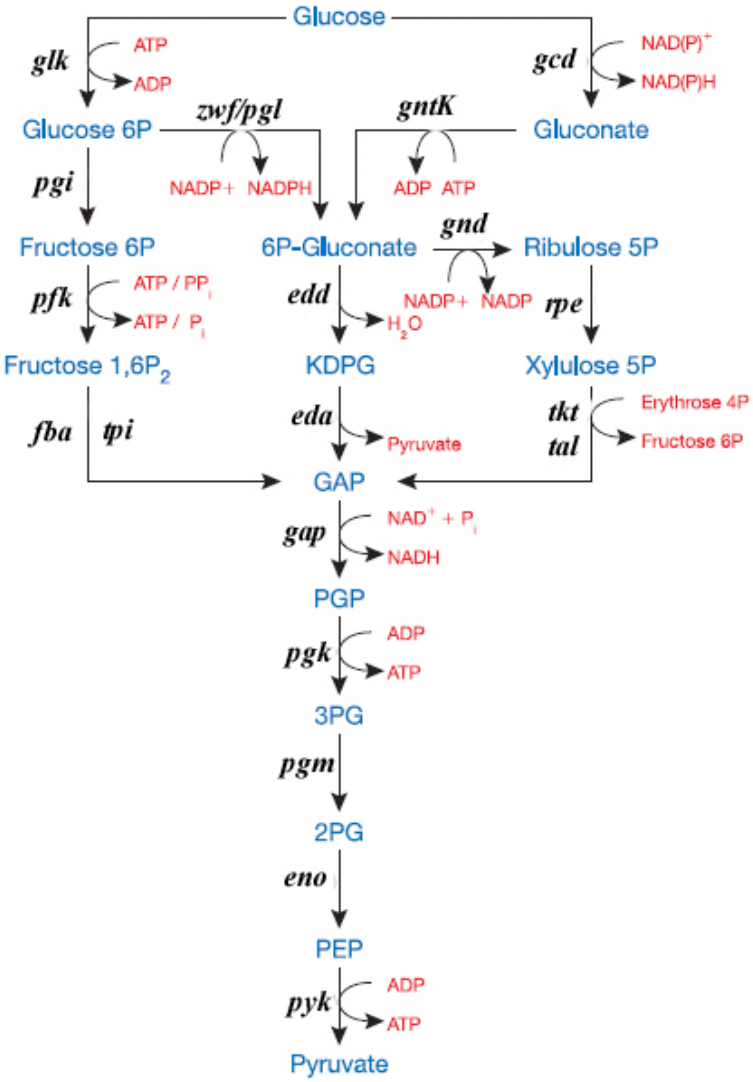


Figure 2.1. Routes of glucose degradation. Embden-Meyerhof pathway (EMP) [*E. coli* genes shown]: *glk*, glucose kinase; *pgi*, phosphoglucose isomerase; *pfk*, phosphofructokinase; *fba*, fructose-1,6-bisphosphate aldolase; *tpi*, triosephosphate isomerase; *gap*, glyceraldehyde 3-phosphate dehydrogenase; *pgk*, phosphoglycerate kinase; *pgm*, phosphoglycerate mutase; *eno*, enolase; *pyk*, pyruvate kinase. Glucose-6-phosphate can be converted to gluconate or 6-phosphogluconate, which can enter the pentose phosphate pathway (PPP) or the Entner-Doudoroff pathway (EDP). *gcd*, glucose dehydrogenase; *gntK*, gluconate kinase; *zwf*, glucose 6-phosphate dehydrogenase; *pgl*, 6-phospho-gluconolactonase; *edd*, 6-phosphogluconate dehydrogenase; *eda*, 2-keto-3-deoxy-6-phosphogluconate aldolase; *gnd*, 6-phosphogluconate dehydrogenase; *rpe*, ribulose-5-phosphate epimerase; *tkt*, transketoase; *tal*, translaldolase. Ribose-5-phosphate and *rpi,* ribose-5-phosphate isomerase not shown. Abbreviations: KGDP, 2-keto-3-deoxy-phosphogluconate; GA, glyceraldehyde; GAP, glyceradehyde-3-phosphate; PGP, glycerate-1,3-bisphosphate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate. Modified from Verhees *et al.* (2003) and Fraenkel (1996).

aldolase (*fba*). Triosephosphate isomerase (*tpi*) interconverts the two three-carbon compounds. In the lower branch of the glycolytic pathway, glyceraldehyde-3-phosphate is ultimately oxidized to one molecule of pyruvate, with the concomitant generation of two ATP molecules, by the enzymes glyceraldehyde-3-phosphate dehydrogenase (*gap*), phosphoglycerate kinase (*pgk*), phosphoglycerate mutase (*pgm*), enolase (*eno*), and pyruvate kinase (*pyk*).

The pentose phosphate pathway (PPP) can also generate glyceraldehyde-3-phosphate, and consists of two branches for production of ribose-5-phosphate, a precursor for the synthesis of nucleotides and nucleic acids. The first is the "oxidative branch" which reduces two molecules of NADP+. The second forms pentoses from hexose and triose phosphates via non-oxidative reactions, which are catalyzed by transketolase and transaldolase (Horecker, 2002). The study of the PPP began with the discovery of NADP by Otto Warburg as the coenzyme required for the oxidation of glucose-6-phosphate to 6-phosphogluconate. Since the role of NAD^+ in glycolysis was well established at the time, an alternative pathway for oxidation of phospho-hexoses was sought and their conversion to phosphopentoses was discovered. The PPP fulfills two important functions: it generates ribose-5-phosphate for biosynthesis of nucelotides, and it provides reducing power in the form of NADPH. Ribose-5-phosphate can also be formed from the hexose fructose-6-phosphate and the triose glyceraldehyde-3-phosphate by non-oxidative rearrangements catalyzed by transketalose-

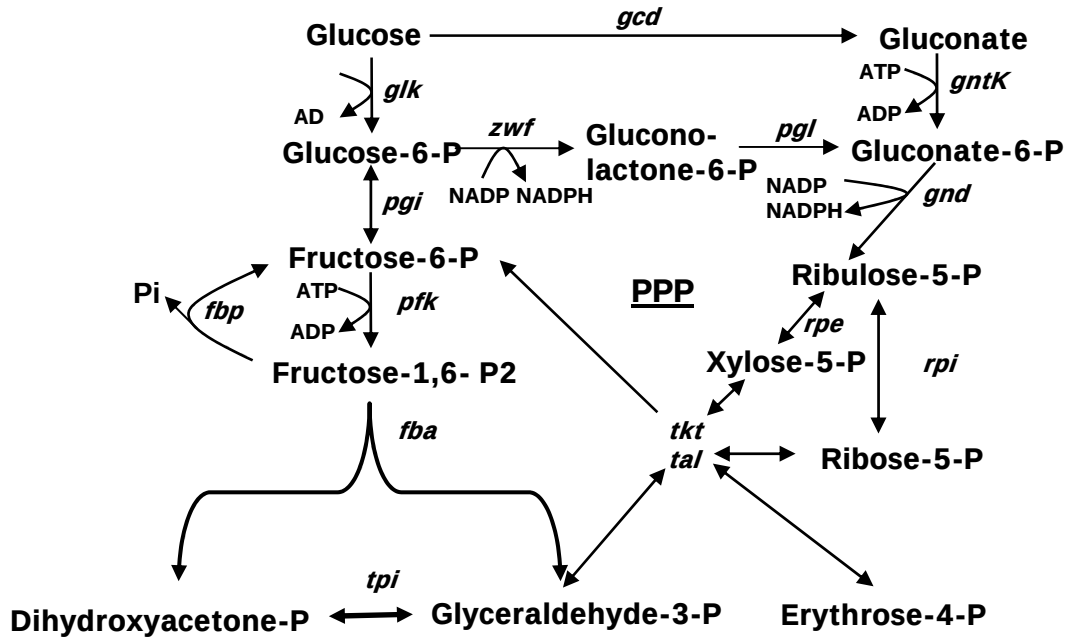


Figure 2.2. The pentose phosphate pathway (PPP). Genes: *gcd*, glucose dehydrogenase; *gntK*, gluconate kinase; *gnd,* 6-phosphogluconate dehydrogenase; *zwf*, glucose 6-phosphate dehydrogenase; *pgl*, 6-phosphogluconolactonase; *rpe,* ribulose 5-phosphate epimerase; *rpi,* ribose 5-phosphate isomerase; *tkt,* transketalose; *tal*, transaldolase. Modified from Fraenkel (1996).

transaldolase (*tkt-tal*) to form the four-carbon erythrose-4-phosphate (Horecker, 2002). *M. tuberculosis* has complete glycolytic and pentose phosphate pathways; transposon mutants in *tkt* and *tal* have been shown to be impaired for growth *in vitro* (Sassetti *et al.,* 2003).

6-phosphogluconate can be oxidized to glyceradehyde-3-phosphate and pyruvate via the Entner-Doudoroff pathway (EDP). The two unique EDP genes are *edd*, encoding 6-P-gluconate hydratase, and *eda*, coding for 2-keto-3-deoxy-6-P-gluconate aldolase (Fraenkel, 1996; Peekhaus and Conway, 1998). The combined action of the two enzymes converts a phosphate hexose into pyruvate and phosphotriose. Recent examination of hyperthermophilic archaea and bacteria has led Romano and Conway (1996) to conclude that the EDP is an older route of sugar dissimilation than the EMP. Based on the presence of gluconeogenesis via a "reversed" EMP route, but the lack of the traditional EMP in ancient organisms, they suggest that the EDP was the original catabolic pathway, with the "reversed" EMP originally serving an anabolic role; the EMP become a catabolic pathway with the evolution of phosphofructokinase. In pathogenic mycobacteria, however, there is no evidence for the existence of the EDP as homologs of *edd* and *eda* are missing (Muñoz-Elías and McKinney, 2006).

Figure 2.3 shows the pathways for glucose degradation in archaea, which occurs via variants of the EMP and EDP (Verhees *et al.*, 2003). As depicted, the EDP is present in an altered form from the EDP in *E. coli*; the reduction of glyceraldehyde-3-phosphate to 3-phosphoglycerate can be accomplished in a

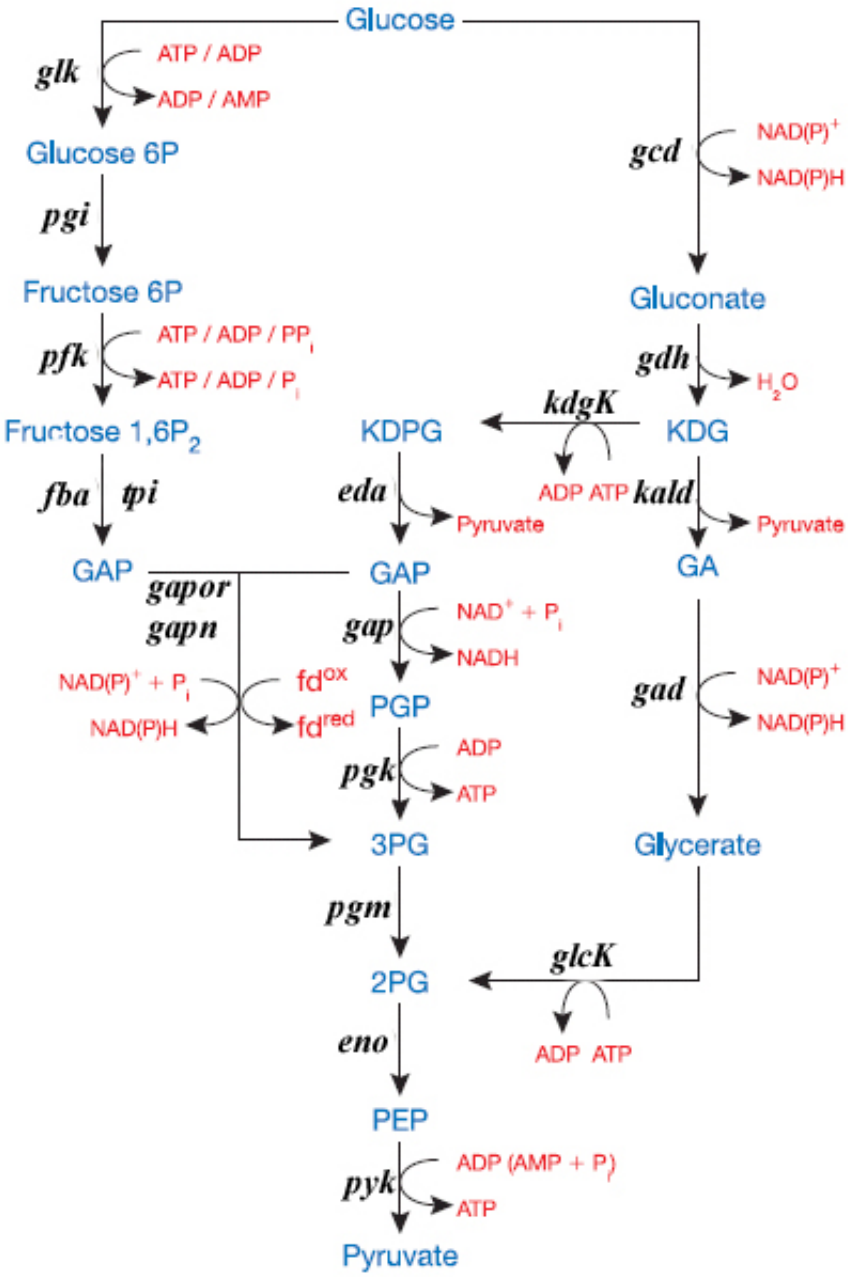


Figure 2.3. Glucose degradation in archaea via variants of the Embden-Meyerhof pathway (EMP) and Entner-Doudoroff pathway (EDP). Unique genes are: *gapor*, glyceraldehyde-3-phosphate ferredoxin oxidoreductase; *gapn*, non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (NAD+ dependent); *gdh*, gluconate dehydratase; *kald*, 2-keto-3-deoxy-gluconate aldolase; *kgdK*, 2-keto-3-deoxy-gluconate kinase; *gad*, glyceraldehyde dehydrogenase; *glcK*, glycerate kinase. GAPN catalyzes the phosphate-independent, single-step oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate, thus differing from GAPDH. KGDP, 2-keto-3-deoxy-phosphogluconate; KGD, 2-keto-3-deoxy-gluconate; GA, glyceraldehyde; GAP, glyceradehyde-3-phosphate; PGP, glycerate-1,3-bisphosphate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate. Modified from Verhees *et al.* (2003).

reaction catalyzed by GAPOR (glyceraldehyde-3-phosphate ferredoxin oxidoreductase) and GAPN (non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase, NAD+ dependent enzyme). Also, gluconate is converted to 2-keto-3-deoxy-phosphogluconate by the enzymes GDH and KGDK in a two-step mechanism, obviating the need for the reaction catalyzed in eubacteria by EDD. KGD can separately be oxidized to 2-phosphoglycerate as well.

Unlike the variant EMP and EDP pathways used by archaea and eubacteria/eukarya, the gluconeogenic pathway is conserved in both (Figure 2.4), consistent with the idea that gluconeogenesis is an ancient pathway, while the catabolic role of the EMP is more recent (Romano and Conway, 1996). Two enzymes unique to the gluconeogenic pathway are fructose-1,6-bisphosphatase (*fbp*) and PEP synthase (*pps*). *E. coli* mutants lacking *fbp* cannot grow on gluconeogenic substrates such as acetate and dicarboxylic acids (Fraenkel and Horecker, 1965). In *M. tuberculosis*, the gluconeogenic dephosphorylation of fructose-1,6-bisphosphate is catalyzed by the class II fructose-1,6-bisphosphatase encoded by the *glpX* gene (Movahedzadeh *et al.,* 2004).

M. tuberculosis does not possess a *pps* homolog; instead, pyruvate phosphate dikinase (*ppdK*) catalyzes the same conversion. *M. smegmatis* has a homolog of *pps*, but not *ppdK.* Aside from phosphorylation of pyruvate, phosphoenolpyruvate (PEP) can be generated by decarboxylation of oxaloacetate, catalyzed by PEP carboxykinase (PCK). *E.* coli strains lacking PCK cannot grow on the TCA cycle intermediates malate, fumarate, or oxaloacetate

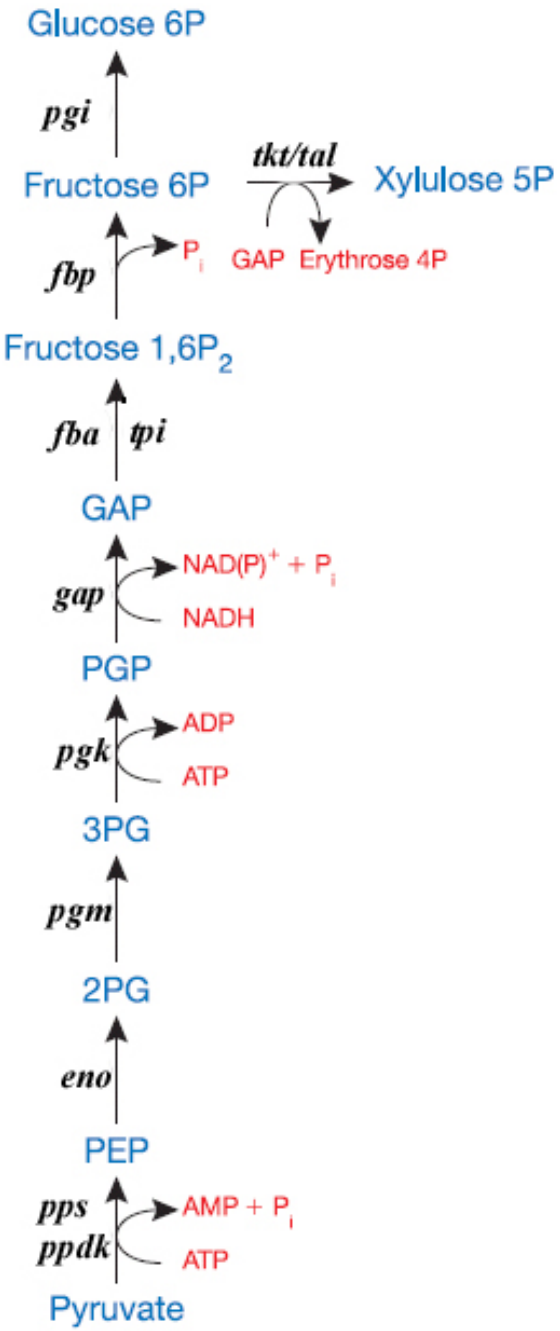


Figure 2.4. Gluconeogenesis. Found in both bacteria/eukarya and archaea, this pathway reduces pyruvate to fructose-6-phosphate, which can be further converted to glucose-6-phosphate or xululose-5-phosphate (or ribose-5-phosphate) via the pentose phosphate pathway (PPP). Pentose sugars are used for nucleotide biosynthesis. Enzymes unique to gluconeogenesis are: *fbp*, fructose-1,6-bisphosphatase; *pps*, phosphoenolpyruvate synthase; *ppdk*, pyruvate phosphate dikinase.
Modified from Verhees *et al*. (2003).

(Anderson and Wood, 1969). PCK is encoded by the *pckA* gene in *M. tuberculosis*; its importance for mycobacterial survival *in vivo* was recently suggested by the demonstration that *pckA*-deficient *M. bovis* BCG is attenuated for virulence in mice and macrophages (Liu *et al.,* 2003). *M. smegmatis* also possesses a single copy of *pckA*. We have isolated *pckA*-deficient transposon mutants with PCK deficiency, which fail to grow on acetate or palmitate (an even-chain fatty acid) or succinate (TCA cycle intermediate) as the sole carbon substrates (L. Merkov and A. Upton, unpubl. observations).

2.3. Fatty acid degradation

When fatty acids are used as the carbon source, they are converted to acetyl-CoA via the β-oxidation cycle (Figure 2.5). Each turn of the β-oxidation cycle generates one molecule of $FADH_2$ and one molecule of NADH, which can be used for production of ATP, making growth on fatty acids more energetically favorable than growth on acetate alone (Clark and Cronan, 1996). While *E. coli* possesses a small number of genes involved in fatty acid degradation, all mycobacterial genomes exhibit an extensive degree of duplication of the *fad* genes, as shown in Fig. 2.5. This gene expansion is not limited to pathogenic species, as *M. smegmatis* contains a comparable number of *fad* genes.

The *E. coli fadL* gene, which is essential for long chain (> C12) fatty acid transport (Black, 1988; 1991), has no annotated homologs in mycobacteria. This may be explained by the fact that FadL resides in the outer membrane of *E. coli*, and is responsible for transport of long chain fatty acid in the periplasmic space, while medium-length and short chain fatty acids may diffuse freely through the outer membrane. FadD then activates fatty acids by linking them to coenzyme-A (CoA), for which it requires free pools of both CoA and ATP; the activation of fatty acids renders their transport unidirectional (DiRusso and Black, 2004). Mycobacteria, on the other hand, have no outer membranes; thus, fatty acids may simply diffuse to, and through, the plasma membrane, where they would be activated by FadD proteins. The lipid-rich outer wall of mycobacteria might be permissive for diffusion of fatty acids toward the plasma membrane.

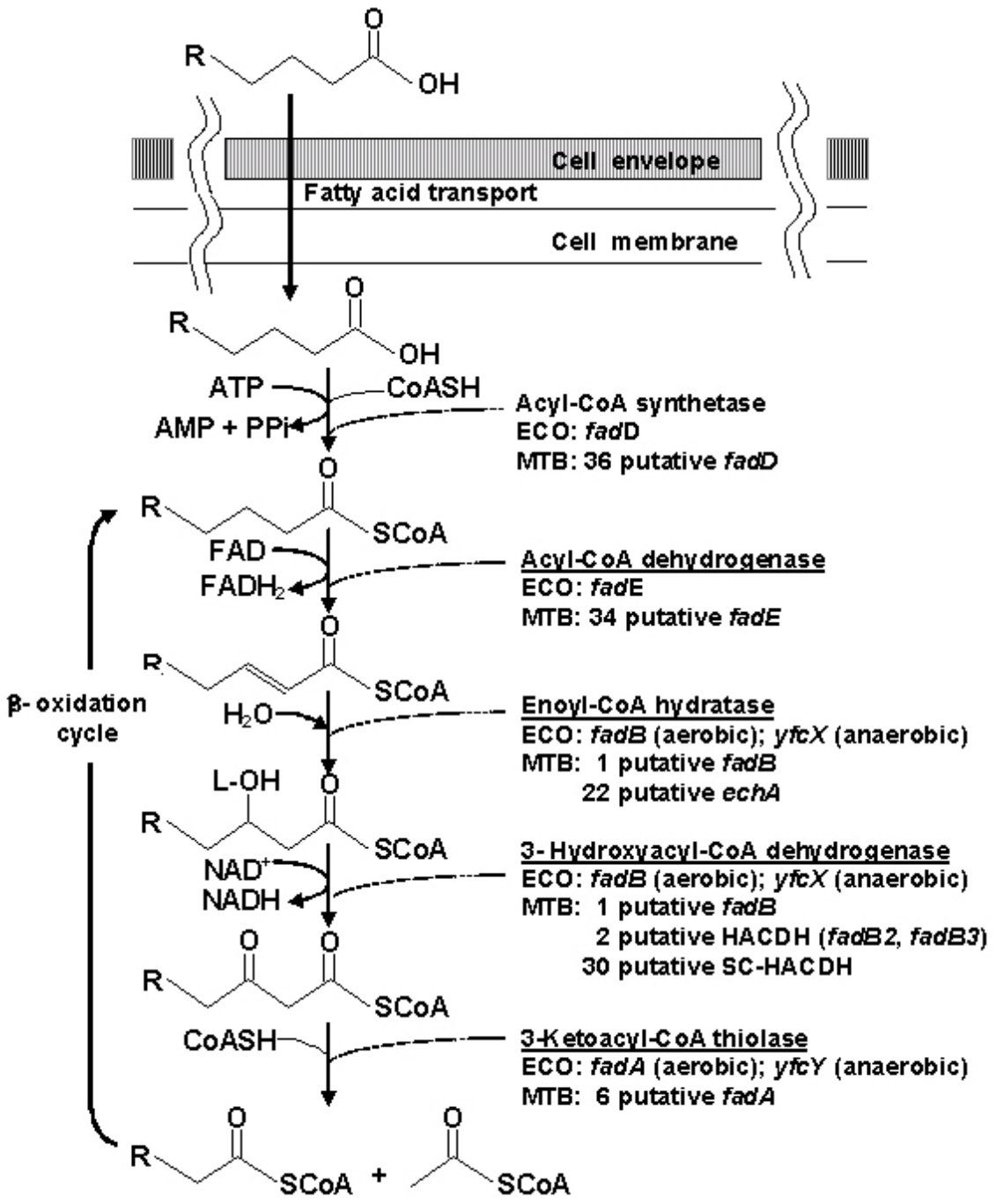


Figure 2.5. Fatty acid β-oxidation cycle. *E. coli* (ECO) and *M. tuberculosis* (MTB) genes (italicized) and enzymes are listed. Reproduced from Muñoz-Elías and McKinney (2006).

The fatty acid CoA synthase (FACS) encoded by *fadD* was originally identified by Overath *et al.* (1969). FACS catalyzes the synthesis of fatty-acyl CoA through the formation of a fatty acyl-AMP intermediate and the hydrolysis of ATP to yield pyrophosphate. All FACS proteins contain two canonical regions, the ATP-AMP motif, and the FACS motif. While FadL is specific for the transport of long-chain fatty acids, the *E. coli* FACS activates fatty acids varying in length from $C_{6:0}$ to $C_{20:4}$, with the highest specificity for $C_{14:0}$ to $C_{18:0}$ (Black and DiRusso, 1993).

The products of the 36 *M. tuberculosis fadD* genes can function as FACL (fatty acyl-CoA ligases) or FAAL (fatty acyl-AMP ligases) (Trivedi *et al.*, 2004). While the FACL form fatty acyl-CoAs of different length, the FAAL form fatty acyl-AMPs. The genes encoding FAAL are located next to polyketide synthesis (PKS) genes; FAAL activate fatty acids as acyl-adenylates and then transfer them onto the PKS. Disruption of *fadD28* (FAAL), in *M. tuberculosis* blocked the synthesis and export of a cell wall-associated lipid, phthiocerol dimycocerosate (PDIM); the PDIM-deficient mutants were attenuated in mice (Cox *et al.,* 1999).

The FACL proteins demonstrate remarkable substrate tolerance in generating fatty acyl-CoAs from the corresponding fatty acids (Arora *et al.*, 2005). The combination of the large number of genes encoding FACL and the substrate promiscuity of individual FACL enables mycobacteria to activate a tremendous variety of fatty acids for biosynthesis or catabolism. These acyl-CoAs can enter the β-oxidation cycle as substrates for acyl-CoA dehydrogenase (FadE), which carries out the first repetitive step of β-oxidation (Figure 2.5).

The enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase activities of the β-oxidation cycle in *E. coli* are carried out by the product of *fadB. M. tuberculosis* and *M. smegmatis* possess over twenty genes predicted to encode enzymes with one of the two activities. The *fadA* genes, whose products catalyze the last step of the cycle, the removal of a molecule of acetyl-CoA, comprise the smallest group of cycle participants. A shortened fatty acyl-CoA molecule then reenters the cycle by binding to FadE, while acetyl-CoA enters the TCA cycle.

Short-chain fatty acids (C4-C6) are assimilated differently by *E. coli. E. coli* FACS cannot activate fatty acids shorter than C6 so the induction of a new set of genes, the *atoADBC* operon, is required for growth on acetoacetate and butyrate (Figure 2.6). The *atoADBC* operon is repressed by FadR; derepression, as well as the expression of a positive regulator encoded by *atoC*, results in the expression of the structural genes *atoADB* (Pauli and Overath, 1972). FadR repression is relieved through binding of an activated fatty acyl-CoA molecule and thus requires FACS activity (DiRusso and Black, 2004; Clark and Cronan, 1996). DNA binding by FadR is strongly inhibited by fatty acyl-CoA esters, with C16-C18 acyl-CoAs showing the lowest K_i values. FadR dimers act as both repressors and activators, inhibiting transcription of the *fad* and *ato* genes, while activating the expression of the fatty acid biosynthesis genes *fabA* and *fabB*. FadR also activates *iclR*, which encodes a transcriptional repressor of the glyoxylate shunt genes; FadR can thus inhibit the expression of both the β-oxidation genes, which produce acetyl-CoA, and the glyoxylate shunt genes (Gui *et al*., 1996).

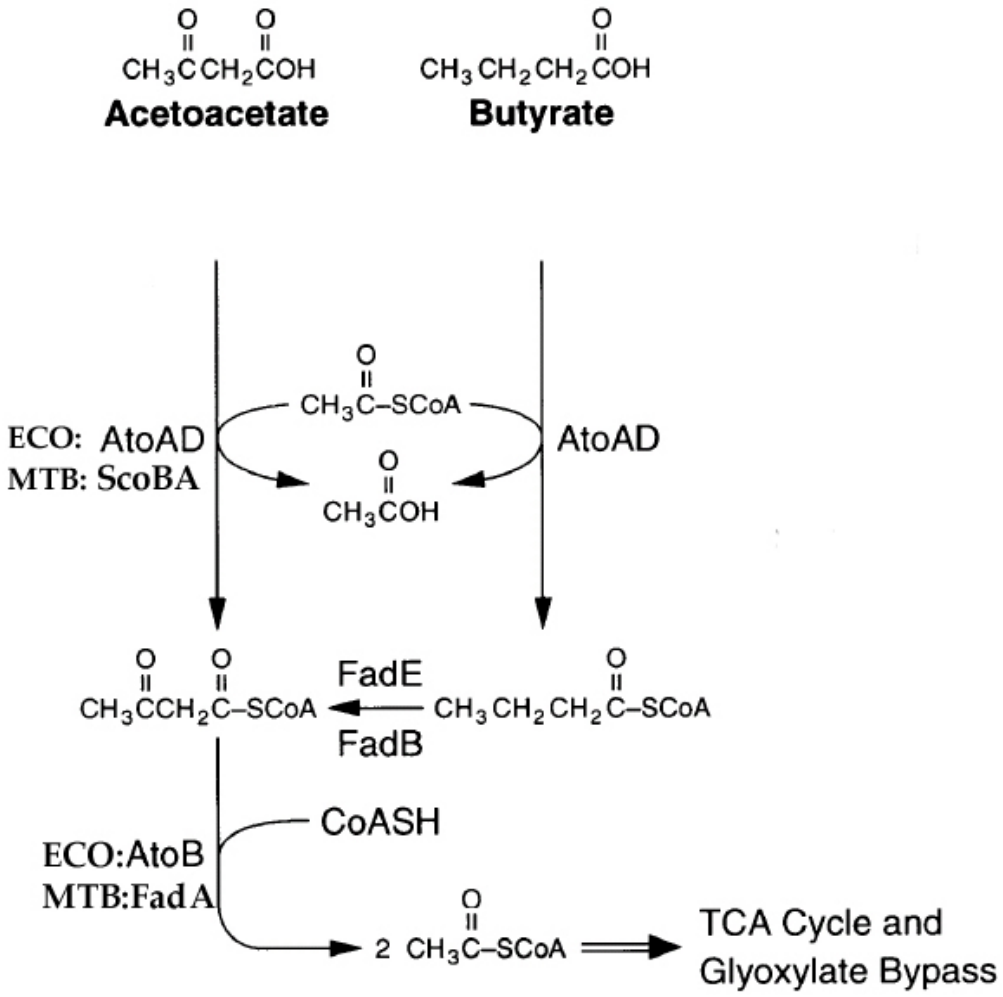


Figure 2.6. Pathways for assimilation of short-chain fatty acids. The *atoABCD* operon is required for catabolism of C4-C6 fatty acids in *E. coli*. The transcription of *atoA*, *atoB*, and *atoD* is under the control of regulators encoded by *atoC* (activator) and *fadR* (repressor). The mycobacterial *scoBA* genes have some homology to *atoAD*, while *fadA* genes have some similarity to *atoB,* which suggests they may play a role similar to the *ato* genes when mycobacteria are grown on C4-C6 fatty acids (butyrate, valerate and hexanoate). Modified from Clark and Cronan (1996).

The *scoBA* genes of *M. tuberculosis* have some homology to *atoAD*, while *fadA* genes have similarity to *atoB.* It is tempting to speculate that the *scoBA* genes participate in the metabolism of short-chain fatty acids; however, there is as yet no functional evidence to supports this hypothesis. On the other hand, neither the *M. tuberculosis* nor the *M. smegmatis* genome contains an obvious homolog of *fadR*. In fact, it is not known if, and how, *fad* genes are regulated in mycobacteria. We have consistently observed a slight lag in the growth of *M. smegmatis* in minimal media containing butyrate or hexanoate as sole carbon sources; this lag was not seen in the growth of *M. smegmatis* in media containing acetate, propionate, or valerate as carbon sources; these data will be presented in detail in the Results section (Chapter 3).

