Characterization of Gene Expression During Biofilm Development in *Mycobacterium smegmatis*, and Genetic Analysis of a Surface Translocation-Defective Transposon Mutant

by

Amrita Balachandran

Bachelor of Science, Mississippi University for Women, 2005

Submitted to the Graduate Faculty of

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This dissertation was presented

by

Amrita Balachandran

It was defended on

January 19, 2012

and approved by

Jeffrey L. Brodsky, Ph.D., Biological Sciences, University of Pittsburgh

Roger W. Hendrix, Ph.D., Biological Sciences, University of Pittsburgh

Gerard J. Nau, M.D., Ph.D., Microbiology and Molecular Genetics, University of Pittsburgh

Jeffrey G. Lawrence, Ph.D. Biological Sciences, University of Pittsburgh

Dissertation Advisor: Graham F. Hatfull, Ph.D., Biological Sciences, University of Pittsburgh

Characterization of Gene Expression During Biofilm Development in *Mycobacterium smegmatis*, and Genetic Analysis of a Surface Translocation-Defective Transposon Mutant

Amrita Balachandran, Ph.D.

University of Pittsburgh, 2012

Tuberculosis is the leading cause of death due to a single infectious agent, and over one-third of the global population is infected with *Mycobacterium tuberculosis*, the etiologic agent of human tuberculosis. In recent years, the likelihood of biofilm-based infections contributing toward bacterial persistence and increased drug tolerance, has gained some recognition. *M. tuberculosis* and its non-pathogenic fast-growing relative, *M. smegmatis*, have been found to form biofilms that harbor bacteria that are more resistant to anti-tuberculosis agents than free-living cells.

Biofilm formation involves the development of several distinct morphological structures, with associated physiological features. Bacteria growing within biofilms exist as heterogeneous populations, dependent on the distinct micro-environments within the complex community structure. Consequently, gene expression profiles differ between sub-populations of cells, and also during distinct stages of biofilm development. Gene expression in biofilms also differs greatly from the gene expression profiles of planktonically growing cells of the same species.

M. smegmatis biofilms can serve as a model for other mycobacterial biofilms. Transcriptome analyses of *M. smegmatis* biofilms have led to the identification of several genes that were induced in a biofilm-specific pattern.

Iron plays an essential role in mycobacterial growth, metabolism and infection. Taken together with its significance in biofilm development, the detailed profiling of the expression of

iron acquisition genes in mature biofilms would provide insight into the physiological state of the bacterial cells within these structures. Using stable fluorescent reporter constructs, we have provided a detailed profile for the expression of the intramembrane-associated siderophore, mycobactin. Our results suggest that mycobactin biosynthesis is differentially induced in biofilms and in liquid cultures. In mature biofilms, a significant proportion of cells induce mycobactin biosynthesis in spite of the availability of iron-rich conditions. Our analyses also attempt to sort out subsets of cells within the biofilm that differentially induce mycobactin biosynthesis.

In a related study undertaken to understand the relationship between biofilm formation and surface translocation in *M. smegmatis*, we have isolated a transposon mutant that is defective in sliding motility, but proficient in biofilm formation. This mutant suggests that biofilm formation in *M. smegmatis* does not depend on sliding motility.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My first experience in a research lab was all thanks to Graham! He offered me a spot in his lab while I came to Pitt as a summer research student, during the summer before my senior year in college. Having gone to a small liberal arts school for my undergraduate work, Graham's lab was my first exposure to research science, and it was truly a treat. My summer in the Hatfull lab was a deciding factor in my decision to pursue a PhD, and I am very grateful to Graham for giving me the opportunity to continue in his lab as a graduate student. Graham's excitement for science and his ability to put forth the 'big picture' has been an inspiration to me, as I grew as a student and scientist here. Graham you have also been an incredibly kind and supportive advisor, and I am very thankful to you, for everything.

The Hatfull lab has been a great place to work in, and I consider myself pretty lucky to have been a part of this vibrant community. There are so many people to thank, and I'll have to do some of that outside of this page; there are some names that do have to go down on paper:
Anil Ojha was my first scientific mentor, and I am deeply indebted to him for teaching me all the small things, and all things biofilm! He has also been an important part of my dissertation work - he began the biofilm work in the Hatfull lab, and continues to keep in touch with my work. Thank you for your suggestions and help, whenever I've needed it. I also would like to thank Julia van Kessel and Laura Marinelli for taking me under their wings when I first joined the lab, and for being such excellent role models. I still miss having you both around in the lab. Mridula

Anand was a great bay-mate, and her jokes made any day a fun day in lab. I have to especially thank a few people who have come to mean a lot to me over my time in the lab: Kim Payne, one of the most industrious people that I have ever met, has come to become a good friend, and I am thankful to her for always being helpful and always being honest! She, along with Shweta Singh and Pallavi Ghosh are an integral part of my time in the lab, and my experience here would never be as enriched, without them having been a part of it. I also want to thank Bekah Dedrick for being the sweetest person that I meet at work. You make the lab a sunny place to work in!

I am also grateful to my thesis committee, for your support over the years. I have enjoyed my interaction with each of you, and greatly appreciate your role in shaping my dissertation, and in making me think harder about my science. Roger's love for all things phage, has infected us all – and I think that you set a wonderful example of the love for science. Jeff Brodsky has always encouraged me to think harder about my project, and I am very grateful for that. I also appreciate your mentorship while I rotated with you. Jeffrey Lawrence has been the committee member with whom I have perhaps discussed my work the most. Your help with my experimental design and analysis are truly invaluable. Thank you. Jerry Nau is the first committee member that I met in Pittsburgh, when Anil and I went up to talk to him about microarrays, while I was rotating in the Hatfull Lab. Thank you for your helpful discussions about my work, especially at the early stages of the flow cytometry.

My teaching mentors, Melanie Popa and Nancy Kaufmann, have been wonderful guides and good friends to me over my years in the department. I am deeply grateful to both of you for your support and encouragement toward my teaching. I will miss you both! I also thank Alison Slinskey-Legg and Valerie Oke for their guidance and mentoring.

I have saved my thank you's to friends and family for last, but they feature as primary on my list of people to acknowledge; words can never express the full extent of my gratitude to each of the people that I mention here. Graduate school and life in Pittsburgh would never have been the same without my dear friends Priya Kannan, Aarti Sahasranaman and Shruthi Vembar. I could not have done this without the three of you, and I've spent some of my happiest times in Pittsburgh with you. Thank you!!

My friends Rosemary and Buky from college have always been supportive and have stayed abreast of all the seminars and exams and possible graduation timelines, and I thank you for your constant vote of confidence! Sujit, Reshmi, Roshni, Vidya, Mani, Anubhav, Debu - a special thank you to each of you.

But I really would not be writing this as a final note in my dissertation document, if it wasn't for the constant and unconditional love and support of my family. My grandmothers, my uncles and aunts in Kodakkat House and Plano, my in-laws, and my dear brother, Sunil – thank you seems too short a phrase, but it will have to do for now, until I see you all in person.

My race to the finish line would be meaningless if it wasn't for the wonderful opportunity to share this with my parents - Ranjini and Balachandran. I will never ever be able to thank you enough for always being there for me, and for being supportive always.

And finally, most importantly, to my beloved husband and my best friend, Bala - I am thrilled to be able to finish this final lap with you by my side. Thank you.

1.0 INTRODUCTION

Tuberculosis (TB) is the second leading cause of death by an infectious agent, following HIV, and the leading cause of disease by a single infectious agent [1]. An estimated one-third of the total global population is infected with *Mycobacterium tuberculosis* - the causative agent of human pulmonary and extra-pulmonary tuberculosis. In 2010 alone, the WHO recorded 8.5-9.2 million cases of TB, and the disease caused approximately 1.2-1.5 million deaths worldwide [1]. TB poses an even more staggering burden when co-incident with HIV, and is the primary cause of death in the global HIV-positive population [1]. Anti-tuberculosis drugs have been available for chemotherapy for over 50 years [1]; however, these drug regimens are quite extensive, and it takes 6-9 months for complete eradication of an active infection [2]. At a global level, non-compliance to drug intake and regimen guidelines, coupled with socio-economic factors that bring about inconsistencies in chemotherapy have led to an increase in drug-resistance in *M. tuberculosis*. The past decade has seen an additional facet to the burden of TB, with an increasing incidence of multiple and extensively drug resistant tuberculosis, or MDR and XDR TB [1]. Furthermore, cases of TB that are resistant to all available antibiotic treatment options have also been recorded [3].

Latency of disease further contributes to the burden of this disease, and may be linked to the problem of drug resistance in *M. tuberculosis*. Latency refers to the establishment of a

chronic infection with persisting bacterial cells that can activate years past the initial exposure [4].

While most front-line antibiotics used against TB kill *M. tuberculosis* within two weeks of administration, as demonstrated by Jindani *et al* in 2003 [5], there remain cells within the population that survive the effect of the antibiotics. This might be due to the inability of the drug to gain access to these cells, or more likely due to dormancy states stochiastically achieved by a small percentage of the population. Growth, energy metabolism, and biosynthetic genes are downregulated in dormant cells, and at times so is the translational machinery that is required for the maintenance of active targets required for bactericidal antibiotics to act upon [6]. In the absence of the expression of these active targets, bactericidal antibiotics fail to kill these cells, thereby leading to persisters than can then repopulate once the antibiotic dosage is removed [7].

Bacterial biofilms have been implicated in conferring resistance to a number of antibacterial threats, including bacteriophage infection, predation by amoebae, biochemicals, antimicrobial agents, and antibiotics [8]. A large number of bacterial diseases are associated with the formation of biofilms, including several important medical conditions involving bacteria as the etiologic agent. Cystic fibrosis caused by *Pseudomonas aeruginosa*, and urinary tract infections caused by *Escherichia coli* are examples of biofilm-associated diseases. These biofilms are found to harbor persister cells, and when antibiotic dosages are removed or reduced, these persisters can repopulate the site of infection as illustrated in Figure 1 [9].

The prolonged treatment regimen required for TB and the persistent nature of this infection are both consistent with the possibility of biofilms playing a role in this disease. As early as 1956, Darzins and Fahr observed that mycobacterial strains that were lethal or

pathogenic in small animal models such as guinea pigs and mice, showed a cord-like pellicle at the liquid-air interface when growing in liquid media in the absence of surfactants [10].

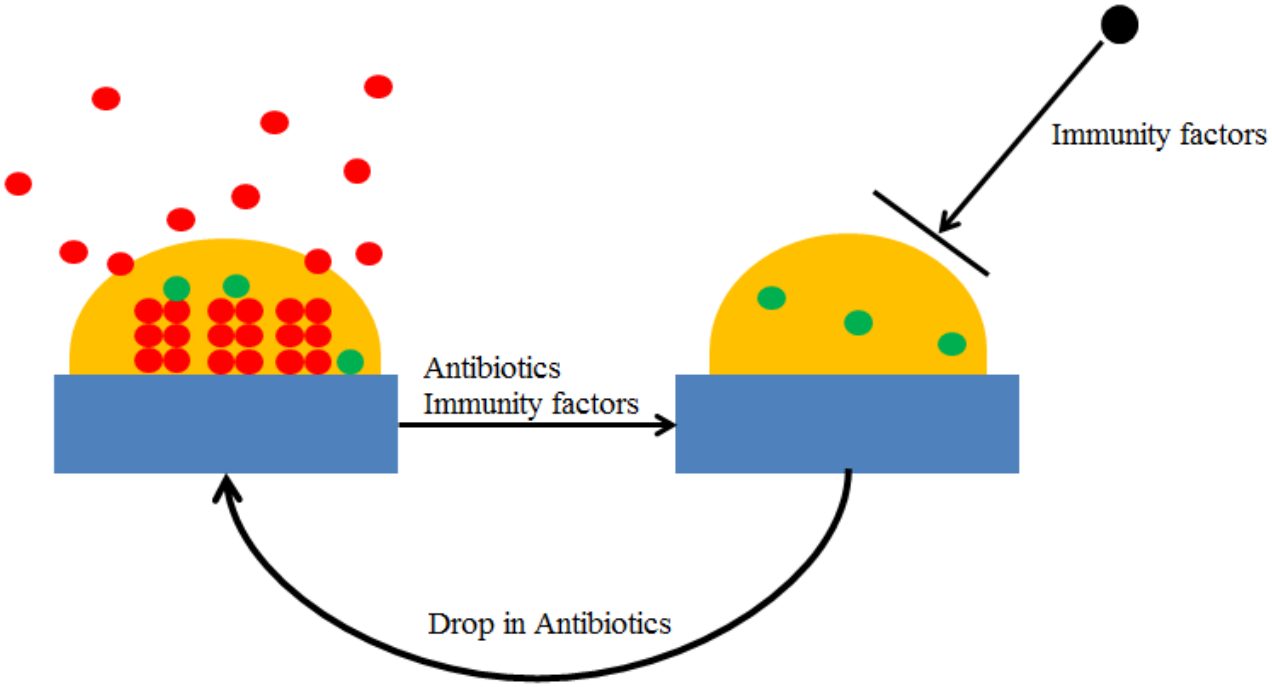


Figure 1. Persisters in Biofilm Infections

Figure 1: Persisters arising within biofilms survive antibiotics and components of the immune system. When levels of antibiotics drop, they can repopulate the biofilm. This figure, adapted from [9], illustrates drug tolerance in bacterial biofilm-based infections.

Studies published by Ojha *et al* in 2008 showed that *M. bovis* BCG and *M. tuberculosis* H37Rv form robust pellicle-like biofilms under specific conditions of media and growth environment. These biofilms are genetically and phenotypically distinct from their planktonic counterparts. Also, *M. tuberculosis* biofilms are observed to harbor drug tolerant persister cells that can withstand anti-tuberculosis agents at much higher concentrations than the minimum inhibitory concentration (MIC) for planktonic cells [11].

The link between biofilms and TB infections has not been explored in any significant depth by the field of clinical TB. Recent studies on heightened drug tolerance in *M. tuberculosis* biofilms, such as the aforementioned study, and the shared characteristics displayed by tuberculosis and infections caused by the biofilms of other pathogens, suggest a likely link between the two. The relevance of biofilms in *M. tuberculosis* infections is also exemplified by observations made by Lenaerts *et al* in 2007, in a study investigating the location of persister cells in a guinea pig model of infection. Following six weeks of chemotherapy, almost all bacilli in the primary and secondary lesions of the infection were cleared; however persisting cells were observed as an acellular rim around the primary granuloma. These drug-tolerant persisters were found as clusters in the extracellular environment of the granuloma - a feature that is reminiscent of bacterial biofilms [12]. Moreover, another recent survey of global isolates of *M. tuberculosis*, undertaken to identify the genes involved in biofilm formation, also observed that these clinical isolates all possessed the ability to form pellicle-like biofilms [13]. Taken together, the recent surge of work on *M. tuberculosis* biofilms, as well as clinical observations of TB infections, indicate a highly likely link between *M. tuberculosis* biofilms and infection. To better address the continued problem of persisiting TB infections, especially in the face of the global threat posed by drug-resistant tuberculosis, modern-day drugs must target and disrupt biofilm development

pathways as a means to disarm *M. tuberculosis*. To do this, it is important to understand the genetic machinery that supports biofilm formation in the mycobacteria, and dissect gene expression profiles within these biofilms to identify optimal targets for chemotherapy.

1.1 BACTERIAL BIOFILMS

Over the past few decades, the concept of bacterial communities and their prevalence and significance in the environment, in industry, and in clinical applications, have been widely accepted. Biofilms exhibit great distinction from laboratory-grown planktonic cells; their architectural and metabolic complexity and community-based lifestyle liken them to differentiated tissues found in multi-cellular organisms. This section of the introductory chapter provides a brief general overview of bacterial biofilms.

1.1.1 An Historical Perspective

Bacteria were considered to be free-living entities until the 1970s, when Costerton *et al* first reviewed bacteria "sticking" together and to surfaces by means of secreted polysaccharides called "glycocalyx" [14]. The first reports of the observation of polysaccharide material encasing mixed populations of bacteria and serving as adhesive to a surface, come from studies in the 1960s on *Streptococcus mutans* and *Streptococcus salivarius*. These bacteria are the causative agents of human dental plaque - a classic example of bacterial biofilms [14]. Pioneering studies on biofilm structure were performed in the 1980s and 1990s, and revealed the complex

architecture of bacterial biofilms [8]. One such investigation involved the analysis of liquid flow within biofilms, showcasing the presence of channels that facilitate nutrient circulation within heterogeneous biofilm structures [15]. Another study introduced scanning confocal laser microscopy (SCLM) to image sections through hydrated biofilms of *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, and *Vibrio parahaemolyticus*. These studies showed significant distinction in the biofilm architecture of these species, thus also highlighting the diversity of bacterial communities [16].

1.1.2 Biofilm Development

Biofilm formation involves distinct stages of development, involving initial attachment, spreading, maturation, and matrix synthesis [17], as illustrated in Figure 2. Non-sessile cells first attach to a solid surface, and then form a monolayer or microcolonies. These microcolonies then differentiate into mature biofilms that are enriched with extracellular matrix material [8]. The finer molecular events that govern these stages of biofilm formation and development differ between Gram-positive and Gram-negative bacterial biofilms [17].