The lack of an obvious FadR homolog in mycobacteria does not preclude the possibility that some transcriptional factors are under the negative or positive control of activated fatty acids, such as fatty acyl-CoAs or fatty acyl-AMPs. The presence of a significant number of hypothetical genes predicted to encode for transcriptional regulators, as well as the existence of numerous *fad* and *fab* promoter regions to analyze computationally, suggests that any potential regulators of the *fad* and *fab* genes will have to be discovered experimentally.

2.4. Acetyl-CoA metabolism

The discovery of coenzyme-A (CoA) and the elucidation of its role as a carrier for acyl groups is arguably one of the most significant biochemical discoveries of the previous century (Bentley, 2000). As early as the 1870s, “reactive” C2 units were implicated in fatty acid biosynthesis; in 1907, Raper proposed a role for active C2 units by stating: “The formation of fatty acids in animals, from carbohydrates, and the occurrence of natural fats such as butter, of all the fatty acids containing an even number of carbon atoms, from two to twenty, suggest that their fatty acids are produced by the condensation of some highly reactive substance containing two carbon atoms and formed in the decomposition of sugar” (Raper, 1907). Rittenberg and Bloch (1945) demonstrated the utilization of acetate carbons for fatty acid synthesis in tissues and suggested that “fatty acids are synthesized by condensation of C2 units” with a $CO \rightarrow CH_2$ conversion.

Investigating lipid catabolism in 1904, Knoop fed dogs fatty acids carrying a terminal phenyl group and analyzed the compounds excreted in the urine; such “tracer” experiments led him to conclude that “fatty acid catabolism required the oxidation at the β carbon with loss of C2 units” (quoted in Bentley, 2000). The β-oxidation theory was confirmed when Schoenheimer and Rittenberg (1937) observed in animals the conversion of C18 stearic acid to C16 palmitic acid to C14 myristic acid and C12 lauric acid. They proposed that the product was acetic acid or a functional derivative, but the structure of the “reactive C2” unit would remain a mystery it was solved in the 1940s by Fritz Lippman (Lippman, 1954).

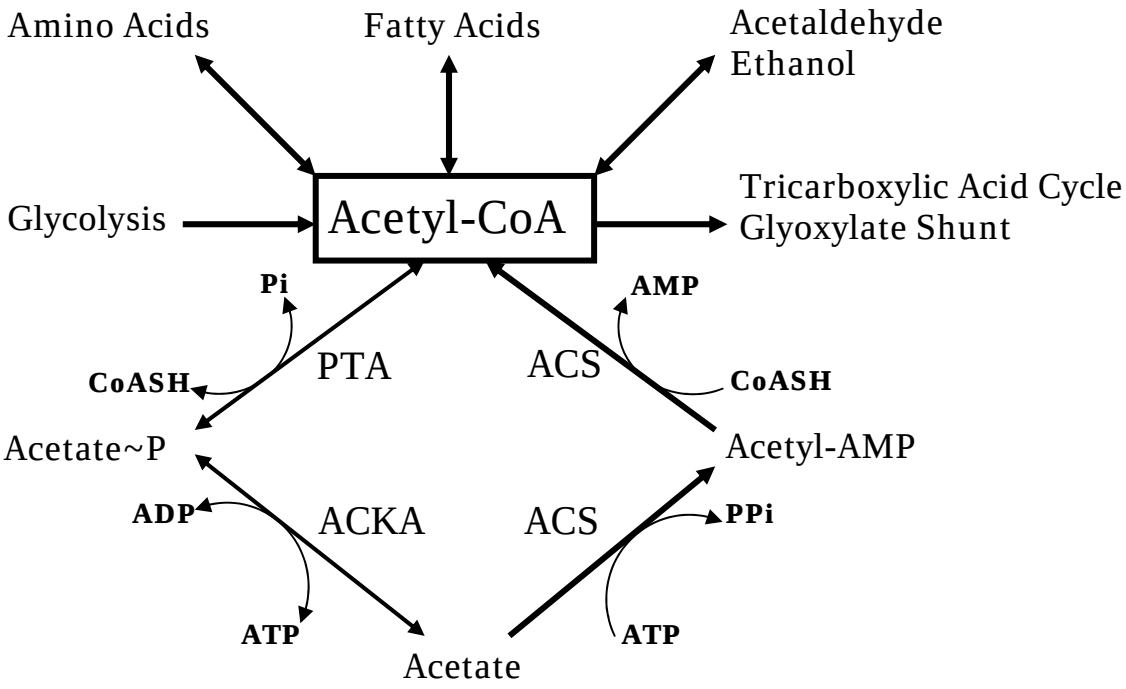


Figure 2.7. Acetyl-CoA in intermediary metabolism. The reactions that activate acetate to acetyl-CoA are detailed. Abbreviated enzymes: PTA, phosphotransacetylase; ACKA, acetate kinase; ACS, AMP-forming acetyl-CoA synthase. Modified from Wolfe, 2005.

Acetyl-CoA can be generated from pyruvate, the end product of glycolysis, via the action of pyruvate dehydrogenase (PDH) (Figure 2.7). As mentioned in the previous subsection, each turn of the β-oxidation cycle releases a molecule of acetyl-CoA. Acetyl-CoA can be generated directly from acetate, by the two-step phosphotransacetylase (PTA) and acetate kinase (ACK) pathway, or directly by acetate CoA synthase (ACS). ACK phosphorylates the acetate molecule, and then PTA replaces the reactive phosphate group with coenzyme A. ACS forms the acetyl-CoA molecyle via an acetyl-AMP intermediate (Wolfe, 2005).

Depending on the organism's metabolic needs, acetyl-CoA can be oxidized by the TCA cycle, where it is converted into two CO_2 molecules and water. When *E. coli* cells are grown in high levels of glucose, they secrete acetate in the media through the PTA-ACKA pathway. When the supply of glucose is exhausted, they "switch" to metabolizing acetate back to acetyl-CoA using ACS (Holms, 1996). Deletion of the *pta-ackA* genes still allows *E. coli* to grow on acetate using ACS; a mutant lacking both pathways, however, is completely unable to utilize acetate (Kumari *et al.,* 1995). *M. tuberculosis* and *M. smegmatis* have homologs of all three genes. Transposon mutants in the *acs* gene of *M. smegmatis* are unable to grow on agar plates containing a low concentration (0.1%) of acetate as the sole carbon. However, when the acetate concentration is increased to 0.5% or 1.0%, the *acs* mutants grow well, implying that *acs* is more efficient at scavenging low amounts of acetate, which cannot be accomplished by PTA-ACK (L. Merkov, unpubl. observations).

2.5. Tricarboxylic Acid Cycle

Hans Adolf Krebs discovered three metabolic cycles: the ornithine cycle, the tricarboxylic acid (TCA) cycle and the glyoxylate cycle (Kornberg, 2000). During his work on the TCA cycle, Krebs built on early experiments done by a Hungarian biochemist, Albert Szent-Gyorgyi, who had studied the processes through which carbohydrates, proteins, and fats were oxidized to CO_2 and water. Szent-Gyorgyi had discovered several C4 salts that were oxidized in tissues—succinate to fumarate, fumarate to malate, and malate to oxaloacetate—but he did not look for a circular pathway of oxidation (Kornberg, 2000).

Krebs envisioned a circle, however, when he realized that adding pyruvate to tissues generated succinate, and that citrate could be formed from oxaloacetate if pyruvate was added. Krebs proposed the cycle in a paper that was rejected by *Nature*; as acetyl-CoA was not known at the time, he suggested that a triose was formed from oxaloacetate that reacted with pyruvate to form citrate (Krebs and Johnson, 1937). Lippman's discovery of acetyl-CoA about a decade later suggested to Krebs the answer to his conundrum (Figure 2.8).

The TCA cycle has two roles: the oxidation of acetyl-CoA to carbon dioxide and water, and the production of precursors for amino acid, porphyrin, and cofactor biosynthesis (Cronan and LaPorte, 1996). The oxidation of acetyl-CoA generates two molecules of NADH, one NADPH, and one $FADH_2$, which in turn are used for ATP production via respiration. The two carbons that enter the TCA cycle from acetyl-CoA are sequentially lost as two carbon dioxide molecules.

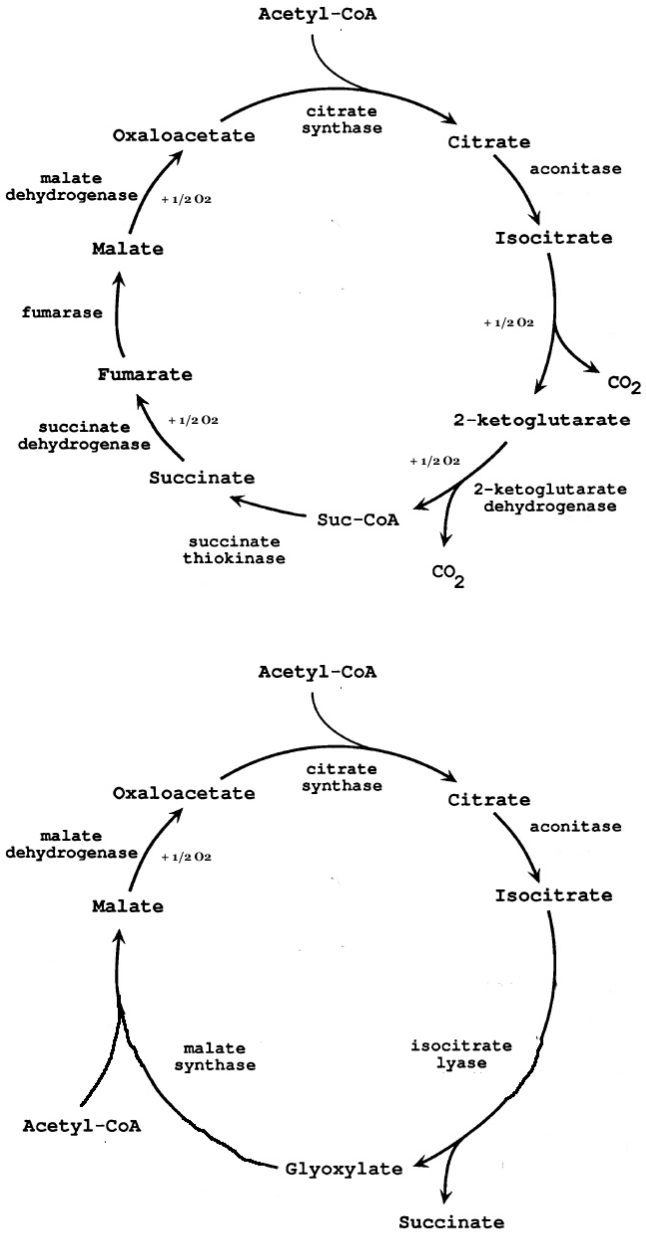


Figure 2.8. Tricarboxylic acid (TCA) cycle (top) and glyoxylate cycle (bottom). The net effect of one turn of the TCA cycle is conversion of one molecule of acetate and two molecules of O_2 into two CO_2 molecules. The glyoxylate cycle bypasses the two CO_2 release steps, thus converting two acetates and half an O_2 molecule into succinate. Modified from Cronan and LaPorte (1996).

The highest induction of the TCA cycle genes in *E. coli* is observed during aerobic growth on acetate or fatty acids. Such conditions, however, necessitate the induction of the glyoxylate shunt genes *aceA* and *aceB* encoding the *E. coli* isocitrate lyase and malate synthase, respectively, so that intermediates lost to biosynthesis can be replenished (Figure 2.8, bottom). The glyoxylate shunt is required for anaplerosis when acetate or fatty acids are the sole carbon sources; partitioning of some of the isocitrate through the shunt allows the regeneration of the four-carbon malate from glyoxylate and acetyl-CoA; this reaction replenishes four-carbon intermediates used for biosynthetis purposes.

Acetyl-CoA enters the cycle in a reaction catalyzed by citrate synthase that yields a molecule of citrate from condensation of oxaloacetate and acetyl-CoA. *M. tuberculosis* has at least three citrate synthase genes (*gltA1, gltA2*, *citA*) while *M. smegmatis* may have at least four (see Table 2.1). Interestingly, an *M. smegmatis* transposon mutant of *gltA2* has a slight growth defect on plates containing acetate as the sole carbon (L. Merkov, unpubl. observations). Aconitase converts citrate to isocitrate; in *E. coli*, the fate of isocitrate is tightly controlled by the phosphorylation status of isocitrate dehydrogenase (IDH). If isocitrate is converted to α-ketoglutarate, a molecule of CO_2 is lost; a second CO_2 molecule is lost when α-ketoglutarate is oxidized to succinyl-CoA by the α-ketoglutarate dehydrogenase complex (KGDH). Succinate thiokinase (*sucCD*) converts succinyl-CoA to succinate, and generates a molecule of ATP. The KGDH genes (*sucAB*) and *sucCD* form an operon in *E. coli.*

Gene	Function	H37Rv	MSMEG
pyk	Pyruvate kinase	1617	3237 0148
pca	Pyruvate carboxylase	2967c	2410 6609
*ppc**	PEP carboxylase	(none)	3107
pck	PEP carboxykinase	0211	0246
ppdK	Pyruvate phosphate dikinase	1127	(none)
*pps**	PEP synthase	(none)	3938
mez	Malic enzyme	2332	(none)
gltA1	Citrate synthase	1131	6608 4040
gltA2	Citrate synthase	0896	5650 5654
acn	Aconitase	1475c	3151
icl1	Isocitrate lyase 1	0467	0904
icl2	Isocitrate lyase 2	1915/16	3713
*prpB**	Methylisocitrate lyase	(none)	6607 6818
glcB	Malate Synthase	1837c	3646
kgd	α-Ketoglutarate carboxylase	1248c	5037
sucB	Pyruvate dehydrogenase E2	2215	4286
sucC	Succinyl-CoA synthase, beta	0951	5503
sucD	Succinyl-CoA synthase, alpha	0952	5502
sdhA	Succinate dehydrogenase A	3318	1667
sdhB	Succinate dehydrogenase B	3319	1666
sdhC	Succinate dehydrogenase C	3316	1669
sdhD	Succinate dehydrogenase D	3317	1668
sdhA2	Succinate dehydrogenase A2	0248c	0408
sdhB2	Succinate dehydrogenase B2	0247c	0407
frdA	Fumarate reductase A	1552	(none)
frdB	Fumarate reductase B	1553	(none)
frdC	Fumarate reductase C	1554	(none)
frdD	Fumarate reductase D	1555	(none)
fum	Fumarase	1098c	5225
mdh	Malate dehydrogenase	1240	(none)
mqo	Malate-quinone oxidoreductase	2852c	2614
*sfcA**	Malic enzyme, NAD linked	(none)	(none)
*maeB**	Malic enzyme, NADP linked	(none)	5043
gcvB	Glycine dehydrogenase	1832	3648

Table 2.1. Homologs of *M. tuberculosis* genes in *M. smegmatis*. Metabolic pathway genes from H37Rv were compared to the recently annotated *M. smegmatis* genome on TIGR's website (http://tigrblast.tigr.org/cmr-blast/). If a gene was not found in H37Rv, the sequence of the *E. coli* protein was used for searches; that gene is indicated in the table with (*). *sucA* (*Rv1248c*) has been renamed *kgd*, per Tian *et al.* (2005).

M. tuberculosis has a different mechanism of generating succinate from α-ketoglutarate (Tian *et al.,* 2005). KGDH activity is missing; α-ketoglutarate is first converted to succinic semialdehyde in a reaction catalyzed by α-ketoglutarate decarboxylase (KGD, encoded by *rv1248c*), and succinic semialdehyde is further oxidized to succinate by two succinic semialdehyde dehydrogenases (SSADH) encoded by *gabD1* (*rv0234*) and *gabD2* (*rv1731*). KGD appears to be important for *in vitro* growth (Sassetti *et al.*, 2003). *M. tuberculosis* would need to generate succinyl-CoA, needed for heme biosynthesis, either from succinate by succinyl-CoA synthetase, or from propionyl-CoA by propionyl-CoA carboxylase and methylmalonyl-CoA mutase (Munoz-Elias and McKinney, 2006).

Interestingly, Tian *et al.* (2005) did not detect KGDH activity in *M. bovis* or *M. smegmatis.* This suggests that the KGDH bypass by the combined activities of KGD and SSADH may be a general feature of mycobacteria. Analysis of the *M. smegmatis* genome revealed that *M. smegmatis* does possess a homolog of *kgd* (*rv1248c*), annotated as MSMEG5037. When testing protein extracts from *M. tuberculosis, M. bovis*, and *M. smegmatis* for TCA cycle enzyme activities, Tian *et al.* (2005) detected malate dehydrogenase (MDH) activity in *M. tuberculosis* and *M. bovis*, but did not detect it in *M. smegmatis*. As shown in Table 2.1, searches with *M. tuberculosis mdh* failed to reveal a homolog in *M. smegmatis*, suggesting the saprophyte may rely solely on MQO, malate-quinone oxidoreductase, to oxidize malate to fumarate. *M. smegmatis* lacks a homolog of the *M. tuberculosis* malic enzyme (*mez*) but has a protein similar to the *E. coli* malic enzyme (*maeB*).

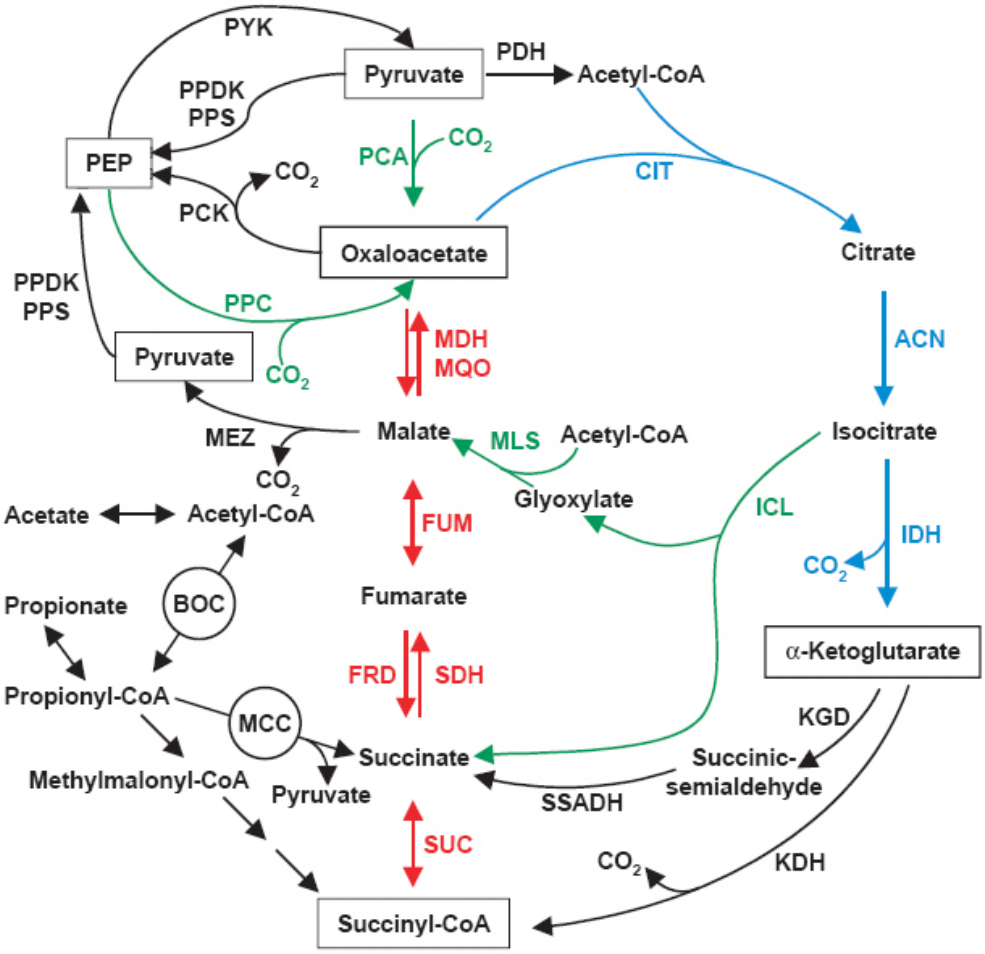


Figure 2.9. Tricaboxylic acid cycle and associated anaplerotic pathways. CIT, ACN and IDH catalyze the oxidative branch (blue arrows), while MDH/MQO, FUM, FRD/SDH and SUC comprise the reductive branch (red arrows) of a bifurcated cycle. Bifurcation occurs when bacteria grow aerobically in excess glucose or anaerobically on any carbon source. Anaplerosis (green arrows) is achieved by PPC or PCA during growth on sugars or by ICL and MLS during growth on acetate or fatty acids. TCA cycle and glycolysis intermediates that are also used for biosynthetic purposes are boxed. See Table 2.1 for gene names. BOC, β-oxidation; MCC, methylcitrate cycle. Reproduced from Muñoz-Elías and McKinney (2006)

2.6. Anaplerosis

The term “anaplerosis” was coined by Hans Kornberg to describe the “filling up” reactions that replenish intermediates in central metabolic pathways (Kornberg, 2003). Since TCA cycle intermediates are used as precursors for biosynthetic reactions, bacteria must replenish them through reactions complementary to the TCA cycle (Figure 2.9). The first anaplerotic enzyme discovered by Kornberg was PEP carboxylase (PPC), which produces oxaloacetate from PEP. *E. coli ppc* mutants are auxotrophic for succinate when grown on sugars but can grow on acetate, lactate, or malate without supplementation (Kornberg, 1966). Residual growth of *ppc* mutants on sugars correlates with expression of the glyoxylate shunt (Peng *et al.*, 2004), likely due to suppressor mutations that increase expression of the glyoxylate shunt genes (Sauer and Eikmanns, 2005).

Although *Corynebacterium glutamicum* possesses a functional PPC, carboxylation of pyruvate to oxaloacetate, mediated by PCA, appears to be the main anaplerotic pathway (Peters-Wendisch *et al.*, 1998). Similarly, *Bacillus subtilis pca* mutants are unable to grow on substrates that enter glycolysis upstream of pyruvate unless supplemented by TCA cycle intermediates (Diesterhaft and Freese, 1973). PPC is the more widely distributed anaplerotic enzyme, as few bacteria aside from *C. glutamicum* and *B. subtilis* use PCA exclusively for anaplerosis (Sauer and Eikmanns, 2005).

PCA appears to be the main anaplerotic pathway in *M. tuberculosis*, as *ppc* is missing (Table 2.1), and *pca* mutants grow poorly on plates containing

both dextrose and glycerol as carbon sources (Sassetti *et al.*, 2003). In constrast, *M. smegmatis* has a homolog of the *E. coli ppc* gene and two homologs of the *M. tuberculosis pca* gene. A recent report implicated PCA as the major anaplerotic enzyme in *M. smegmatis* (Mukhopadhyay and Purwantini, 2000), but additional experiments are needed to reconcile the biochemical data in this report with the existence of a *ppc* gene in addition to two *pca* genes. *pca* appears to be the only anaplerotic gene in *M. bovis* and *M. leprae,* as determined by genomic analysis.

When grown on acetate or fatty acids instead of carbohydrates, *E. coli* depends on the glyoxylate shunt enzymes isocitrate lyase and malate synthase for anaplerosis (Kornberg and Krebs, 1957; Kornberg, 1966). The enzymes of the glyoxylate shunt in *E. coli* are encoded by *aceA* and *aceB*, respectively, and they reside in an operon with a third gene, *aceK*, which encodes IDH kinase-phosphatase, which is involved in regulating carbon flow through the shunt by phosphorylation-mediated inactivation of IDH (Cronan and LaPorte, 1996). Expression of the *aceBAK* genes is induced by acetate or fatty acids; it is under the direct negative control of IclR and indirect negative control of FadR; FadR exerts inhibition by enhancing the expression of the *iclR* gene (Gui *et al.,* 1996). IclR inhibition of the *aceBAK* operon is relieved by PEP (Cortay *et al.,* 1991) while FadR repression is obviated by FadR interaction with fatty acyl-CoA molecules. *E. coli* possesses a second malate synthase, encoded by *glcB*, which functions in growth on glyoxylate and glycolate as sole carbon sources (Molina *et al.,* 1994).

Some bacteria can grow on C2 compounds as sole carbon sources independent of isocitrate lyase activity, due to alternative anaplerotic pathways. *Streptomyces collinus* uses crotonyl-CoA reductase (CCR), which catalyzes the conversion of two acetyl-CoA molecules to butyryl-CoA; *ccr* mutants exhibit reduced growth on acetate. Disruption of a second gene, *meaA*, which may encode an enzyme similar to isobutyryl-CoA mutase, also impairs growth of *S. collinus* on acetate. Interestingly, *S. collinus* can use isocitrate lyase to grow on fatty acids but not on acetate, which requires induction of *ccr* and *meaA* (Han and Reynolds, 1997). Recent evidence suggests the existence of a third pathway for assimilation of acetate in *S. cinnamonensis*, which is capable of growth on acetate even if the glyoxylate cycle and butyryl-CoA pathway are mutationally inactivated (Akopiants *et al.*, 2006).

Most methylotrophic bacteria also lack isocitrate lyase and convert acetyl-CoA to glyxolyate through butyryl-CoA and propionyl-CoA intermediates. Korotkova *et al.* (2002) demonstrated that the acetyl-CoA assimilation cycle in *Methylobacterium extorquens* is linked to the polyhydroxybutyrate cycle and the serine cycle for assimilation of C1 compounds. Two acetyl-CoA molecules are converted to crotonyl-CoA $\rightarrow$ propionyl-CoA $\rightarrow$ succinyl-CoA $\rightarrow$ malyl-CoA, which is broken down to glyoxylate and acetyl-CoA. Glyoxylate then enters the serine cycle, which assimilates C1 and C2 compounds. This complex cycle regenerates glyoxylate in the absence of isocitrate lyase.

2.7. Glyoxylate assimilation pathways

The reacton catalyzed by isocitrate lyase produces succinate and glyoxylate from isocitrate. Malate synthase condenses glyxoxylate with acetyl-CoA to produce malate. As mentioned already, *E. coli* has two malate synthases, which are induced by acetate and fatty acids (20 fold induction of *aceB)* or by glycolate (1,000 fold induction of *glcB*) (Clark and Cronan, 1996). Malate synthase A (*aceB*) is needed for growth on acetate as the sole carbon source; malate synthase G (*glcB*) is dispensable for growth on glycolate or glyoxylate, as glyoxylate induces glyoxylate carboligase (*gcl*) (Ornston and Ornston, 1969).

Glyoxylate carboligase activity was initially described by Krakow and Barkulis (1956) as the conversion of glyoxylate to hydroxypyruvate. Biochemical characterization of the enzyme indicated that GCL converts two glyoxylates into a molecule of tartronate semiladehyde, with the release of carbon dioxide (Krakow *et al.,* 1961) (Figure 2.10). The purified enzyme is a FAD flavoprotein requiring thiamine pyrophosphate and magnesium for activity (Gupta and Vannesland, 1964). The tartronate semialdehyde is reduced to d-glycerate, which is converted by glycerate kinase to 3-phosphoglycerate, an EMP intermediate (Kornberg and Sadler, 1960; Kornberg and Gotto, 1961; Hansen and Hayashi, 1962). Tartronate semialdehyde can be converted to hydroxypyruvate by a hydroxypyruvate isomerase, but the reduction to glycerate represents the relevant conversion for anaplerotic purposes. Malate synthase G does not play a role in this pathway (Ornston and Ornston, 1969).

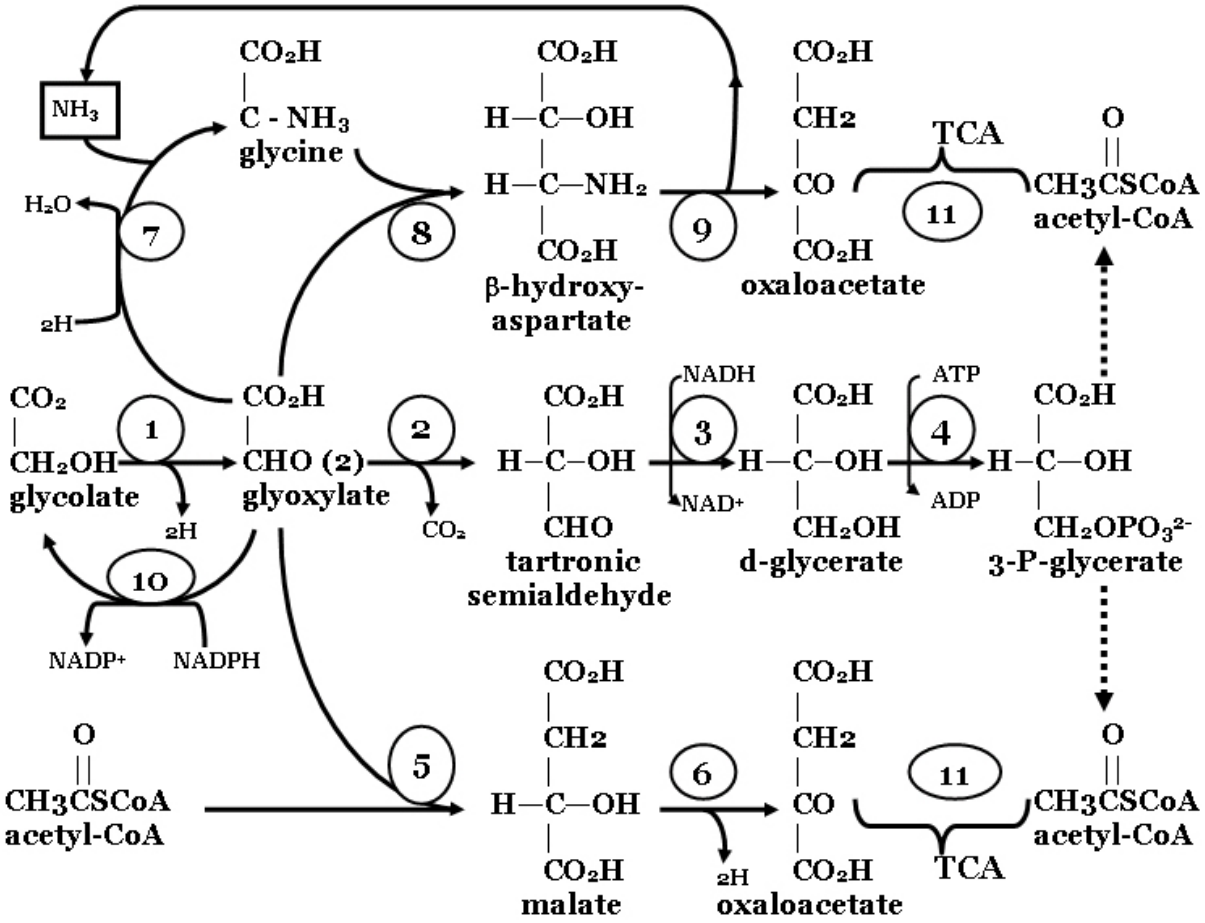


Figure 2.10. Glyoxylate utilization pathways in bacteria. Glyoxylate can be generated from glycolate by glycolate oxidase (GOX, #1 above), from isocitrate by isocitrate lyase (ICL, not shown), or other pathways in ICL-negative bacteria. It can then be utilized via the glyoxylate shunt TCA cycle (5, 6, 11); d-glycerate pathway (2, 3, 4); or *erythro*-β-hydroxyaspartate pathway (7, 8, 9). Numbered enzymes: 2, glyoxylate carboligase; 3, tartronic semialdehyde reductase; 4, glycerate kinase; 5, malate synthase; 6, malate dehydrogenase; 7, glycine dehydrogenase; 8, β-hydroxyaspartate aldolase; 9, β-hydroxyaspartate dehydratase; 10, glyoxylate reductase; 11, citrate synthase. Modified from Kornberg and Morris (1965) and Ornston and Ornston (1969).

The *glcB* gene is unlinked to either the *aceBAK* operon or *gcl* region in *E. coli*. *gcl* maps at min. 12 on the *E. coli* chromosome, while *glcAB* is at min. 64 and *aceBAK* is at min. 90 (Clark and Cronan, 1996). The *gcl* gene was identified by Chang *et al*. (1993); subsequent work by Cusa *et al*. (1999) demonstrated the presence of both *gcl*, *glxR* (encoding the tartronate semialdehyde reductase), and *glxK* (glycerate kinase). These genes are grouped with genes involved in allantoin metabolism; allantoin is a product of purine degradation and can be used as a nitrogen source by *E. coli*. Ashiuchi and Misono (1999) located the hydrohypyruvate isomerase (*hyi*) gene between *gcl* and *glxR.* The *glc* locus, besides the *glcB* gene, also contains *glcE* and *glcD*, encoding glycolate oxidases, *glcF*, an iron-sulfur oxidase, and the positive regulator *glcC* (Pellicer *et al.*, 1996).