1.1.2.1 Initial Attachment

Several studies implicate the role of environmental signals as cues for the bacterial switch from a free-living planktonic existence to sessile biofilms, although the exact cues themselves differ by organism. Some of these cues include growth media, iron levels, osmolarity, oxygen, pH, and temperature [17]. While *E.coli* strain K-12 requires amino acid supplementation of minimal media to form biofilms [18], the *E.coli* O517:H7 strain forms biofilms specifically in minimal

media [19]. Other Gram-negative species such as *P. fluorescens* and *P. aeruginosa* are less finicky about growth conditions, typically forming biofilms under a number of different nutrient conditions [20].

In Gram-positive species such as *Staphylococcus epidermidis* and *S. aureus*, surface attachment begins with cell-surface interaction facilitated by various types of surface proteins [17, 21].

1.1.2.2 Surface Attachment

A large volume of research has focused on understanding the molecular details of surface attachment in Gram-negative bacteria such as *P. aeruginosa*, *P. fluorescens*, *E. coli* and *V. cholerae*. These studies reveal the essentiality of surface structures such as pili (Type I or IV), flagella, and LPS, in mediating attachment to surfaces. Organisms differ in their overall modes of attachment, and also in the modes of attachments to different surfaces [17]. In Gram-positive bacteria such as *S. epidermidis* and *S. aureus*, initial surface attachment is followed by cell-cell attachment, leading to bacterial aggregation and the formation of microcolonies. In both these species, cell-cell attachment is facilitated by the polysaccharide intercellular adhesion (PIA). This observation is illustrated by several lines of evidence in *S. epidermidis*; the intercellular adhesion (*ica*) locus that is required for the synthesis of PIA is also present in *S. aureus*, and is found to be required for biofilm formation in this organism as well [22-26].

1.1.2.3 Maturation

The maturation of most well-studied Gram-negative bacterial biofilms, including those produced by *P. aeruginosa* and *E. coli*, involves the increased production of extracellular polymeric

substances (EPS) as the matrix material. This process is accompanied by the acquisition of antibiotic resistance [17]. Most well-characterized Gram-positive bacterial biofilms also produce EPS. *S. epidermidis* has been shown to do so in response to iron starvation and overall nutrient limitation [27]. The EPS material may be polysaccharides, as well as proteins, glycolipids, glycoproteins, and extrachromosomal DNA (e-DNA) [28, 29]. Recent research has also identified *ica* operon-independent biofilm development in *S. epidermidis* and *S. aureus*, instead involving protein adhesions [28]. *Bacillus subtilis* is another important Gram-positive bacterial species that has been studied extensively with regard to biofilms. The extracellular matrix in *B. subtilis* biofilms is largely composed of an exopolysaccharide encoded by the *epsA-O* operon. The matrix also contains a major protein component, TasA, which is encoded by the *yqxM-sipW-tasA* operon [30]. TasA provides structural integrity to the biofilm by forming amyloid fibers that hold cells together [31].

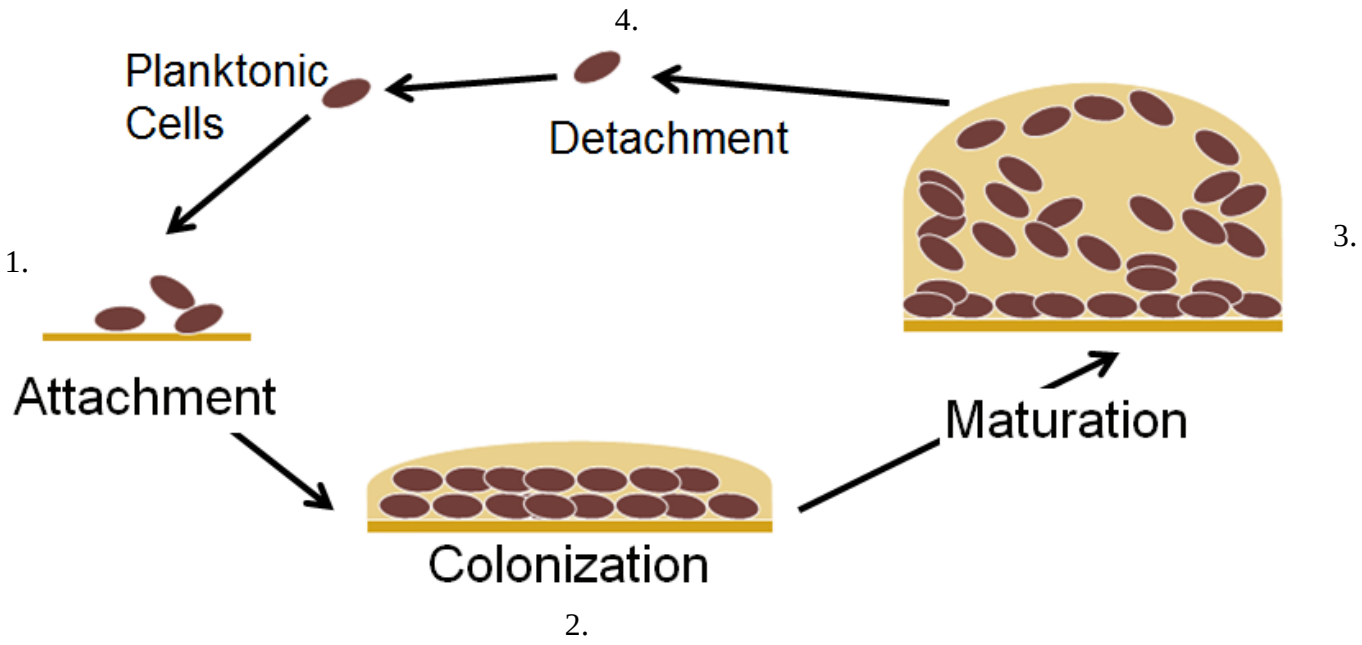


Figure 2. Stages in Biofilm Development

Figure 2: Biofilm formation can be likened to a developmental process, wherein cells respond to various environmental and developmental cues to 1. attach to a surface, 2. colonize the surface, 3. form a mature biofilm containing extracellular matrix material, 4. based on the type of biofilm, respond to detachment signals and revert back to planktonic growth. This schematic, adapted from Parsek and Singh, 2003 [32], integrates species-specific information on the stages of biofilm development.

1.1.2.4 Gene Expression in Biofilms

Bacterial cells within biofilms exist as heterogeneous populations, based on their distinct micro-environments within the complex community structure. Consequently, genes expressed within the biofilm differ between sub-populations of cells, and also during distinct stages of biofilm development [17]. Extensive transcriptome studies on biofilms of *P. aeruginosa* [33-35], *E. coli* [36-38] , *B. subtilis* [39], and *S. aureus* [40], amongst others, confirm that the process of biofilm development involves widespread changes in the gene expression profiles of these organisms, as compared to planktonically grown cultures. Studies such as these also provide evidence for spatial and temporal distinction in gene expression patterns [41].

In addition to transcriptome analyses, transcriptional reporter fusions have also been used to demonstrate spatio-temporal differences in gene expression within biofilms. One such study that was performed on biofilms of *B. subtilis* exquisitely reveals the differential gene expression patterns of distinct types of cells - matrix-producing cells, motile cells, or sporulating cells - within the biofilm structure. The size of these sub-populations as well as the localization of these cells by functional types is temporally dynamic during development [42, 43].

1.2 MYCOBACTERIAL BIOFILMS

A notable number of the non-pathogenic and opportunistically pathogenic mycobacteria, including *M. smegmatis* [44], *M. avium* [45, 46], *M. marinum* [47], and *M. fortuitum* [48], have been reported to form biofilms. These communities are observed to be more resistant to anti-tuberculosis agents than planktonically growing bacteria [44, 49]. *M. smegmatis* in particular has

been shown to form a pellicle-like biofilm at the liquid-air interface, with distinct developmental changes occurring over a period of 3-7 days. The extracellular matrix in *M. smegmatis* biofilms consists of C_{56}-C_{68} fatty acids, generated by a GroEL1-mediated switch in mycolic acid biosynthesis [49].

More recently, *M. tuberculosis* has been shown to form biofilms under specific growth and environmental conditions, *in vitro*. These biofilms have also been demonstrated to contain free mycolic acids and drug-tolerant bacteria [50]. Although the link between *M. tuberculosis* infections and biofilm formation has not yet been conclusively established, as discussed earlier, the characteristic persistent nature of *M. tuberculosis* infections suggests the possible role of biofilms. Moreover, *M. marinum* biofilms as well as biofilms of *M. chelonae* and *M. ulcerans* - etiologic agents of chronic skin cancers - have been implicated in the infections caused by these organisms [47, 51-54].

1.2.1 The Extracellular Matrix in *M. smegmatis* and *M. tuberculosis* Biofilms

Although much is known about the composition of biofilms in several other pathogenic bacteria that are known to form biofilms, such as *P. aeruginosa* or *S. aureus*, the field of mycobacterial biofilms is still relatively nascent. Recent work is beginning to reveal more details about the complex biofilms formed by these bacteria. Some of the earlier investigations of biofilm requisites in the mycobacteria, specifically in *M. smegmatis* and *M. marinum*, demonstrated the role of lipid-based compounds in biofilm formation. The Kolter laboratory established the importance of glycopeptidolipids (GPLs) in biofilm formation [55]. The Liu laboratory reported that the small DNA-binding protein, Lsr2, mediates the biosynthesis of compounds containing

mycolic acids, known as mycolyl-diacylglycerols (MDAGs). Disruption of the *lsr2* gene abrogates biofilm formation, and generates smooth colonies on plates [56]. The Liu lab also demonstrated the requirement of lipooligosaccharides (LOSs), which are antigenic glycolipids found in a few of the mycobacteria. Loss of LOS biosynthesis compromises biofilm formation and sliding motility in *M. marinum* [57] .

A distinct feature of mycobacterial biofilms is that they are composed of a lipid-rich extracellular matrix, in contrast to the exopolysaccharide-based matrices seen in several other biofilm-forming bacteria. As mentioned earlier, *M. smegmatis* and *M. tuberculosis* biofilms were found to contain large amounts of free mycolic acids, and in *M. smegmatis*, the biofilm-specific GroEL1 chaperone interacts with the type II fatty acid synthase system to regulate mycolic acid synthesis [49, 58]. Recent work that continued to study these biofilms elucidated that the release of free mycolic acids (FM) as matrix material is facilitated by a serine carboxyesterase encoded by *MSMEG_1529*. This enzyme catalyzes the hydrolysis of a precursor lipid, trehalose di-mycolate (TDM), into FM [59]. Figure 3 presents a schematic model for biofilm formation and development in *M. smegmatis* [59, 60].

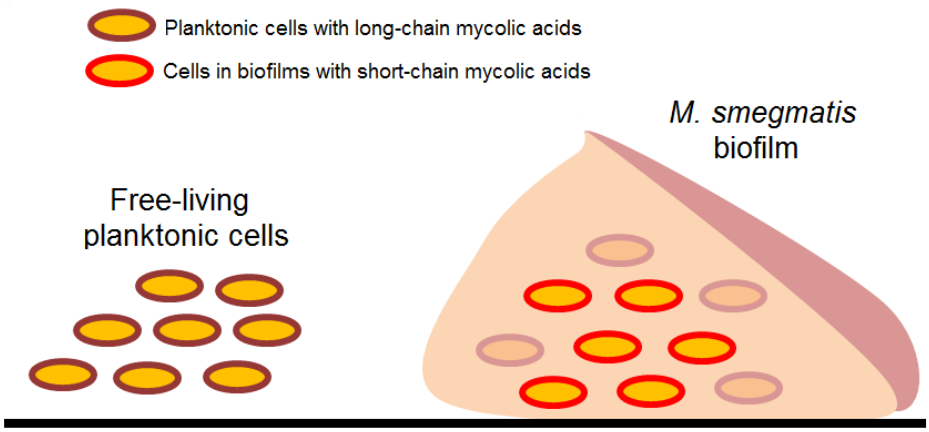


Figure 3. Biofilm Formation in *M. smegmatis*

Figure 3: Model for *M. smegmatis* biofilm formation, adapted from Zambrano and Kolter, 2005 [60], based on Ojha *et al*, 2005 [49]. When planktonic cells commit to biofilms, they modify their mycolic acids from the long chain to a shorter chain variant. A portion of these short-chain mycolates are secreted to form the extracellular biofilm matrix.

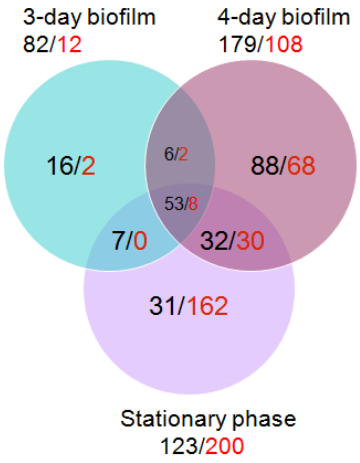


Figure 4. Differential Gene Expression in M. smegmatis Biofilms and Planktonic Cells

Figure 4: Gene expression patterns revealed by microarray analyses of three-day biofilms, four-day biofilms, and stationary phase planktonic cells. Numbers in black font represent up-regulated genes; numbers in red font represent down-regulated genes, in each case. Figure recreated from Ojha and Hatfull, 2007 [61].

1.2.2 Gene expression in *M. smegmatis* biofilms

In a study undertaken to understand the transcription profile of *M. smegmatis* biofilms, Ojha and Hatfull, in their 2007 report, performed microarray analyses to compare the gene expression profile of planktonic exponential phase cultures of *M. smegmatis* to three-day and four-day biofilms [61]. This study also compared expression profiles of planktonic exponential phase cultures to planktonic stationary phase cultures. Significant subsets of genes were up-regulated in a biofilm-specific pattern. Similarly, there were subsets of genes that were down-regulated,

specifically in biofilms. Figure 4 illustrates the number of genes up-regulated or down-regulated in each condition, relative to their expression in exponential phase planktonic cultures [61].

In spite of the presence of 2 μM iron in the initial biofilm growth medium, a notable number of iron-acquisition genes were induced in a biofilm-specific pattern [61]. These genes will be discussed further in Chapter 2. The addition of 2 μM iron to biofilm growth media is the standard practice for *M. smegmatis* biofilms [49]. The global gene expression profile observed during growth as biofilms also includes other functional genes that are up-regulated in a biofilm-specific manner [61]. The most striking amongst these genes are - *aceA*, encoding an isocitrate lyase, and the transcriptional regulators *lexA*, *uvrD*, u*vrC*, *recA* and *recX*. *aceA* is the *M. smegmatis* homolog for the gene encoding isocitrate lyase (*icl*) in *M. tuberculosis*. Icl is an important player in the glyoxylate cycle and is essential for growth on certain carbon sources such as acetate [62]. In the context of *aceA* up-regulation in biofilms, it has been reported that in *M. tuberculosis*, the enzymatic activity of the glyoxylate cycle is increased in low oxygen environments [63]. Biofilms consist of gradients of oxygen tensions, and there are likely to be several pockets of cells that are exposed to low oxygen within their micro-environment [15].

Another notable set of genes that were up-regulated in biofilms and in stationary stage cells were stress response genes; these include the universal stress response or SOS genes, *lexA* and *radA* [61].

1.3 BACTERIAL SURFACE TRANSLOCATION

In 1972, Jorgen Hendrichsen published a survey of bacterial motility on a surface, testing hundreds of strains that represented forty species of bacteria. His detailed study identified six different forms of bacterial surface motility: swimming, swarming, gliding, twitching, darting and sliding [64, 65]. Since then, the mechanisms of swimming, swarming, twitching and gliding motility have been characterized in some detail in model organisms that exhibit these forms of surface translocation.

Table 1 lists the known motive forces that are responsible for generating the different forms of bacterial motility. Swimming and swarming motility, as exhibited by bacteria such as *E. coli* or *Salmonella typhimurium* [66], and by *Proteus mirabilis* [67], respectively, have been found to be dependent on flagella. Twitching motility, characterized extensively in *P. aeruginosa*, is flagella-independent, and instead dependent on Type IV pili [68]. While some forms of gliding motility, such as "social gliding" seen in *Myxococcus xanthus* require Type IV pili, the mechanism of other forms such as "adventurous gliding" in *M. xanthus*, and gliding seen in filamentous cyanobacteria, are still uncharacterized [69]. Social gliding refers to gliding motility that is dependent on pili, and may be considered to be similar in mechanism to twitching motility. Adventurous gliding refers to gliding on a surface in the absence of pili. The mechanism responsible for this form of gliding is still unknown; however, it is likely to require outer-membrane lipoproteins that could potentially act as pumps that allow for bacterial propulsion on a surface [65]. Sliding motility, as exhibited in *M. smegmatis* and *M. avium* [70], as well as spreading motility, as seen in *Serratia marcescens* [71], are flagella-independent

passive forms of surface translocation. Bacterial surface motility plays an important role in bacterial surface colonization abilities, as described in Table 1 [65].

Table 1. Surface Translocation in Bacteria

Type of Motility	Motive Force	Function	Example(s)
Swimming/ Swarming	Flagella; cells move individually	Surface colonization	*Escherichia coli, Proteus mirabilis*
Gliding	Generated by individual cells; move together	Surface colonization	*Myxococcus xanthus, Anabaena variabilis*
Twitching	Generated by individual cells; move separately	Surface colonization, biofilm formation, phage infection, conjugation	*Pseudomonas aeruginosa*
Darting	Generated by cell community; tension forces	Surface colonization	*Vibrio cholerae*
Sliding	Generated by cell community; force by expansion	Surface colonization	*Mycobacterium smegmatis*

1.3.1 Sliding Motility in *M. smegmatis*

M. smegmatis has been reported to exhibit a spreading phenotype on moist, solid surfaces, by means of a sliding mechanism which accompanies expansive forces generated by division and growth on a surface [70]. This form of sliding motility has been found to require a class of cell wall-associated proteins known as glycopeptidolipids (GPLs) [72]. Previously isolated mutants for sliding motility have mapped to genes involved in GPL biosynthesis; these mutants have also been reported to be defective in biofilm formation [55].

1.4 PREVIEW TO CHAPTERS 2 AND 3

The Ojha and Hatfull study [61] on overall gene expression in *M. smegmatis* biofilms, identified a specific class of genes - iron acquisition genes - as being significantly induced during biofilm development. This study reported the up-regulation of twenty-nine iron-responsive genes, composing a total of nine operons, in biofilms and stationary-phase cultures [61].

Iron plays an essential role in the growth and pathogenesis of the mycobacteria, and these organisms have evolved different classes of molecules, known as siderophores, to secure iron from their environment. The genes involved in iron acquisition are tightly regulated in the mycobacteria, and are up-regulated under conditions of iron limitation [73]. The importance of many of these genes in *M. tuberculosis* infections has already been established. Approximately 150 genes in *M. tuberculosis* respond to changes in the concentration of iron in the environment, and one-third of these genes are controlled by the iron-dependent repressor or IdeR [74, 75].

IdeR is a DNA-binding protein that belongs to the DtxR family of proteins; it is an iron-responsive protein that controls the transcription of genes involved in iron acquisition, storage, and macrophage survival, in the case of *M. tuberculosis* [76]. Siderophore biosynthesis in *M. smegmatis* is also under the control of the *M. smegmatis* IdeR homolog [77]. Ojha and Hatfull also established an essential role for iron in *M. smegmatis* biofilm development, and revealed a correlation between iron and the fatty acid component of the extracellular matrix [61].

A portion of my dissertation work is derived from observations made in the Ojha and Hatfull, 2007 study, and specifically focuses on the expression of iron acquisition genes in *M. smegmatis* biofilms. My dissertation research primarily sought to study spatial and temporal patterns of gene expression in biofilms, using fluorescent reporters. The strong induction of iron acquisition genes during biofilm development made this group of genes ideal candidates to profile during biofilm development. Chapter 2 illustrates a detailed analysis of the expression profile of the mycobactin siderophore biosynthesis gene, *MSMEG_4515* or *mbtB*, in biofilms of *M. smegmatis*. We provide microscopy-based, as well as quantitative flow cytometry-based data to provide a comprehensive profile of the induction of mycobactin biosynthesis during biofilm development.

A second focus of my dissertation research developed during the search for solutions to address technical difficulties experienced during optimization of fluorescence microscopy for the aforementioned gene expression studies. As an alternative to using *M. smegmatis* pellicle-like biofilms that form at the liquid-air interface, we considered using samples that were inoculated and grown on semi-solid plates, for ease of technical manipulation for confocal microscopy. As introduced in Section 1.3, *M. smegmatis* exhibits a passive form of surface translocation when grown on semi-solid nutrient surfaces. Previous work on *M. smegmatis* surface translocation has

remained limited to the identification of key genes involved in facilitating this process. Mutants lacking the ability to surface translocate were also found to be deficient in biofilm formation. Before proceeding to explore a solid-surface translocation assay as a surrogate for traditional biofilm culture methods, we chose to further investigate the correlation between surface translocation and biofilm formation. Chapter 3 describes the results of the transposon mutagenesis screen that was undertaken to test this link. We identified a transposon insertion mutant that is severely defective in sliding on a moist agarose surface, and performed genetic analysis to try and identify the gene disruption that was responsible for this defect. This chapter also discusses preliminary bioinformatic analyses on the protein product of the gene that when disrupted is predicted to be responsible for the translocation defect.

2.0 CHARACTERIZATION OF THE EXPRESSION OF THE MYCOBACTIN BIOSYNTHESIS GENE, *MSMEG_4515*, IN *M. SMEGMATIS* BIOFILMS

2.1 INTRODUCTION

Iron is an important co-factor in many bacteria, and several pathogenic bacteria face iron starvation in their host environments, especially if these environments are within mammalian tissues, where iron is sequestered away by host proteins [78]. Iron is also extremely important for biofilm development in several pathogenic bacteria. Banin *et al* have illustrated the importance of iron acquisition molecules - siderophores - for the formation of *P. aeruginosa* biofilms, as illustrated by experiments that measured biofilm formation under varying conditions of iron, using mutants in the two primary siderophore genes [79]. Other studies have demonstrated that iron starvation prevents biofilm formation, and instead promotes twitching motility which leads to a non-sessile lifestyle. On the other hand, sufficient iron in the media acts as a signaling molecule for biofilm development [80, 81]. Similarly, biofilm formation by urinary tract isolates of *E. coli* is abrogated by the addition of Zn(II) and Co(II), on account of their higher affinity for the Fur protein - the master regulator for iron uptake [82]. *M. smegmatis* as well as *M. tuberculosis* biofilms require iron to grow. Efficient growth of *M. smegmatis*

biofilms require greater than 1 μM iron in the surrounding media [61], and *M. tuberculosis* biofilms require iron levels greater than 2 μM [11].

The *M. tuberculosis* genome encodes at least 40 enzymes that require iron as a co-factor, including enzymes that function in the electron transport chain and DNA synthesis; iron is an absolute requisite for the growth of these bacteria [83]. DNA microarray analysis of the *M. tuberculosis* transcription profile in wild-type strains, strains mutated for *ideR*, and mutant strains complemented for *ideR*, revealed a significant number of genes that are regulated by iron. These genes encode a variety of functions, including siderophore synthesis, storage of iron, virulence, transcriptional regulation, lipid metabolism, amongst others that are regulated by iron [74].

Ojha and Hatfull were the first to report a comprehensive gene expression profile for *M. smegmatis* biofilms, in 2007 [61]. Of approximately 100 genes found to be up-regulated specifically in three-day or four-day stages of biofilm growth, indicating initial attachment and maturation, respectively, a sizable subset were involved in iron acquisition in the mycobacteria. This subset included genes involved in the biosynthesis of siderophores - both the cell-associated mycobactin, and the extracellular exochelin, as well as genes encoding for iron uptake proteins, *fxuA*, *fxuB*, *fxuC*, and *fxbA* [61, 84]. The proteins encoded by these genes are homologous to FepG, FepC and FepD, the iron permease proteins in *E. coli* [85]. Table 2 provides a detailed list of the 29 putative iron acquisition genes that are induced during *M. smegmatis* biofilm formation, in media containing 2 μM iron [61].

Since no prior work had been performed to investigate the spatial and detailed temporal expression of individual sets of genes within *M. smegmatis* biofilms, we decided to pursue this question as a derivative of the microarray profile published in Ojha and Hatfull [61]. The prominent expression of iron acquisition genes during biofilm growth, and the well-characterized

importance of iron in mycobacterial growth made this group of genes an ideal subset to profile within *M. smegmatis* biofilms. Over the course of my work in this area, I chose to concentrate specifically on the M. smegmatis gene *MSMEG_4515*, which is the *M. tuberculosis mbtB* homolog, involved in mycobactin biosynthesis. The reasons for this choice are three-fold:

First, although *M. smegmatis* expresses both classes of siderophores - mycobactins as well as exochelins, *M. tuberculosis* only employs the mycobactin homologs [73]. Moreover, an *M. tuberculosis mbtB* mutant lacking a portion of the *mbtB* gene was shown to be avirulent in macrophage infections. This observation conclusively illustrated that this siderophore, and more importantly, iron acquisition, plays an important role in *M. tuberculosis* infections [86].

Second, the biofilm transcription profile reported by Ojha and Hatfull, shows greater induction of the mycobactin biosynthesis genes than the exochelin genes, at the three-day and four-day stages of biofilm development [61].

Third, since this was our first use of reporter fusions in studying gene expression in *M. smegmatis* biofilms, for technical reasons and imaging considerations, we hypothesized that a well-induced promoter would have a better likelihood of being detected. Our work with P_{hsp60}, a strong mycobacterial promoter, provided significantly lower fluorescence signals than a similar fusion carried on an extrachromosomal multicopy plasmid. This observation held true in planktonically grown cells as well as in biofilms.

Table 2. Putative iron-acquisition genes induced during *M. smegmatis* biofilm formation

List of iron-responsive genes induced during biofilm development at 2 μM Fe^{2+}, as reported by Ojha and Hatful [61]. Microarray analyses used exponential phase planktonic cultures grown in 2 μM Fe^{2+} as the standard. Functional annotations and *M. tuberculosis* homologs were compiled from the JCVI Comprehensive Microbial Resource database [109].

GENE(S)	Function	Homolog in *M. tuberculosis* strain H37Rv
MSMEG_0011-0013; MSMEG_0015	Exochelin uptake	Rv2895c
MSMEG_0014;MSMEG_0016-0019	Exochelin biosynthesis	None; RV1406,Rv2377c,Rv0194,Rv0101
MSMEG_4508-4516	Mycobactin biosynthesis	Rv2377c-Rv2384
MSMEG_2130-2132	Mycobactin acylation	Rv1346-Rv1344
MSMEG_6062-6064	Iron ABC Transporter	None
MSMEG_6553-6554	Metal ABC Transporter (putative)	Rv1349-Rv1348
MSMEG_6453	FurA-family regulator	Rv1707
MSMEG_5039	Iron utilization protein (putative)	Rv1348

2.2 CONSTRUCTION OF MODULAR FLUORESCENT REPORTER VECTORS

2.2.1 The rationale for using integrative fluorescent reporter constructs

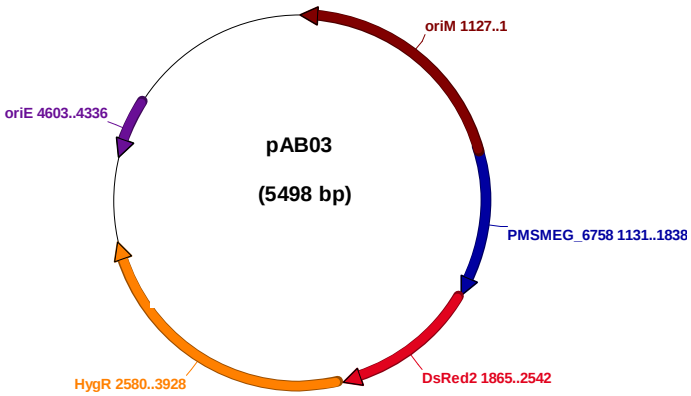
To investigate the expression patterns of individual genes within *M. smegmatis* biofilms, the approach adopted was to construct promoter-reporter fusions, and to use fluorescence microscopy to efficiently monitor the differential expression of genes within the biofilm structure.

Although multi-copy plasmids would provide significant amplification of even low fluorescent signals from weaker promoters, they also present the increased likelihood for generating artificial expression levels due to the presence of multiple copies of the reporter fusions [87]. Furthermore, to maintain selection pressure for the retention of extra-chromosomal plasmids within cells, especially for extended durations such as in the three-seven day biofilm assays, biofilms would have to be grown in the presence of antibiotics. Standard biofilm assay conditions avoid the use of antibiotics. We have found that the addition of antibiotics alters/delays the biofilm development profile, even when the strain carries the appropriate resistance cassette (data not shown).

In our earliest efforts to compare the effectiveness of multi-copy and integration-proficient vector systems for promoter-reporter fusion studies, we used a multi-copy pJL37-based vector system to generate a P_{MSMEG_6758}-DsRed2 fusion construct (Figure 5A), and a P_{MSMEG_0615}-eGFP fusion integrated at the L5 integration site as the single-copy sample.

Additionally, for these initial experiments, instead of imaging promoter expression in biofilms or planktonic cells grown in liquid media, we used colonies growing on plates containing biofilm media reinforced with a small amount of agarose. The core of the colony was considered to be similar to stationary phase cells, while the leading edges were representative of newer, exponentially growing cells. Ojha and Hatfull showed that *M. smegmatis* stationary phase cells exhibit similar gene expression profiles as biofilm samples [61]. In light of this observation, we used colonies while developing an optimal reporter system, to allow for a quicker and easier assay for fluorescence than biofilms grown at a liquid-air interface. *MSMEG_0615* encodes a putative AAA family ATPase, and shows an approximately eight-fold increase in expression in stationary phase planktonic cultures and biofilms, as compared to exponential phase planktonic cultures [61]. *MSMEG_6758* encodes a putative membrane transport protein; compared to cells in exponential phase of planktonic growth, this gene is approximately 32-fold down-regulated in stationary phase planktonic cells and biofilms [61]. In the colony assay, we expected to see P_{MSMEG_0615}-eGFP expressed at the core of the colony, and P_{MSMEG_6758}-DsRed2 at the growing edges. Figure 5B shows homogenous expression of P_{MSMEG_0615}-eGFP at the center of the colony; however, only certain patches at the edge of the colony expressed P_{MSMEG_6758}-DsRed2. Results obtained from the integrative approach were more robust and repeatable than those obtained from experiments using pAB03. We therefore chose to develop a more dependable system for promoter expression analysis by constructing a modular set of integration-proficient reporter constructs.

A.



B.

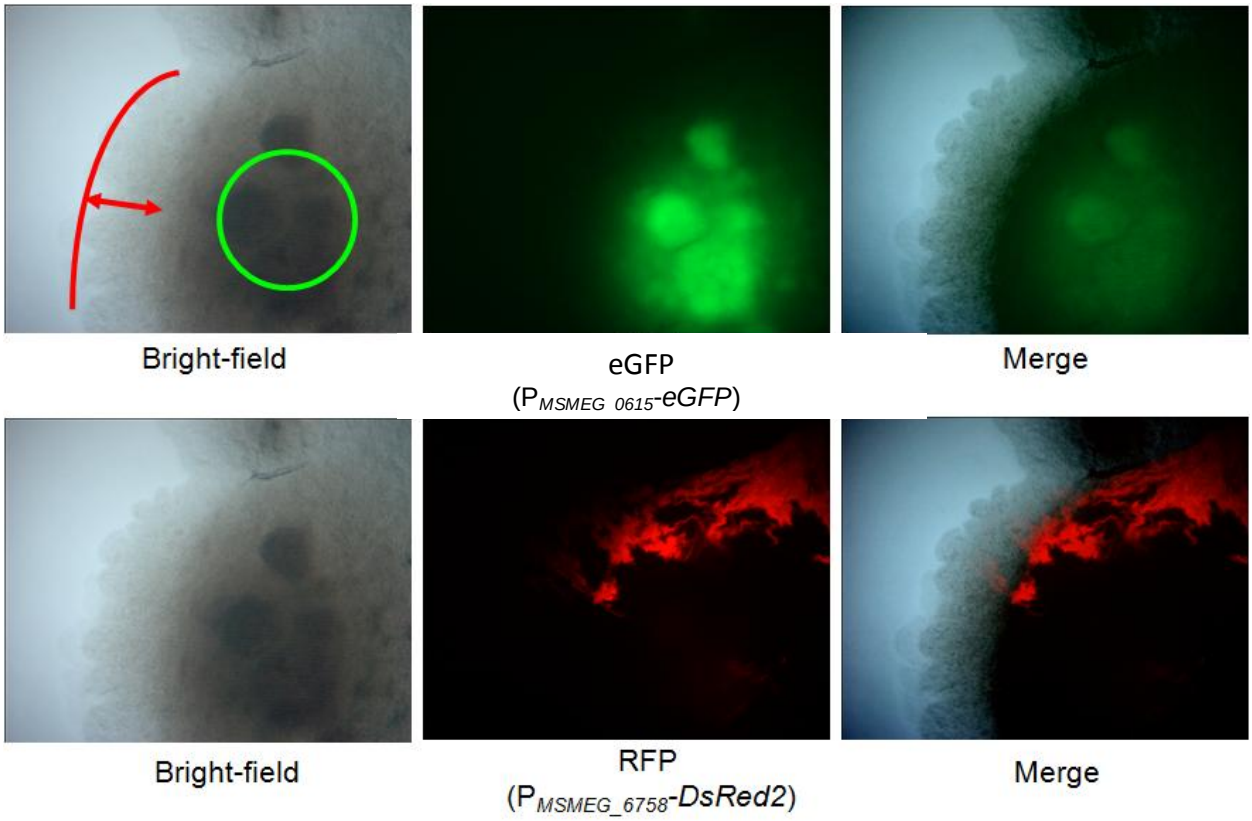


Figure 5. Extrachromosomal vs. Integration-Proficient Fusion Vectors

Figure 5: A. Schematic representation of pAB03, a pJL37-based replicative plasmid carrying a P_{MSMEG_6758}-DsRed2 fusion. Graphical representation of the plasmid was generated using MacPlasmap. **B.** Colony assay performed with a strain co-expressing P_{MSMEG_0615}-eGFP as a single-copy fusion integrated at the L5 integration site, and pAB03. The illustrative green and red lines in the first panel represent expected zones of eGFP and DsRed2 expression, respectively.