Micrococcus denitrificans does not possess glyoxylate carboligase activity but can grow on glycolate or glyoxylate as the sole carbon source by using the β-hydroxyaspartate pathway (Kornberg and Morris, 1965). A molecule of glyoxylate is first reduced to glycine by glycine dehydrogenase; then β-hydroxyaspartate aldolase condenses one glyoxylate and one glycine to form a C4 unit, β-hydroxyaspartate, which is next converted to oxaloacetate by β-hydroxyaspartate dehydratase (Figure 2.10). Oxaloacetate can be converted to PEP by the PEP carboxykinase (PCK) to generate C3 units and acetyl-CoA. While this pathway completely bypasses the d-glycerate pathway, it has been described only in *M. denitrificans* and in *Paracoccus denitrificas,* which also possesses a D-3-hydroxyaspartate aldolase (Liu *et al.*, 2003).

2.8. The *in vivo* importance of the glyoxylate cycle

The first study that in retrospect would suggest a role for the glyoxylate cycle in the ability of intracellular bacteria to survive in the tissues of infected hosts (Segal and Bloch, 1956) was published one year before the glyoxylate cycle was described in bacteria by Kornberg and Krebs (1957). Using a novel method to isolate pathogenic mycobacteria from the lungs of infected mice, Segal and Bloch (1956) compared their ability to respire carbohydrates, benzoates, and fatty acids to the respiratory profile of bacteria grown *in vitro*. Importantly, they noticed that while the *in vitro*-grown bacteria rapidly oxidized glucose, glycerol, pyruvate and acetate, as well as sodium salicylate and fatty acids, bacteria recovered from infected tissues had a respiratory response only to salicylic acid and fatty acids. The long-chain oleic acid (C18) was oxidized more readily than heptanoic acid (C7) or octanoic acid (C8) (Segal and Bloch, 1956). These observations suggested that the metabolism of bacteria in the lungs was adapted for catabolism of fatty acids.

Kanai and Kondo (1974) measured fatty acid oxidation in mycobacteria grown *in vivo* as well. Kondo *et al*. (1970) demonstrated an association between lipids and bacilli harvested from the lungs of mice, by developing an *in vitro* system of lechitin-cholesterol liposomes to study lipid-mycobacteria interactions and then showing that bacilli could metabolize liposomes and release fatty acids.

In vitro models of non-replicating persistence (NRP), which aimed to imitate the chronically infected lung environment, also detected the induction of

enzymes to utilize alternative energy sources (Wayne and Sohaskey, 2001). Murthy *et al*. (1973) reported a decrease in the activity of TCA cycle enzymes in aging *M. tuberculosis* cultures, combined with a five-fold induction of isocitrate lyase activity. Wayne and Lin (1982) saw a comparable induction of isocitrate lyase (ICL) activity in the NRP model, but no change in malate synthase (MLS) activity. Instead, they observed a ten-fold induction in glycine dehydrogenase activity, which catalyzes the reductive amination of glyoxylate to glycine while oxidizing NADH to NAD^+ (Goldman and Wagner, 1962). Wayne and Lin (1982) proposed that the main purpose of glyoxylate $\rightarrow$ glycine reduction was the replenishment of NAD^+ to support microaerobic metabolism. The gene encoding this glycine dehydrogenase activity has not been identified; a likely candidate, *gcvB*, appears to be essential in *M. tuberculosis* (Sassetti *et al.,* 2003).

The importance of ICL for the *in vivo* survival and persistence of *M. tuberculosis* was demonstrated by genetic studies of bacteria lacking *icl1* or *icl2* or both genes (McKinney *et al*., 2000; Munoz-Elias and McKinney, 2005). An *M. tuberculosis* Δ*icl1* mutant successfully established infection in the lungs of mice, but failed to persist in the face of the host immune response as the number of Δ*icl1* bacteria in the lungs decreased over time (McKinney *et al*., 2000). An *M. tuberculosis* Δ*icl1*Δ*icl2* mutant failed to replicate in the lungs of mice, and was cleared from the lungs by two weeks post-infection. Δ*icl1*Δ*icl2* bacteria were also avirulent in mice lacking the macrophage-activating cytokine interferon-γ, which

are exquisitely sensistive to wild-type *M. tuberculosis*, demonstrating that ICL is absolutely required for *in vivo* survival (Munoz-Elias and McKinney, 2005).

ICL activity has been implicated in the virulence of a number of other bacterial and fungal pathogens (Lorenz and Fink, 2002; Idnurm and Howlett, 2002; Wang *et al.*, 2003; Fang *et al.*, 2005; Wall *et al.*, 2005). Similarly, loss of MLS activity has been shown to attenuate pathogenicity in *Rhodococcus fascians* (Vereecke *et al.*, 2002) and *Stagonospora nodorum* (Solomon *et al.*, 2004). Thus, either glyoxylate shunt enzyme, ICL or MLS, could be a viable target for drug development. The existence of two ICL isoenzymes in *M. tuberculosis* necessitates a dual-specific inhibitor that is effective against both. Muñoz-Elías and McKinney (2005) showed that 3-nitropropionate inhibits both ICLs, implying that the development of one drug targeting both enzymes might be possible. Availability of the X-ray crystal structures of ICL1 (Sharma *et al.,* 2000) and MLS (Smith *et al.*, 2003) should facilitate drug development.

The crystal structure of *M. tuberculosis* MLS suggests that it might be a more attractive drug target than ICL1/ICL2. The active site in MLS is at the end of ~ 15 angstrom tunnel which accommodates coenzyme-A; this tunnel appears to be an attractive target for inhibitor binding (Smith *et al.*, 2003). Attempts to delete the *M. tuberculosis glcB* gene encoding MLS have been unsuccessful so far, suggesting that *glcB* might be essential (Muñoz-Elías, 2005). We decided to evaluate the role of MLS in mycobacterial metabolism using the saprophyte *M. smegmatis,* a non-pathogenic relative of *M. tuberculosis*, as a model system.

CHAPTER 3

Studies of the Glyoxylate Shunt in *M. smegmatis*

3.1. Deletion of the malate synthase gene (*glcB*) in *M. smegmatis*

Mycobacteria, unlike *E. coli,* possess only one gene encoding malate synthase (MLS); the protein encoded by it is more homologous to malate synthase G than malate synthase A. The MLS protein [2.3.3.9] of *M. smegmatis* is encoded by the *glcB* gene and is annotated as MSMEG3646 in The Institute of Genomic Research (TIGR) Comprehensive Microbial Resource (CMR) database. The *M. smegmatis* MLS has 732 amino acids, 9 fewer than the *M. tuberculosis* MLS; the two proteins share 80 percent identity and 86 percent similarity.

The *M. smegmatis glcB* gene was deleted by homologous recombination, using a two-step counter-selection method (Figure 3.0). An internal 1.9 kb MscI fragment was removed to create an in-frame deletion (Figure 3.1), and the deletion was confirmed by Southern blot. The mutant was complemented with an integrative plasmid carrying the *M. tuberculosis glcB* gene.

An ethanemethylsulfonate (EMS)-induced mutant carrying a point mutation in the *icl1* gene was unable to grow on agar plates containing acetate as the sole carbon source (McKinney *et al.*, 2000). Thus, we tested the ability of the Δ*glcB* mutant to grow on minimal media plates containing glucose, short-chain fatty acids (C2-C5), and a long-chain fatty acid, methylpalmitate (C16). Wild-type *M. smegmatis* forms colonies on glucose or propionate (C3) in 3 days, acetate (C2) in 4 days, and butyrate (C4) and valerate (C5) in 5 days. Growth of

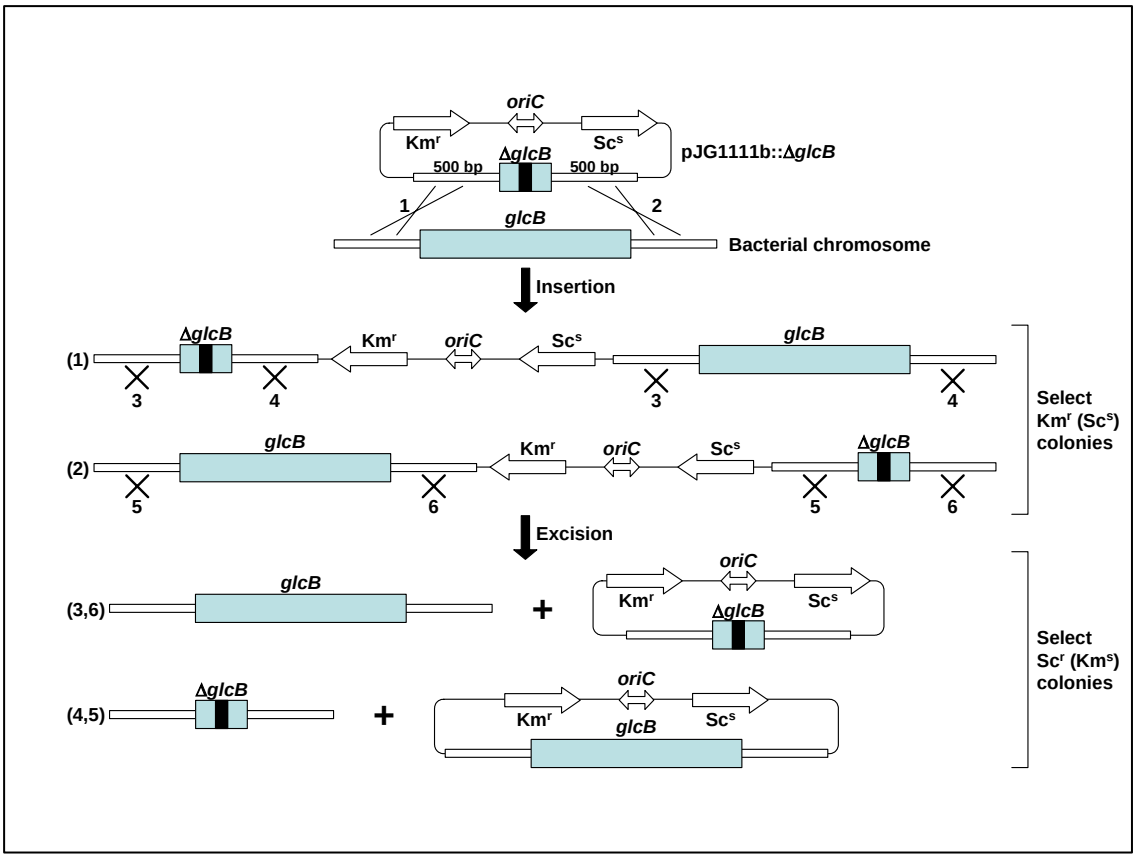


Figure 3.0. Construction of the Δ*glcB* mutant. The knockout vector contains the *aph* kanamycin resistance marker (Km^r), the *sacB* sucrose sensitivity marker (Sc^s), and *oriC* for replication in *E. coli*. The *glcB* knockout plasmid contains the Δ*glcB* allele with an in-frame deletion (black) in the *glcB* ORF (grey) and ~ 500 bp of 5'/3' flanking sequences (white). The Δ*glcB* knockout vector is electroporated into *M. smegmatis*; transformants are selected on 7H10 agar + 25 µg ml-1 Km. The knockout vector does not contain a mycobacterial *ori*; Km^r transformants can only arise by insertion of the plasmid in the chromosome. Recombination between the Δ*glcB* allele and the chromosome upstream (**1**) or downstream (**2**) of the internal *glcB* deletion results in plasmid insertion and merodiploid formation with the configurations shown. Outgrowth of Km^r colonies in the absence of Km allows accumulation of Km^s cells in which plasmid excision from the chromosome occurred by reverse recombination. The Km^s progeny retain wild-type *glcB* (**3,6**) or Δ*glcB* (**4,5**) in the chromosome. The excised plasmid, which cannot replicate, is lost during successive rounds of cell division. The reverse recombinants are selected by plating on 7H10 agar + 5% sucrose, which kills cells expressing the Sc^s marker. Individual Km^s Sc^r colonies are cloned and PCR-screened to identify Δ*glcB* mutants, which are confirmed by Southern blot. Figure was kindly provided by Dr. John McKinney.

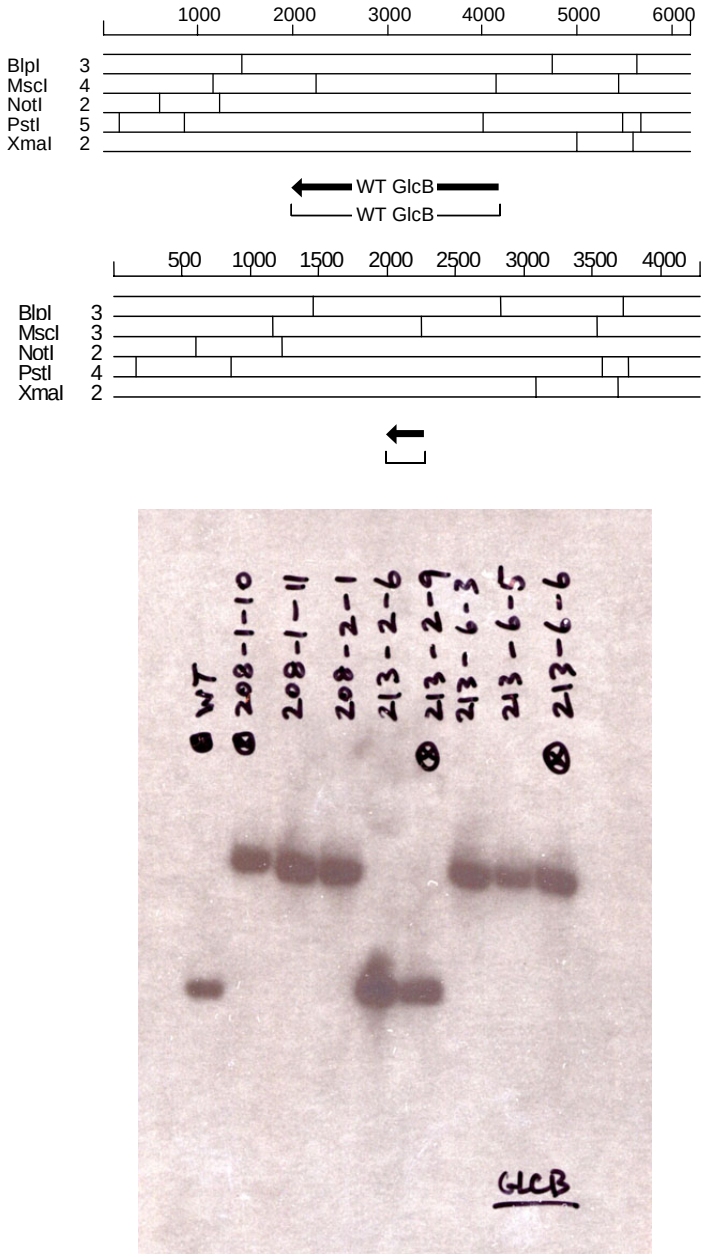


Figure 3.1. Southern blot to confirm the deletion of the *glcB* gene in wild type *M. smegmatis.* Selected restriction sites of the genomic region are shown for the wild-type (top) and mutant (bottom) alleles. Genomic DNA was digested with PstI for the Southern blot (right), and the BlpI-MscI fragment downstream of the gene was used as a probe. Expected bands: 1,464 bp for the wild-type, 2,708 bp for the *glcB* mutant. Strain 208-1-10 is a Δ*glcB* single mutant.

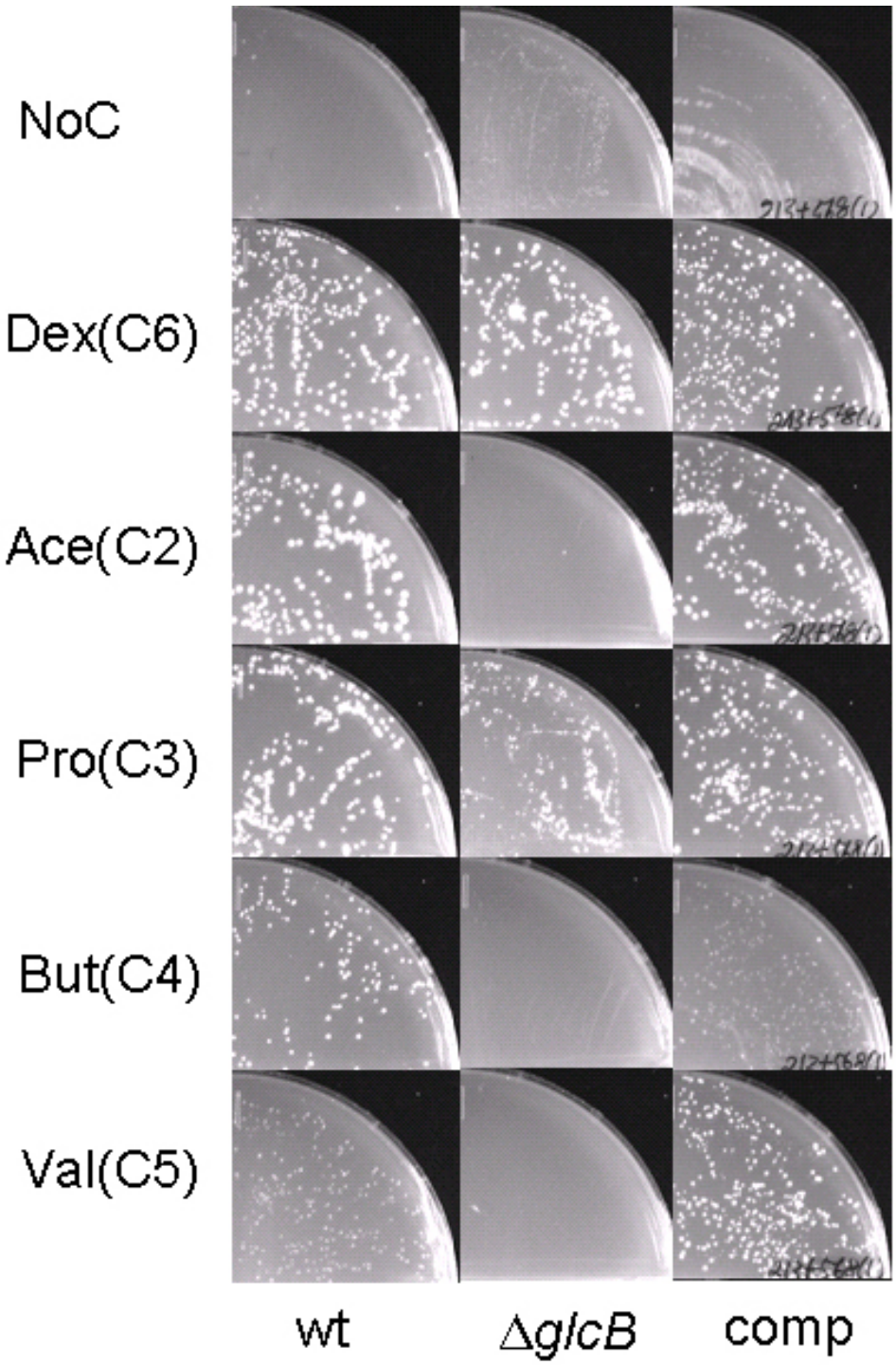


Figure 3.2. Phenotype of the Δ*glcB* mutant of *M. smegmatis* after 5 days on minimal media plates, with the indicated compounds (glucose, acetate, propionate, butyrate or valerate) added as the sole carbon sources. Growth of wild type *M. smegmatis* (wt) and the complemented mutant (comp) is also shown. The "comp" strain is the Δ*glcB* mutant complemented with the *M. tuberculosis glcB* gene carried on a single-copy integrative plasmid.

the Δ*glcB* mutant on plates with various carbon substrates was therefore inspected at day 5 (Figure 3.2). As expected, by 5 days wild type (wt) *M. smegmatis* had formed colonies on all carbon sources, as had the complemented (comp) Δ*glcB* mutant. The Δ*glcB* mutant formed colonies on agar plates containing glucose and propionate, as we had observed with the *icl1* EMS mutant, confirming that the glyoxylate shunt is dispensable for growth of mycobacteria on carbohydrates, and that propionate assimilation in *M. smegmatis* does not require the glyoxylate shunt, in contrast to *M. tuberculosis* (Muñoz-Elías *et al.*, 2006). However, at day 5, the Δ*glcB* strain had failed to form colonies on plates containing acetate or the short-chain fatty acids (SCFA) butyrate and valerate as the sole carbon sources.

However, when the plates were inspected after 10 days of incubation, small colonies of the Δ*glcB* mutant had appeared on the acetate and valerate containing plates (Figure 3.3). After 15 days, the Δ*glcB* mutant formed colonies on all substrates. Furthermore, the Δ*glcB* mutant could grow, albeit not robustly, on solid media containing methylpalmitate as the sole carbon source. Thus, unlike *icl1* in *M. smegmatis*, *glcB* is not required for growth on solid media containing acetate or SCHA as the sole carbon source, although *glcB* is required for optimal growth.

MLS was also dispensable for growth on solid media containing glyoxylate as the sole carbon source. In fact, the Δ*glcB* mutant grew just as well as the wild type strain on glyoxylate plates (Figure 3.4), suggesting the existence of another,

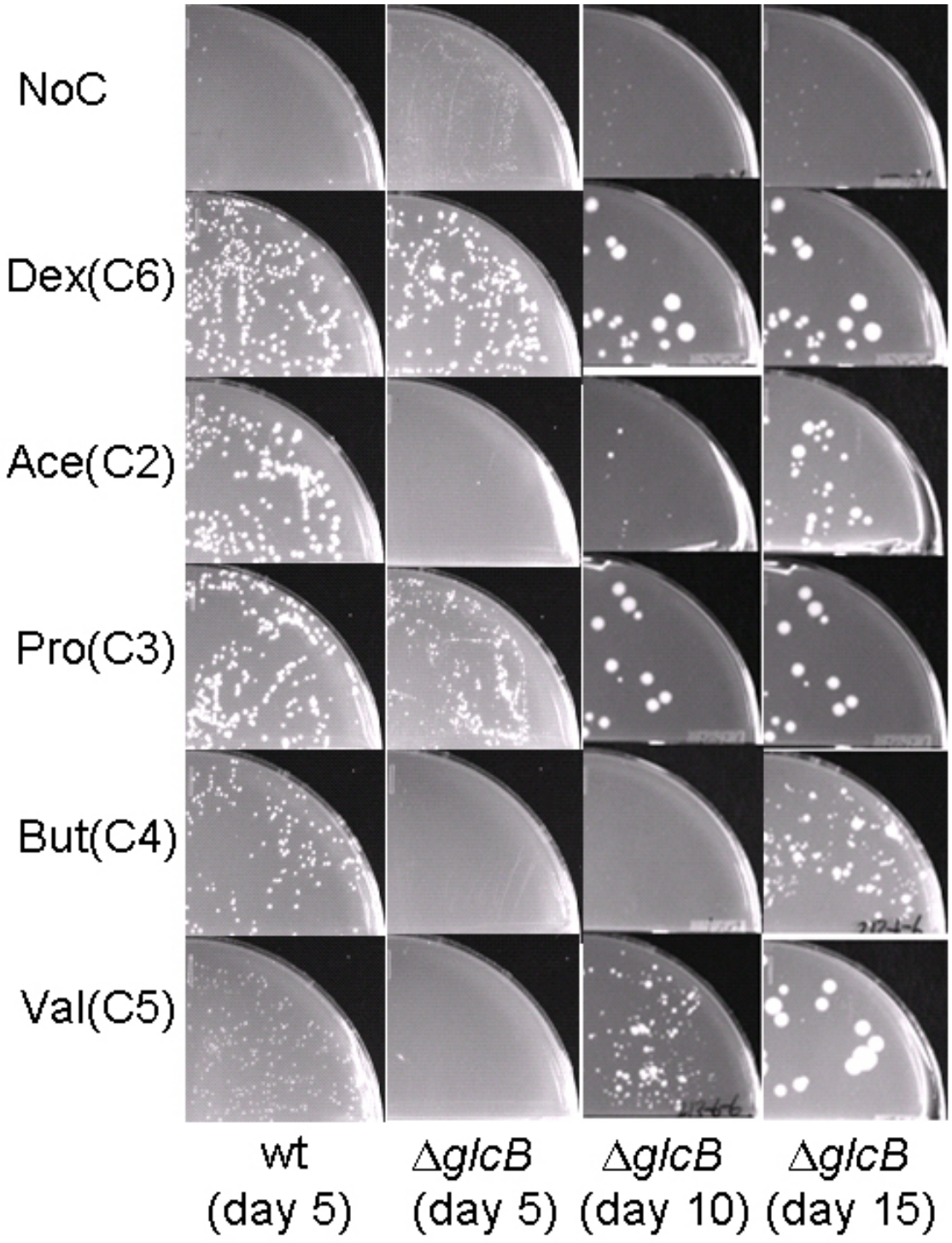


Figure 3.3. MLS is not essential for growth of *M. smegmatis* on acetate and short-chain fatty acids (butyrate and valerate); however, MLS is required for optimal growth on these compounds as sole carbon sources. The Δ*glcB* mutant grows fastest on plates with propionate, then glucose, then valerate, then acetate and then butyrate.

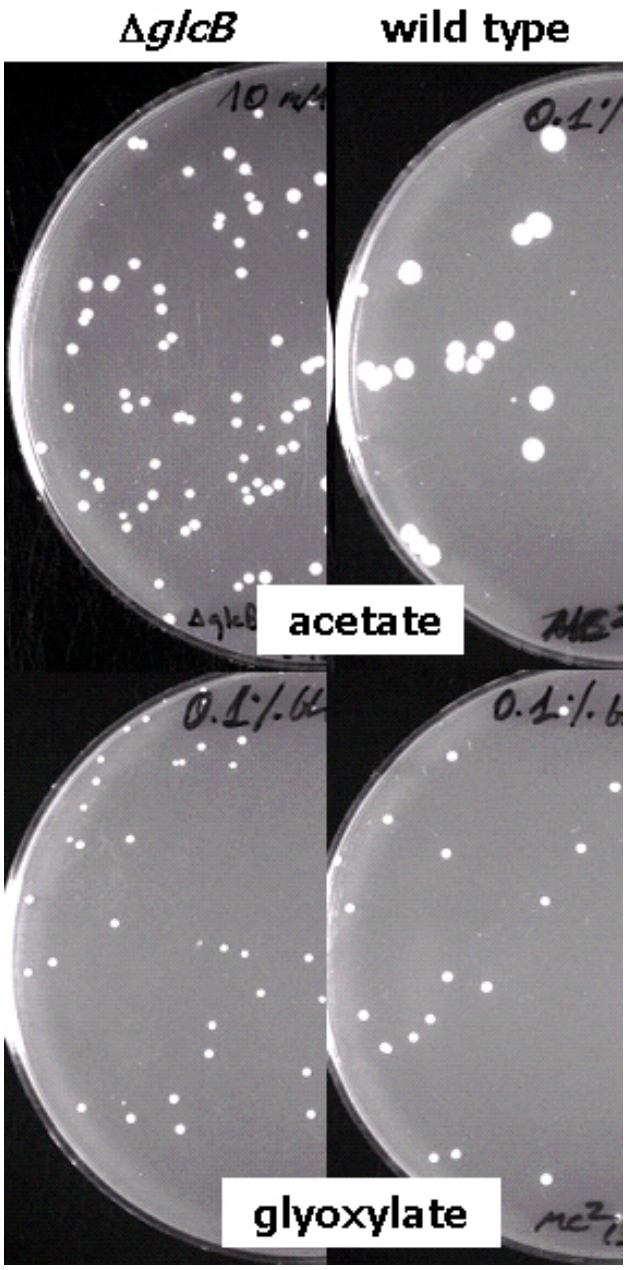


Figure 3.4. MLS is dispensable for growth on acetate (top) and glyoxylate (bottom) in *M. smegmatis*. While the Δ*glcB* mutant (left) does not grow as well as wild type on acetate, it is indistinguishable from wild type (right) on glyoxylate.

MLS-independent, pathway for the assimilation of glyoxylate. This alternative pathway appears to be the dominant pathway for growth on glyoxylate as the sole carbon source and MLS does not seem to play an essential role in it. We considered the possibility that a second MLS might exist in *M. smegmatis*, although we were not able to detect it by Southern blots. To test this possibility, we performed enzyme assays on cell-free extracts (Figure 3.5).

The wild type, Δ*glcB* mutant, and complemented strains were grown in minimal liquid media containing glucose, acetate, propionate, or short-chain fatty acids as sole carbon sources, and the MLS activity of cell-free extracts was measured using the method of Smith *et al.* (2003). In wild type *M. smegmatis*, a two- to three-fold induction of MLS activity by acetate and short-chain fatty acids, was observed. The smallest induction was observed with hexanoic acid (C6). In contract, the MLS activity of the complemented strain (controlled by the *M. tuberculosis glcB* promoter) was not induced by any carbon source.

These data agree with Smith *et al.* (2003), who measured the activity of MLS in cell-free extracts of *M. tuberculosis* and saw no induction of enzyme activity by acetate, palmitate, or glucose. Smith *et al*. (2003) observed a two-fold increase in activity when glycolate was the sole carbon source. It appears that *M. smegmatis* may regulate the expression of MLS depending on the available carbon source, which is not the case in pathogenic *M. tuberculosis*. Importantly, the cell-free extracts from Δ*glcB* bacteria had no detectable MLS activity, irrespective of the amounts of substrates or protein used in the assays.

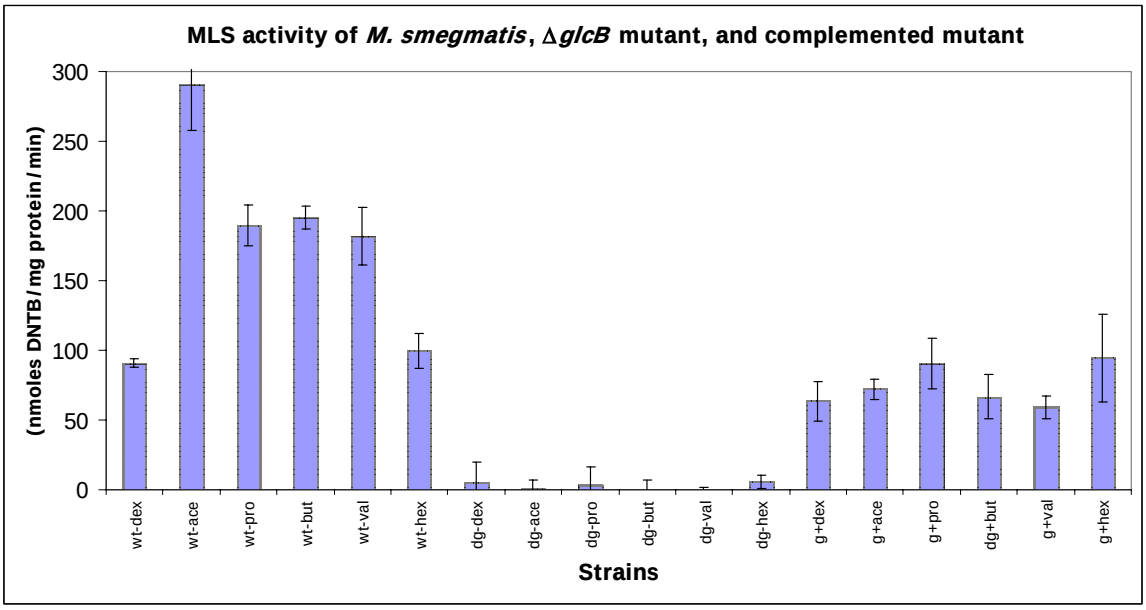


Figure 3.5. MLS activity in cell-free extracts of wild-type *M. smegmatis* (wt), Δ*glcB* mutant (Δg), and the Δ*glcB* mutant complemented with the *M. tuberculosis glcB* gene (g+). Bacteria were grown to early log phase in M9 minimal media with 0.1 % glucose (dex), acetate (ace), propionate (pro), butyrate (but), valerate (val), or hexanoate (hex) as the sole carbon source, and MLS activity of the cell-free extracts was measured. Three cultures were grown for each strain in each medium, and duplicate measurements of MLS activity were obtained from each extract.

3.2. Deletion of the isocitrate lyase genes (*icl1*, *icl2*) in *M. smegmatis*

The unexpected finding that MLS activity is not absolutely required for growth of *M. smegmatis* on minimal media with acetate or fatty acids as the sole carbon source led us to reconsider the anaplerotic role of the glyoxylate shunt of *M. smegmatis*. Prior genetic studies of the gloxylate shunt in our laboratory were done with a mutant obtained by chemical mutagenesis; sequencing of the *icl1* gene of the mutant revealed a nonsense mutation in the coding sequence, indicating the *icl1* gene was likely inactive (A. Upton, personal communication). However, to make certain that our studies were done with a truly ICL-null strain of *M. smegmatis*, we deleted the *icl1* and *icl2* genes, individually and in combination.

The *icl1* gene was deleted by homologous recombination. An internal 855 bp SfoI fragment was removed, giving rise to a shortened protein lacking 285 amino acids and missing the ICL catalytic KKCGH motif (Sharma *et al.,* 2000; Figure 3.6). The *icl1* gene was deleted in wild-type *M. smegmatis* and in the Δ*glcB* strain, which generated a mutant deficient in both steps of the glyoxylate cycle (Δ*icl1* Δ*glcB*). Ernesto Muñoz-Elías in the laboratory had generated a Δ*icl2* strain of *M. smegmatis* and he used the *icl1* knockout construct to create a double mutant (Δ*icl1* Δ*icl2*). All of the mutated alleles were constructed as unmarked in-frame deletions, using the strategy depicted in Figure 3.0, in order to avoid polar effects on expression of downstream genes.

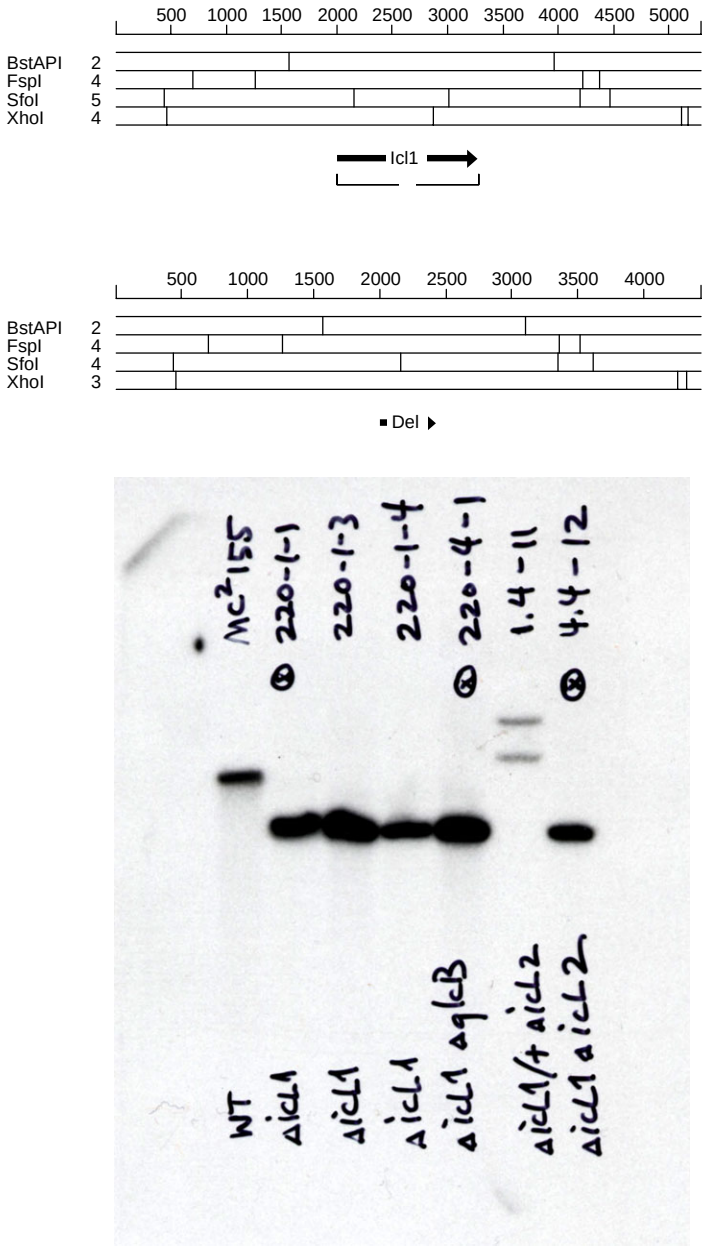


Figure 3.6. Southern blot to confirm the deletion of the *icl1* gene in wild type *M. smegmatis*, in the Δ*glcB* strain, and in the Δ*icl2* strain. Selected restriction sites of the genomic region are shown for the wild type (top) and mutant (bottom) alleles. The 855 bp SfoI site was removed to create an in-frame delection. Genomic DNA was digested with FspI for the Southern blot (right), and a 677 bp region upstream of the gene was used as a probe. Expected bands: 2,958 bp for the wild type, 2,103 bp for the mutant. Strains 220-1-1,3,4 are Δ*icl1* single mutants; 220-4-1 is the Δ*icl1*Δ*glcB* double mutant; and strain 4.4-12 is the Δ*icl1*Δ*icl2* double mutant.

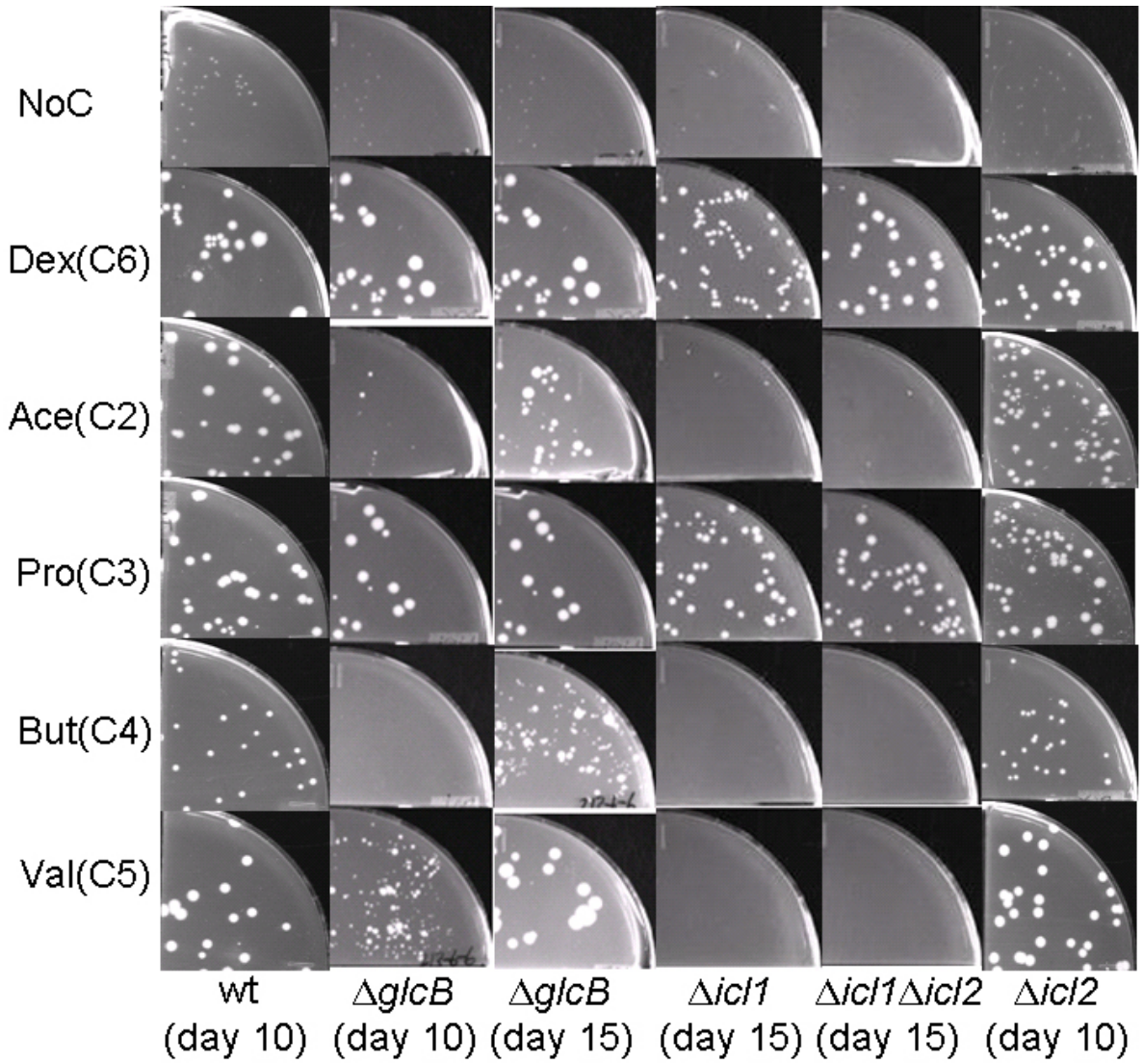


Figure 3-7. The *glcB* and *icl2* genes are dispensable for growth on acetate and short chain fatty acids (C4-C6). The Δ*icl2* mutant is indistinguishable from wild type on all substrates, while the Δ*glcB* strain has a growth delay on acetate, butyrate and valerate. The Δ*icl1* and Δ*icl1*Δ*icl2* mutant grow in liquid media supplemented with valerate (C5) as the sole carbon source (not shown), but grow poorly on plates with valerate.

The wild-type, Δ*icl1*, Δ*icl2*, Δ*icl1* Δ*icl2*, and Δ*glcB* strains were directly compared for growth on minimal solid media containing glucose, short-chain fatty acids, or the long chain fatty acid methylpalmitate as sole carbon sources (Figure 3.7). The Δ*glcB* single mutant was compared to the *icl* mutants in order to dissect the role of the individual glyoxylate cycle enzymes in the anaplerotic growth of *M. tuberculosis* on C2 substrates. Several aspects of the glyoxylate cycle in *M. smegmatis* become apparent from Figure 3.7. First, the *icl2* gene is dispensable for bacterial growth on any carbon substrate; on solid media and in liquid media, the Δ*icl2* mutant behaves like the wild-type strain. Second, the presence of a single functional copy of the *icl2* gene under the control of its native promoter fails to rescue the *icl1* deficiency. However, overexpression of the *icl2* gene on a multi-copy epidomal plasmid did enable growth of the Δ*icl1* and Δ*icl1* Δ*icl2* mutants on acetate and fatty acids (Ernesto Muñoz-Elías, personal communication), suggesting that most likely the endogenous *icl2* gene is functional but is not expressed at sufficient levels to compensate for the loss of the *icl1* gene.

Third, the ability of *M. smegmatis* to grow on acetate and on fatty acids is dependent on the *icl1* gene. The Δ*icl1* Δ*glcB* double mutant cannot grow on C2 substrates, similarl to the Δ*icl1* and Δ*icl1* Δ*icl2* mutants (Figure 3.8). However, the Δ*icl1* Δ*glcB* mutant will grow on glucose or propionate, confirming that the glyoxylate shunt is dispensable in *M. smegmatis* when sugars or propionate are utilized as sole carbon sources. Lastly, it appears that *M. smegmatis* uses a MLS-

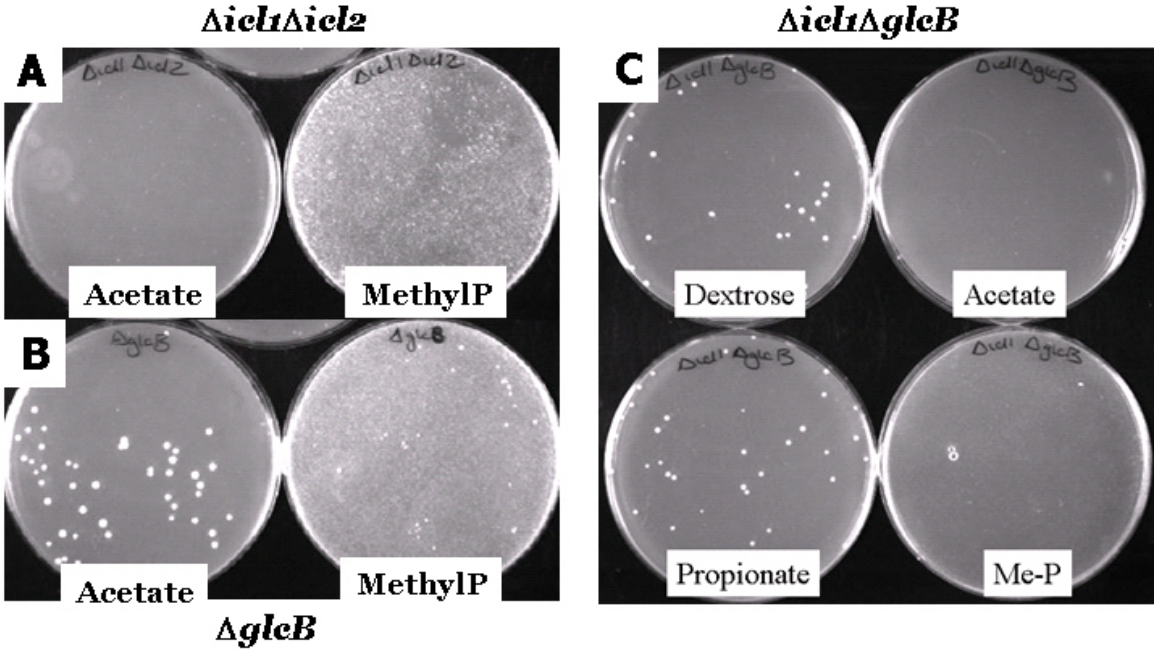


Figure 3.8. The Δ*icl1* Δ*icl2* mutant (**A**) cannot grow on plates containing acetate or the fatty acid methylpalmitate as carbon sources, while the Δ*glcB* mutant (**B**) does grow, albeit slowly. As expected, the Δ*icl1* Δ*glcB* mutant cannot grow on acetate or methylpalmitate, but grows on gloucose and propionate (**C**).

independent but ICL-dependent pathway for anaplerotic growth on acetate and fatty acids. Such a pathway has not been described in mycobacteria, but it is logical that this other gene, or set of genes, will assimilate glyoxylate. Searches of the published mycobacterial genomes did not reveal any homologs of known glyoxylate utilization genes, although *M. tuberculosis* possesses at least one homolog of the *E. coli* glycolate oxidases, which convert glycolate to glyoxylate.

3.3. Deletion of the alanine dehydrogenase (*ald*) gene in *M. smegmatis*

The concomitant induction of glycine dehydrogenase activity and ICL activity in aging or non-replicating cultures of *M. tuberculosis* (Murthy *et al.* 1973; Wayne and Lin, 1982) suggested that glycine dehydrogenase might be responsible for ICL-dependent, MLS-independent growth of *M. smegmatis* on acetate and fatty acids. An obstacle to testing this hypothesis was the fact that the gene(s) encoding glycine dehydrogenase activity had not been identified in mycobacteria. The *M. tuberculosis* genome contains at least three putative glycine dehydrogenase (*gcv*) genes; the product of the most likely candidate gene, *gcvB*, is thought to preferentially catalyze the decarboxyation of glycine rather than the amination of glyoxylate (Cole *et al.,* 1998; Wayne and Sohaskey, 2001).

Alanine dehydrogenase of *M. smegmatis* was reported to have preference for the reductive amination of pyruvate to alanine; furthermore, it was induced in *M. smegmatis* during dormancy, similarly to the glycine dehydrogenase activity in the "Wayne model" of non-replicating persistence (Hutter and Dick, 1998). Usha *et al*. (2002) purified the glycine dehydrogenase activity from *M. smegmatis* grown under microaerobic conditions, and reported that the purified protein could catalyze glyoxylate and pyruvate amination and L-alanine deamination. They suggested that the alanine dehydrogenase (ALD) of *M. smegmatis* might be responsible for metabolizing glyoxylate. Searches of the TIGR databases revealed one *ald* gene in *M. smegmatis*, encoding a 371 aa protein. The *ald* gene was

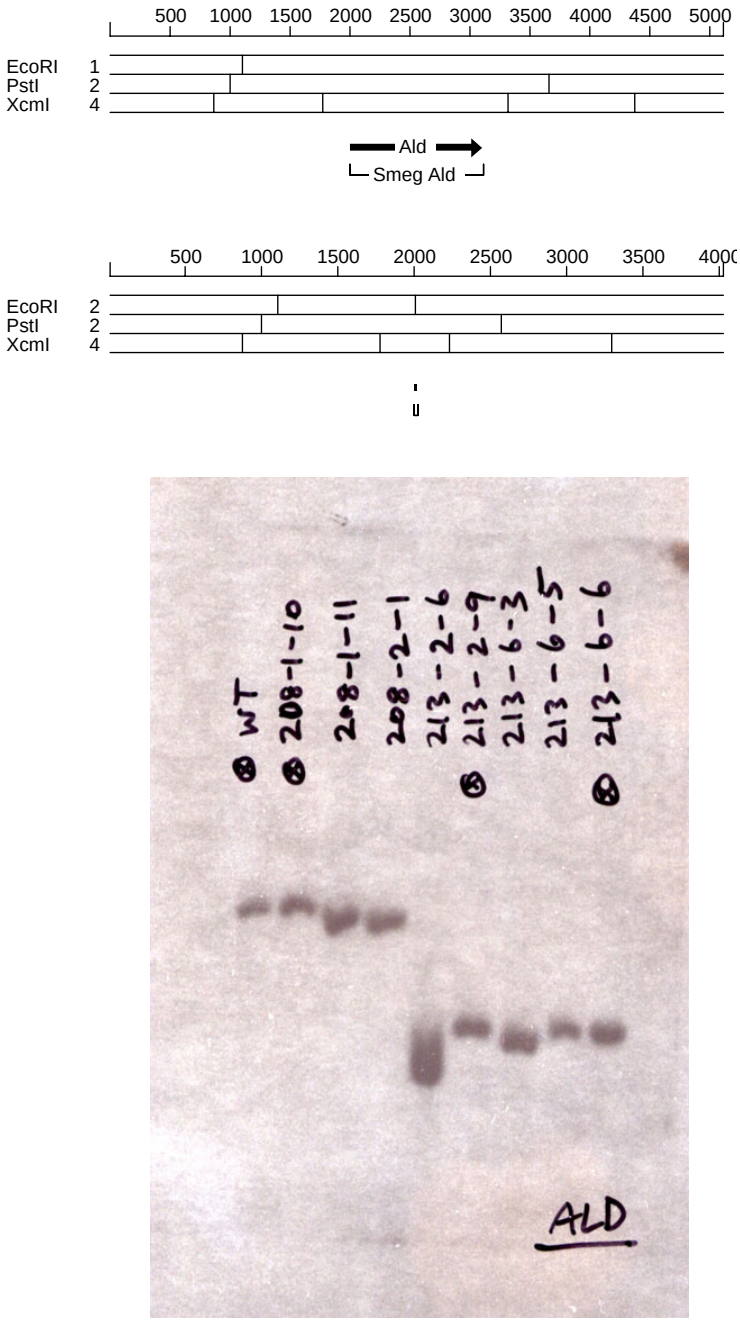


Figure 3.9. Southern blot to confirm the deletion of the *ald* gene in wild-type *M. smegmatis* and in the Δ*glcB* strain. Selected restriction sites of the genomic region are shown for the wild type (top) and mutant (bottom) alleles; the deletion introduced an EcoRI site in the mutant strain. Genomic DNA was digested with PstI for the Southern blot and the 800 bp region upstream of the gene was used as a probe. Expected bands: 2,664 bp for wild-type, 1,575 bp for Δ*ald* mutant. Strain 208-1-10 is the Δ*glcB* single mutant; strain 213-2-9 is the Δ*ald* single mutant; strain 213-6-6 is the Δ*glcB* Δ*ald* double mutant.

deleted by creating an unmarked in-frame deletion that removed all but the first 6 and last 6 amino acids (Figure 3.9). The *ald* gene was deleted in wild type *M. smegmatis* and in the Δ*glcB* strain. If ALD represented the ICL-dependent, MLS-independent activity, which enabled the Δ*glcB* mutant to grow on acetate and glyoxylate, then the Δ*glcB* Δ*ald* double mutant should not grow on solid media containing acetate or fatty acids as the sole carbon source. However, the Δ*glcB* Δ*ald* double mutant grew as well as the parental Δ*glcB* strain on media contaioning acetate (Figure 3.10) or glyoxylate (not shown). It is thus clear that, even if the *M. smegmatis* ALD could catalyze glyoxylate deamination, the product of the *ald* gene is not essential for glyoxylate assimilation.

At the time the Δ*ald* and Δ*glcB* Δ*ald* mutants were generated in our lab, a published report described the disruption of the *ald* gene in *M. smegmatis* (Feng *et al.,* 2002). In this study, the *ald* mutant had a defect in utilizing alanine as the sole nitrogen source, and grew poorly under microaerobic conditions; however, its glycine dehydrogenase activity remained at wild type levels, indicating that another *M. smegmatis* enzyme was responsible for the reductive amination of glyoxylate. Although the report by Feng *et al*. (2002) made it unnecessary to test the glycine dehydrogenase activity in our Δ*ald* and Δ*glcB* Δ*ald* mutants, it provided no indication as to which gene was encoding glycine dehydrogenase activity and, more importantly, which other gene(s) encoded the ICL-dependent, MLS-dependent pathway for growth on acetate and fatty acids.

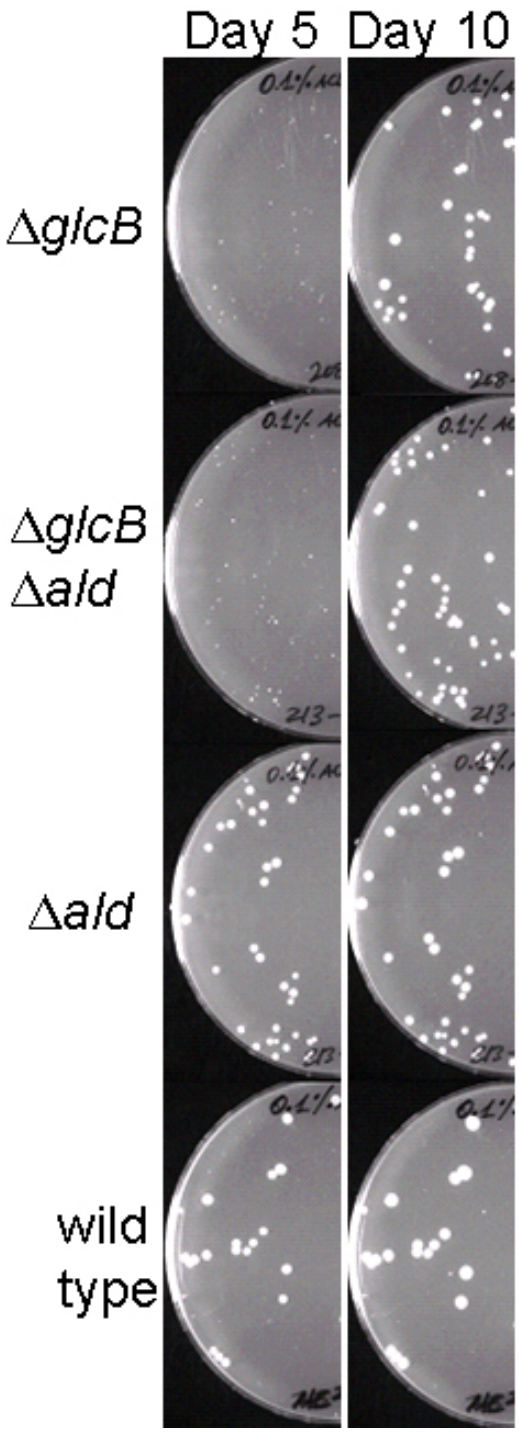


Figure 3.10. Alanine dehydrogenase (ALD) is not required for growth on acetate (or glyoxylate or methylpalmitate, not shown). Thus, ALD does not seem to be required for the assimilation glyoxylate, acetate, or fatty acids by *M. smegmatis*.

3.4. Screen for C2 mutants in MLS-deficient *M. smegmatis*

To identify mutants that were defective in the putative ICL-dependent, MLS-independent pathway for growth on acetate and fatty acids, we created a library of transposon mutants in the Δ*glcB* strain of *M. smegmatis*. Since the Δ*glcB* mutant grew on acetate and fatty acids, albeit slowly, we screened for transposon mutants that were completely unable to form colonies on agar plates containing acetate and methylpalmitate as sole carbon sources. We then transformed these Ace^- Mep^- mutants individually with a plasmid containing the *M. tuberculosis glcB* gene, and screened for complementation of growth on acetate and methylpalmitate. If we could restore a mutant's ability to grow on acetate and fatty acids by complementation with *glcB*, this would indicate that the mutant was defective in the putative ICL-dependent, MLS-independent pathway we were seeking (Figure 3.11).

We constructed a library of ~ 50,000 transposon mutants in the *glcB*-deficient strain by using the transposon donor phagemid ϕMycoMarT7 (Sassetti *et al*., 2003), which was a generous gift from Dr. Eric Rubin (Harvard School of Public Health). More than 4,000 transposon mutants were screened to identify mutants that were capable of grow on glucose but not on acetate or methylpalmitate. Of the 4,000 mutants screened, more than 50 were retested to confirm their phenotypes; 24 bred true on retesting, and these were transformed individually with a plasmid carrying a copy of the *M. tuberculosis glcB* gene. The sites of transposon insertion in the 24 mutants were identified by subcloning the

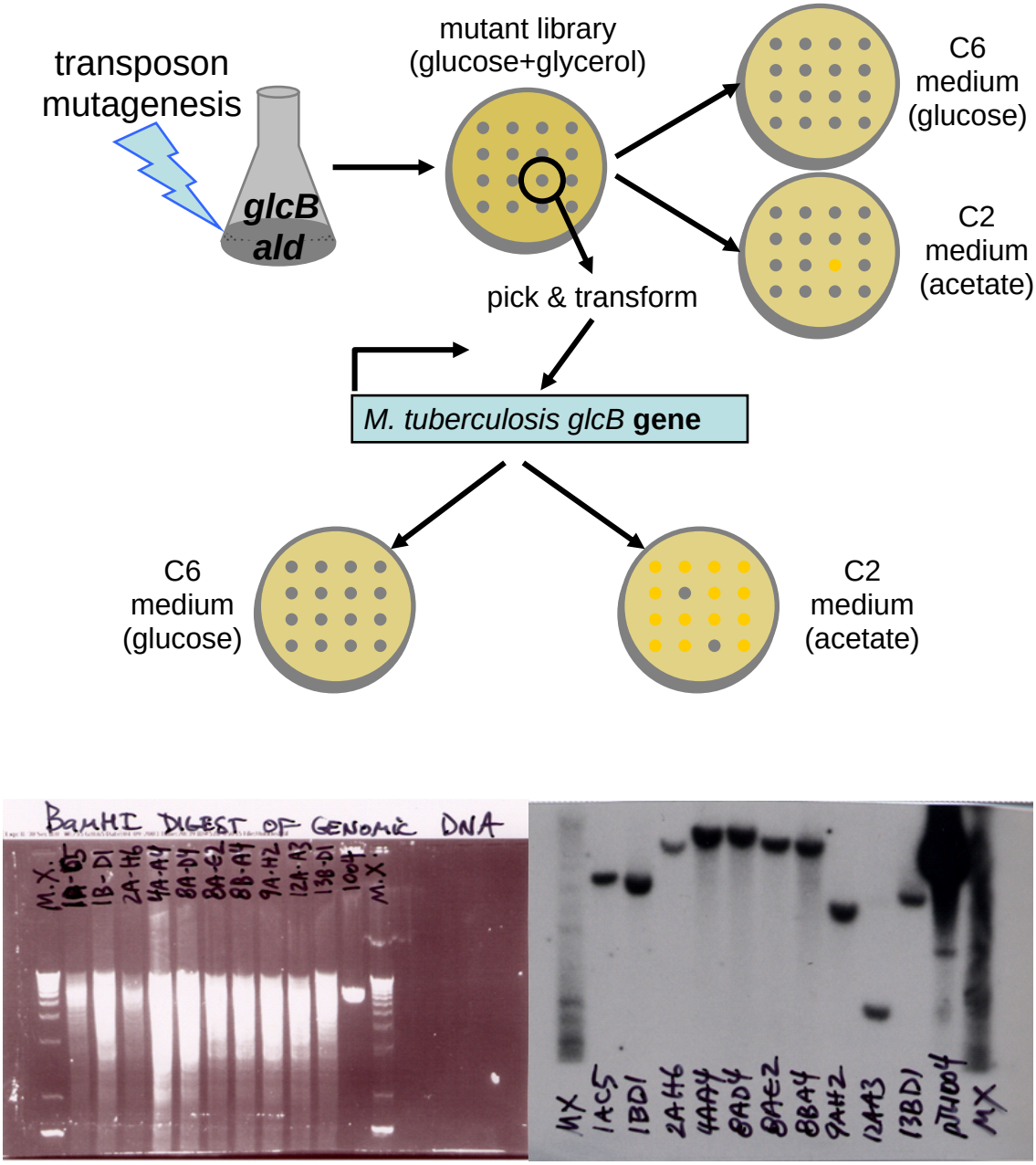


Figure 3.11. Top: Strategy to identify genes in *M. smegmatis* that are essential for the function of the putative ICL-dependent, MLS-independent pathway for growth on acetate and fatty acids as the sole carbon source. Figure provided by Dr. John McKinney. Bottom: Southern blot of selected transposon mutants. Genomic DNA was digested with BamHI or ApaLI (neither enzyme cuts within the phiMycoMarT7 transposon) and probed with the kanamycin resistance cassette from the transposon.

transposon, along with flanking genomic DNA sequences, and sequencing the genomic DNA by using a primer that hybridizes at the end of the kanamycin resistance cassette internal to the transposon. Genomic DNA from the 24 mutants selected for complementation was subjected to sequencing and Southern blotting to confirm that each mutant contained only one insertion (Figure 3.11). Several of the transposon insertions were in the *icl1* gene, which was expected; all of the *icl1* mutants failed to grow on both acetate and methylpalmitate, and they were not complemented by *glcB*. Another transposon insertion was in *pckA* encoding the gluconeogenic PEP carboxykinase; this mutant also failed to grow on both acetate and palmitate, and was not complemented by *glcB*. A large number of insertions mapped to the *acs* gene, which encodes acetyl-CoA synthase; all of the *acs* mutants grew on methylpalmitate but not on 0.1% acetate, the concentration of acetate used in the plates for the screen. Several of the transposon insertions were in genes that had no homologs in *M. tuberculosis*; several of the genes that did have clear homologs in *M. tuberculosis* encoded hypothetical proteins of unknown function. The identifies of the mutants generated in this screen and a screen to identify C2 transposon mutants in wild type *M. smegmatis*, are showin in Appendix A and B.

Importantly, 3 of the 24 mutants were unable to grow on acetate or methylpalmitate, and were complemented by transformation with *glcB* (Figure 3.12). The first mutant, 2AH6 (Figure 3.12, F) grew rather poorly on glucose, and did not grow at all on acetate or methylpalmitate. Addition of *glcB* improved

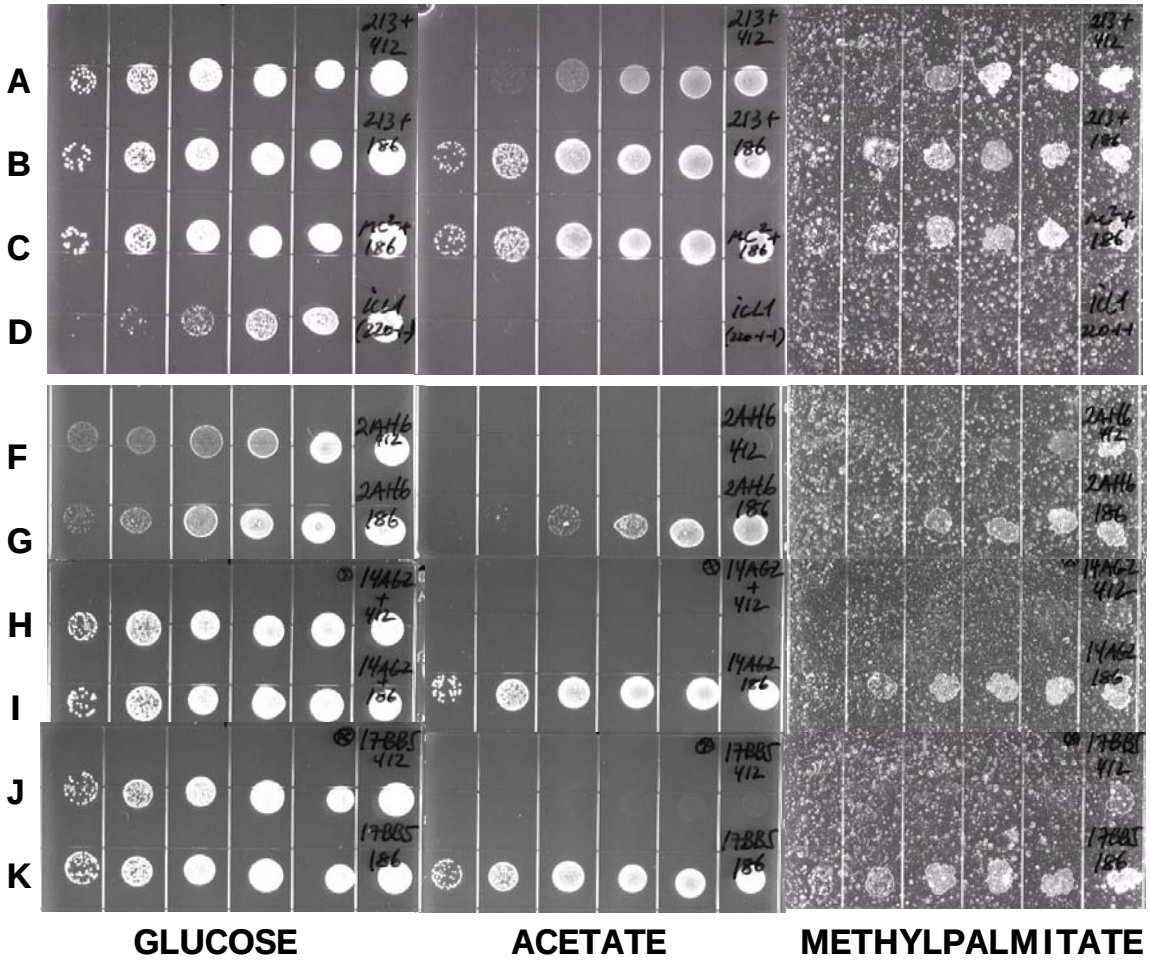


Figure 3.12. Transposon mutants 2AH6 (**F**), 14AG2 (**H**), and 17BB5 (**J**) grow on glucose (left) but cannot grow on acetate (middle) or methylpalmitate (right) as sole carbon source; these mutants are complemented by *glcB*. Serial dilutions of late log-phase cells were spotted on agar plates with 0.1% glucose or acetate, or 0.5% methyl-palmitate. **A**, Δ*glcB* plus vector; **B**, Δ*glcB* plus *M. tuberculosis glcB*; **C**, wild-type plus *glcB*; **D**, Δ*icl1*; **F**, mutant 2AHG plus vector; **G**, 2AHG plus *glcB*; **H**, mutant 14AG2 plus vector; **I**, 14AG2 plus *glcB*; **J**, mutant 17BB5 plus vector; **K**, 17BB5 plus *glcB*.

the growth of this mutant on all three carbon sources but the complementation was partial (Figure 3.12, G). The second and third mutants, 14AG2 and 17BB5 (H and J in Figure 3.12, respectively), grew as well as the Δ*glcB* strain on glucose but did not grow on acetate or methylpalmitate unless they were complemented with the *glcB* gene (I and K in Figure 3.12, respectively). As expected, transformation of the *M. tuberculosis glcB* gene into the Δ*glcB* strain complemented its moderate growth defect on solid media containing acetate or methylpalmitate (A and B in Figure 3.12).