2.2.2 Integration-proficient modular fluorescent reporter vectors

The first step toward analyzing gene expression using single-copy integrated promoter fusions was to create a modular set of integration-proficient fluorescent reporter vectors. These reporter constructs include mycobacteriophage integration cassettes that allow for the promoter-reporter fusions to be integrated into the *M. smegmatis* chromosome at specific sites of recombination. We have developed two sets of reporter constructs, each set derived from a distinct integration-proficient plasmid. The integration cassettes used in developing these two distinct sets of reporter constructs come from mycobacteriophage Tweety and mycobacteriophage Giles, and have been previously described [88, 89]. Each of these vectors carry the following common features: an origin of replication in *E. coli* - *oriE*, an antibiotic resistance cassette - Kan^R cassette in pTTP1B-derived plasmids or Hyg^R in pGH1000A-derived plasmids, the respective integrase gene and *attP* site, *eGFP*, *eCFP*, or *eYFP* reporter gene, fused to the *hsp60* promoter. The *hsp60* promoter can be removed by restriction enzyme digestion at an upstream restriction site which forms part of a multiple-cloning site, and *Nde*I at the promoter-reporter junction (Figure 6). Promoters can be cloned into or excised from these reporter vectors by using *Nde*I and *Kpn*I sites that are included for ease of cloning.

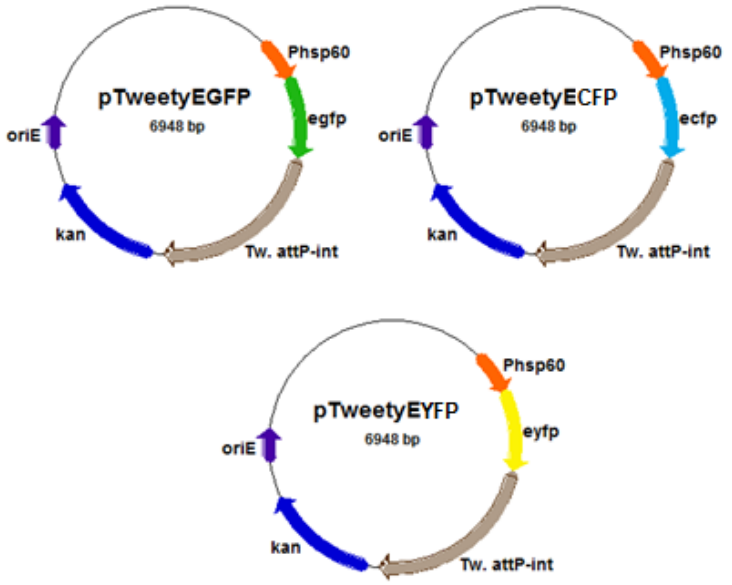


Figure 6. Tweety-based Modular Integration-Proficient Fluorescent Vectors

Figure 6: Schematic map showing integration-proficient modular vectors that contain P_{hsp60}-eGFP fusions. These modular vectors integrate into the host chromosome at the Tweety integration site. The constitutive *hsp60* promoter serves as a positive control; P_{hsp60} can be excised and candidate promoters can be efficiently cloned into these vectors at standardized multiple cloning sites provided in each construct. The construction of modular vectors containing distinct integration sites, coupled with different fluorescent reporters, allows for the simultaneous analysis of more than one candidate promoter during future studies.

2.3 DIFFERENTIAL GENE EXPRESSION PROFILES IN BIOFILMS AND PLANKTONIC CULTURES

2.3.1 Iron-responsive expression of P_{MSMEG_4515}-eGFP in biofilms and planktonic cultures

2.3.1.1 Microscopy-based analysis of P_{MSMEG_4515}-eGFP induction in planktonic cells

Iron is an important pre-requisite for mycobacterial growth, and in the case of *M. tuberculosis*, for virulence [73]. The acquisition of iron in the intracellular milieu of the host poses a challenge for these organisms, since this iron is tightly sequestered for use by the host itself. To overcome this challenge, the mycobacteria, like other microorganisms, have evolved strategies to acquire sufficient iron to facilitate growth. *M. tuberculosis* uses the class of siderophores known as the mycobactins to acquire iron, while *M. smegmatis* employs mycobactins as well as another class of siderophores known as the exochelins [73]. In choosing specific iron-acquisition genes to profile during biofilm development, we chose the *M. smegmatis* mycobactin biosynthesis gene, *MSMEG_4515*, which is homologous to the *M. tuberculosis mbtB* gene. *M. tuberculosis mbtB* has been conclusively reported to be involved in mycobactin production. An *M. tuberculosis mbtB* mutant strain was unable to produce siderophore, and this mutant was also found to be avirulent in macrophages [86]. Previous microarray experiments suggest that *MSMEG_4515*, which forms part of the mycobactin biosynthesis operon, is up-regulated in stationary phase cells, as well as in four-day biofilms grown in 2 μM Fe^{2+} [61].

To profile *MSMEG_4515* induction at the cellular level, we generated a P_{MSMEG_4515}-eGFP fusion on the Tweety-derived fluorescent vector, and tracked eGFP expression in planktonic cultures and four-day biofilms by phase contrast and confocal fluorescence microscopy.

Phase contrast fluorescence microscopy on planktonically grown cells showed that as expected during planktonic growth, the induction of P_{MSMEG_4515}-eGFP is tightly regulated by the amount of Fe^{2+} in the media. Cells were grown in the presence of Tween, and 2-5 μL of culture material were used to prepare slides for microscopy, as described in Section 5.5 of Materials and Methods. eGFP expression is significant at low concentrations of Fe^{2+}, such as when no additional Fe^{2+} is added to the media, and is down-regulated in the presence of high amounts (50 μM) of Fe^{2+}, even at high cell densities (Figure 7).

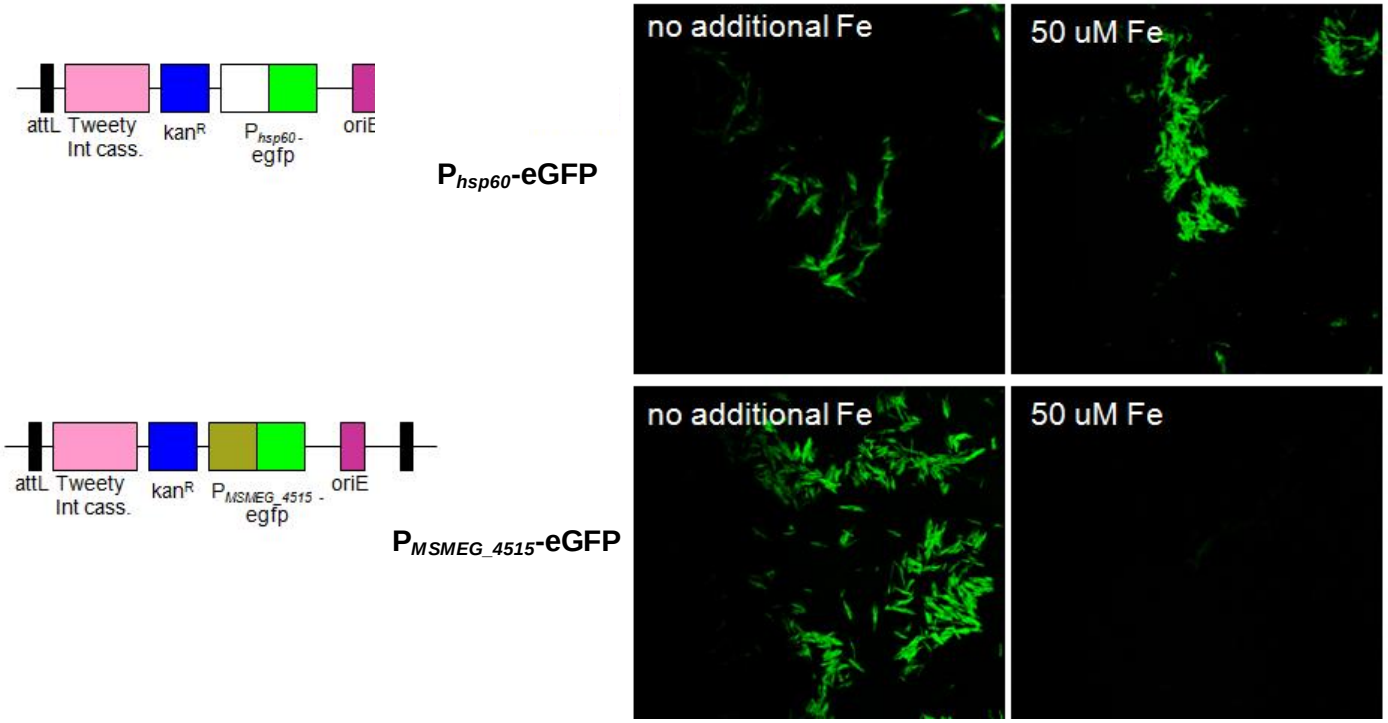


Figure 7. Iron-dependent expression of P_{MSMEG_4515} in planktonic cells

Figure 7: Expression of P_{MSMEG_4515}-eGFP in planktonic cultures grown under low and high-iron conditions. *No additional iron* indicates that the only iron present in the media is residual iron from the addition of other reagents. Cells were grown to an OD_{600} of 2. P_{hsp60}-eGFP serves as a constitutively active control.

2.3.1.2 Confocal microscopy-based analysis of P_{MSMEG_4515}-eGFP induction in biofilms

In contrast to the tight down-regulation of *MSMEG_4515* that was observed in early stationary phase planktonic cells grown in high iron, analysis of four-day biofilms by confocal microscopy showed that cells within these biofilms induce P_{MSMEG_4515}-eGFP even in the presence 50 μM Fe^{2+} (Figure 8). For confocal microscopy-based experiments, biofilms were grown on glass cover-slips that were placed at a 45-degree angle within biofilm media, as described in Section 5.5 of Materials and Methods. Cover-slips retrieved from the growth media were placed on a glass slide, and overlayed with a second cover-slip prior to imaging. The uppermost plane of the biofilm represents the liquid-air interface, and the cover-slip associated plane represents the media-exposed lower surface of the biofilm pellicle. Sections of the biofilm were scanned in increasing order of depth, ranging from the liquid-air interface to the lower surface of the film which was exposed to the media. eGFP expression was detected at each depth that was analyzed (Figure 8).

To further investigate the population of cells that induce *MSMEG_4515* within these biofilms, we repeated the above studies using a strain that constitutively expresses a single copy of the P_{hsp60}-mCherry fusion at the Giles integration site. This mCherry construct is codon optimized for the mycobacteria, and includes a strong ribosome binding site from gp9 in mycobacteriophage TM4 (cassette originally obtained from the Eric Rubin laboratory, Harvard School of Public Health; modified by Mariana Piuri in the Hatfull laboratory, unpublished). These features make this reporter particularly strong in expression strength and therefore to sensitivity of detection. Our results show that within mature four-day biofilms, almost all cells in

the population induce *MSMEG_4515* and express eGFP (Figure 9). These results suggest that unlike in the case of planktonic growth, while growing in a biofilm, a majority of cells induce mycobactin production even at high concentrations of iron. We do observe some cells in the field of view that lack eGFP expression, and a few that lack mCherry expression (Figure 9C, insets). Further expression studies performed using flow cytometry address this observation later in this chapter.

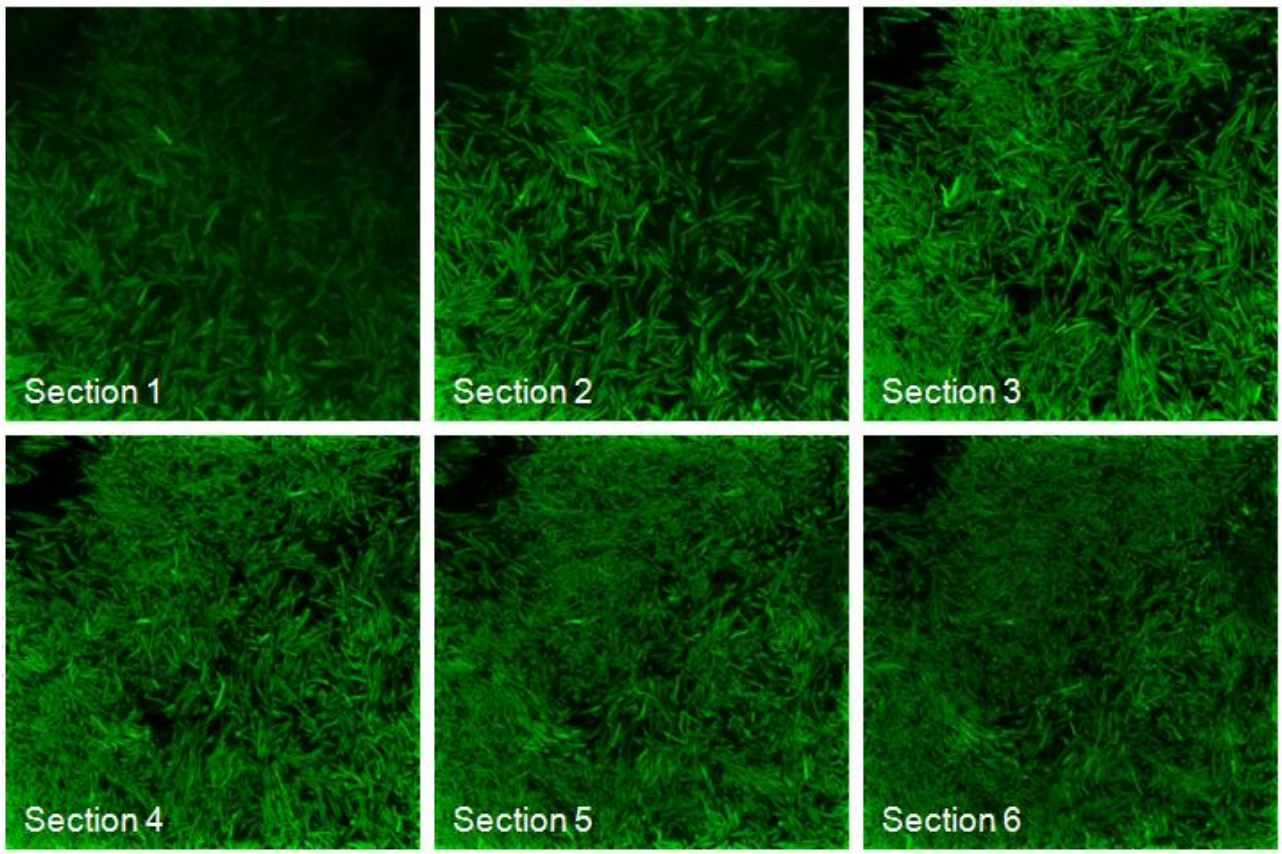


Figure 8. P_{MSMEG_4515} expression in four-day biofilms in iron-rich media

Figure 8: P_{MSMEG_4515}-eGFP expression in cells in a four-day biofilm grown in 50 μM Fe^{2+}, as seen under a confocal microscope. Sections 1-6 represent increasing depth through the biofilm.