3.5. Genomic characterization of the Δ*glcB* C2 transposon mutants

The identity of the three mutants of interest, 14AG2, 17BB5, and 2AH6, was determined by sequencing the DNA immediately downstream of the transposon insertion sites. The gene disrupted in mutant 14AG2 encoded a protein that had homology to a group of acetolactate synthases in the *M. tuberculosis* genome; a BLAST search against the non-redundant protein databases revealed that the protein was homologous to glyoxylate carboligases in other species. Several kb of DNA sequence upstream and downstream of the insertion site were downloaded from TIGR's databases and assembled with Vector NTI. Putative ORFs were identified and the conceptual translation products were compared to protein databases to deduce the identities of the genes in the region. Glyoxylate carboligase appears to be in an operon along with genes encoding three enzymes and a transcriptional regulator (Figure 3.13, A). The transposon insertion in mutant 17BB5 mapped to a homolog of *M. tuberculosis fadD1*. The *M. smegmatis fadD1* gene lies upstream of an operon encoding ribosomal proteins, and downstream of a gene cluster for threonine biosynthesis (Figure 3.13, B). The transposon insertion in 2AH6 mapped to an approximately 100 bp intergenic region between two divergently transcribed genes predicted to encode enzymes for the production of thiamine. As thiamine pyrophosphate is an essential cofactor of several important enzymes, including pyruvate dehydrogenase, transketolase, acetolactate synthase, pyruvate decarboxylase, and glyoxylate carboligase (Schorken and Sprenger, 1998), it seems unlikely that

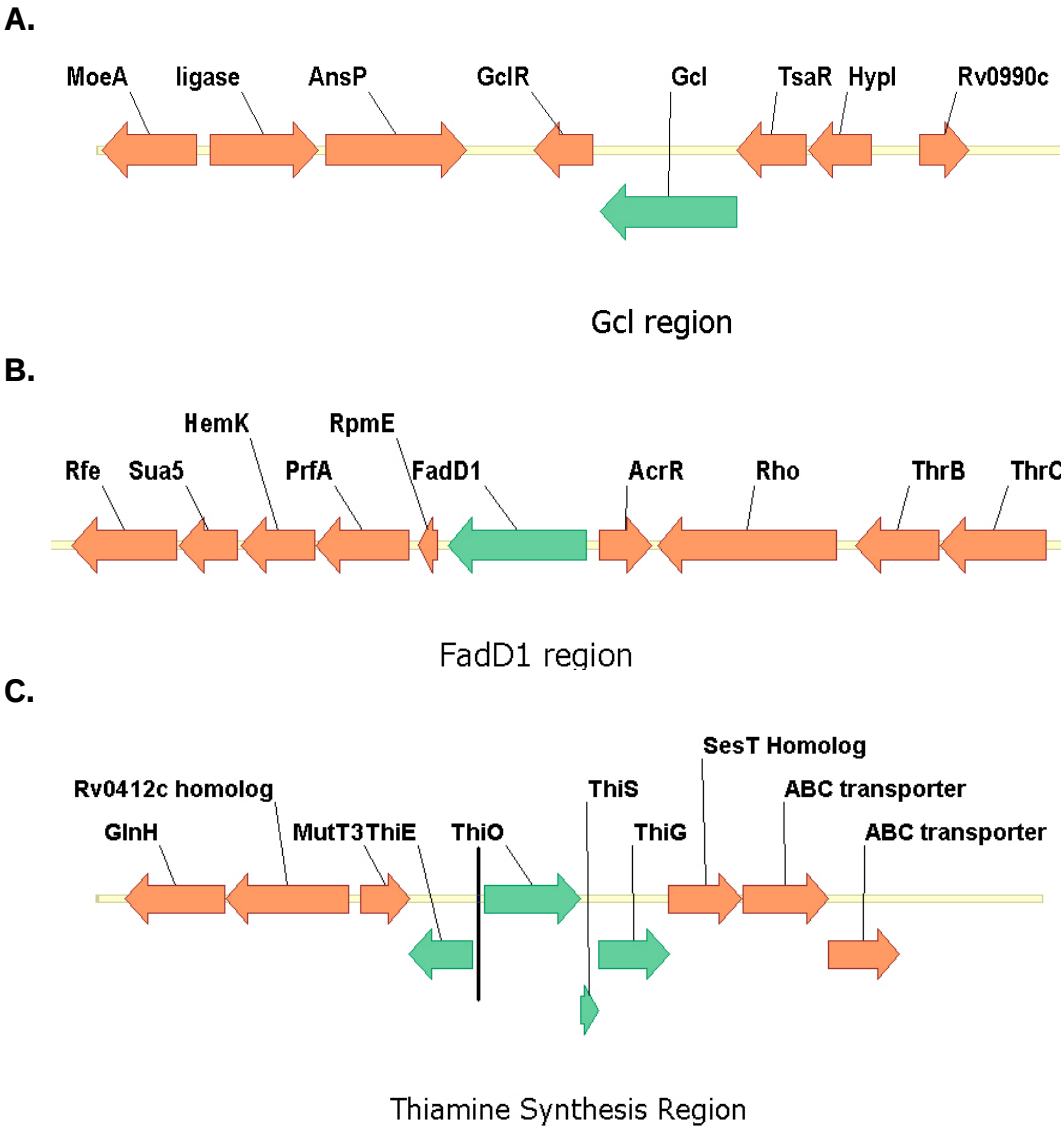


Figure 3.13. Genomic structure of the (**A**) glyoxylate carboligase (GCL), (**B**) FadD1, and (**C**) thiamine biosynthesis (ThiE/O) regions in *M. smegmatis.* Annotation of the regions was done after sequencing the sites of transposon insertions in the mutants. About 5-7 kb of sequence upstream and downstream of the insertion site were obtained from TIGR's unfinished genomes database, and BLAST searches were done with the putative ORFs against the *M. tuberculosis* genome and against NCBI's protein databases. The names of the closest homologues are provided. Black line in **C** indicates the site of transposon insertion in the intergenic region between *thiE* and *thiO*. The transposon inserted in the coding regions of *gcl* and *fadD1.*

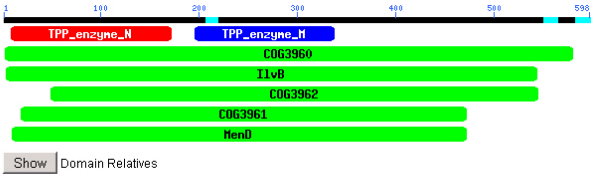
2AH6 is a complete thiamine auxotroph, and we suspect that the mutant may be impaired (but not null) for thiamine production by the insertion (Figure 3.13, C).

Upon annotating the genes surrounding *gcl*, it became apparent that *gcl* forms a hypothetical operon with the preceding two genes. The stop codon of *hypI* is 30 basepairs upstream of the translation start site of *tsaR*; the stop codon of *tsaR* overlaps the start site of *gcl*. A 107 bp intergenic region separates the 3' end of the *gcl* gene and the start site of *gclR*. The 27 bp palindromic sequence ccactattcc acaccgc ggaatagtgg is located at nucleotides -27 to -53 upstream of the *gclR* start codon; the 19 bp palindromic sequence ttttccg tattg cggaaaa is found at nt -24 to -42 of *hypI*. Palindromic sequences upstream of genes can be binding sites for the helix-turn-helix (HTH) transcriptional regulators of the AraC/XylS family; a few representative members of this family are AraC, Crp, GntR, IclR, LuxR and TetR (Callegos *et al.*, 1997). It thus appears that the hypothetical operon might be under the control of HTH regulators.

The *gcl* gene encodes for a putative protein of 599 amino acids. Searches with the hypothetical *M. smegmatis* GCL sequence against the NCBI's Conserved Domains databases revealed that it contains two thiamine pyrophosphate (TPP) binding sites in its N-terminal and central regions; it aligns closely with COG3960 (glyoxylate carboligase) from other species and has more limited similarity to other TPP-requiring enzymes such as acetolactate synthase and pyruvate carboxylase (Figure 3.14, A). The hypothetical product of *tsaR* is 305 aa in length and has a hypothetical NAD^+ binding domain in its N-terminus. TSAR appears to

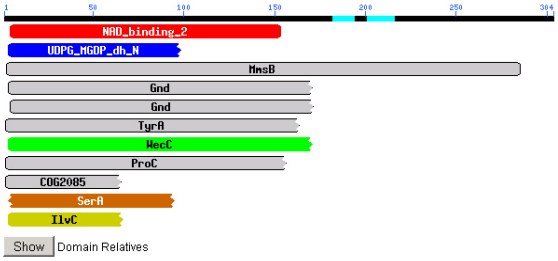
be a β-hydroxyacid dehydrogenase and is homologous to the well-characterized enzymes β-hydroxyisobutyrate dehydrogenase, a requisite enzyme for valine catabolism, and 6-phosphogluconate dehydrogenase, a $NADP^+$-dependent enzyme that plays a role in the pentose phosphate pathway by oxidizing 6-phosphogluconate. The tartronate semialdehyde reductase of *E. coli* is a β-hydroxyacid dehydrogenase with homology to β-hydroxyisobutyrate dehydrogenase and 6-phosphogluconate dehydrogenase (Njau *et al.*, 2000). Given the location of the *M. smegmatis tsaR* gene just upstream of *gcl* and the homology between the hypothetical *tsaR* product and known TSAR proteins (Figure 3.14, B), it is likely that *tsaR* encodes a tartronate semialdehyde reductase.

The gene product just upstream of *tsaR* has homology to hydroxypyruvate isomerases (COG3622). Makato and Misuno (1999) demonstrated that the *E. coli hyi* gene, located immediately downstream of *gcl* and upstream of *glxR* encodes a 258 aa hydroxypyruvate isomerase protein, which catalyzes the conversion of tartronic semialdehyde to hydroxypyruvate. *M. smegmatis hypI* encodes a 278 aa protein, most likely a hydroxypyruvate isomerase (Figure 3.14, C). Finally, the *gclR* gene encodes a 255 aa peptide with an N-terminal HTH motif typical of the IclR family of transcriptional regulators, and has significant overall homology to IclR (Figure 3.14, D).



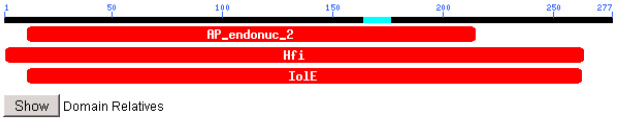
PSSMs producing significant alignments:

		Score (bits)	E value
gnl|CDD|25959	pfam02776, TPP_enzyme_N, Thiamine pyrophosphate enzyme, N-term...	168	1e-42
gnl|CDD|25441	pfam00205, TPP_enzyme_M, Thiamine pyrophosphate enzyme, centra...	101	2e-22
gnl|CDD|13266	COG3960, COG3960, Glyoxylate carboligase [General function pre...	905	0.0
gnl|CDD|9904	COG0028, IlvB, Thiamine pyrophosphate-requiring enzymes [aceto...	401	2e-112
gnl|CDD|13268	COG3962, COG3962, Acetolactate synthase [Amino acid transport ...	131	2e-31
gnl|CDD|13267	COG3961, COG3961, Pyruvate decarboxylase and related thiamine ...	119	1e-27
gnl|CDD|10884	COG1165, MenD, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carbo...	50.3	8e-07



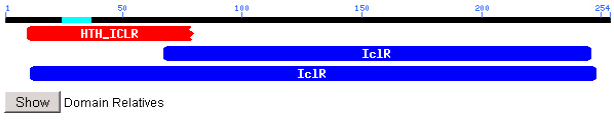
PSSMs producing significant alignments:

		Score (bits)	E value
gnl|CDD|5204	pfam03446, NAD_binding_2, NAD binding domain of 6-phosphogluco...	53.4	3e-08
gnl|CDD|23479	pfam03721, UDPG_MGDP_dh_N, UDP-glucose/GDP-mannose dehydrogena...	36.7	0.004
gnl|CDD|11792	COG2084, MmsB, 3-hydroxyisobutyrate dehydrogenase and related ...	240	1e-64
gnl|CDD|10751	COG1023, Gnd, Predicted 6-phosphogluconate dehydrogenase [Carb...	87.3	2e-18
gnl|CDD|10236	COG0362, Gnd, 6-phosphogluconate dehydrogenase [Carbohydrate t...	79.9	4e-16



PSSMs producing significant alignments:

		Score (bits)	E value
gnl|CDD|25722	pfam01261, AP_endonuc_2, AP endonuclease family 2...	51.2	1e-07
gnl|CDD|12946	COG3622, Hfi, Hydroxypyruvate isomerase [Carbohydrate transpor...	161	1e-40
gnl|CDD|10804	COG1082, IolE, Sugar phosphate isomerases/epimerases [Carbohyd...	78.0	1e-15



PSSMs producing significant alignments:

		Score (bits)	E value
gnl|CDD|7347	smart00346, HTH_ICLR, helix_turn_helix isocitrate lyase regula...	57.5	2e-09
gnl|CDD|2151	pfam01614, IclR, Bacterial transcriptional regulator. This fam...	125	6e-30
gnl|CDD|11129	COG1414, IclR, Transcriptional regulator [Transcription].	172	4e-44

Figure 3.14. Alignment of proteins in the *gcl* region with NCBI's Conserved Domains databases. GCL (**A**) belongs to the family of TPP (thiamine pyrophosphate)-dependent enzymes. TSAR (**B**) is related to 3-hydroxyisobutyrate dehydrogenases and 6-phosphogluconate dehydrogenases. HYPI (**C**) appears to be a hydroxypyruvate isomerase, while the GclR protein (**D**) belongs to the IclR family of regulators.

A.

```
Sequences producing significant alignments:                          (bits) Value

gi|22984089|ref|ZP_00029247.1|  hypothetical protein [Burkho...   814   0.0
gi|27378277|ref|NP_769806.1|  gcl [Bradyrhizobium japonicum]...   813   0.0
gi|16763897|ref|NP_459512.1|  glyoxylate carboligase [Salmon...   810   0.0
gi|22980710|ref|ZP_00026413.1|  hypothetical protein [Ralsto...   808   0.0
gi|15800243|ref|NP_286255.1|  glyoxylate carboligase [Escher...   806   0.0
gi|1773188|gb|AAB40260.1|  glyoxylate carboligase [Escherich...   806   0.0
gi|26246516|ref|NP_752555.1|  Glyoxylate carboligase [Escher...   805   0.0
gi|16265136|ref|NP_437928.1|  putative glyoxylate carboligas...   805   0.0
gi|16759494|ref|NP_455111.1|  glyoxylate carboligase [Salmon...   805   0.0
gi|30061999|ref|NP_836170.1|  glyoxylate carboligase [Shigel...   804   0.0
gi|24111883|ref|NP_706393.1|  glyoxylate carboligase [Shigel...   804   0.0
gi|15596699|ref|NP_250193.1|  glyoxylate carboligase [Pseudo...   803   0.0
gi|21224525|ref|NP_630304.1|  putative glyoxylate carboligas...   797   0.0
gi|23060519|ref|ZP_00085419.1|  hypothetical protein [Pseudo...   797   0.0
gi|13470522|ref|NP_102091.1|  glyoxylate carboligase [Mesorh...   793   0.0
gi|26990988|ref|NP_746413.1|  glyoxylate carboligase [Pseudo...   788   0.0
gi|17547998|ref|NP_521400.1|  PROBABLE GLYOXYLATE CARBOLIGAS...   785   0.0
gi|29828570|ref|NP_823204.1|  putative glyoxylate carboligas...   777   0.0
gi|18152916|gb|AAL61905.1|  putative glycoxylate carboligase...   741   0.0
```

B.

```
Query= Smegmatis Glyoxylate Carboligase Region
         (2760 letters)

Database: M. tuberculis proteins
           3996 sequences; 1,337,453 total letters

Searching.....................................................done

                                                                    Score     E
Sequences producing significant alignments:                         (bits)  Value

M. tuberculis bacteria|Rv3003c|ilvB1 PROBABLE ACETOLACTATE SYNT...   294  3e-80
M. tuberculis bacteria|Rv0118c|oxcA PROBABLE OXALYL-COA DECARBO...   118  2e-27
M. tuberculis bacteria|Rv1820|ilvG Probable Acetolactate syntha...   114  3e-26
M. tuberculis bacteria|Rv3470c|ilvB2 PROBABLE ACETOLACTATE SYNT...    90  1e-18
M. tuberculis bacteria|Rv1719|Rv1719 PROBABLE TRANSCRIPTIONAL R...    65  4e-11
M. tuberculis bacteria|Rv0853c|pdc PROBABLE PYRUVATE OR INDOLE-...    54  5e-08
M. tuberculis bacteria|Rv3509c|ilvX PROBABLE ACETOHYDROXYACID S...    46  1e-05
M. tuberculis bacteria|Rv0751c|mmsB PROBABLE 3-HYDROXYISOBUTYRA...    36  0.013
M. tuberculis bacteria|Rv2839c|infB PROBABLE TRANSLATION INITIA...    34  0.050
M. tuberculis bacteria|Rv1832|gcvB Probable glycine dehydrogena...    32  0.25


>M. tuberculis bacteria|Rv3003c|ilvB1 PROBABLE ACETOLACTATE SYNTHASE
          (LARGE SUBUNIT) ILVB1 (ACETOHYDROXY-ACID SYNTHASE)
         Length = 618

 Score =  294 bits (744), Expect = 3e-80
 Identities = 175/552 (31%), Positives = 271/552 (48%), Gaps = 3/552 (0%)
 Frame = -3
```

C.

```
Rv0751c mmsB (mmsB) (Mycobacterium tuberculo...   331  4.4e-39   2
NT02MT0880 3-hydroxyisobutyrate dehydrogenase ( ) (Mycoba...   331  4.4e-39   2
Rv0770 hypothetical protein Rv0770 (NULL) (M...   266  1.1e-24   1
NT02MT0902 conserved hypothetical protein ( ) (Mycobacter...   266  1.1e-24   1
Rv1844c gnd (gnd) (Mycobacterium tuberculosi...   166  1.7e-11   1
NT02MT2165 6-phosphogluconate dehydrogenase ( ) (Mycobact...   166  1.8e-11   1
Rv1122 gnd2 (gnd2) (Mycobacterium tuberculos...   115  3.5e-06   2
NT02MT1315 gnd ( ) (Mycobacterium tuberculosis H37Rv (lab...   115  3.5e-06   2
```

D.

```
Rv0498 hypothetical protein Rv0498 (NULL) (M...    75  0.15      1
NT02MT0569 conserved hypothetical protein ( ) (Mycobacter...    75  0.15      1
Rv0378 hypothetical protein Rv0378 (NULL) (M...    51  0.69      1
Rv1405c hypothetical protein Rv1405c (NULL) ...    66  0.81      1
NT02MT1659 spore germination protein c2 ( ) (Mycobacteriu...    66  0.81      1
Rv0581 hypothetical protein Rv0581 (NULL) (M...    48  0.98      1
NT02MT0669 hypothetical protein (Mycobacterium tuberculos...    48  0.98      1
NT02MT0206 hypothetical protein (Mycobacterium tuberculos...    61  0.992     1
NT02MT0531 hypothetical protein (Mycobacterium tuberculos...    64  0.997     1
Rv3354 hypothetical protein Rv3354 (NULL) (M...    56  0.997     1
NT02MT3925 conserved hypothetical protein ( ) (Mycobacter...    56  0.997     1
```

Figure 3.15. (**A**) tBLASTX of *M. smegmatis* GCL against NCBI's databases. (**B-D**) BLASTP of GCL (**B**), TSAR (**C**) and HYPI (**D**) proteins against *M. tuberculosis* genomes. *mmsB* (**C**) is annotated as a 3-hydroxyisobutyrate dehydrogenase in the H37Rv genome.

The gene names used here were assigned on the basis of results from protein homology searches and comparison of the regions flanking the transposon insertion in *M. smegmatis* to the *gcl* operon in the *E. coli* chromosome (Cusa *et al.*, 1999). In TIGR's databases, these genes are assigned MSMEG numbers 5454 (*gclR*), 5455 (*gcl*), 5456 (*tsaR*) and 5457 (*hypI*); MSMEG5455 is annotated as glyoxylate carboligase, while the other three proteins have not been assigned functions.

The protein sequences of the *gcl* locus were aligned against the genomes of pathogenic mycobacteria on TIGR's website (http://tigrblast.tigr.org/cmr-blast/). The *gcl* gene is absent from the genomes of all sequenced pathogenic species (*M. avium, M. bovis, M. leprae,* and *M. tuberculosis*). The *ilv* genes, encoding acetolactate synthases, were the closest homologs of *gcl* in *M. tuberculosis* (Figure 3.15, B). Similarly, the best *tsaR* matches were the *M. tuberculosis* genes encoding β-hydroxyisobutyrate dehydrogenase and 6-phosphogluconate dehydrogenase (Figure 3.15, C). No hits were obtained for HYPI, suggesting that this protein is also missing in *M. tuberculosis*.

The gene disrupted in mutant 17BB5, on the other hand, had an obvious homolog in *M. tuberculosis* (Figure 3.16). It is 62 percent identical to the FadD1 protein, with lesser homology to FadD17 and FadD6. Both FadD17 (Trivedi *et al.*, 2004) and FadD6 (Arora *et al*, 2005) were shown to synthesize the formation of fatty acyl-CoA and not fatty acyl-AMP; FadD6 also showed broad specificity for the length of its fatty acid substrates. These data indicate that FadD1 is probably

A.

```
Database: M. tuberculis proteins
          3996 sequences; 1,337,453 total letters

Searching.................................................done

                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

M. tuberculis bacteria|Rv1750c|fadD1 POSSIBLE FATTY-ACID-COA LI...   643  0.0
M. tuberculis bacteria|Rv3506|fadD17 POSSIBLE FATTY-ACID-COA SY...   431  e-121
M. tuberculis bacteria|Rv1206|fadD6 PROBABLE FATTY-ACID-COA LIG...   107  3e-24
M. tuberculis bacteria|Rv0099|fadD10 POSSIBLE FATTY-ACID-COA LI...    95  2e-20
M. tuberculis bacteria|Rv0270|fadD2 PROBABLE FATTY-ACID-COA LIG...    90  8e-19
M. tuberculis bacteria|Rv2505c|fadD35 PROBABLE FATTY-ACID-COA L...    89  1e-18
M. tuberculis bacteria|Rv1427c|fadD12 POSSIBLE LONG-CHAIN-FATTY...    89  2e-18
M. tuberculis bacteria|Rv1193|fadD36 PROBABLE FATTY-ACID-COA LI...    88  3e-18
M. tuberculis bacteria|Rv1058|fadD14 PROBABLE MEDIUM CHAIN FATT...    85  2e-17
M. tuberculis bacteria|Rv0214|fadD4 PROBABLE FATTY-ACID-COA LIG...    84  5e-17
M. tuberculis bacteria|Rv0119|fadD7 PROBABLE FATTY-ACID-COA LIG...    84  5e-17
M. tuberculis bacteria|Rv0166|fadD5 PROBABLE FATTY-ACID-COA LIG...    82  2e-16
M. tuberculis bacteria|Rv0551c|fadD8 PROBABLE FATTY-ACID-COA LI...    82  2e-16


>M. tuberculis bacteria|Rv1750c|fadD1 POSSIBLE FATTY-ACID-COA LIGASE
           FADD1 (FATTY-ACID-COA SYNTHETASE) (FATTY-ACID-COA
           SYNTHASE)
          Length = 532

 Score =  643 bits (1640), Expect = 0.0
 Identities = 321/516 (62%), Positives = 371/516 (71%), Gaps = 6/516 (1%)
 Frame = -1
```

B.

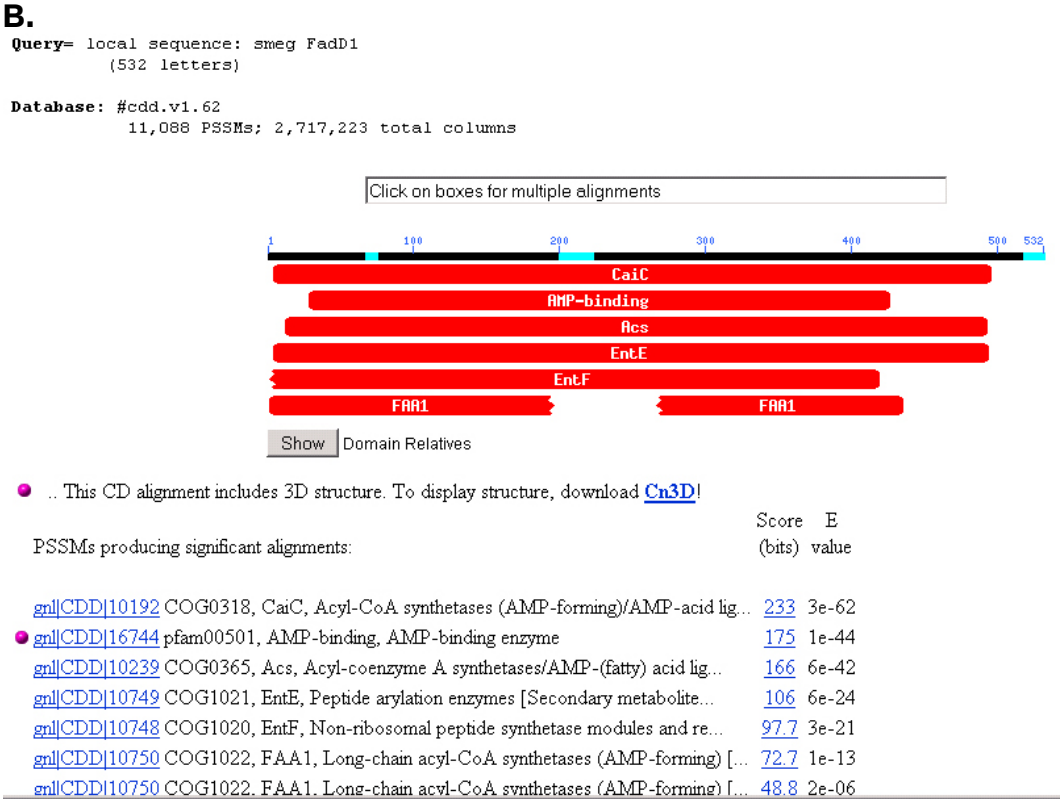


Figure 3.16. tBLASTX of the *M. smegmatis fadD1* locus against the H37Rv genome (**A**), and the NCBI Conserved Domains databases (**B**). FADD1's significant alignment against COGs 0318, 0365, and 1022 suggests that it is a fatty acyl-CoA synthase.

a fatty acyl-CoA ligase and does not play a role in polyketide synthesis.

The location of *fadD1* in the genome of *M. smegmatis* suggests that it is transcribed independently of surrounding genes (Figure 3.13, B). Interestingly, in *M. tuberculosis*, the threonine biosynthesis (*thr*) gene cluster resides immediately upstream of the ribosomal peptide genes. The order and orientation of the *thr* and ribosomal synthesis genes are conserved in *M. smegmatis*, but *fadD1* and the putative transcriptional regulator gene *AcrR* are inserted in the middle.

3.6. Metabolic characterization of the Δ*glcB* C2 mutants

The carbon requirements of the Δ*glcB gcl* and Δ*glcB fadD1* mutants were more fully characterized. The 2AH6 mutant, hypothesized to be partially defective for synthesis of thiamine pyrophosphate, grew poorly on all carbon sources, including glycerol and glucose, and complementation with *glcB* only partially restored growth on all substrates (Figure 3.19). As thiamine pyrophosphate is a cofactor necessary for the enzymatic activity of glyoxylate carboligase (Krakow *et al.,* 1961; Schorken and Sprenger, 1998) as well as other enzymes, 2AH6 was not characterized as fully as the Δ*glcB gcl* and Δ*glcB fadD1* mutants in subsequent experiments.

The Δ*glcB,* Δ*glcB gcl*, and Δ*glcB fadD1* mutants, as well as the corresponding complemented strains carrying the *M. tuberculosis glcB* gene on a plasmid, were assayed for growth in liquid minimal media containing glucose or glyoxylate as carbon sources (Figure 3.17). All strains grew as well as wild type *M. smegmatis* on glucose; however, only the wild type, Δ*glcB*, and complemented Δ*glcB* strains grew at all on glyoxylate. This demonstrated that, similarly to *E. coli* (Ornston and Ornston, 1969), the activity of malate synthase is dispensable for growth on glyoxylate, and that glyoxylate carboligase is absolutely required for assimilaiton of glyoxylate into the general metabolism of *M. smegmatis*. Surprisingly, the Δ*glcB fadD1* strain also failed to grow on glyoxylate, regardless of whether malate synthase activity was present or not (Figures 3.17, 3.18). Given that *fadD1* appears to encode a fatty acyl-CoA ligase

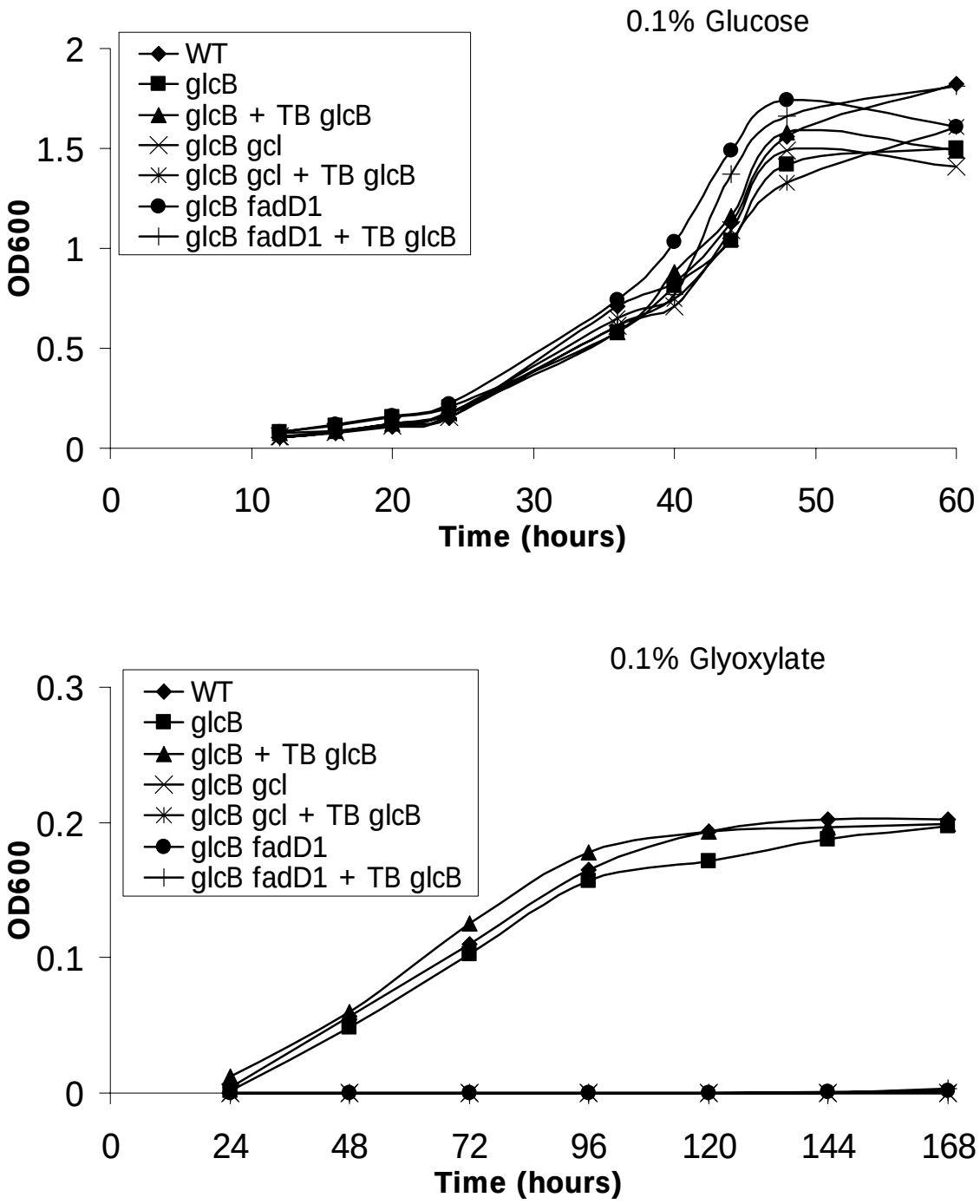


Figure 3.17. Growth of wild type *M. smegmatis*, mutant strains, and mutant strains complemented with MTB *glcB* in M9 minimal media containing either 0.1% dextrose (top panel) or 0.1% glyoxylate (bottom panel) as sole carbon sources. The growth studies were done at least three times for each strain and carbon source; a representative experiment is shown.

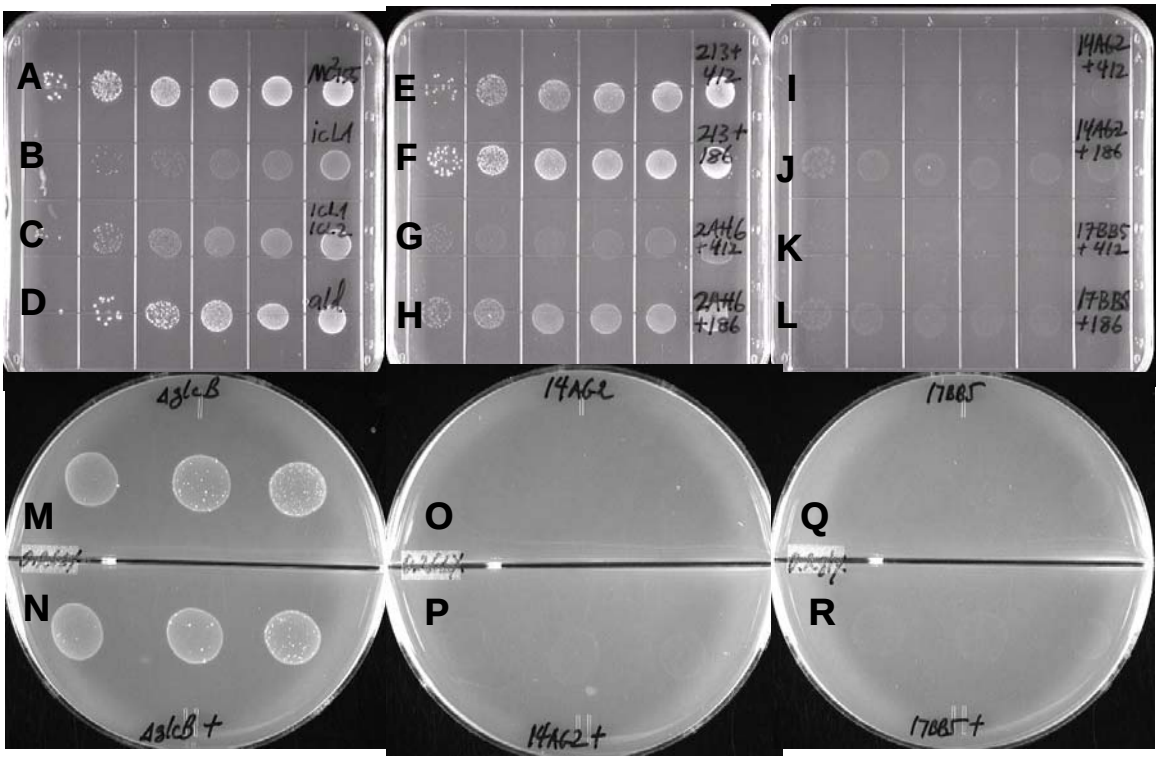


Figure 3.18. Growth of *M. smegmatis* glyoxylate shunt mutants on glyoxylate as the sole carbon source. Serial dilutions of early log-phase cultures of wild-type *M. smegmatis* (**A**) and mutants were spotted on solid plates containing agar (top, square plates) or agarose (bottom, round plates) as the solidifying agent, and 0.2% glyoxylate as the sole carbon source. The Δ*ald* (**D**), Δ*glcB* (**E, M**) mutants, and the Δ*glcB* mutant complemented with the MTB *glcB* gene (**F, N**) grew as well as wild-type bacteria, indicating that *ald* and *glcB* are dispensable for growth on glyoxylate. Interestingly, the Δ*icl1* (**B**) and Δ*icl1* Δ*icl2* (**C**) mutants appeared to have a slight growth defect on glyoxylate plates, which was not the case with the Δ*icl2* mutant (not shown). The Δ*glcB gcl* (**I, O**) and Δ*glcB fadD1* (**K, Q**) strains did not grow on glyoxylate, even when transformed with a plasmid containing the MTB *glcB* gene (**J, P; L, R**).

(FACL), and FACLs have not been shown to play a role in the assimilation of glyoxylate, it is unclear why the Δ*glcB fadD1* strain is unable to grow on glyoxylate media.

The mutant strains were further tested for their ability to grow on agar plates or liquid minimal media supplemented with short-chain fatty acids as the sole carbon sources. Agar plates containing C2, C4, C5, and C6 compounds were checked for growth after one and three weeks of incubation at 37˚C. The Δ*glcB* mutant had formed microscopic colonies on C2 and C6 after a week, and formed large colonies on all plates after three weeks (Figure 3.19, left-most column). The Δ*glcB gcl* mutant did not grow on C2, C4, or C6 fatty acids at any time; restoration of malate synthase activity enabled it to assimilate these even-chain fatty acids. Interestingly, the Δ*glcB gcl* strain did form colonies on valerate (C5) after three weeks (Figure 3.19, left middle).

The Δ*glcB fadD1* mutant behaved similarly to the Δ*glcB gcl* mutant, with a few exceptions: it formed noticeable colonies on C5 sooner and it formed small but discernable colonies on C4 and C6 after three weeks. Thus, the Δ*glcB fadD1* mutation appeared to be slightly but perceptibly leaky on butyrate and hexanoic acid. However, the Δ*glcB fadD1* mutant, like the Δ*glcB gcl* strain, absolutely failed to grow on C2 (acetate) unless complemented with *glcB* (Figure 3.19, right middle column). Lastly, the 2AH6 mutant (Δ*glcB thi*) eventually grew on all fatty acids, regardless of whether it had malate synthase activity or not; in fact, complementation enhanced its growth only on acetate, but rather minimally.

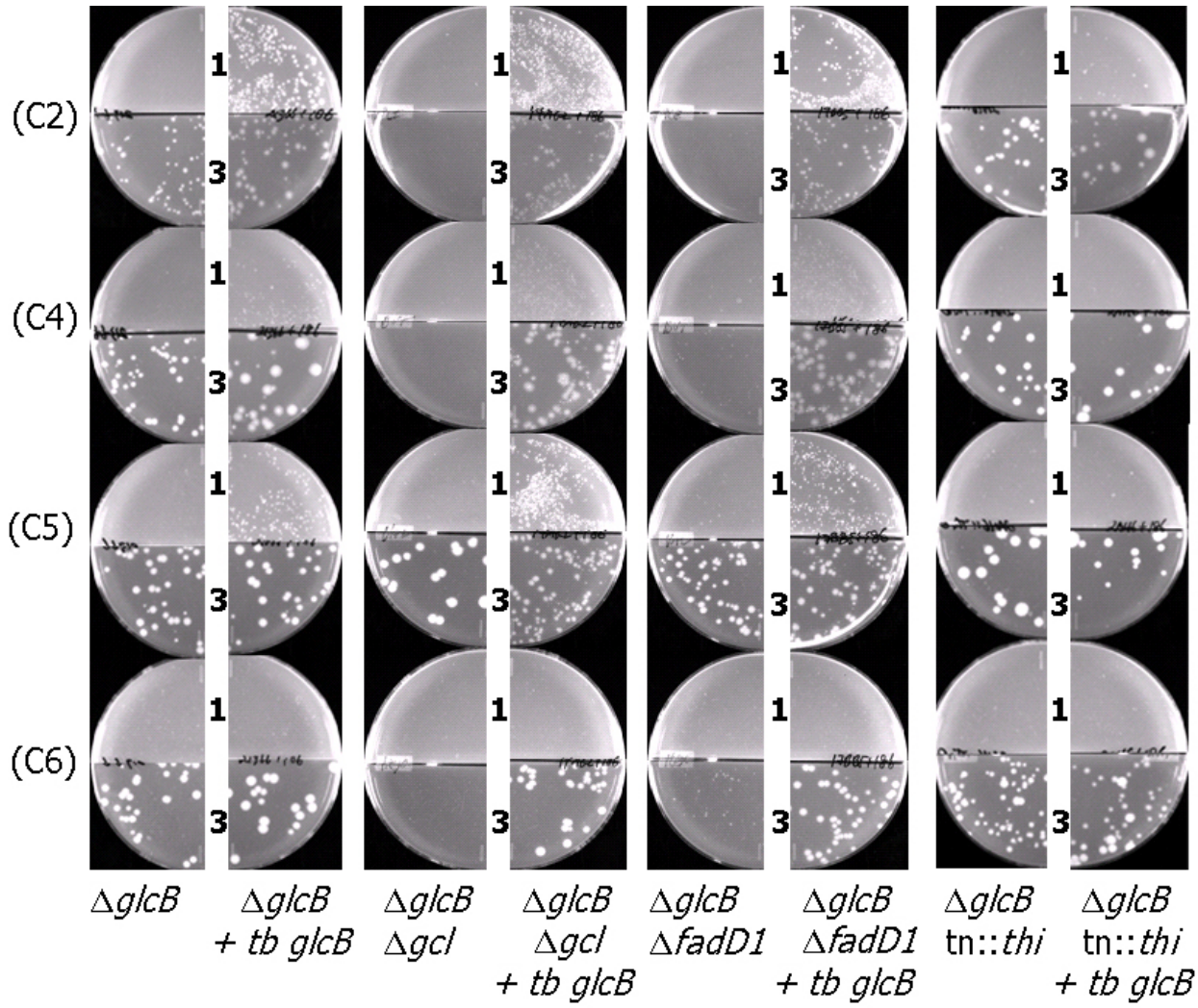


Figure 3.19. Complementation of Δ*glcB gcl* and Δ*glcB fadD1* mutants with the MTB *glcB* gene (panels marked "+ tb glcB") restores wild type growth on even-chain fatty acids (acetate, C2; butyrate, C4; hexanoate, C6) as sole carbon sources. The Δ*glcB thi* mutant can grow on media containing these substrates as sole carbon sources, albeit poorly. The Δ*glcB gcl* mutant fails to grow on even-chain fatty acids even after 3 weeks, while Δ*glcB fadD1* forms small, but discernable, colonies on C4 (butyrate) and C6 (hexanoate) after 3 weeks. 2AH6 (Δ*glcB thi*) grows slowly on all carbon sources, including glucose (not shown); complementation with the MTB *glcB* gene enhances growth of the Δ*glcB thi* mutant on acetate. "1" = one week of growth at 37˚C; "3" = three weeks of growth at 37˚C.

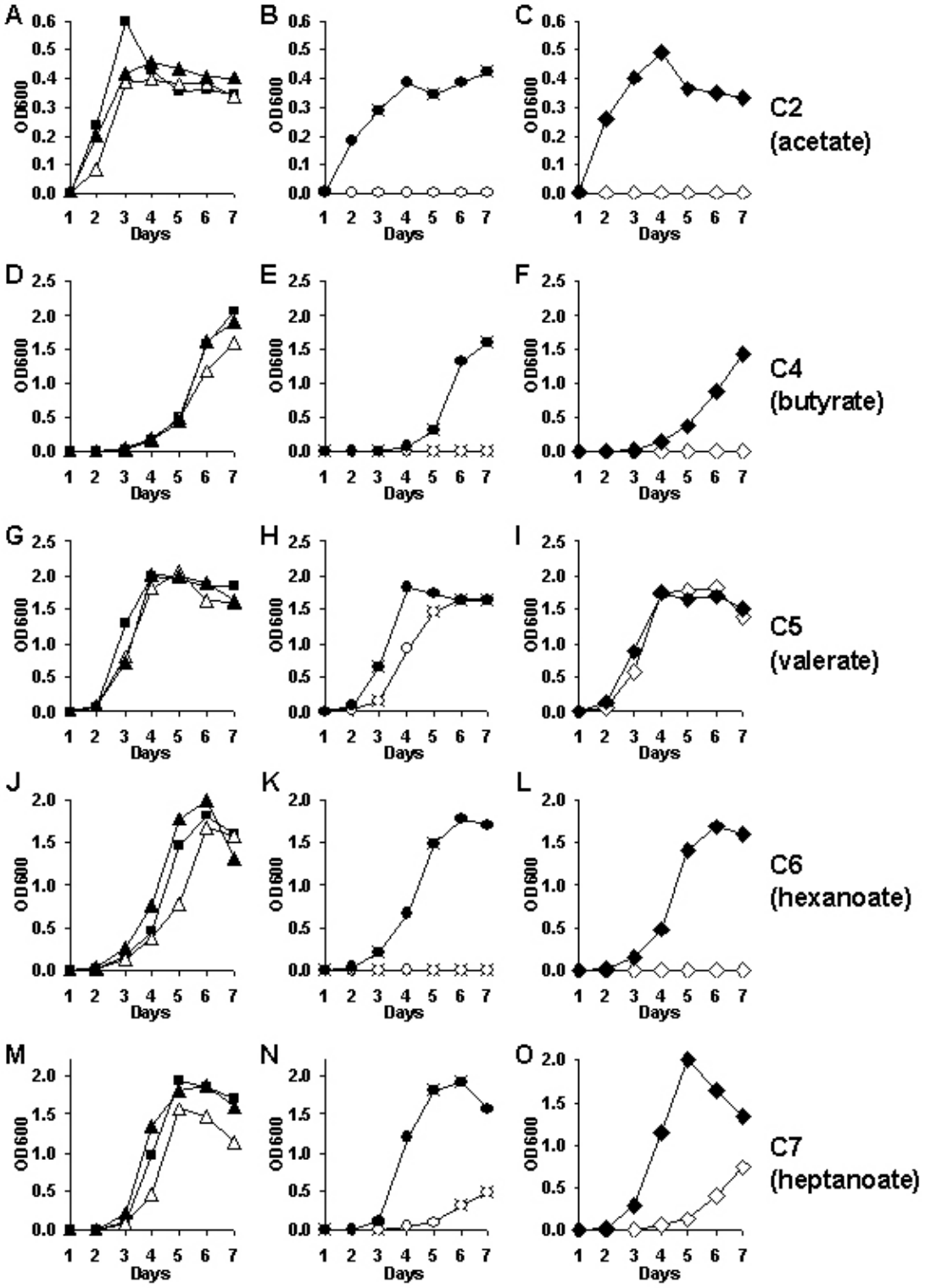


Figure 3.20. Growth of *M. smegmatis* in M9 liquid media containing 0.1% of the indicated C2, C4, C5, C6, and C7 compounds as the sole carbon source. Strains: (**A**,**D**,**G**,**J**,**M**) wild-type (squares), Δ*glcB* (empty triangles), and compl. Δ*glcB* (filled triangles); (**B**,**E**,**H**,**K**,**N**) Δ*glcB gcl* (empty circles) and compl. Δ*glcB gcl* (filled circles); (**C**,**F**,**I**,**L**,**O**) Δ*glcB fadD1* (empty diamonds) and compl. Δ*glcB fadD1* (filled diamonds).

These results suggested that the Δ*glcB gcl* and Δ*glcB fadD1* strains might have different growth kinetics on fatty acids, and that they behave differently on even- vs. odd-chain fatty acids. To address this issue, we grew these strains in M9 minimal liquid media supplemented with C2-C7 fatty acids (Figure 3.20). In M9 liquid media, the growth delay of the Δ*glcB* mutant was less pronounced than on solid media, but nonetheless significant and reproducible (Figure 3.20A,D,G,J,M). The Δ*glcB* strain lagged behind wild-type bacteria when grown on acetate (Figure 3.20A), which was not due to adaptation problems, as the lag was not eliminated by pre-adaptation of Δ*glcB* cells to acetate (not shown). A clear lag was also observed when the Δ*glcB* strain was cultured in C4 (Figure 3.20D), C6 (Figure 3.20J), or C7 (Figure 3.20M) media, but was very slight when cells were grown in C5 medium (Figure 3.20G).

The Δ*glcB gcl* and Δ*glcB fadD1* strains failed to grow in media containing even-chain fatty acids of C2 (Figure 3.20B,C), C4 (Figure 3.20E,F), or C6 (Figure 3.20K,L) chain length. However, both strains were able to grow on media containing C5/valerate (Figure 3.20H,I) and, to a lesser degree, C7/heptanoate (Figure 3.20N,O). The Δ*glcB fadD1* mutant consistently grew better than the Δ*glcB gcl* strain on C5 and C7 substrates; in fact, growth of the Δ*glcB fadD1* (Figure 3.20I) and wild-type (Figure 3.20G) strains on C5/valerate was similar.

In media containing odd-chain fatty acids, the Δ*glcB gcl* and Δ*glcB fadD1* strains grew better on C5/valerate than C7/heptanoate. A molecule of C7/heptanoate is degraded, via β-oxidation, to yield two acetyl-CoA (C2) units

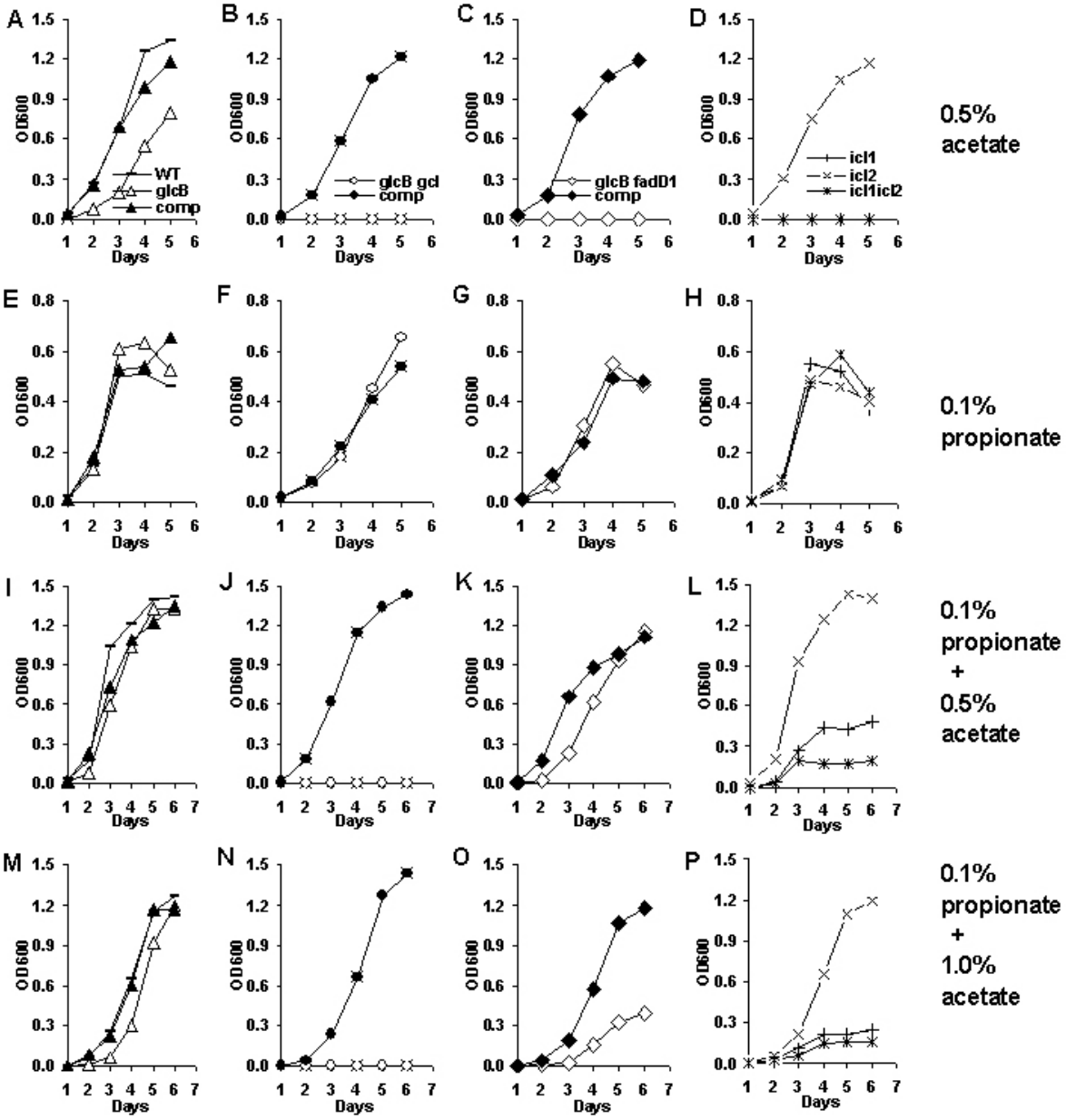


Figure 3.21. Growth of *M. smegmatis* in liquid media containing acetate, propionate, or a mixture of acetate and propionate as the sole carbon sources, as indicated. Strains: (**A**,**E**,**I**,**M**) wild-type, Δ*glcB*; comp. Δ*glcB*; (**B**,**F**,**J**,**N**) Δ*glcB gcl*, comp. Δ*glcB gcl*; (**C**,**G**,**K**,**O**) Δ*glcB fadD1*, comp. Δ*glcB fadD1*; (**D**,**H**,**L**,**P**) Δ*icl1*, Δ*icl2*, Δ*icl1*Δ*icl2*. Note: "comp" indicates that the corresponding mutant strain was complemented with MTB *glcB* to restore malate synthase activity.

and one propionyl-CoA (C3) unit, whereas β-oxidation of C5/valerate yields one acetyl-CoA unit and one propionyl-CoA unit. We considered the possibility that, in the absence of a functional glyoxylate shunt, the Δ*glcB gcl* and Δ*glcB fadD1* strains might oxidize acetyl-CoA via the TCA cycle for energy generation and assimilate propionyl-CoA via the methylcitrate cycle for anaplerosis. However, our observations suggested that acetyl-CoA might exert a dominant inhibitory effect in the absence of an intact glyoxylate shunt. We propose that catabolism of longer odd-chain fatty acids would result in higher acetyl-CoA:propionyl-CoA ratios in the cells, thereby increasing growth inhibition in proportion to chain length. We tested this hypothesis by growing the Δ*glcB gcl* and Δ*glcB fadD1* strains in M9 minimal media containing mixtures of propionate (C3) and acetate (C2) in different ratios. We also tested the strains that lacked one, or both, of the isocitrate lyase (*icl*) genes in these conditions. As shown in Figure 3.21D, the Δ*icl1* and Δ*icl1* Δ*icl2* mutants, but not the Δ*icl2* mutant, failed to grow on acetate, as did the Δ*glcB gcl* (Figure 3.21B) and Δ*glcB fadD1* (Figure 3.21C) strains. The Δ*glcB* mutant grew on 0.5 percent acetate (Figure 3.21A), but the lag was more pronounced than when it was grown on 0.1 percent acetate (Figure 3.21E). As expected, all strains grew well in media containing 0.1 percent propionate (Figure 3.21E,F,G,H).

Interestingly, addition of acetate inhibited the ability of the Δ*icl1* and Δ*icl1* Δ*icl2* mutants to grow on media containing propionate (Figure 3.21L,P). There are several possible explanations for this phenomenon: a) acetate is toxic

to cells unless they can assimilate it via the glyoxylate cycle; b) high concentrations of acetate may inhibit some enzymes, or the transcription of some genes, which are involved in propionyl-CoA metabolism; c) ACKA converts acetate to acetyl-phosphate, which is a global regulator of gene expression in *E. coli* (Wolfe, 2005), and acetyl-phosphate-mediated changes in gene expression could be growth-inhibitory in mycobacteria; d) acetate is converted to acetyl-CoA, which, at high levels, can be inhibitory to a number of important enzymes. Also of note is the fact that the presence of a functional *icl2* gene can partially compensate for the loss of *icl1* at moderate concentrations of acetate in the media (Figure 3.21L) but not at high acetate concentrations (Figure 3.21P). It appears that *icl1* can effectively buffer loss of *icl2*, inasmuch as deletion of *icl2* alone had no effect on bacterial growth on any substrates or substrate mixtures examined (Figure 3.21D,H,L,P).

The Δ*glcB* mutant grew well in mixtures of 0.5 percent acetate and 0.1 percent propionate (Figure 3.21I), suggesting that GCL can effectively metabolize the glyoxylate that is produced by isocitrate lyase at this acetate concentration. However, in mixtures of 1.0 percent acetate and 0.1 percent propionate (Figure 3.21M), growth of the Δ*glcB* strain was paritally inhibited, perhapse due to accumulation of glyoxylate or acetyl-CoA under these conditions.

The Δ*glcB gcl* mutant was unable to grow on 0.1 percent propionate in the presence of 0.5 percent (Figure 3.21J) or 1.0 percent (Figure 3.21N) acetate,

despite being capable of normal growth on media containing 0.1 percent propionate alone (Figure 3.21F). We hypothesize that isocitrate lyase converts acetate to glyoxylate, which accumulates to inhibitory levels in the absence of GlcB and GCL. Consistent with this hypothesis, glyoxylate-mediated growth inhibition has been demonstrated in a malate synthase mutant of *Rhodococcus fascians* (Vereecke *et al.*, 2002).

Interestingly, the Δ*glcB fadD1* mutant is capable of growth on propionate alone (Figure 3.21G) as well as mixtures of acetate and propionate (Figure 3.21K,O), despite being completely unable to utilize acetate alone (Figure 3.21C). Since either malate synthase (Figure 3.21J,N) or glyoxylate carboligase (Figure 3.21I,M) appears to be sufficient for growth in media containing acetate/propionate mixtures, the Δ*glcB fadD1* mutant may induce glyoxylate carboligase activity when provided with propionate plus acetate, or with odd-chain fatty acids, as the carbon source. We propose a model in which FadD1 is required to upregulate (probably indirectly) the expression or activity of glyoxylate carboligase during growth on acetate (or even-chain fatty acids) as the sole substrate; this requirement would be partially bypassed during growth on mixtures of acetate and propionate (or odd-chain fatty acids), via an unknown mechanism.

3.7. FadD1-mediated regulation of the *gcl* operon

To examine the possible regulation of the *gcl* operon by FadD1, cell-free extracts were prepared from wild-type bacteria grown on glucose, valerate (C5), or a 1:1 mixture of acetate (C2) and propionate (C3), and the activity of TSAR (tartronate semialdehyde reductase) was measured spectrophotometrically in a coupled enzyme assay. Each molecule of the reduced tartronic semilaldehyde oxidizes a molecule of NADH to NAD^+, so the conversion of tartronic semialdehyde to d-glycerate can be quantified by measuring the disappearance of NADH at 340 nm (Hansen and Hayashi, 1962). Tartronic semialdehyde was generated *in situ* from glyoxylate by GCL upon the addition of $MgCl_2$ and TPP (see Schematic in Figure 3.22A)

Cell-free extracts of wild-type *M. smegmatis* grown on glucose contained no appreciable GCL-TSAR activity (Figure 3.22B, filled diamonds), but activity was readily detected in extracts from bacteria grown in media containing valerate (Figure 3.22B, crosses) or acetate+propionate mixtures (Figure 3.22B, filled triangles). Production of tartronic semialdehyde was apparently dependent on glyoxylate carboligase activity, because reactions in which thiamine pyrophosphate (TPP) and $MgCl_2$ were omitted were negative for NADH oxidation (Figure 3.22B, empty squares). Extracts from Δ*glcB fadD1* bacteria grown in valerate or acetate+propionate mixtures contained substantially reduced GCL-TSAR activity (Figure 3.22B, righthand panel), as compared to extracts from wild-type bacteria grown under the same conditions (Figure 3.22B, lefthand

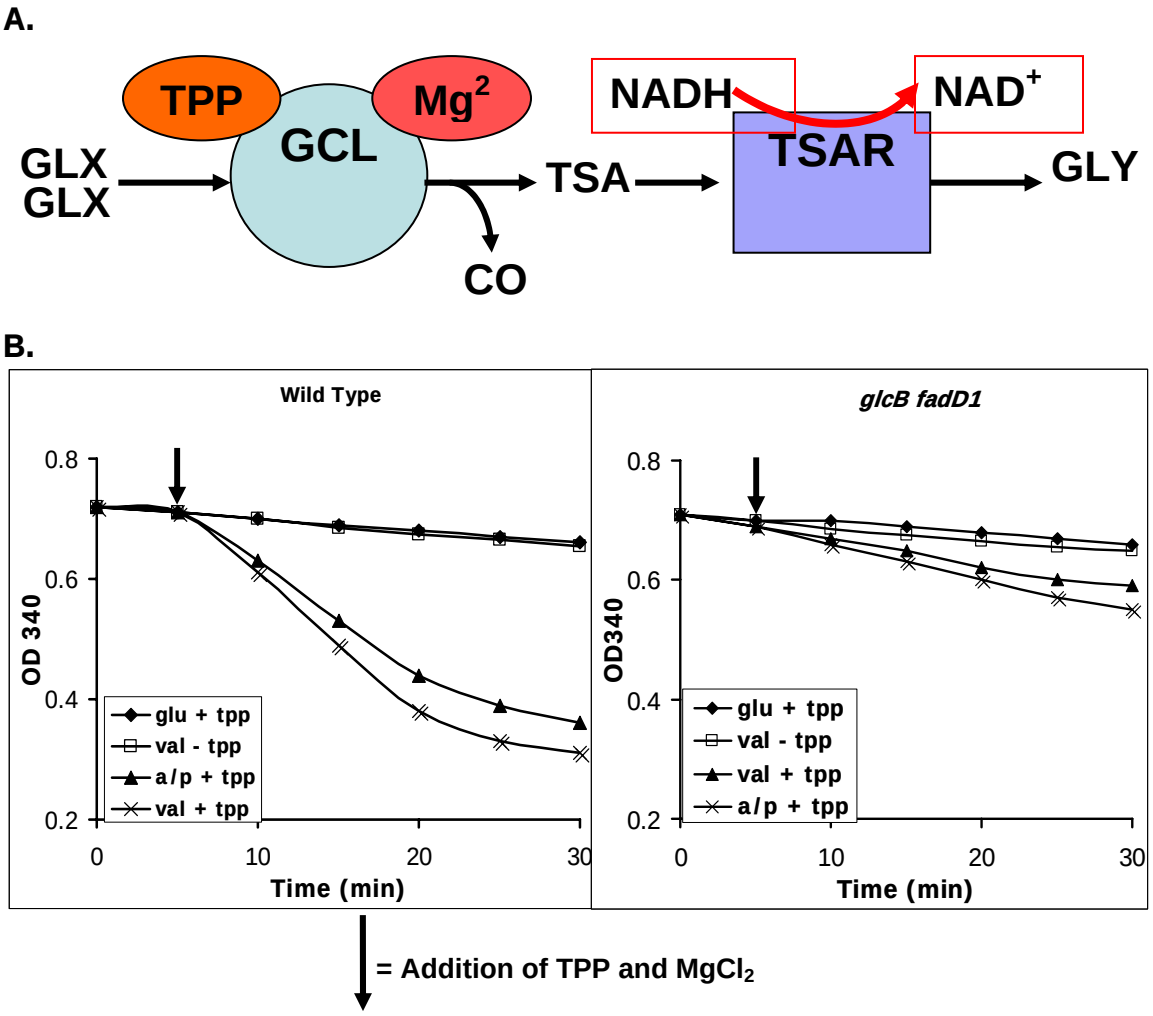


Figure 3.22. A functional copy of *fadD1* is required for the full induction of GCL-TSAR activity in *M. smegmatis*. (**A**) Schematic of the coupled GCL-TSAR assay in cell-free extracts. The reaction mixture contains glyoxylate as the substrate; upon the addition of TPP and $MgCl_2$, two molecules of glyoxylate are condensed to one tartronic semialdehyde (by GCL), which is reduced to glycerate with concomitant oxidation of NADH to NAD^+ (by TSAR). NADH oxidation is measured spectrophotometrically at 340 nm. (**B**) In wild-type *M. smegmatis* (left panel), GCL-TSAR activity was not detected in cultures grown on glucose (glu) or when TPP/$MgCl_2$ were not added to the reaction mix (- tpp). GCL-TSAR activity was strongly induced in cultures grown on acetate/propionate (a/p) or valerate (val). Extracts from Δ*glcB fadD1* bacteria grown on acetate/propionate or valerate contained reduced, but still detectable, GCL-TSAR activity (right panel), suggesting that a FadD1-independent mechanism for induction of GCL-TSAR activity might exist in *M. smegmatis*.

panel). Nonetheless, TSAR activity was still detectable in extracts from the Δ*glcB fadD1* bacteria grown in valerate or in acetate+propionate, suggesting that a second, FadD1-independent mechanism for regulation of the *gcl* operon might exist in *M. smegmatis*. This mechanism might involve activation by a C3 unit, which could be generated from metabolism of propionate or odd-chain fatty acids.