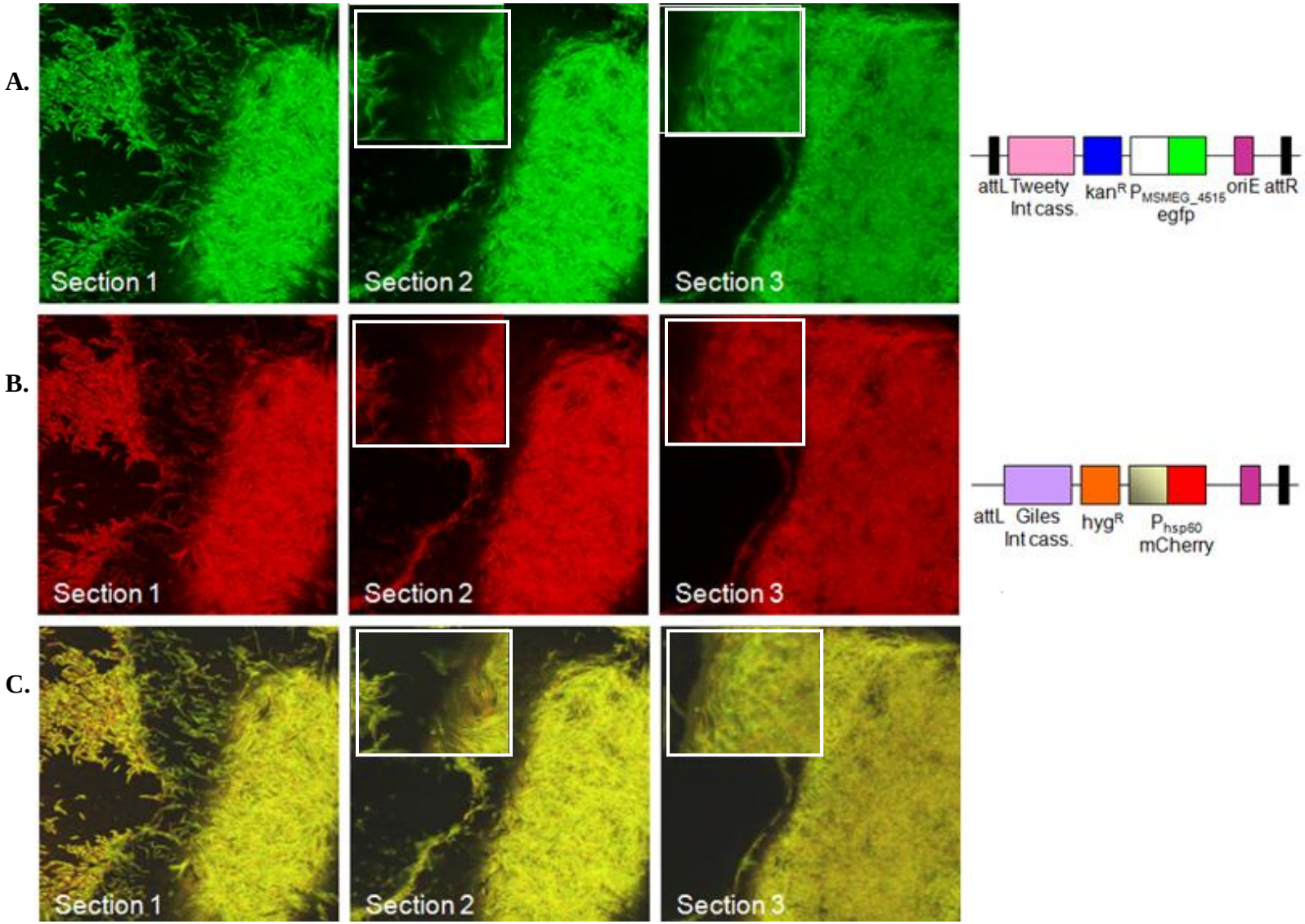


Figure 9. P_{MSMEG_4515} and P_{hsp60} expression in four-day biofilms in high iron

Figure 9: Expression of P_{MSMEG_4515}-eGFP and P_{hsp60}-mCherry in four-day biofilms grown in 50 μM Fe^{2+}, as seen under a confocal microscope. **A.** Cells inducing P_{MSMEG_4515} in four-day biofilms at high levels of iron, express eGFP (green fluorescence). **B.** Cells in the same field of view that induce P_{hsp60} in the same biofilm, express mCherry (red fluoresence). **C.** Merged view of A and B. Sections 1-3 represent increasing depth through the biofilm. Insets show zoom-in views from a region of the same section.

2.3.1.3 Flow cytometry-based analysis of P_{MSMEG_4515} and P_{hsp60} in planktonic cultures and mature four-day biofilms

Microscopy on biofilms and planktonic cultures provides us with insights into P_{MSMEG_4515} induction during the various growth states that were tested; however, these studies do not allow for a read-out of the relative levels of promoter induction in response to growth conditions or iron availability. To provide quantitative corroboration for our microscopy-based data, we performed flow cytometry and analyses on strains carrying reporter fusions for P_{MSMEG_4515}. Each strain described in Table 3 was grown as planktonic cultures and as biofilms in varying concentrations of iron, and then assayed for fluorescence by flow cytometry. Planktonic cultures were grown with agitation in reconstituted biofilm media that contained Tween-80 as a detergent. For flow cytometry-based experiments, biofilms were grown in 48-well plastic culture dishes, each well containing 1 mL of reconstituted biofilm media.

Samples were prepared for flow cytometry as described in Section 5.6 of Materials and Methods. Strains mc^2155 carrying P_{hsp60}-eGFP or P_{hsp60}-mCherry alone, were used to compensate for spillover of fluorescence signals across channels as described in Section 5.6. P_{hsp60} was used as the experimental standard for promoter expression. In each growth state (planktonic or biofilm), strains were grown in varying concentrations of iron to test for iron-responsive regulation of mycobactin biosynthesis.

Compensation for fluorophore signal spillover across channels, and the following quantitation of mean eGFP and mCherry fluorescence for each experiment was performed as described in detail in Section 5.6 of Materials and Methods.

Histograms of fluorescence signal distribution of cells that express eGFP and/or mCherry in samples grown as planktonic cultures are shown in Figures 10-15. Figure 10A and 10C show the raw fluorescence signal distribution obtained for control strains mc^2155 with P_{hsp60}-mCherry and mc^2155 with P_{hsp60}-eGFP, respectively. These histograms show events that have been filtered past pre-determined thresholds for size (forward scatter) and exclusion of debris (side scatter), but have not been filtered for fluorescence or compensated for signal spillover across channels. Figures 10B and 10D show histograms of fluorescence distribution in cells, after compensating for signal spillover across channels.

Overall, Figure 10 provides fluorescence distribution information for single-stain controls grown under iron-deficient conditions. Figures 11-13 use the same compensation matrices as those derived from and applied to raw data as seen in Figure 10, for planktonic growth in 0 μM, 2 μM and 4 μM Fe^{2+}, respectively. The compensation matrices used for Figure 10 B and D, are applied to data shown in Figure 11 A and C, and used to derive fluorescence signal distribution in the dual-fluorescent strains mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-Cherry and mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (Figure 11 B and D). Similarly, compensation matrices obtained for single-fluorescent control strains grown in 2 μM and 4 μM Fe^{2+}, are used to derive fluorescent signal distribution in dual-fluorescent strains grown in the corresponding iron conditions (Figures 12 and 13). Panels A in Figures 11, 12 and 13 show raw, uncompensated fluorescence signal distribution for strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-Cherry grown in 0 μM, 2 μM and 4 μM Fe^{2+}, respectively. Panels B in Figures 11-13 show compensated fluorescence distribution under these increasing concentrations of iron. Panels C and D show raw and compensated fluorescence signal distribution for strain mc^2155 with P*hsp60*-eGFP + P*hsp60*-Cherry, respectively.

Figure 14 shows single fluorescent protein-tagged controls, similar to Figure 10, but this time for samples grown in iron-rich conditions (50 μM Fe^{2+}). Compensation matrices derived using these single fluorescent protein-tagged control strains were used to obtain compensated fluorescence signal distributions for strains mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry and mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry grown in 50 μM Fe^{2+} (Figure 15 B and D).

Mean fluorescence values for eGFP and mCherry in planktonic samples grown under each iron condition were recorded as an indication of overall P_{MSMEG_4515} and P_{hsp60} expression. Comparing eGFP fluorescence signal distribution across Panels B in Figures 11, 12, 13 and 15 shows an overall decrease in the mean fluorescence detected for P_{MSMEG_4515}-eGFP as the concentration of iron increases. Fluorescence signal distribution for P_{hsp60}-eGFP and P_{hsp60}-mCherry does not change significantly, relative to changes in iron concentration (Panels D in Figures 11-15).

Figure 16 shows mean eGFP fluorescence values from two independent planktonic growth experiments for strains expressing P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP under increasing iron availability. Our analysis of the mean fluorescence values obtained for P_{MSMEG_4515}-eGFP expression in exponential phase planktonic cultures shows a sharp decrease, as the concentration of iron in the media increases (Figure 16). This result is consistent with the hypothesized iron-dependent regulation of *MSMEG_4515*, and with our observations by microscopy. The difference in eGFP fluorescence levels is statistically significant when comparing cells grown in iron-deficient media (0 μM Fe^{2+}) to cells grown in iron-supplemented (4 μM Fe^{2+}) and iron-rich (50 μM Fe^{2+}) media (Table 4). Expression levels observed in 2 μM Fe^{2+} are not found to be significantly different than the levels observed in 0 μM Fe^{2+} (Table 4).

As in the case of our confocal microscopy-based studies, we used the mycobacterial *hsp60* promoter as a control; we hypothesized that P_{hsp60} would not be extensively affected by changes in iron concentration. We also used this promoter to provide a read-out of promoter activity in cells, based on standard laboratory usage of P_{hsp60} as a strong, constitutively active promoter. Flow cytometry-based quantitation of P_{hsp60}-eGFP mean fluorescence levels in exponential phase planktonic cultures does not show much change in eGFP levels between samples grown in different concentrations of iron. Nevertheless, the difference in expression levels observed between cells grown in iron-deficient media (0 μM Fe^{2+}) and in iron-rich (50 μM Fe^{2+}) media is found to be statistically significant (Table 5).

Figure 17 shows raw fluorescence signal distributions for single-fluorescent control strains grown as biofilms in iron-deficient media. The histograms represent events from samples processed from mature biofilms grown for a period of four days. Compensation matrices derived using standard curves represented in Panels A and C of Figure 17 were used to then compensate for signal spillover across channels. Compensated single-fluorescent fluorescence signal distributions are shown in Panels B and D of Figure 17. These compensation values were also used to derive fluorescence signal distributions for dual-fluorescent strains mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry and mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry grown in 0 μM Fe^{2+} (Panels B and D, Figure 18).

Single-fluorescent strains grown in 2 μM, 4 μM and 50 μM Fe^{2+} were used to derive compensation matrices for signal spillover across channels. These compensation values were applied to raw fluorescence signal distributions obtained for the dual-fluorescent strains, to generate compensated fluorescence signal distributions for these strains. Figures 19, 20 and 21 show fluorescence signal distributions for strains mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-

mCherry and mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry grown as biofilms in media containing 2 μM, 4 μM and 50 μM Fe^{2+}, respectively. Comparison of the fluorescence signal distribution and mean fluorescence values for P_{MSMEG_4515}-eGFP across Panels B in Figures 18-21 does not show a decrease in P_{MSMEG_4515}-eGFP fluorescence with an increase in the concentration of iron in the media.

Compiled mean fluorescence values from three experimental repeats, for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP expression in four-day biofilms grown in increasing amounts of iron, are shown in Figure 22. P_{MSMEG_4515}-eGFP expression in four-day biofilms does not show significant differences in the total mean fluorescence observed as the concentration of iron in the media increases (Figure 22A; Table 6). Interestingly, we find that although P_{hsp60}-eGFP expression levels at 0, 2 and 4 μM Fe^{2+} remain similar, expression increases at 50 μM Fe^{2+} (Figure 22B). This observed increase in P_{hsp60}-eGFP expression at high iron levels is found to be statistically significant (Table 7).

Table 3. Strains used for flow cytometry-based analyses

Reporter strains used for flow cytometry, including single fluorescent protein-tagged control strains and dual fluorescent protein-tagged strains that express P_{hsp60}-mCherry as a basal reporter for *M. smegmatis* cells.

Strains Used	Reporter(s)	Integration Site(s)
mc^2155 with P_{hsp60}-eGFP	eGFP	Tweety
mc^2155 with P_{hsp60}-mCherry	mCherry	Giles
mc^2155 with P_{MSMEG_4515}-eGFP and P_{hsp60}-Cherry	eGFP; mCherry	Tweety; Giles
mc^2155 with P_{hsp60}-eGFP and P_{hsp60}-mCherry	eGFP; mCherry	Tweety; Giles

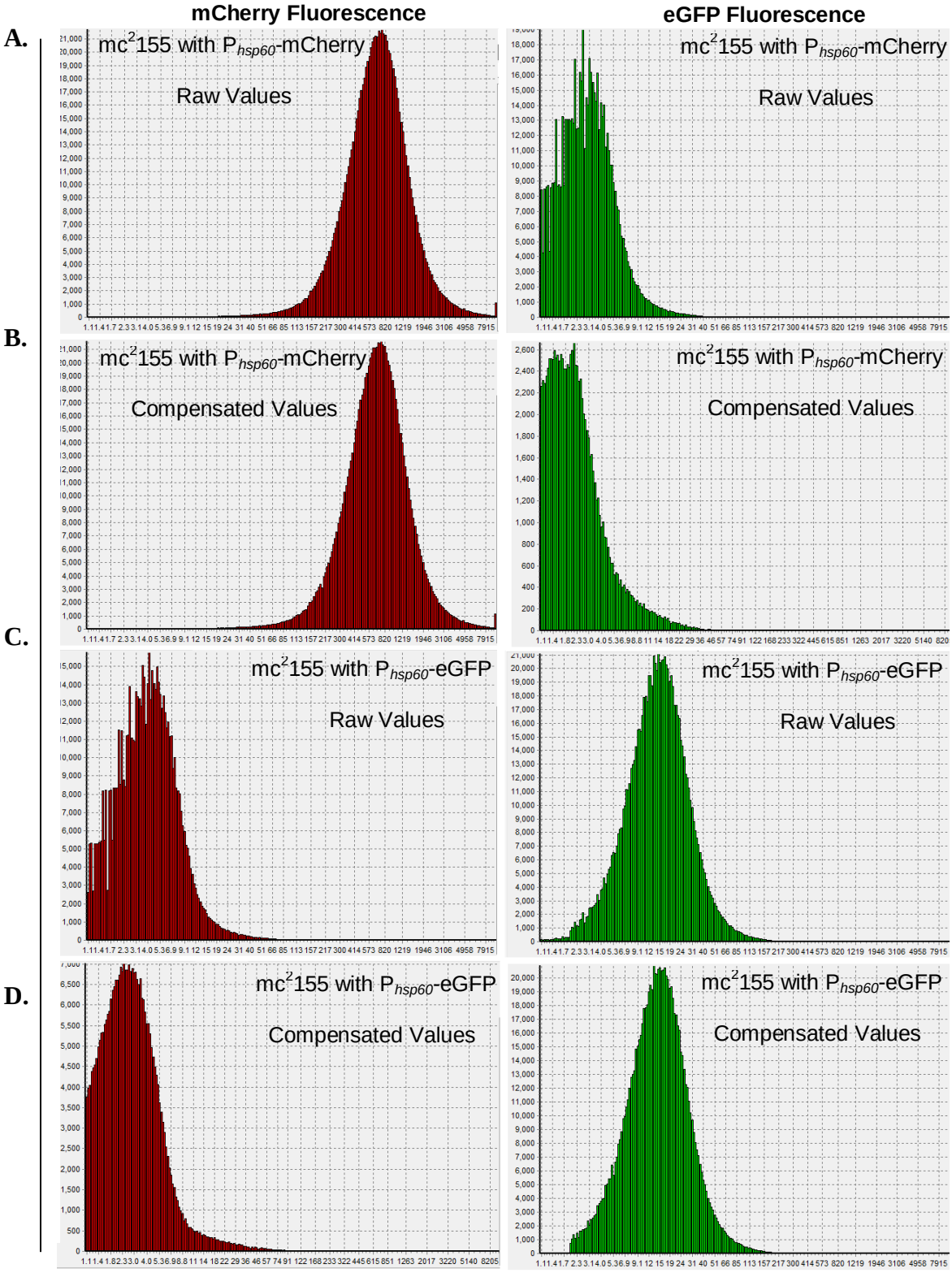


Figure 10. Fluorescence distribution in planktonic controls grown in 0 μM Fe^{2+}

Figure 10: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP (**C** and **D**), each expressing single-color fluorophores. Strains were grown in iron-deficient (0 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

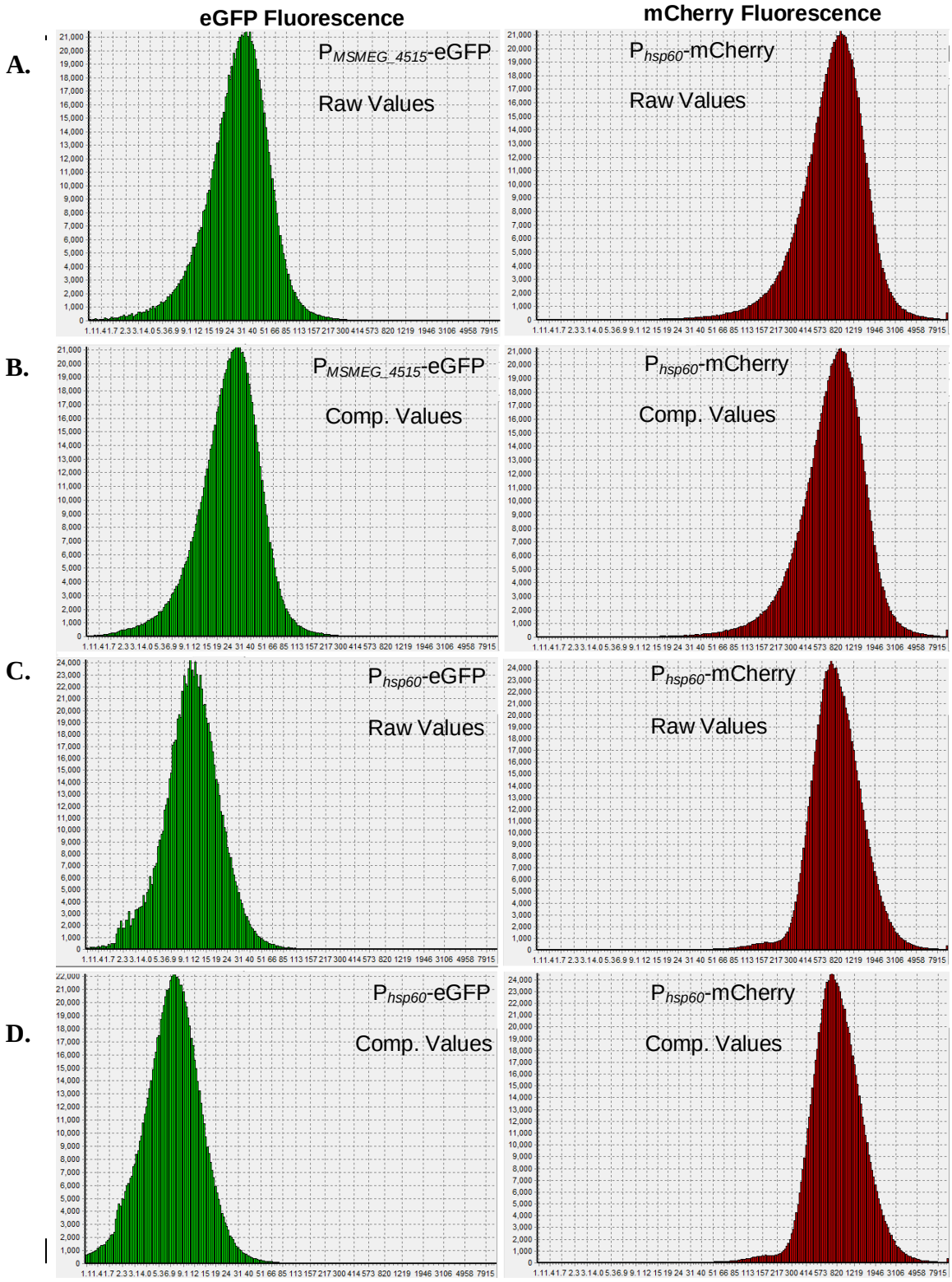


Figure 11. Fluorescence distribution for P_{MSMEG_4515} and P_{hsp60} in cultures in 0 μM Fe^{2+}

Figure 11: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-deficient (0 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

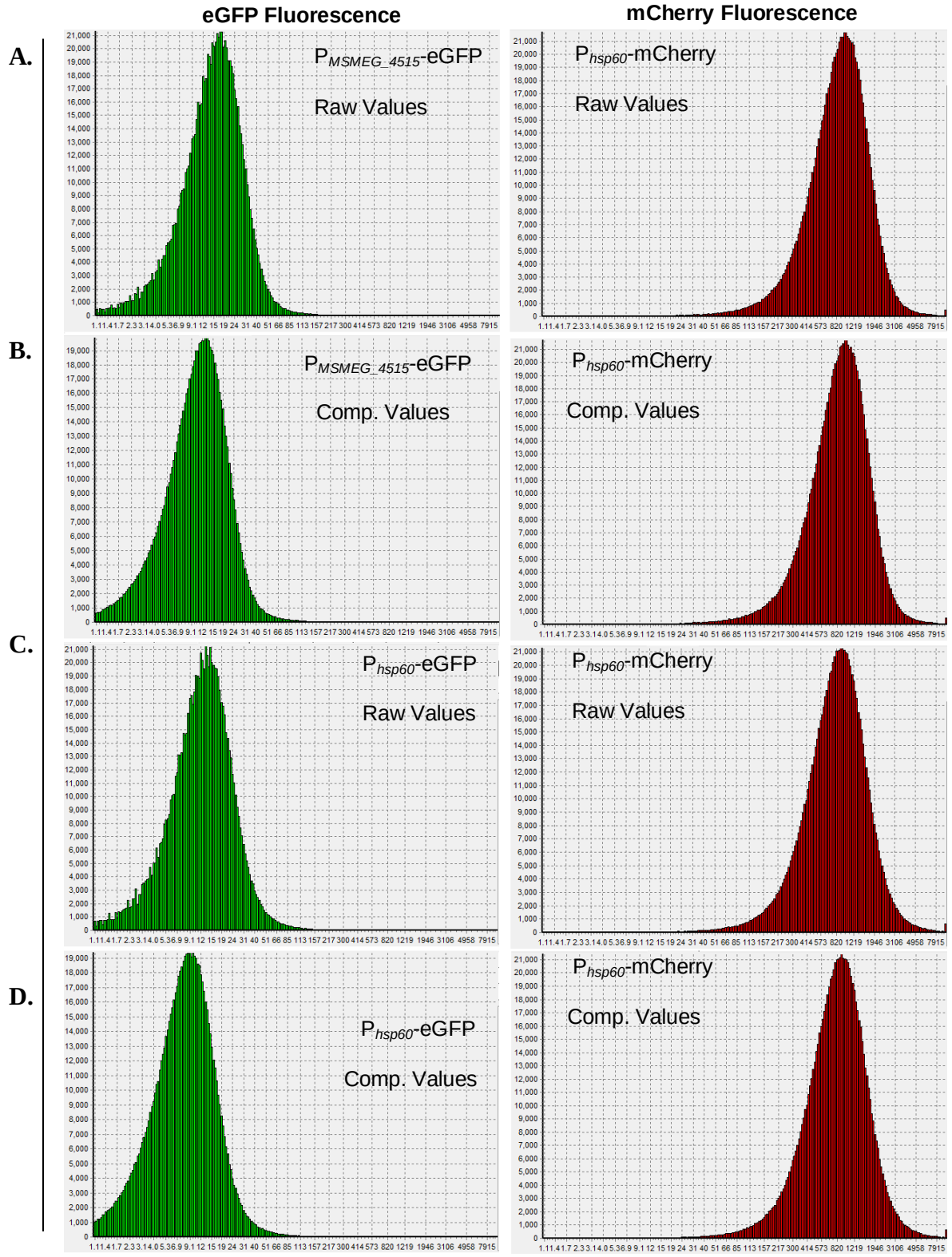


Figure 12. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP in planktonic samples in 2 μM Fe^{2+}

Figure 12: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-supplemented (2 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

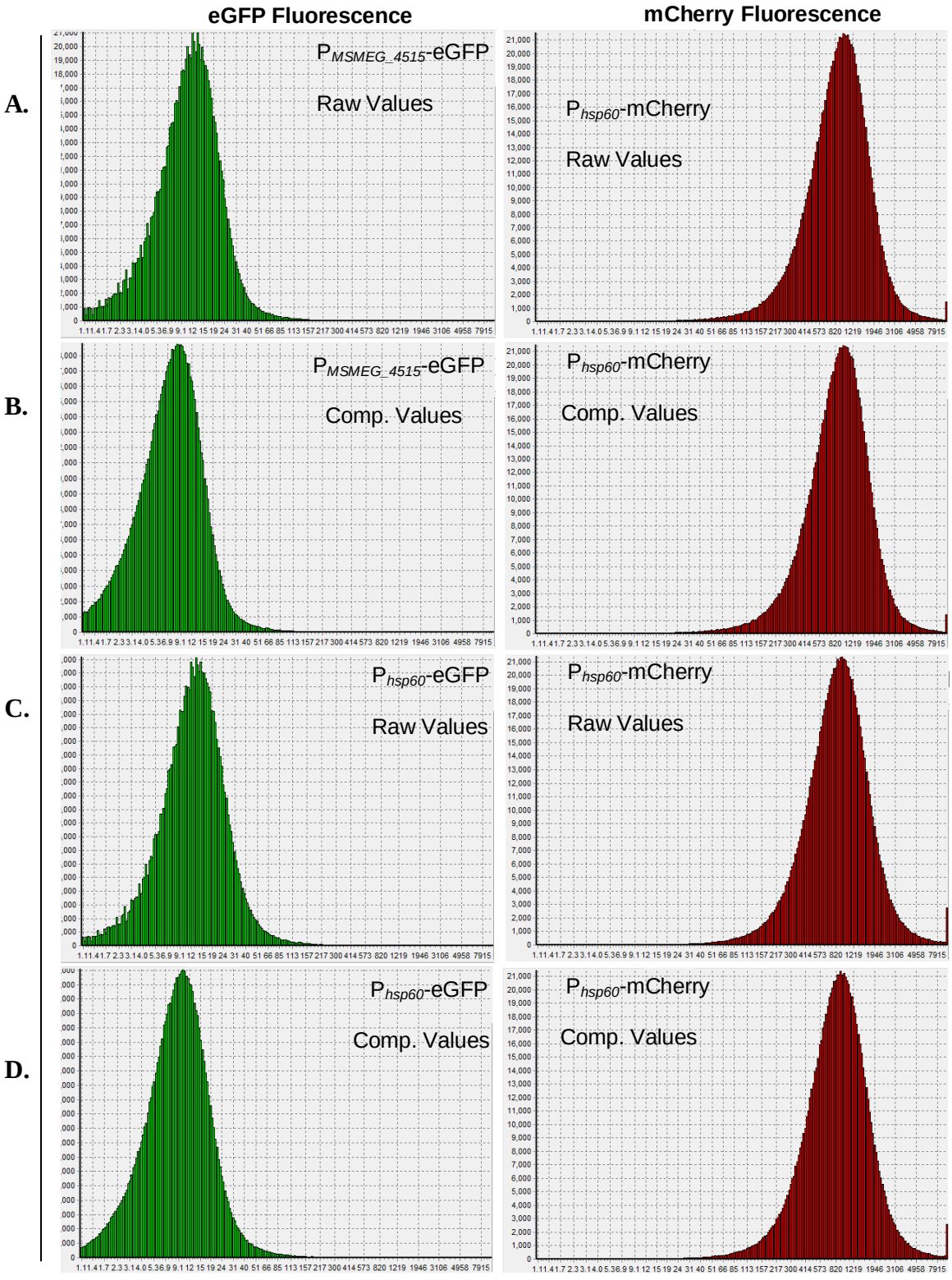


Figure 13. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP in planktonic samples in 4 μM Fe^{2+}

Figure 13: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-supplemented (4 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

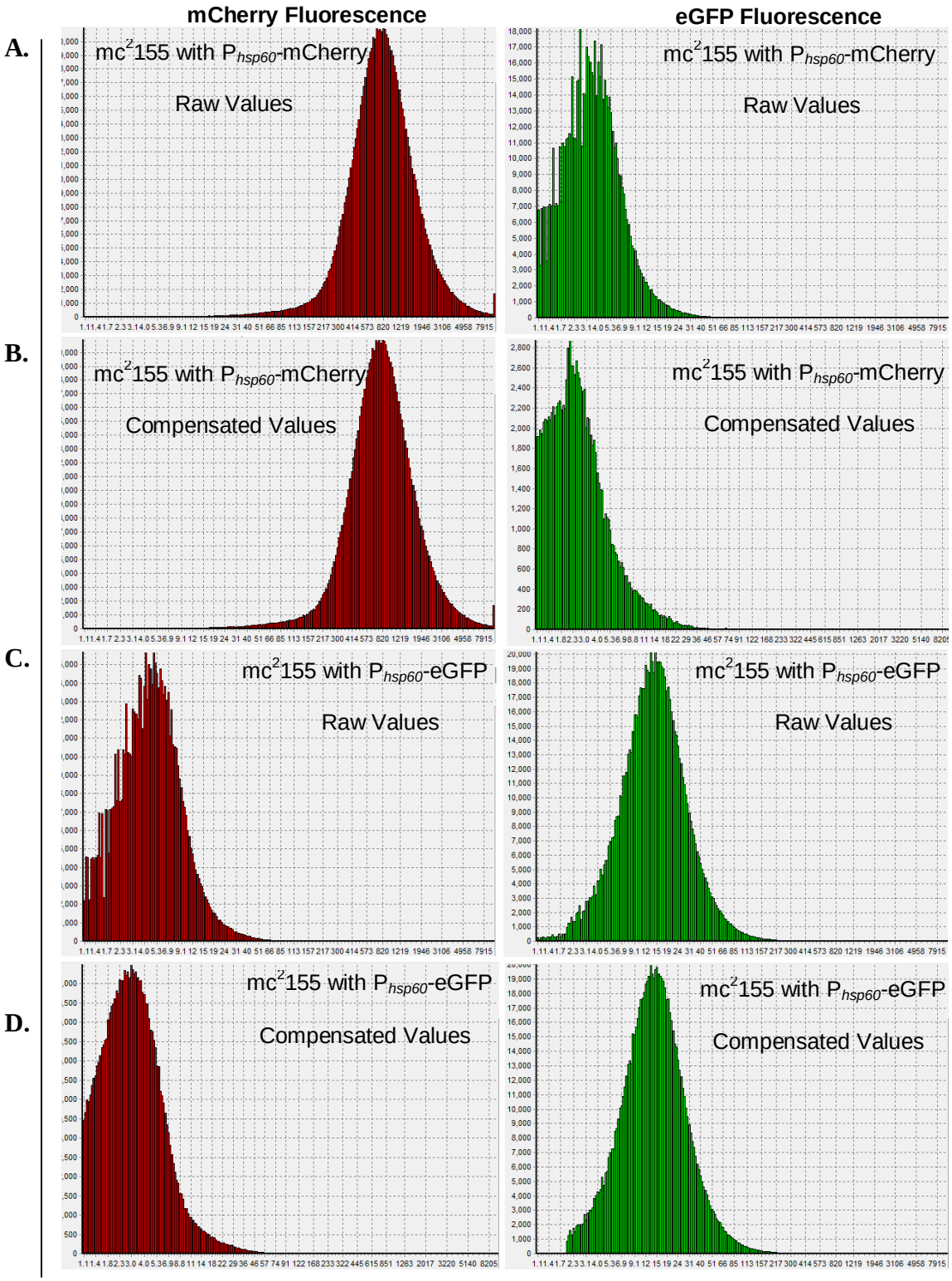


Figure 14. Fluorescence distribution for control strains in planktonic samples in 50 μM Fe^{2+}

Figure 14: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP (**C** and **D**), each expressing single-color fluorophores. Strains were grown in iron-rich (50 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