3.8. In-frame deletions of *gcl* and *fadD1* in *M. smegmatis*

The experiments described so far were carried out in strains of *M. smegmatis* that harbored in-frame deletions in *glcB* and *ald*, as well as tranposon insertions in glyoxylate carboligase (*gcl*), fatty acyl-CoA ligase 1 (*fadD1*), and the intergenic space between *thiO* and *thiE.* To rule out the possibility of polar effects of the transposon insertions on genes downstream of the transposon insertion sites, we generated strains that carried unmarked, in-frame deletions in either *gcl* or *fadD1* on both the wild type background and the Δ*glcB* background. These four strains (Δ*gcl*, Δ*fadD1*, Δ*glcB*Δ*gcl*, and Δ*glcB*Δ*fadD1*) were then plated on agar plates containing glucose, glyoxylate, or various chain-length fatty acids (from C2/acetate to C5/valerate) as the sole carbon substrates.

None of these four strains could grow on glyoxylate plates, confirming that, in *M. smegmatis*, both GCL and FadD1 activities are necessary for utilization of glyoxylate as the sole carbon source (not shown). The single Δ*gcl* and Δ*fadD1* mutants grew as well as wild-type bacteria on all other carbon sources, including acetate and the short chain fatty acids butyrate (C4) and valerate (C5), confirming that malate synthase is the dominant anaplerotic pathway for assimilation of the glyoxylate produced by isocitrate synthase when *M. smegmatis* grows on acetate or fatty acids. Finally, similarly to what we observed with the transposon-induced double mutants, the Δ*glcB*Δ*gcl* and Δ*glcB*Δ*fadD1* strains could not grow on acetate or butyrate as the sole carbon sources, but could form colonies on valerate after prolonged incubation (not shown). The

phenotypes of the Δ*glcB gcl* and Δ*glcB fadD1* transposon mutants were thus solely due to loss of GCL and FadD1 activities.

CHAPTER 4

Understanding Glyoxylate Metabolism In Mycobacteria

4.1. The role of malate synthase in pathogenic mycobacteria

Muñoz-Elías and McKinney (2005) reported that the two isocitrate lyases (ICL) of *M. tuberculosis* are jointly required for virulence in the lungs of infected mice. They also demonstrated that the two enzymes were necessary for bacterial growth on acetate (C2), butyrate (C4), laureate (C12) and palmitate (C16). Thus, it appears that, in *M. tuberculosis*, the ability to assimilate fatty acids (or at least even-chain fatty acids) through the glyoxylate shunt may correlate with the ability of the tubercle bacillus to grow and persist *in vivo*.

This correlation between the ability to carry out β-oxidation in the lungs of infected hosts and the ability to cause disease is consistent with evidence presented by Segal and Bloch (1956), Kanai and Kondo (1974) and Bharadwaj *et al.*, (1987), which suggested that lipid, rather than carbohydrate, catabolism was of crucial importance to the *in vivo* survival of pathogenic mycobacteria. Similarly, Raynaud *et al.*'s (2002) report that exported phospholipase C isoforms may be required for virulence lends further support to the idea that β-oxidation may be important *in vivo*. The results presented by Muñoz-Elías and McKinney (2005) suggest that novel drugs that can target both ICLs of mycobacteria could be effective new therapeutics for treating active mycobacterial infections.

However, despite the unambiguous importance of ICLs in mycobacterial pathogenesis, it is not clear if the other enzyme of the glyoxylate shunt, malate

synthase (MLS), is equally important for *M. tuberculosis* virulence. So far, for reason unknown, it has not been possible to isoalte a strain of *M. tuberculosis* harboring a null mutation in the *glcB* gene (Muñoz-Elías, 2005, Ph.D. thesis). It is possible that MLS activity is essential in pathogenic mycobacteria, and that a Δ*glcB* strain will never be obtained. In that case, MLS could still be a viable target for novel drug development. The potential essentiality of *glcB* needs to be demonstrated experimentally, however; experiments to that end are ongoing in our laboratory.

It is also plausible that the importance of ICL1/ICL2 *in vivo* is due to their participation in a pathway that does not involve MLS. One such pathway is the methylcitrate cycle, which appears to be the dominant pathway for metabolism of propionyl-CoA in bacteria (Figure 4.1; Textor *et al.*, 1997; Horswill and Escalante-Semerena, 1999; Claes *et al.*, 2002). The putative involvement of *M. tuberculosis* ICL, but not MLS, in the methylcitrate cycle came from the unexpected observation that ICL1/ICL2 were jointly required for *in vitro* growth of *M. tuberculosis* on propionate, and from the fact that *M. tuberculosis* lacks a *prpB* homolog encoding methylisocitrate lyase (Muñoz-Elías and McKinney, 2005; Muñoz-Elías and McKinney, 2006). Recently, however, Muñoz-Elías *et al.* (2006) reported that, although ICL1/ICL2 in fact do have have methylisocitrate lyase activity and participate in the methylcitrate cycle, the methylcitrate cycle (and perhaps, by extension, propionate metabolism) is not important for the ability of *M. tuberculosis* to establish and maintain infection in the lungs of mice. Thus, the

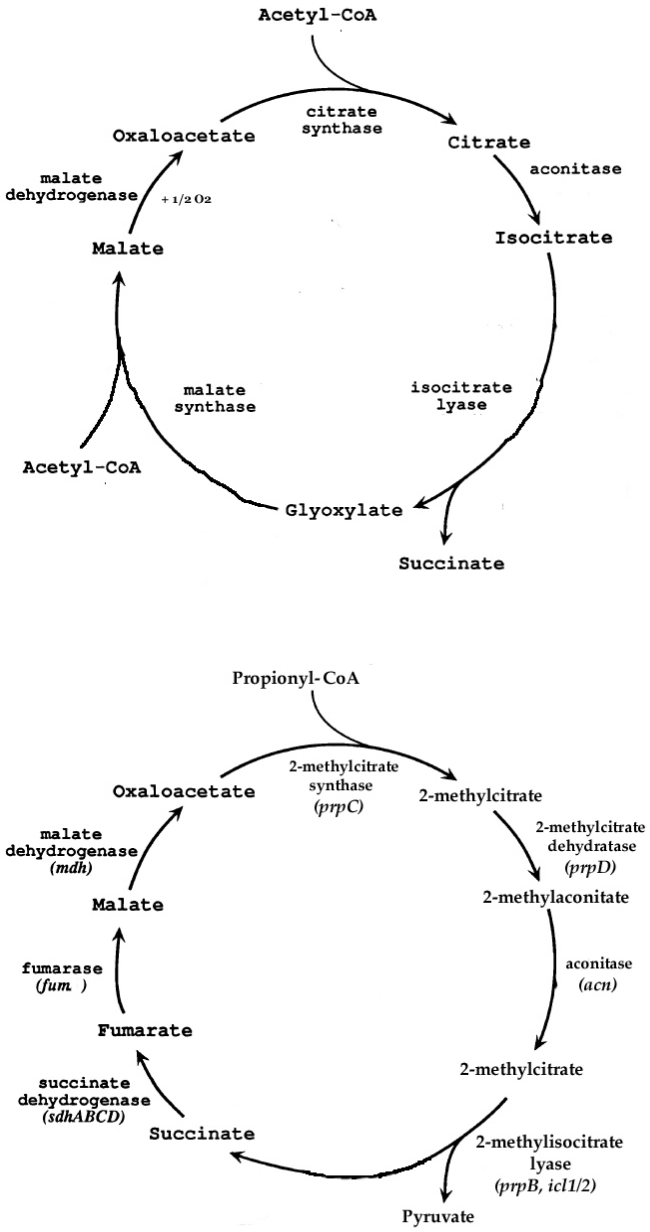


Figure 4.1. Glyoxylate cycle (GC), top; methylcitrate cycle (MCC), bottom. Diagrams were modified from Clark and Cronan (1996).

importance of ICL1/ICL2 *in vivo* seems to be restricted to, or in addition to, their role in the glyoxylate cycle. A scenario could be envisaged, however, in which despite the essentiality of ICLs *in vivo*, it is the glycine dehydrogenase activity reported by Wayne and Lin (1982), and not the MLS activity of *glcB,* that would be critical for the ability of the tubercle bacillus to survive and persist in host tissues. Thus, Boshoff and Barry (2005) propose that ICLs may be crucial as members of a pathway to regenerate reducing equivalents when no external electron acceptors (such as oxygen) are available, rather than as anaplerotic / gluconeogenic activities. They suggest that the importance of glyoxylate generation by ICL might lie in the fact that glyoxylate reduction by glycine dehydrogenase would reoxidize one molecule of NADH. Under these circumstances, MLS activity might not be essential.

Such questions will not be answered conclusively until strains lacking MLS or glycine dehydrogenase activities are generated and tested for *in vivo* survival and virulence. As discussed already, creating a strain of *M. tuberculosis* lacking glycine dehydrogenase is could be problematic since it is still not known exactly which gene encodes this activity. The most likely candidate, *gcvB*, might encode a protein that selectively decarboxylates glycine (Cole *et al.*, 1998; Wayne and Sohaskey, 2001). Another potential confounder for genetic studies is that *gcvB* might be an essential gene (Sassetti *et al.*, 2003). Therefore, generating a strain of *M. tuberculosis* lacking MLS activity, and then evaluating its properties *in vivo*, could be the only way to ascertain, unambiguously, the reason why ICL activity

is essential *in vivo:* for generating reducing equivalents, or for achieving anaplerosis (and initiating gluconeogenesis) during growth on fatty acids as the major available carbon source in the lungs of infected mammals.

4.2. The role of malate synthase in non-pathogenic *M. smegmatis*

We report here that malate synthase (MLS) activity is non-essential in *M. smegmatis*. The ICL (*icl1, icl2*) and MLS (*glcB*) genes can be deleted in *M. smegmatis* without affecting growth on glucose, glycerol, or propionate as the sole carbon source. While *icl1* is essential for *M. smegmatis* growth on acetate and even-chain fatty acids, *glcB* is dispensable for growth on these substrates, although it is required for optimal growth under these conditions. The *icl2* gene appears to be completely dispensable for growth on all substrates tested.

The Δ*glcB* strain has a significant phenotype on acetate and even-chain fatty acids. Compared to wild-type *M. smegmatis*, the Δ*glcB* mutant grows more slowly on solid media containing acetate or even-chain fatty acids as the sole carbon source, and it exhibits a reproducible growth lag in M9 liquid media conatinaing acetate (C2), hexanoate (C6), and, to a lesser extent, butyrate (C4) (Figure 3.20). Furthermore, increased concentrations of acetate exacerbate the growth defect of the Δ*glcB* strain, even when a metabolizable carbon source such as propionate is provided (Figure 3.21).

An intact copy of *glcB* clearly improves the ability of *M. smegmatis* to grow on acetate and fatty acids; indeed, MLS is the dominant anaplerotic enzyme for glyoxylate assimilation during growth on these substrates. Expression of MLS activity completely masks the loss of glyoxylate carboligase activity on all carbon substrates tested, with the exception of glyoxylate. ICL-MLS thus represents the main anaplerotic pathway on acetate or fatty acid substrates in *M. smegmatis.*

4.3. The role of glyoxylate carboligase in *M. smegmatis*

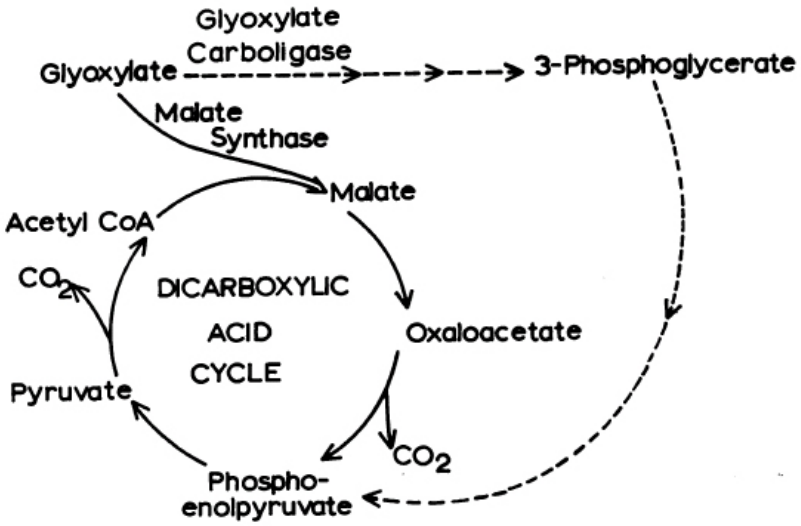
This thesis provides the first idenfication of mycobacterial genes encoding an operon for the utilization of glyoxylate as the sole carbon source. Glyoxylate carboligase activity was first reported in *E. coli* (Krakow and Barkulis, 1956) and in *Pseudomonas* sp. (Kornberg and Sadler, 1960). This “d-glycerate pathway” comprises the enzymes glyoxylate carboligase, tartronate semialdehyde reductase, and glycerate kinase, which convert two molecules of glyoxylate (C2) into one molecule of d-glycerate (C3) and then 3-phosphoglycerate (Ornston and Ornston, 1969; see Figure 2.10). The significance of this pathway to the physiology of *M. smegmatis* is that it enables the organism to grow on poor carbon substrates like glyoxylate. *M. smegmatis* also appears to have at least two homologs of *E. coli* glycolate oxidases, which can convert glycolate to glyoxylate, so *M. smegmatis* might also be capable of growth on glycolate.

Unlike in *E. coli*, the *M. smegmatis gcl* operon represents a second pathway, besides MLS, for assimilation of the glyoxylate that is produced by ICL during growth on acetate or fatty acids. In this context, the d-glycerate pathway can perform an anaplerotic role similar to the glyoxylate shunt. The anaplerotic role of glyoxylate carboligase during growth on acetate or fatty acids would normally be masked by the presence of MLS. The reaction catalyzed by MLS produces malate, a four-carbon molecule, from two C2 units; glyoxylate carboligase produces a three-carbon molecule from two C2 units, so this reaction is not as efficient at assimilating carbon, unless glyoxylate is the sole substrate.

The experiments performed here allow us to revisit the work of Ornston and Ornston (1969) on the fate of the 3-phosphoglycerate molecule produced by the d-glycerate pathway. Kornberg and Sadler (1960) had proposed the existence of a dicarboxylic acid cycle, in which MLS activity was required for the oxidation of glyoxylate by *E. coli*. Hansen and Hayashi (1962) argued that glyoxylate was eventually oxidized by the tricarboxylic acid cycle, obviating the need for MLS and the glyoxylate shunt (Figure 4.2). Our experiments with *M. smegmatis* argue that glyoxylate is indeed oxidized through the TCA cycle, as the Δ*glcB* mutant grows on glyoxylate as well as wild-type bacteria. Therefore, the dicarboxylic acid cycle, although it may operate in *M. smegmatis*, is apparently not essential for the oxidation of glyoxylate. This conclusion was also reached by Ornston and Ornston (1969), when they obtained *E. coli* mutants lacking malate synthase G activity.

However, the TCA cycle oxidation model of Ornston and Ornston (Figure 4.2B) needs to be updated, as it implies the existence of an anaplerotic step for the formation of malate, which would be catalyzed by malate synthase A. While that may be the case in *E. coli*, where two malate synthases exist, it cannot be the case in *M. smegmatis*, as deletion of the sole MLS gene (*glcB*) still allows growth on glyoxylate. Most likely, 3-phosphoglycerate is converted to PEP or pyruvate, followed by conversion of PEP to oxaloacetate by PEP carboxylase, or by conversion of pyruvate to oxaloacetate by pyruvate carboxylase (Figure 4.3). *M. smegmatis* has one *ppc* and two *pca* genes in its genome.

A.



B.

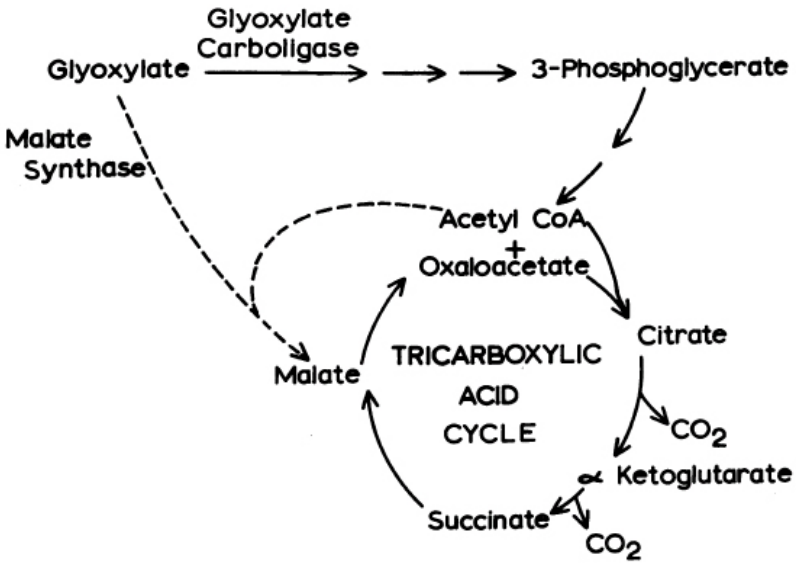


Figure 4.2. Pathways for oxidation of glyoxylate in *E. coli*. Solid lines stand for oxidative reactions, dashed lines represent anaplerotic reactions. (**A**) According to the dicarboxylic acid cycle mechanism (Kornberg and Sadler, 1960), malate synthase is an oxidative enzyme. (**B**) In the tricarboxylic acid cycle diagram, glyoxylate carboligase is an oxidative enzyme, and anaplerosis is achieved by malate synthase A (Hansen and Hayashi, 1962). Drawings were reproduced from Ornston and Ornston (1969).

4.4. The Role of FadD1 in *M. smegmatis*

The deletion of both the *glcB* and *gcl* genes resulted in a mutant that behaves exactly like the Δ*icl1* Δ*icl2* mutant: it cannot grow on acetate or even-chain fatty acids. It thus appears that the d-glycerate pathway is the ICL-dependent, MLS-independent anaplerotic pathway that we set out to discover. However, disruption of *fadD1* resulted in a phenotype very similar to that obtained by disruption of *gcl*, inasmuch as the Δ*glcB gcl* and Δ*glcB fadD1* mutants could not grow on acetate or even-chain fatty acids. More importantly, the double mutants could not use glyoxylate as the sole carbon source, while the Δ*glcB* single mutant could. These results suggest that FadD1 might activate the expression of the *gcl* operon. The assays shown in Figure 3.22 suggests that the activity of FadD1 is partially required to induce the *gcl* operon during growth on valerate or acetate/propionate, indicating that another activator of the *gcl* operon may exist.

FadD1 is most likely a fatty acyl-CoA ligase, and fatty acyl-CoAs have been shown to inactivate the regulator FadR in *E. coli* (Black and DiRusso, 2003). When FadR binds the fatty acyl-CoA molecule, it dissociates from the operators in its target promoters and allows for transcription of the downstream genes. It is tempting to suggest a similar mechanism for FadD1. However, mycobacteria do not have an obvious FadR homolog, and little is known about how gene regulation works in mycobacteria. The hypothetical product of the GclR gene has homology to the IclR transcription factor in *E. coli*. It is known that IclR represses the *aceBA* genes in *E. coli*, and this repression is relieved by the

binding of PEP to IclR (Cortay *et al.*, 1991). FadR also activates expression of *iclR* (Gui *et al*, 1996).

It is currently unknown whether a similar transcriptional control mechanism exists in *M. smegmatis*. The *gcl* operon has two palindromic sites, which could serve as binding sites for transcriptional activators/repressors: one upstream of *hypI*, and one upstream of *gclR*. GclR itself may be involved in the regulation of the expression of the *gcl* operon, but the mode of regulation needs to be determined experimentally. Since little is known about the modulation of gene expression in mycobacteria, regulation of the *gcl* operon could involve a novel mechanism.

4.5. The malate synthase of pathogenic mycobacteria

The malate synthase activity in *M. smegmatis* is dispensable for growth on C2 units because of the presence of the d-glycerate pathway. As it was pointed out, however, the genes of the d-glycerate pathway are not found in any pathogenic *Mycobacterium* species. The closest homologs of glyoxylate carboligase that are found in the *M. tuberculosis* genome are the acetolactate synthases. In *E. coli*, it was shown that acetolactate synthase could catalyze condensation of two glyoxylate molecules to form tartronate semialdehyde; however, this reaction was so inefficient that it would be probably be physiologically meaningless (Chang *et al.,* 2003). *M. smegmatis* possesses a number of genes encoding acetolactate synthase homologs, yet these genes apparently cannot compensate for the loss of glyoxylate carboligase activity.

The Δ*glcB* and Δ*glcB gcl* strains of *M. smegmatis* complemented with a copy of the *M. tuberculosis glcB* gene could be useful tools for testing potential inhibitors of the *M. tuberculosis* MLS. Should MLS be essential for survival of *M. tuberculosis in vivo,* as has been shown for ICL1/ICL2, MLS inhibitors could serve as leads for the development of novel drugs for treatment of tuberculosis.

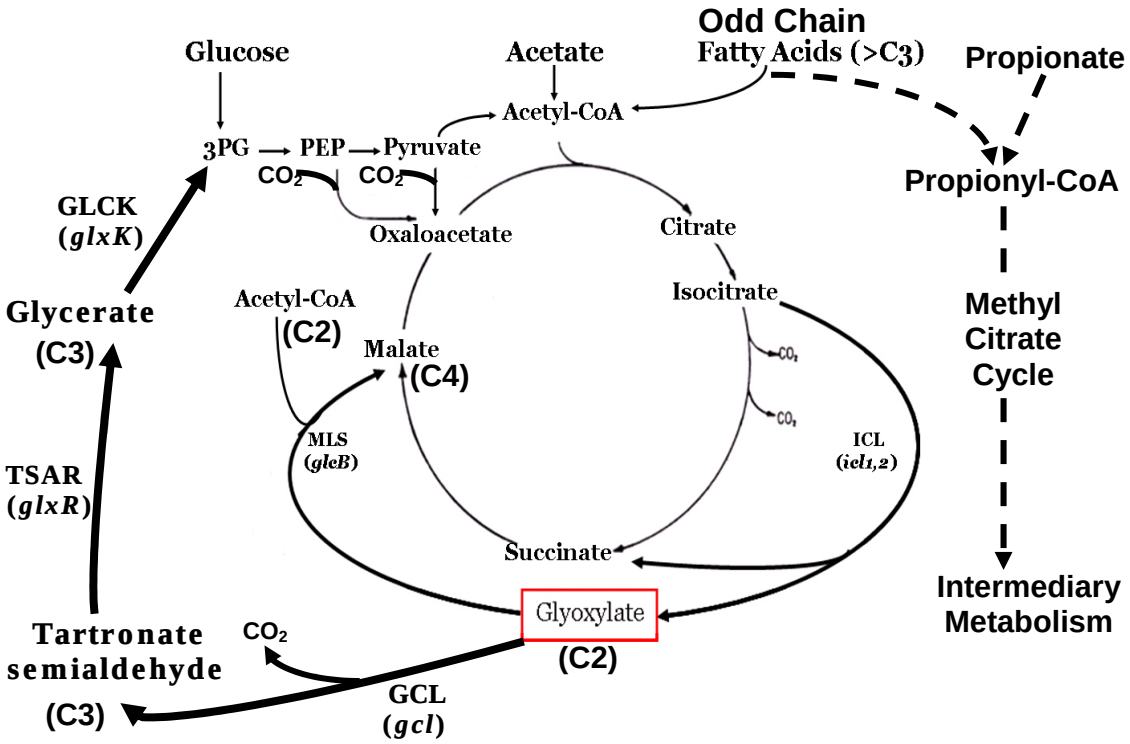


Figure 4.3. Glyoxylate assimilation in *M. smegmatis*. Abbreviations: ICL, isocitrate lyase; MLS, malate synthase; GCL, glyoxylate carboligase; TSAR, tartronic semialdehyde reductase; GLXK, glycerate kinase.

CHAPTER 5

Materials and Methods

5.1. Bacterial strains and media

M. smegmatis mc^2155 was stored at -80C in 15% glycerol.

Bacteria were grown at 37˚C; in Middlebrook 7H9 broth (DIFCO) with 10% ADS (DIFCO), 0.5% glycerol, and 0.05% Tween-80; on Middlebrook 7H10 agar (DIFCO) containing 10% ADS and 0.5% glycerol.

Antibiotics: hygromycin (50 µg/ ml), kanamycin (25 µg /ml) (Sigma)

M9 minimal media: M9 salts (DIFCO), 0.1 mM $CaCl_2$, 2mM $MgSO_4$; agar was added at 15 grams per liter. Carbon substrates (from Sigma) were added at (w/v) 0.1%, 0.2%, or 0.5% (for methylpalmitate).

5.2. Growth curves in minimal media/growth on agar plates

M. smegmatis mc^2155 and mutant strains were grown in Middlebrook 7H9 broth to OD_{600} of 1.0, then diluted to final OD_{600} in M9 media + 0.1% glucose, grown again to OD_{600} of 1.0, then diluted in M9 media + 0.1% or 0.2% of the indicated carbon sources. For plating on M9 solid media, cells grown in M9 media plus 0.1% dextrose were serially diluted in PBS + 0.1% Tween 80, plated on solid media, and incubated at 37˚C for the indicated times

5.3. Construction of transposon libraries

The ϕMicoMarT7 transposon donor phagemid was provided by Dr. Eric Rubin (Harvard School of Public Health). *M. smegmatis* mc^2155 was grown to OD_{600} of 1.0 in 7H9 media, washed twice with MP buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM $MgSO_4$, 2 mM $CaCl_2$) and resuspended in $1/10^{th}$ original volume in MP buffer. 10^{10} PFU (plaque-forming units) of phage per ml of original culture was added for 3 hrs at 37˚C. Infected cells were plated on 7H10 agar with 10% ADS, 0.5% glycerol, 0.5% glucose, and 25 µg/ml kanamycin. After 3 days, distinct colonies were individually picked with sterilized toothpicks in 96 well plates containing 7H9 plus 0.5% glucose for 2 days, then spotted with a multi-channel pippetor on M9 agar plates containing 0.1% dextrose, 0.1% acetate, or 0.5% methylpalmitate as sole carbon source. Plates were incubated at 37˚C for 5 days. Original 96 well master plates were stored at –80˚C. Mutants with the desired phenotypes were selected, grown in 7H9 media, and then tested for growth in M9 liquid or M9 agar.

5.4. Generation of deletion mutants

5.4.1. Deletion of *glcB* gene

A 3.8 kb NotI-XmaI genomic fragment containing the *M. smegmatis glcB* gene was obtained from a cosmid and subcloned in pBSKS (pBluescript) digested with NotI and XmaI. The new construct was cut with MscI and the 1.9 kb (smaller) fragment was removed. The larger fragment containing upstream and downstream regions was purified and re-ligated. The truncated gene was cut out

with NotI and XmaI, subcloned into p2NIL digested with NotI and XmaI, then ligated into pGOAL19 as described (Parish and Stoker, 2000) to give the *glcB* knockout vector. p2NIL and pGOAL19 have been described previously (Parish and Stoker, 2000); they were a generous gift from Dr. Tanya Parish. The *glcB* knockout plasmid was electroporated in *M. smegmatis*, and mutant strains were selected as described in Figure 3.0. The *glcB* complementing integrative vector was pEM186, provided by Ernesto Muñoz-Elías; it contains a 10,107 base pair genomic region from H37Rv including the full-length *glcB* gene. In particular, it contains nucleotides 2,081,482 to 2,091,588 of the H37Rv genome, as it is annotated on Tuberculist (http://genolist.pasteur.fr/TubercuList/). The cosmid contains genes *rv1836c*, *glcB (rv1837c)*, *rv1838c, rv1839c,* and *rv1840c*. It was electroporated into the Δ*glcB* strain and into individual Δ*glcB* transposon mutants and transformants were selected on 7H10 agar containing hygromycin.

5.4.2. Deletion of *icl1* gene

A 2.3kb fragment containing the entire *icl1* gene of *M. smegmatis* was amplified by PCR from genomic DNA using primers (the XhoI binding site is underlined):

smicl1upxhoI: 5-CTCGAGCTTCGACCACATGAACAACG-3;

smicl1dnxhoI: 5-CTCGAGGATCTTCATGATCGGGATGC-3.

The PCR product was subcloned into pCR2.1 and sequenced. A 2.3kb EcoRI fragment containing the PCR product was subcloned into EcoRI-digested pBSKS (pBluescript) then digested with SfoI to remove an 855 bp fragment internal to

the *icl1* gene (the deletion removes 285 amino acids from the ICL protein, including the catalytic site, and results in an in-frame deletion in the ORF). The other (5.4kb) SfoI fragment was re-ligated, and then digested with XhoI to excise the 1.5kb band containing the truncated *icl1* gene. This 1.5kb fragment was subcloned into XhoI-digested pJG1100 vector (a gift from Dr. James Gomez) to create the *icl1* knockout plasmid. It was electroporated in wild type *M. smegmatis* and the Δ*icl2* strain and mutants were isolated by the two-step counterselection method described in Figure 3.0.

5.4.3. Deletion of the *icl2* gene

The *icl2* gene was deleted by Ernesto Muñoz-Elías. For disruption of *icl2*, a 3.9kb AgeI fragment from cosmid pEM353 (the *M. smegmatis* cosmid genomic library was a gift of B. Subramanian, AstraZeneca Research Foundation) carrying *icl2* was cloned into pSL301 to generate pEM701, which was digested with EcoRV to obtain a 3.5kb fragment carrying *icl2*, which was then cloned into the PmeI site of pJG1001 to generate pEM902. pEM902 was BstEII-digested to remove a 660-bp fragment, which generated an unmarked in-frame deletion after re-ligation into pEM903. The deletion encompasses amino acids 307-526 of the *icl2* ORF. pEM903 was electroporated into wild-type *M. smegmatis* and mutants were isolated by the two-step counter-selection method described in Figure 3.0.

5.4.4. Deletion of the *ald* gene

The upstream and downstream regions of *ald* were PCR amplified using primers:

Ald-up-F: TTAATTAAGAGGGCTCGGCCATCTCG (underlined is the PacI site)

Ald-up-R: GAATTCCGGGATTCCGACGAGCAT (underlined is the EcoRI site)

Ald-dn-F; GAATTCGCACAGTTCCTGGCGTAA (underlined is the EcoRI site)

Ald-dn-R: GGCCGGCCGTCCTTGAGGACGACGGT (underlined is the AscI site)

The upstream and downstream regions (about 800 bp each) were PCR amplified and subcloned in pCR2.1 and sequenced. Correct clones were digested with PacI - EcoRI (Ald-Up) and AscI - EcoRI (Ald-Dn) and subcloned (3 fragment ligation) in PacI-AscI-digested pJG1004 to create the *ald* knockout plasmid. pJG1004, an earlier version of pJD1100, was a gift from Dr. James Gomez. The *ald* KO plasmid was electroporated into *M. smegmatis* and *ald* was deleted by the two-step counter-selection method described in Figure 3.0.

5.4.5. Deletion of the *gcl* gene

The upstream and downstream regions of *gcl* were PCR amplified using primers:

Gcl-up-L1: TTAATTAACCAGCTGATCGTGGCGGG (PacI site underlined)

Gcl-up-R1: CCTAGGCATGCGGGTCATGGCCGC (AvrII site underlined)

Gcl-dn-L1: CCTAGGGCGCTGTTCGATTAGTGA (AvrII site underlined)

Gcl-dn-R1: GGCGCGCCAGATCGCGGACGTCGTTG (AscI site underlined)

The upstream PCR product (up primers) was ~400 basepairs, while the downstream PCR product (dn primers) was ~600 basebairs. PCR products were subcloned in pCR2.1 using the TopoTA kit (Invitrogen) and several clones were

sequenced. Fragments with the correct sequence were subcloned in pJG1100 in a stepwise fashion: first the upstream region was excised with PacI-AvrII, and subcloned into PacI-AvrII-digested pJG1100; then the downstream region was excised with AscI-AvrII and subcloned into AscI-AvrII-digested vector. Δ*gcl* and Δ*glcB*Δ*gcl* mutants were isolated by the two-step counter-selection method described in Figure 3.0.