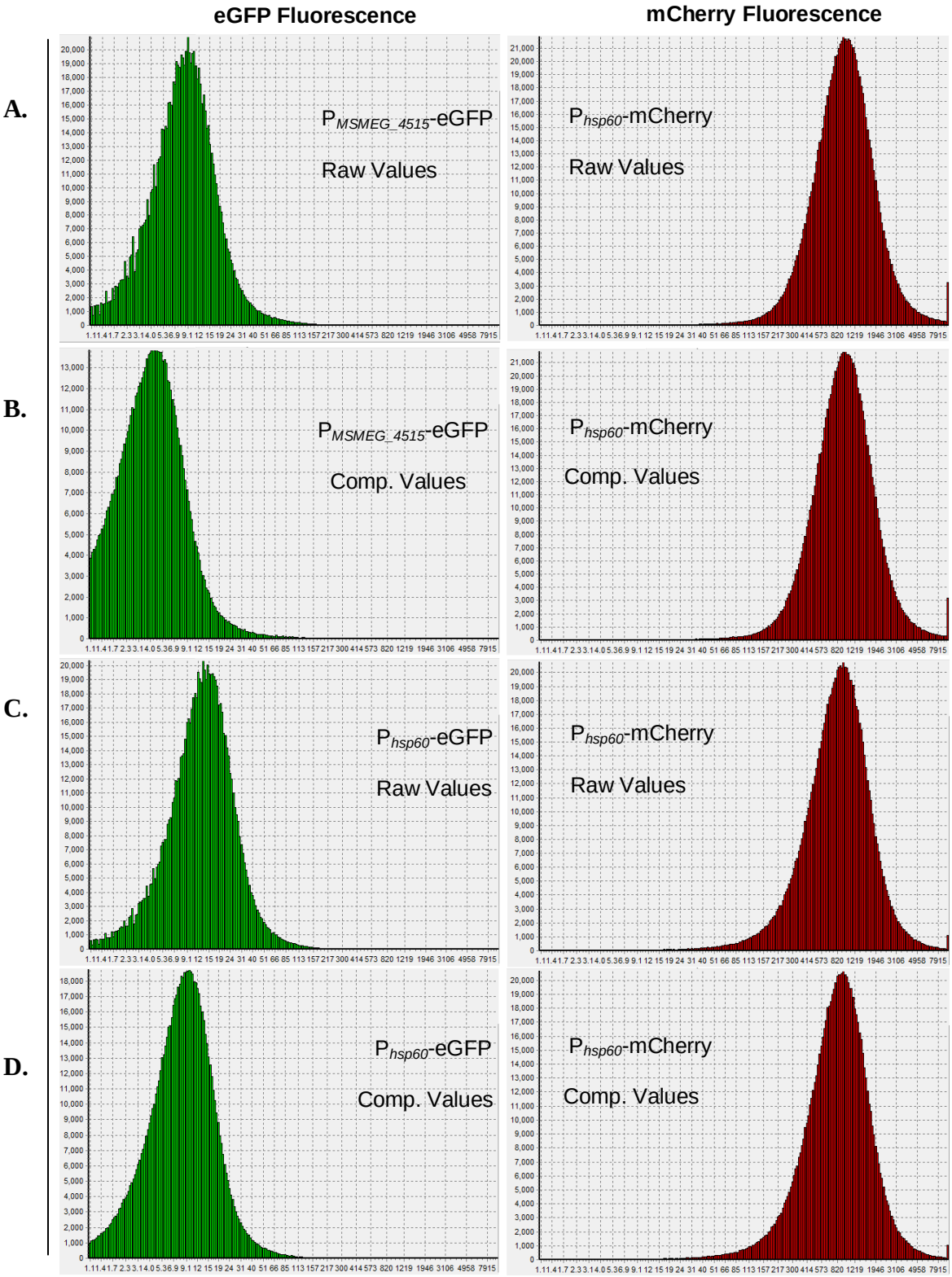
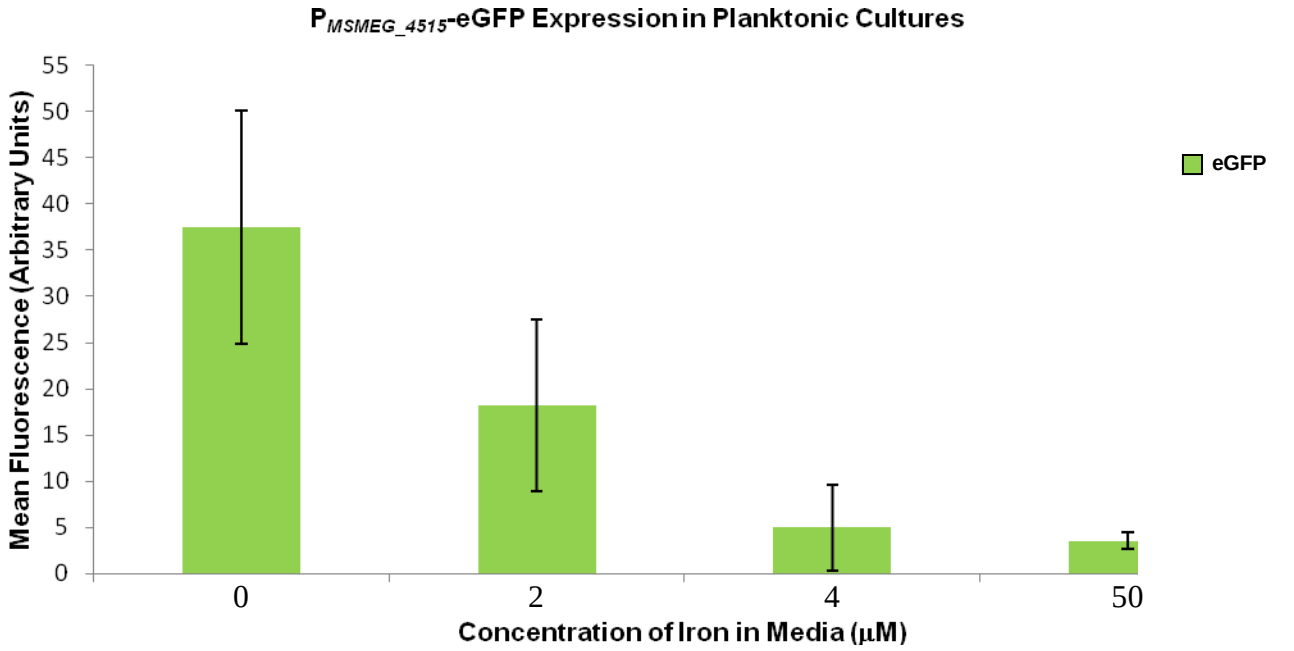


Figure 15. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP in planktonic samples in 50 μM Fe^{2+}

Figure 15: Fluorescence distribution curves obtained using Ferdinand, for planktonically grown samples of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-rich (50 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

A.



B.

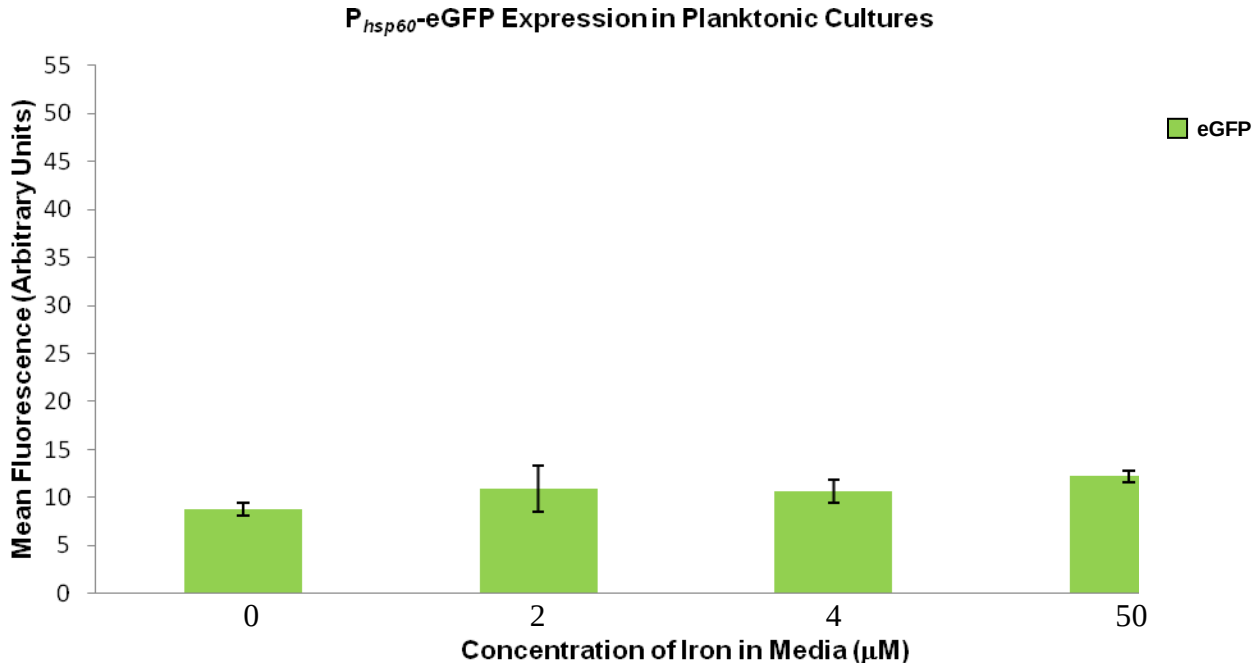


Figure 16. P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP expression in planktonic cultures

Figure 16: Compensated mean fluorescence values for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP in planktonic cells. P_{MSMEG_4515}-eGFP expression **(A)** and P_{hsp60}-eGFP expression **(B)** in exponential phase planktonic cells grown in 0, 2, 4 and 50 μM Fe^{2+}. Data are averages of two experimental replicates, and indicate standard deviation from the mean.

Table 4. Statistical significance for P_{MSMEG_4515}-eGFP fluorescence in planktonic cultures

Mean fluorescence values for P_{MSMEG_4515}-eGFP obtained under each Fe^{2+} concentration were tested against the mean values obtained from each of the other three Fe^{2+} conditions, in a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. Differences in eGFP expression values across Fe^{2+} conditions were found to be statistically significant when comparing fluorescence between 0 and 4 μM Fe^{2+} and 0 and 50 μM Fe^{2+}. P values ≤ 0.10 are represented in bold.

Experimental Condition	0 μM	2 μM	4 μM	50 μM
0 μM	1	0.226	**0.08**	**0.06**
2 μM	0.23	1	0.21	0.16
4 μM	**0.08**	0.21	1	0.71
50 μM	**0.06**	0.16	0.71	1

Table 5. Statistical significance for P_{hsp60}-eGFP fluorescence in planktonic cultures

Similar methodology as employed for data presented in Table 4, using mean fluorescence values obtained for P_{hsp60}-eGFP under each Fe^{2+} concentration. Differences in eGFP expression values across Fe^{2+} conditions were found to be statistically significant when comparing fluorescence between 0 and 50 μM Fe^{2+}. P values ≤ 0.05 are represented in bold.

Experimental Condition	0 μM	2 μM	4 μM	50 μM
0 μM	1	0.35	0.19	**0.03**
2 μM	0.35	1	0.91	0.53
4 μM	0.19	0.91	1	0.24
50 μM	**0.03**	0.53	0.24	1

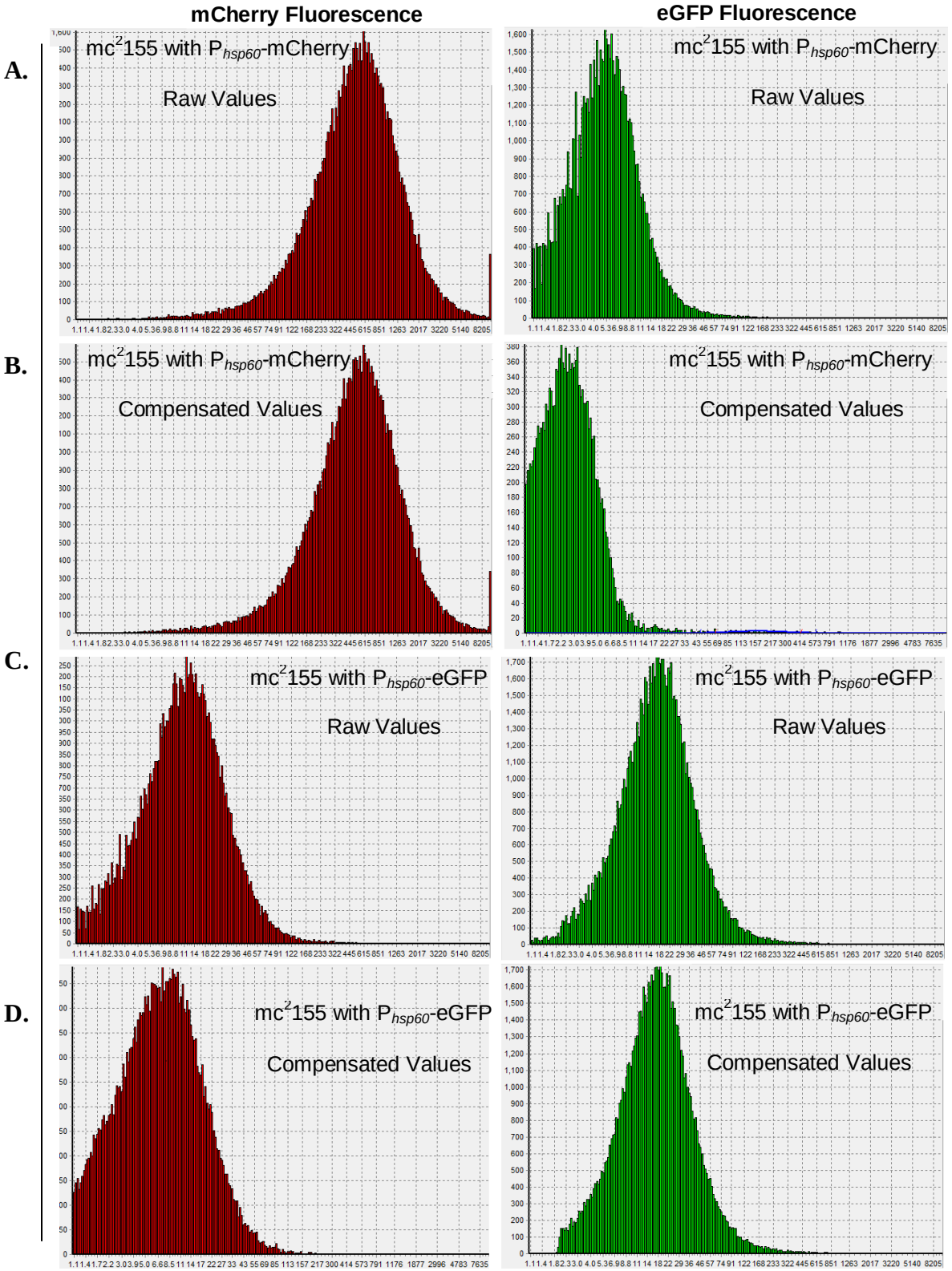


Figure 17. Fluorescence distribution in four-day biofilms of single-color controls in 0 μM Fe^{2+}

Figure 17: Fluorescence distribution curves obtained using Ferdinand, for four-day biofilms of strain mc^2155 with P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP (**C** and **D**), each expressing single-color fluorophores. Strains were grown in iron-deficient (0 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

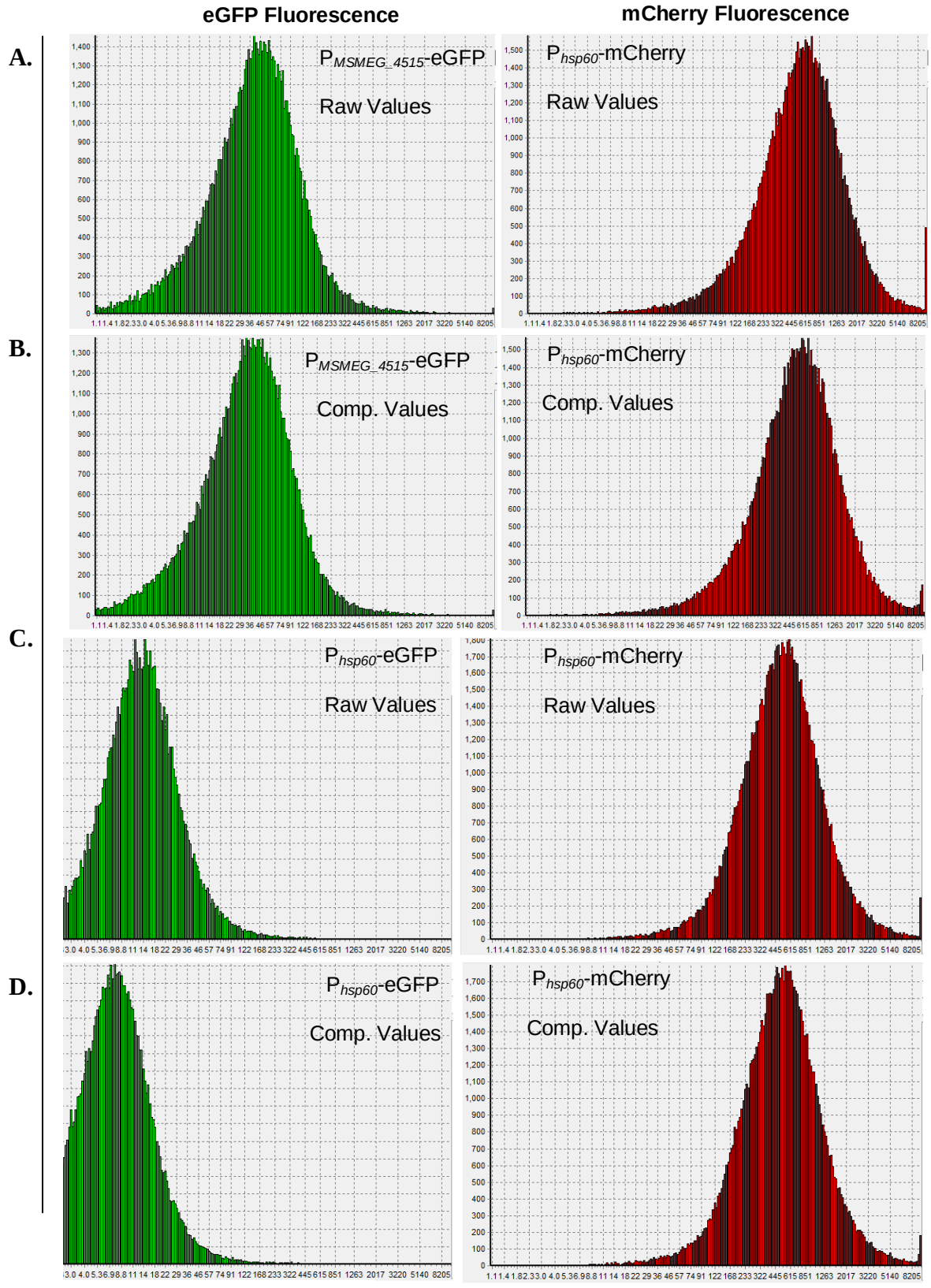


Figure 18. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP biofilms in 0 μM Fe^{2+}

Figure 18: Fluorescence distribution curves obtained using Ferdinand, for four-day biofilms of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-deficient (0 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