5.4.6. Deletion of the *fadD1* gene

The upstream and downstream regions of *fadD1* were PCR amplified using primers:

FadD1-up-L1: TTAATTAAACTCCTCGGGTTCGTCGA (PacI site underlined)

FadD1-up-R1: CCTAGGCTGCAGCGTATCGGCCAT (AvrII site underlined)

FadD1-dn-L1: CCTAGGGCGCCAGGAATATCCCCT (AvrII site underlined)

FadD1-dn-R1: GGCGCGCCGAGCTCGCGGGCCGCCTC (AscI site underlined)

The upstream region PCR fragment (~500 basepairs) and downstream PCR fragment (~700 basepairs) were subcloned into pCR2.1 using the TopoTA kit (Invitrogen) and sequenced. Correct clones were subcloned into pJG1100 the same way the *gcl* knockout plasmid was generated, and Δ*fadD1* mutants were isolated using the two-step counter-selection method described in Figure 3.0.

5.5. DNA manipulation

5.5.1. Identification of transposon mutants

M. smegmatis genomic DNA was isolated as described (Muñoz-Elías *et al.*, 2005) and digested with BamHI or ApaLI. One half of the digested DNA was used for Southern blotting, the other was ligated with T4 ligase (NEB) overnight and transform in Pir1 competent *E. coli* cells (Invitrogen). Kanamycin-resistant colonies were isolated, and plasmid DNA was isolated and sequenced using the primer cttctgagcgggactctgggg, which hybridizes near one end of the ϕMycoMarT7 transposon.

5.5.2. Southern blots

5 micrograms of genomic DNA was digested with the indicated restriction enzymes, separated on 1% TBE gels, transferred to Hybond N membrane (Amersham), and probed with ^{32}P-labeled DNA using the random primer labeling kit (Boehringer Mannheim) and hybridized from 2 hours to overnight. The hybridized membrane was sequentially washed with 2X SSC, 0.1% SDS and 0.1X SSC, and 0.1% SDS before exposing the membrane to autoradiography

5.6. Enzyme assays

5.6.1. Malate synthase assay

Cells were harvested, washed three times with PBST (PBS plus 0.05% Tween 80), and resuspended in Tricine buffer (20 mM Tricine pH 7.5, 5 mM $MgCl_2$, 1 mM EDTA) supplemented with protease inhibitors (Sigma). The cells were

disrupted by bead-beating and the cell-free extract was clarified by centrifugation in a tabletop centrifuge for 30 min at 14,000 rpm at 4˚C.

For MLS assays, a protocol modified from Smith *et al.* (2003) was used. In 1 ml final volume, 20 mM Tricine-HCl pH 7.5, containing 5 mM $MgCl_2$, 0.8 mM EDTA, 0.2 mM glyoxylate, and 0.2 mM acetyl-CoA were mixed with 50 microliters of protein extract stored in Tricine-HCl pH 7.5, 5 mM $MgCl_2$, and 0.8 mM EDTA buffer, and the reaction was incubated at room temperature for 30 min. The reaction was stopped by adding DTNB (Sigma) to a final concentration of 2 mM in Tris-HCl pH 8.0. The amount of CoASH released was measured by titrating the free thiol groups with the DTNB and measuring change in absorbance at 412 nm.

5.6.2. Tartronate semialdehyde reductase assay:

Protein extracts were obtained by bead-beating cells in 50 mM KH_2PO_4 buffer (pH 7.0) containing protein inhibitors mix (Sigma) then centrifuging them in a tabletop centrifuge for 30 min at 14,000 rpm at 4˚C. TSAR assays were done essentially as described in Hansen and Hayashi (1962). In a 1 ml cuvette at room temperature, in 100 mM KH_2PO_4 buffer (pH 7.0), 0.33 mM NADH, 0.1 mM glyoxylate, and 50 microliters of protein extract were mixed. Tartronate semialdehyde was generated by adding 10 mM $MgCl_2$ and 0.1 mg TPP at t = 5 minutes. Generation of d-glycerate from tartronic semialdehyde, catalyzed by tartronate semialdehyde reduction, was measured by recording the disappeance of NADH at 340nm.

Appendix A.

MycoMarT7 transposon mutants on the Δ*glcB*Δ*ald* (dg-tn) mutant strain of *Mycobacterium smegmatis* with growth phenotypes on minimal media plates containing glucose (G), acetate (A) or methyl-palmitate (M) as the sole carbon sources.

Mutant Name	Phenotype	Identity of Rv homolog	Annotated Function in TubercuList
dg-tn-1AC5	D+A-M+	*fadD28*	Fatty acyl-CoA ligase
dg-tn-1BD1	D~A+M+	*n.a.*	n.a.
dg-tn-2AH6	D~A-M-	*thiO/thiE*	Thiamine synthesis oxidoreductases
dg-tn-4AA4	D+A-M-	*icl1*	Isocitrate Lyase 1
dg-tn-8AD4	D+A~M~	*fixA*	Electron transfer flavoprotein
dg-tn-8BA4	D+A~M-	*Rv0338c*	Fe-S binding reductase
dg-tn-9AH2	D+A-M+	*acs*	Acetyl-CoA synthase
dg-tn-12AA3	D+A-M-	*icl1*	Isocitrate lyase 1
dg-tn-13BD1	D-A+M+	*dut*	dUTP pyrophosphatase
dg-tn-14AA4	D+A~M~	*Rv2974c*	Conserved hypothetical
dg-tn-14AG2	D+A-M-	*gcl*	Glyoxylate carboligase
dg-tn-15BD3	D+A-M-	*icl1*	Isocitrate lyase 1
dg-tn-15BF2	D~A-M+	*ctaB*	Cytochrome c oxidase
dg-tn-17BA1	D+A-M+	*acs*	Acetyl-CoA synthase
dg-tn-17BB5	D+A-M-	*fadD1*	Fatty acyl-CoA ligase
dg-tn-17BD5	D+A-M-	*icl1*	Isocitrate lyase 1

dg-tn-19AH6	**D+A-M+**	***acs***	**Acetyl-CoA synthase**
dg-tn-20BB5	**D+A-M+**	***n.a.***	**n.a.**
dg-tn-23BH1	**D+A-M+**	***acs***	**Acetyl-CoA ligase**
dg-tn-25AG3	**D+A-M+**	***acs***	**Acetyl-CoA synthase**
dg-tn-54C2	**D+A-M-**	***sdhD***	**Succinate dehydrogenase D**
dg-tn-56B5	**D+A-M-**	***icl1***	**Isocitrate lyase 1**

Approximately 4,000 mutants containing the ϕMycoMarT7 transposon (a kind gift from Dr. Eric Rubin, Harvard School of Public Health) were screened for loss of the ability to grow on glucose, acetate or methyl-palmitate as the sole carbon source in minimal basal agar plates. Mutants with desired phenotypes were then re-screened and the insertion site of the transposon was identified by sequencing the region downstream of the insertion. DNA sequences were aligned against the *Mycobacterium tuberculosis* genome http://genolist.pasteur.fr/TubercuList/. Two mutants did not have clear homologs in *M. tuberculosis.*

All 24 mutants were transformed with an integrative copy of the *M. tuberculosis* malate synthase (*glcB*) gene under the control of the endogenous promoter, and the transformants were tested for complementation of the respective phenotypes by *glcB.* 3 of the 24 mutants were complemented by *glcB*: mutant 2AH6, mutant 14AG2, and mutant 17BB5. These three mutants are underlined in the table.

Appendix B.

Mycobacterium smegmatis **ϕMycoMarT7 transposon mutants with growth phenotype on minimal media plates containing glucose (G), acetate (A) or methyl-palmitate (M) as the sole carbon sources.**

Mutant Name	Phenotype	Identity of Rv homolog	Annotated Function in TubercuList
mc-tn-2D8	D+A-M-	*icl1*	Isocitrate lyase 1
mc-tn-6C3	D+A~M+	*Rv3588c*	Carbonic anhydrase
mc-tn-6C7	D-A+M+	*ppgK*	Polyphosphate glucokinase
mc-tn-7A11	D+A-M-	*Icl1*	Isocitrate Lyase 1
mc-tn-11C7	D+A~M+	*gltA2*	Citrate synthase
mc-tn-11C11	D+A-M+	n.a.	n.a.
mc-tn-11D1	D+A-M+	*acs*	Acetyl-CoA synthase
mc-tn-13E5	D+A-M+	*acs*	Acetyl-CoA synthase
mc-tn-21B6	D+A-M-	*icl1*	Isocitrate lyase 1
mc-tn-22D8	D+A-M+	*acs*	Acetyl-CoA synthase
mc-tn-24D5	D+A-M+	*acs*	Acetyl-CoA synthase
mc-tn-26D2	D+A-M-	*icl1*	Isocitrate lyase 1
mc-tn-31F4	D+A-M+	*Rv3662c*	Conserved hypothetical
mc-tn-33C11	D+A-M+	*acs*	Acetyl-CoA synthase
mc-tn-37B5	D+A-M-	*icl1*	Isocitrate lyase 1
mc-tn-38B8	D+A~M+	*Rv1342c*	Conserved membrane protein

mc-tn-39E7	**D+A~M-**	***Rv0338c***	**Fe-S binding reductase**
mc-tn-43H1	**D+A-M+**	***acs***	**Acetyl-CoA synthase**
mc-tn-45C8	**D~A~M~**	***argH***	**Argininosuccinate lyase**
mc-tn-45F8	**D+A-M+**	***acs***	**Acetyl-CoA synthase**
mc-tn-47G4	**D~A~M~**	***nirA***	**Nitrite reductase**
mc-tn-49E11	**D+A~M-**	***Rv0338c***	**Fe-S binding reductase**
mc-tn-53H5	**D-A-M+**	***Rv2974c***	**Conserved hypothetical**
mc-tn-54D8	**D+A-M+**	***acs***	**Acetyl-CoA synthase**
mc-tn-54H4	**D~A~M~**	***nirA***	**Nitrite reductase**
mc-tn-58B7	**D+A-M-**	***pckA***	**Phosphoenolpyrivate carboxykinase**
mc-tn-60B9	**D+A~M~**	***Rv1841c***	**Conserved hypothetical**
mc-tn-67H8	**D+A~M~**	***glcB***	**Malate synthase**
mc-tn-68C6	**D+A-M+**	***Rv2468c***	**Conserved hypothetical**
mc-tn-69F10	**D+A~M+**	**n.a.**	**n.a.**

Approximately 7,000 mutants containing the fMycoMarT7 transposon (a kind gift from Dr. Eric Rubin, Harvard School of Public Health) were screened for loss of the ability to grow on glucose, acetate or methyl-palmitate as the sole carbon source in minimal basal agar plates. Mutants with desired phenotypes were then re-screened and the insertion site of the transposon was identified by sequencing the region downstream of the insertion. DNA sequences were aligned against the *Mycobacterium tuberculosis* genome at http://genolist.pasteur.fr/TubercuList/.

All but two of the mutants had clear homologs in *M. tuberculosis.* Mutant 11C11 has some homology to D-amino-acylases from various bacterial species; mutant 69F10 has homology to “probable transporter proteins”. The work to study these mutants, along with the entire 7,000+ library that was generated in this genetic screen, was done by Michael Silverman and Lubomir Merkov.

Bibliography

Akopiants, K., G. Florova, C. Li, and K. Reynolds. 2006. Multiple pathways for acetate assimilation in *Streptomyces cinnamonensis*. J. Ind. Microbiol. Biotech. **33:** 141-150.

Anderson, A. B., and W. A. Wood. 1969. Carbohydrate metabolism in bacteria. Annu. Rev. Microbiol. **23:** 539-578.

Arora, P., A. Vats, P. Saxena, D. Mohanty, and R. S. Gokhale. 2005. Promiscuous fatty acyl CoA ligases produce acyl-CoA and acyl-SNAC precursors for polyketide biosynthesis J. Am. Chem. Soc. **127:** 9388-9389.

Bharadwaj, V. P., V. M. Katoch, V. D. Sharma, K. B. Kannan, A. K. Datta, and C. T. Shivannavar. 1987. Metabolic studies on mycobacteria. IV. Assay of isocitrate lyase and malate synthase activity in *M. leprae*. Indian J. Lepr. **59:** 158-162.

Barnes, H. L., and L. R. P. Barnes. 1928. Duration of life in pulmonary tuberculosis with cavity. Am. Rev. Tuberculosis **18:** 412-424.

Bentley, R. 2000. From "reactive C_2 units" to acetyl coenzyme A: a long trail with acetyl phosphate detour. TIBS **25:** 302-305.

Black, P. N. 1988. The *fadL* gene product of *Escherichia coli* is an outer membrane protein required for uptake of long-chain fatty acids and involved in sensitivity to bacteriophage T2. J. Bacteriol. **170:** 2850–2854.

Black, P. N. 1991. Primary sequence of the *Escherichia coli fadL* gene encoding an outer membrane protein required for long-chain fatty acid transport. J. Bacteriol. **173:** 435–442.

Black, P. N., and C. C. DiRusso. 1993. Transmembrane movement of exogenous long-chain fatty acids: Proteins, enzymes, and vectorial esterification. Microbiol. Mol. Biol. Rev. **67:** 454-472.

Boshoff, H. I., and C. E. Barry, 3rd. 2005. Tuberculosis - metabolism and respiration in the absence of growth. Nat. Rev. Microbiol. **3:** 70-80.

Brennan, P. J. 2003. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis **83:** 91-97.

Callegos, M-T., R. Schleif, A. Bairoch, K. Hofmann, and J. L. Ramos. 1997. AraC/ XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. **61:** 393-310.

Chang, Y. Y., A. Y. Wang, and J. E. Cronan Jr. 1993. Molecular cloning, DNA sequencing and biochemical analysis of *Escherichia coli* glyoxylate carboligase. J. Biol. Chem. **268:** 3911-3919.

Claes, W. A., A. Puhler, and J. Kalinowski. 2002. Identification of two *prpDBC* gene clusters in *Corynebacterium glutamicum* and their involvement in propionate degradation via the 2-methylcitrate cycle. J. Bacteriol. **184:** 2728-2739.

Clark, D.P., and J. E. Cronan, Jr. 1996. Two-carbon compounds and fatty acids as carbon sources. In: F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (Ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology, 2nd ed.*, vol. 1. ASM Press, Washington, D.C.

Clark-Curtiss, J. E. and S. E. Haydell. 2003. Molecular genetics of *Mycobacterium tuberculosis* pathogenesis. Annu. Rev. Microbiol. **57:** 517-549.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell, *et al*. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:** 537-544.

Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. Nature **409:**1007-1011.

Comstock, G. W. 1982. Epidemiology of Tuberculosis. Am. Rev. Respir. Dis. **125:** 8-15.

Conway, A. H, and T. Romano. 1996. Evolution of carbohydrate metabolic pathways. Res Microbiol **147:** 448-455.

Cortay, J. C., D. Negre, A. Galinier, B. Duclos, G. Perriere, and A. J. Cozzone. 1991. Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor. EMBO J. **10:** 675-679.

Cox, J.S., Chen, B., McNeil, M., and W. R. Jacobs, Jr. 1999. Complex lipid determines the tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature **402:** 79-83.

Cronan Jr., J. E., and D. LaPorte. 1996. Tricarboxylic Acid Cycle and Glyoxylate Bypass. In: F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (Ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology, 2nd ed., vol. 1*. ASM Press, Washington, D.C.

Cusa, E., N. Obradors, L. Baldoma, J. Badia, and J. Aguilar. 1999. Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli*. J. Bacteriol. **181:** 7479-7484.

Dannenberg Jr., A. M., and G. A. W. Rook. 1994. Pathogenesis of pulmonary tuberculosis: an interplay of tissue-damaging and macrophage-activating immune responses-dual mechanisms that control bacillary multiplication. pp. 459-484. In: B.R. Bloom (Ed.), *Tuberculosis, Pathogenesis, Protection and Control.* ASM Press, Washington D.C.

Diesterhaft, M. D. and E. Freese. 1973. Role of pyruvate carboxylase, phosphoenolpyruvate carboxykinase, and malic enzyme during growth and sporulation of *Bacillus subtilis*. J. Biol. Chem. **248:** 6062-6070.

DiRusso, C. C., and P. N. Black. 2004. Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. J. Biol. Chem. **279:** 49563-49566.

Edwards, W.M., R. S. Sox, J. P. Cooney, and R. I. Crone. 1970. Active pulmonary tuberculosis with cavitation of forty-one years' duration. Am. Rev. Respir. Dis. **102:** 448-455.

Fang, F. C., S. J. Libby, M. E. Castor, and A. M. Fung. 2005. Isocitrate lyase (*aceA*) is required for *Salmonella* persistence but not for acute lethal infection in mice. Infect. Immun. **73:** 2547-2549.

Feng, Z., N. Caceres, G. Sarath, and R. Barletta. 2002. *Mycobacterium smegmatis* L-alanine dehydrogenase (Ald) is required for proficient utilization of alanine as sole nitrogen source and sustained anaerobic growth. J. Bacteriol. **184:** 5001-5010.

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T.,

Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs, W.R., Jr., Venter, J.C., and C.M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. **184:** 5479-5490.

Fraenkel, D. G. 1996. Glycolysis. pp. 189-198. In: F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E.C.C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology, 2nd ed*. ASM Press, Washington, D.C.

Fraenkel, D. G., and B. L. Horecker. 1965. Fructose 1,6-diphosphatase and acid hexose phosphatase of *Escherichia coli*. J. Bacteriol. **90:** 837-842.

Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. Proc. Natl. Acad. Sci. USA **100:** 7877-7882.

Goldman, D. and M. J. Wagner. 1962. Enzyme systems in the mycobacteria. XIII. Glycine dehydrogenase and the glyoxylic acid cycle. Biochim. Biophys. Acta **65:** 197-306.

Gui, L., A. Sunnarborg, and D. C. LaPorte. 1996. Regulated expression of a repressor protein: FadR activates *iclR*. J. Bacteriol. **178**: 4704-4709.

Gupta, N. K. and B. Vennesland. 1964. Glyoxylate carboligase of *Escherichia coli*: a flavoprotein. J. Biol. Chem. **239:** 3787-3789.

Han, L. and K. Reynolds. 1997. A novel alternative anaplerotic pathway to the glyoxylate cycle in *Streptomycetes*. J. Bacteriol. **179:** 5157-5164.

Hansen, R. W., and J. A. Hayashi. 1962. Glycolate metabolism in *Escherichia coli*. J. Bacteriol. **83:** 679-687.

Hernandez-Pando, R., M. Jeyanathan, G. Mengistu, D. Aguilar, H. Orozco, M. Harboe, G. Rook, and G. Bjune. 2000. Persistence of DNA from *Mycobacterium tuberculosis* in superficially normal lung tissue during latent infection. Lancet **356:** 2133-2138.

Holms, H. 1996. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. FEMS Microbiol. Rev. **19:** 85-116.

Horecker, B. L. 2002. The pentose phosphate pathway. J. Biol. Chem. **277:** 47965-47971.

Horswill, A. R., and J. C. Escalante-Semerena. 1997. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the *prp* locus at 8.5 centisomes, *prpR* encodes a member of the sigma-54 family of activators, and the *prpBCDE* genes constitute an operon. J. Bacteriol. **179:** 928-940.

Hutter, B, and T. Dick. 1998. Increased alanine dehydrogenase activity during dormancy in *Mycobacterium smegmatis*. FEMS Microbiol. Lett. **167:** 7-11.

Idnurm, A., and B. J. Howlett. 2002. Isocitrate lyase is essential for pathogenicity of the fungus *Leptosphaeria maculans* to conola (*Brassica napus*). Eukaryot. Cell **1:** 719-724.

Kanai, K. and E. Kondo. 1974. Chemistry and biology of mycobacteria grown *in vivo.* Jpn. J. Med. Sci. Biol. **27:** 135-160.

Kondo, E., K. Kanai, K. Nishimura, and T. Tsumita. 1970. Analysis of host-originated lipids associated with "in vivo grown tubercle bacilli". Jpn. J. Med. Sci. Biol. **23:** 315-326.

Kondo, E., K. Suzuki, K. Kanai, and T. Yasuda. 1985. Liposomes-mycobacteria incubation systems as a partial model of host-parasite interaction at cell membrane level. Jpn. J. Med. Sci. Biol. **38:** 169-180.

Kornberg, H. L., and H. A. Krebs. 1957. Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. Nature **179:** 988-991.

Kornberg, H. L., and J. R. Sadler. 1960. Microbial oxidation of glycollate via a dicarboxylic acid cycle. Nature **185:** 153-155.

Kornberg, H. L. and A. M. Gotto. 1961. The metabolism of C_2 compounds in micro-organisms. VI. Synthesis of cell constituents from glycolate by *Pseudomonas* sp. Biochem. J. **78:** 69-82.

Kornberg, H. L. 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. Biochem. J. **99:** 1-11.

Kornberg, H. L. 2000. Krebs and his trinity of cycles. Nat. Rev. Mol. Cell. Biol. **1:** 225-227.

Kornberg, H. L. 2003. Memoirs of a biochemical hod carrier. J. Biol. Chem. **278:** 9993-10001.

Korotkova, N., L. Chistoserdova, V. Kuksa, and M. Lidstrom. 2002. Glyoxylate regeneration pathway in the methylotroph *Methylbacterium extorquens* AM1. J. Bacteriol. **184:** 1750-1758.

Krakow, G., and S. Barkulis. 1956. Conversion of glyoxylate to hydroxypyruvate by extracts of *Escherichia coli*. Biochim. Biophys. Acta **21:** 593-594.

Krakow, G., S. Barkulis, and H. Hayashi. 1961. Glyoxylic acid carboligase: an enzyme present in glycolate-grown *Escherichia coli*. J. Bacteriol. **81:** 509-518.

Krebs, H. A. and W. A. Johnson. 1937. The role of citric acid in intermediate metabolism in animal tissues. Enzymologia **4:** 148-156.

Kumari, S, R. Tishel, M. Eisenbach and A. Wolfe. 1995. Cloning, characterization, and functional expression of *acs*, the gene which encodes acetyl-CoA synthetase in *Escherichia coli*. J. Bacteriol. **177:** 2878-2886.

Lillebaek, T., A. Dirksen, I. Baess, B. Strunge, V. O. Thomsen, and A. Andersen. 2002. Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. J. Infect. Dis. **185:** 401-404.

Lippman, F. 1954. Development of the acetylation problem, a personal account. Science **120:** 855-865.

Liu, J. Q., T. Dairi, N. Itoh, M. Kataoka, and S. Shimizu. 2003. A novel enzyme, D-3-hydroxyaspartate aldolase from *Paracoccus denitrificans* 13301: purification, characterization, and gene cloning. Appl. Microbiol. Biotechnol. **62:** 53-60.

Liu, K., Y. Jinzhi, and D. G. Russell. 2003. *pckA*-deficient *Mycobacterium bovis* BCG shows attenuated virulence in mice and in macrophages. Microbiology **149:** 1829-1835.

Lorenz, M. C., and G. R. Fink. 2001. The glyoxylate cycle is required for fungal virulence. Nature **412:** 83-86.

Makato, A., and H. Misuno. 1999. Biochemical evidence that *Escherichia coli hyi* (orf b0508, *gip*) gene encodes hydroxypyruvate isomerase. Biochim. Biophys. Acta **1435:** 153-159.

McKinney, J., W. R. Jacobs, and B. R. Bloom. 1998. Persisting problems in tuberculosis. pp. 51-146. In: R. Krause, J. I. Gallin, and A.S. Fauci (Ed.): *Emerging Infections*. Academic Press, New York, NY.

McKinney, J. D., K. Honer zu Bentrup, E. J. Munoz-Elias, A. Miczak, B. Chen, W. T. Chan, D. Swenson, J. C. Sacchettini, W. R. Jacobs, Jr., and D. G. Russell. 2000. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme *isocitrate lyase*. Nature **406:** 735-738.

Molina, I., M. T. Pellicer, J. Badia, J. Aguilar, and L. Baldoma. 1994. Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. Eur. J. Biochem. **224:** 541-548.

Movahedzadeh, F., S. C. Rison, P. R. Wheeler, S. L. Kendall, T. J. Larson, and N. G. Stoker. 2004. The *Mycobacterium tuberculosis* Rv1099c gene encodes a GlpX-like class II fructose 1,6-bisphosphatase. Microbiology **150:** 3499-3505.

Mukhopadhyay, B. and E. Purwantini. 2000. Pyruvate carboxylase from *Mycobacterium smegmatis*: stabilization, rapid purification, molecular and biochemical characterization and regulation of the cellular level. Biochim. Biophys. Acta **1475:** 191-206.

Muñoz-Elías, E. J. 2005. The role of the glyoxylate cycle in the pathogenesis of *Mycobacterium tuberculosis*. PhD thesis. The Rockefeller University, New York, NY.

Muñoz-Elías, E.J., J. Timm, T. Botha, W. T. Chan, J. E. Gomez, and J. D. McKinney. 2005. Replication dynamics of *Mycobacterium tuberculosis* in chronically infected mice. Infect. Immun. **73:** 546-551.

Muñoz-Elías, E. J., and J. D. McKinney. 2005. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. Nat. Med. **11:** 638-644.

Muñoz-Elías, E.J., and J. D. McKinney. 2006. Carbon metabolism of intracellular bacteria. Cell. Microbiol. **8:** 10-22.

Muñoz-Elías, E. J., A. Upton, J. Cherian, and J. D. McKinney. 2006. Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. Mol. Microbiol. (in press).

Murthy, P. S., M. Sirsi, and T. Ramakrishnan. 1973. Effect of age on the enzymes of tricarboxylic acid and related cycles in *Mycobacterium tuberculosis* H37Rv. Am. Rev. Respir. Dis. **108:** 689-690.

Njau, R. K., C. A. Hendron, and J. W. Hawes. 2000. Novel β-hydroxyacid dehydrogenases in *Escherichia coli* and *Haemophilus influenzae*. J. Biol. Chem. **275:** 38780-38786.

Ornston, L. N., and M. K. Ornston. 1969. Regulation of glyoxylate metabolism in *Escherichia coli* K-12. J Bacteriol **98:** 1098-108.

Overath, P., G. Pauli, and H. Schairer. 1969. Fatty acid degradation in *Escherichia coli.* An inducible acyl-CoA synthetase, the mapping of *old*-mutations, and the isolation of regulatory mutants. Eur. J. Biochem. **7:** 559-574.

Parish, T., and N. Stoker. 2000. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis thyA plcABC* mutant by gene replacement. Microbiology **146:** 1969-1975.

Pauli, G., and P. Overath. 1972. *ato* operon: A highly inducible system for acetoacetate and butyrate degradation in *Escherichia coli.* Eur. J. Biochem. **29:** 553-562.

Peekhaus, N., and T. Conway. 1998. What's for Dinner? Entner-Doudoroff Metabolism in *Escherichia coli*. J. Bacteriol. **180:** 3495-3502.

Pellicer, M. T., Badia, J., Anguilar, J., and L. Baldoma. 1996. *glc* locus of *Escherichia coli*: characterization of genes encoding the subunits of glycolate oxidase and the *glc* regulator protein. J. Bacteriol. **178:** 2051-2059.

Peng, L., M. J. Arauzo-Bravo, and K. Shimizu. 2004. Metabolic flux analysis for a *ppc* mutant *Escherichia coli* based on ^{13}C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. FEMS Microbiol. Lett. **235:** 17-23.

Peters-Wendisch, P., C. Kreutzer, J. Kalinowski, M. Patek, H. Sahm, and B. Eikmanns. 1998. Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. Microbiology **144:** 915-927.

Raper, H.S. 1907. The condensation of acetaldehyde and its relation to the biochemical synthesis of fatty acids. J. Chem. Soc. 1831-1838.

Raynaud, C., C. Guilhot, J. Rauzier, Y. Bordat, V. Pelicic, R. Manganelli, I. Smith, B. Gicquel, and M. Jackson. 2002. Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis.* Mol. Microbiol. **45:** 203-217.

Rittenberg, D., and K. Bloch. 1945. The utilization of acetic acid for the synthesis of fatty acids. J. Biol. Chem. **160:** 417-424.

Russell, D. G. 2001. *Mycobacterium tuberculosis*: here today and here tomorrow. Nat. Rev. Mol. Cell Biol. **2:** 1-9.

Sassetti, C. M. and E. J. Rubin. 2003. Genetic requirements for mycobacterial survival during infection. Proc. Natl. Acad. Sci. USA **100:** 12989-12994.

Sassetti, C. M., D. Boyd, and E. J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. **48:** 77-84.

Sauer, U. and B. J. Eikmanns. 2005. The PEP-pyruvate-oxaloacetate node at the switch point for carbon flux distribution in bacteria. FEMS Microbiol. Rev. **29:** 765-794.

Schoenheimer, R. and D. Rittenberg. 1937. Deuterium as an indicator in the study of of intermediary metabolism. IX. The conversion of the stearic acid into palmitic acid in the organism. J. Biol. Chem. **120:** 155-165.

Schorken, U., and G. Sprenger. 1998. Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses. Biochem. Biophys. Acta **1385:** 229-243.

Segal, W., and H. Bloch. 1956. Biochemical Differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. J. Bacteriol. **72:** 132-141.

Sharma, V., S. Sharma, K. Hoener zu Bentrup, J. D. McKinney, D. G. Russell, W. R. Jacobs, Jr., and J. C. Sacchettini. 2000. Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. Nat. Struct. Biol. **7:** 663-668.

Smith, C. V., C. C. Huang, A. Miczak, D. G. Russell, J. C. Sacchettini, and K. Höner zu Bentrup. 2003. Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis.* J. Biol. Chem. **278:** 1735-1743.

Solomon, P. S., R. C. Lee, T. J. G. Wilson, and R. P. Oliver. 2004. Pathogenicity of *Stagonospora nodorum* requires malate synthase. Mol. Microbiol. **53:** 1065-1073.

Textor, S., V. F. Wendisch, A. A. De Graaf, U. Muller, M. I. Linder, D. Linder, and W. Buckel. 1997. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. Arch. Microbiol. **168:** 428-436.

Tian, J., R. Bryk, M. Itoh, M. Suematsu, and C. Nathan. 2005. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of α-ketoglutarate decarboylase. Proc. Natl. Acad. Sci. **102:** 10670-10675.

Trivedi, O. A., P. Arora, V. Sridharan, R. Tickoo, D. Mohanty, and R. S. Gokhale. 2004. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature **428:** 441-445.

Usha, V., R. Jayaraman, J. C. Toro, S. E. Hoffner, and K. S. Das. 2002. Glycine and alanine dehydrogenase activities are catalyzed by the same protein in *Mycobacterium smegmatis*: upregulation of both activities under microaerophillic adaptation. Can. J. Microbiol. **48:** 7-13.

Vereecke, D., K. Cornelis, W. Timmerman, M. Jaziri, M. V. Montagu, M. Holsters and K. Goethals. 2002. Chromosomal locus that affects the pathogenicity of *Rhodococcus fascians*. J. Bacteriol. **184:** 1112-1120.

Verhees, C. H., Kengen, S. W. M., Tuininga, J. E., Schut, G. J., Adams, M. W. W., de Vos, W. M., and J. van der Oost. 2003. The unique features of glycolytic pathways in Archaea. Biochem. J. **375:** 231-246.

Wall, D.M., Duffy, P.S., DuPont, C., Prescott, J.F., and W.G. Meijer. 2005. Isocitrate lyase activity is required for virulence of the intracellular pathogen *Rhodococcus equi*. Infect. Immun. **73:** 6736-6741.

Wang, Z.-Y., C. R. Thornton, M. J. Kershaw, L. Debao, and N. J. Talbot. 2003. The glyoxylate cycle is required for temporal regulation of virulence by the plant pathogenic fungus *Magnaporthe grisea*. Mol. Microbiol. **47:** 1601-1612.

Wayne, L. G. 1994. Cultivation of M. tuberculosis for research purposes. pp. 73-83. In: B.R. Bloom (Ed.) *Tuberculosis: Pathogenesis, Protection & Control*. ASM Press, Washington, D.C.

Wayne, L. G., and L. G. Hayes. 1996. An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. Infect. Immun. **64:** 2062-2069.

Wayne, L. G., and Lin, K-Y. 1982. Glyoxylate metabolism and adaptation of *Mycobacterium tuberculosis* to survival under anaerobic conditions. Infect. Immun. **37:** 1042-1049.

Wayne, L. G., and C. D. Sohaskey. 2001. Nonreplicating persistence of Mycobacterium tuberculosis. Annu. Rev. Microbiol. **55:** 139-163.

Wheeler, P., and C. Ratledge. 1994. Metabolism of *Mycobacterium tuberculosis.* pp. 353-85. In: B. R. Bloom (Ed.) *Tuberculosis: Pathogenesis, Protection & Control*. ASM Press, Washington, D.C.

Wolfe, A. J. 2005. The acetate switch. Microbiol. Mol. Biol. Rev. **69:** 12-50.

World Health Organization (WHO). 2005. TB-Global Facts, Sept. 2005. PDF newsletter at www.who.int.