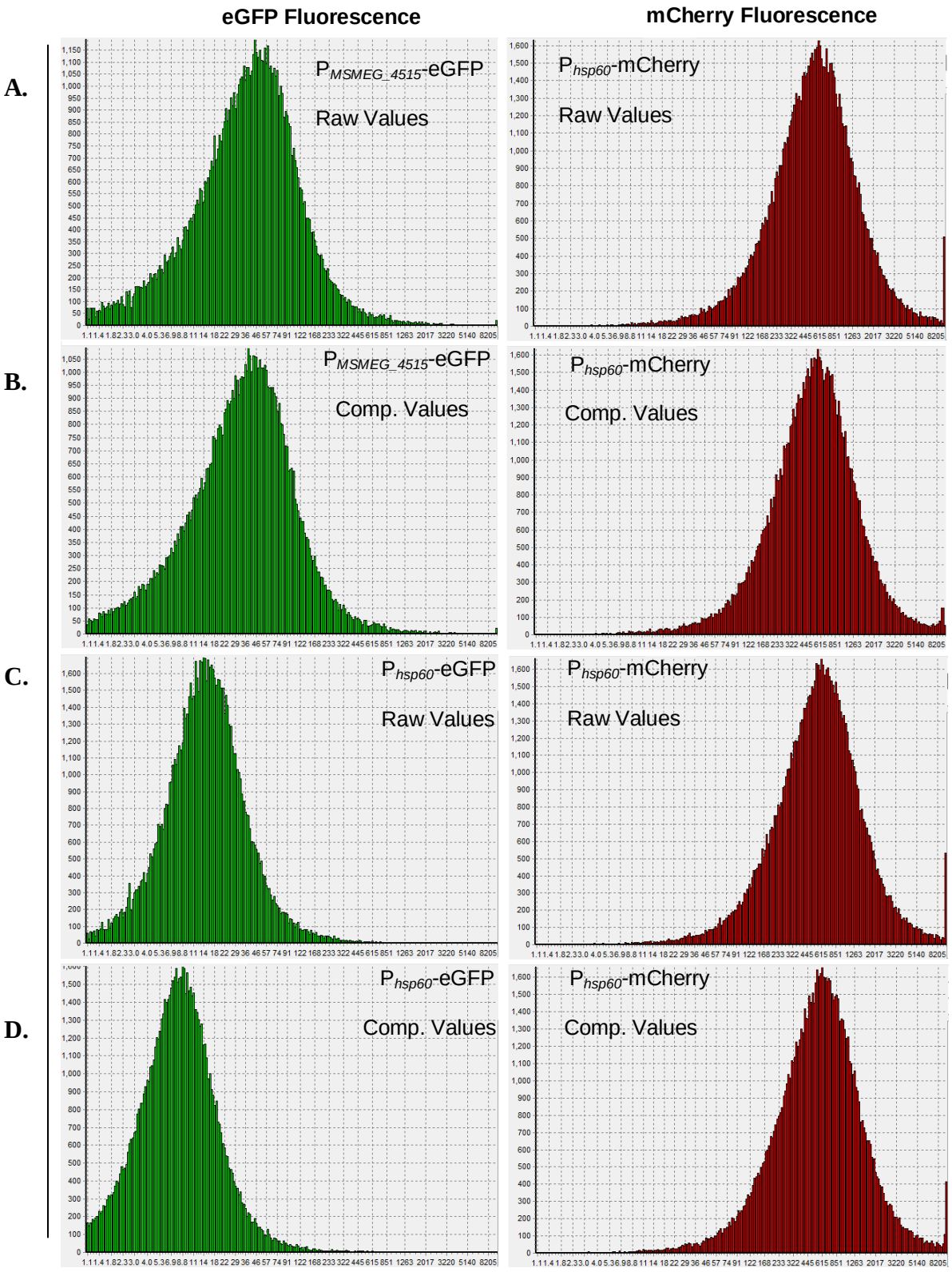


Figure 19. Fluorescence distribution for P$_{MSMEG_4515}$-eGFP and P$_{hsp60}$-eGFP biofilms in 2 μM Fe^{2+}

Figure 19: Fluorescence distribution curves obtained using Ferdinand, for four-day biofilms of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-supplemented (2 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

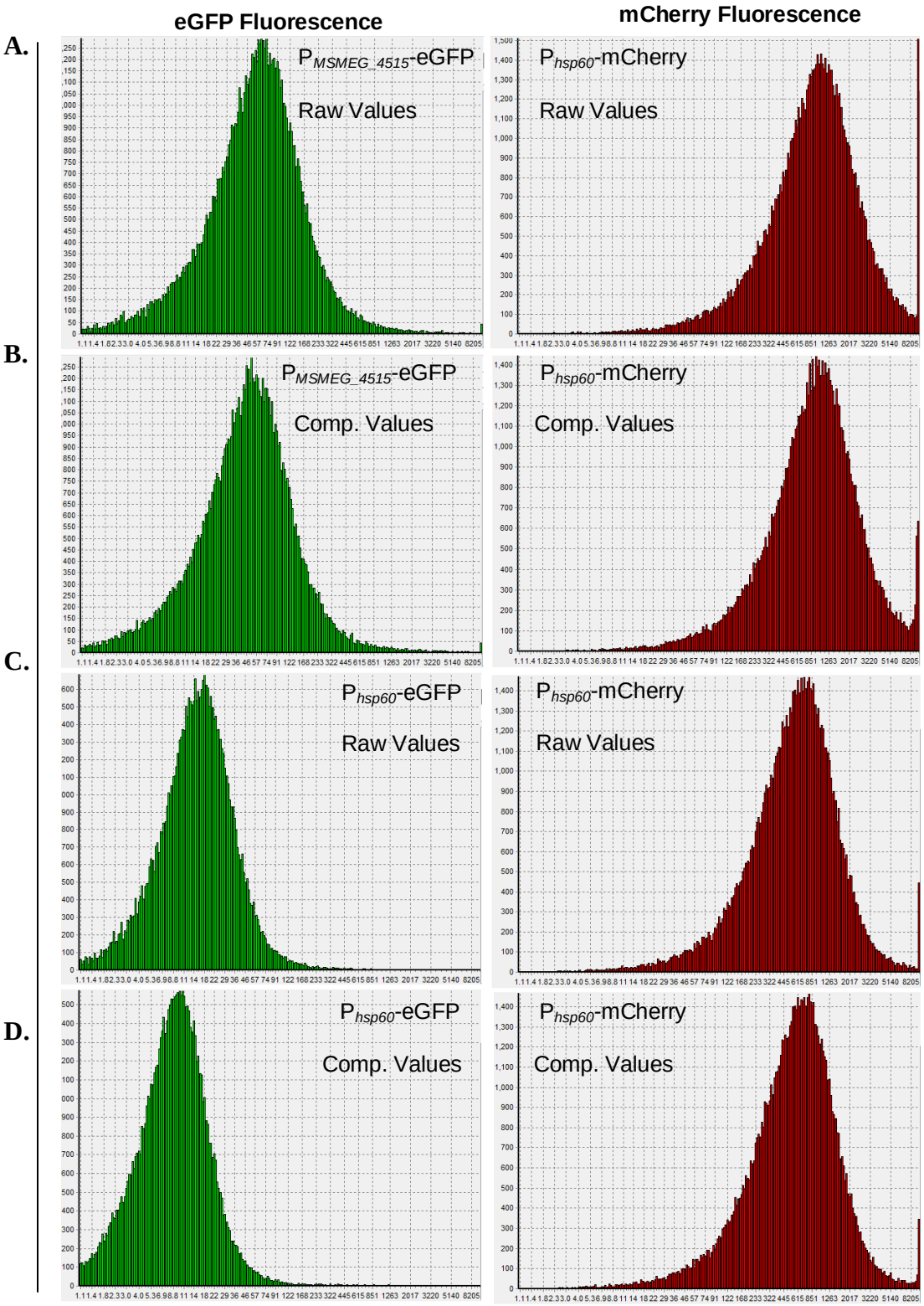


Figure 20. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP biofilms in 4 μM Fe^{2+}

Figure 20: Fluorescence distribution curves obtained using Ferdinand, for four-day biofilms of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-supplemented (4 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

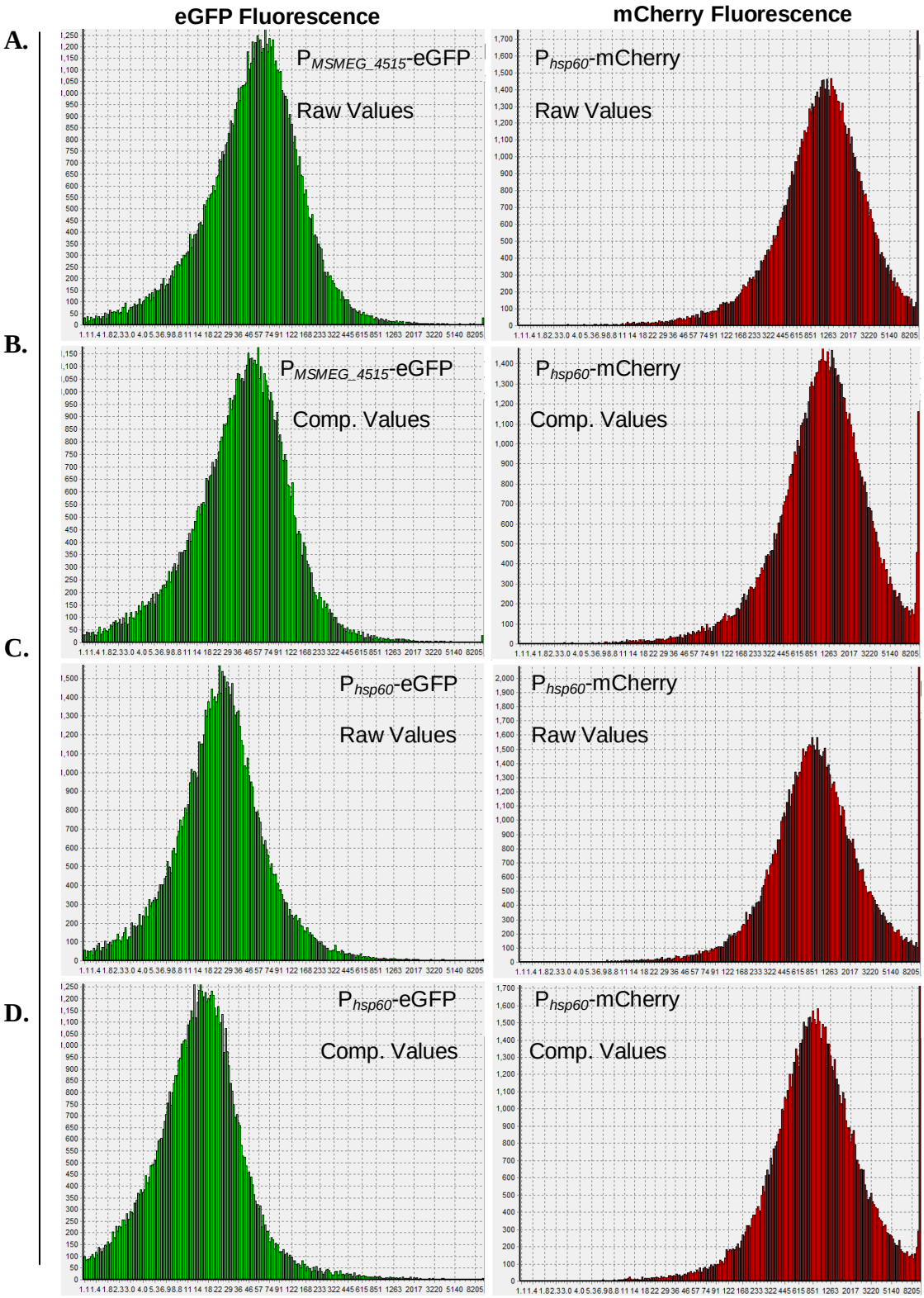


Figure 21. Fluorescence distribution for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP biofilms in 50 μM Fe^{2+}

Figure 21: Fluorescence distribution curves obtained using Ferdinand, for four-day biofilms of strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry (**A** and **B**) and strain mc^2155 with P_{hsp60}-eGFP + P_{hsp60}-mCherry (**C** and **D**), each expressing dual-color fluorophores. Strains were grown in iron-rich (50 μM Fe^{2+}) media. The horizontal axis represents fluorescence values detected for each event that passes through Z-scoring filters for the fluorescent channels. The vertical axis represents the frequency of values across events detected on the horizontal axis. Panels **A** and **C** show raw fluorescence values detected for each event in each strain. Panels **B** and **D** represent the corresponding compensated fluorescence values, following correction for spillover across channels. Data represent one experimental set, and is repeatable. *Note:* X-axis and Y-axis ranges are not conserved across all panels, since raw distribution curves were obtained individually for each strain tested, and Ferdinand does not allow for customized export of distribution curve images.

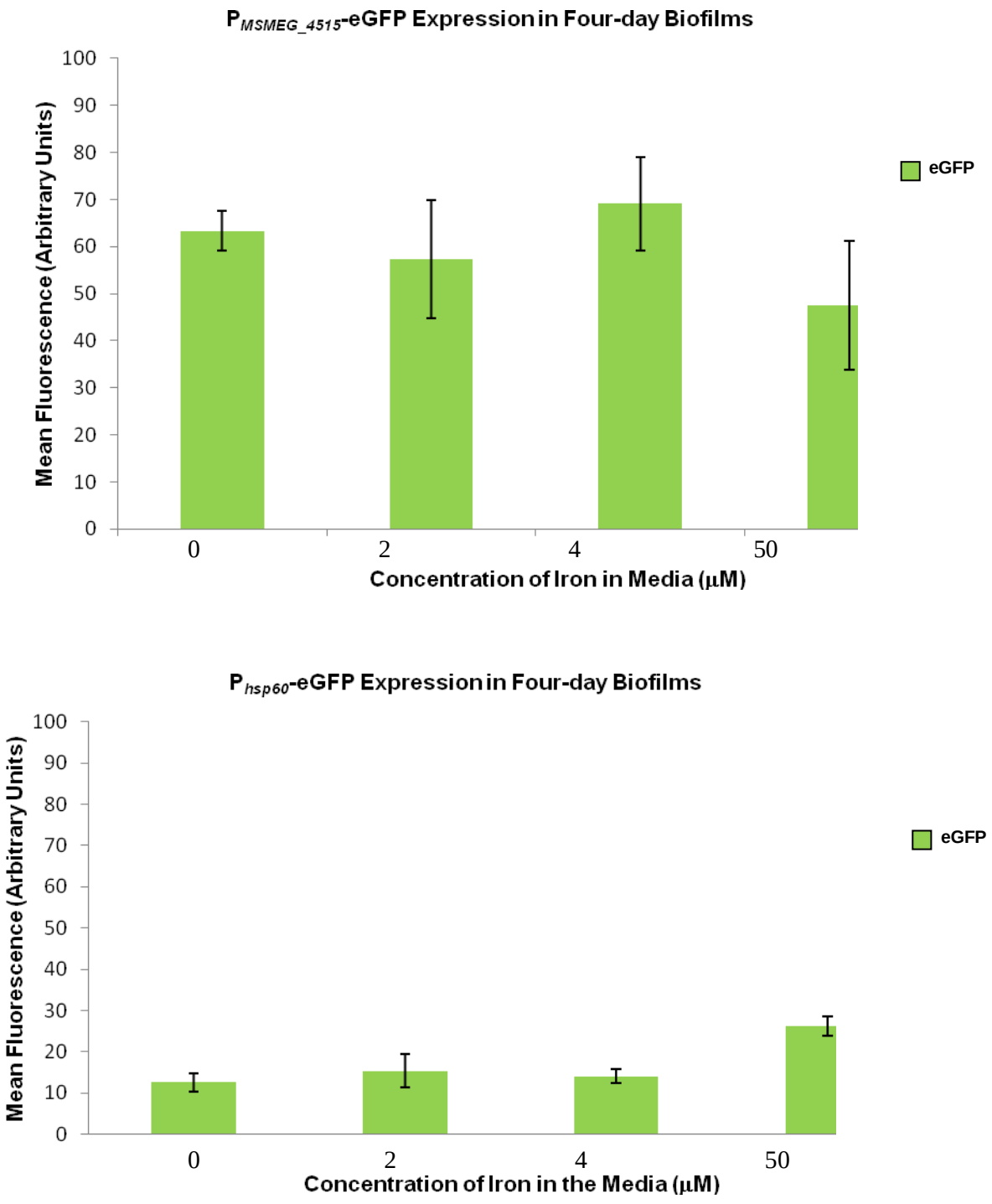


Figure 22. P_{MSMEG_4515}-eGFP and P_{hsp60}-mCherry expression in four-day biofilms

Figure 22: Compensated mean fluorescence values for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP in four-day biofilms. P_{MSMEG_4515}-eGFP expression **(A)** and P_{hsp60}-eGFP expression **(B)** in four-day biofilms grown in 0, 2, 4 and 50 μM Fe^{2+}. Data are mean of three experimental replicates, and indicate standard deviation from the mean.

Table 6. Statistical significance for P_{MSMEG_4515}-eGFP expression in 4-day biofilms

Mean fluorescence values for P_{MSMEG_4515}-eGFP obtained under each Fe^{2+} concentration were tested against the mean values obtained from each of the other three Fe^{2+} conditions, in a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. Differences in eGFP expression values across Fe^{2+} conditions were not found to be statistically significant (P values > 0.05).

Biofilms (Fe^{2+} Conc.)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.474	0.408	0.128
2 μM	0.474	1	0.271	0.411
4 μM	0.408	0.271	1	0.09
50 μM	0.128	0.411	0.09	1

Table 7. Statistical significance for P_{hsp60}-eGFP expression in 4-day biofilms

Similar methodology as employed for data presented in Table 4, 5 and 6, using mean fluorescence values obtained for P_{hsp60}-eGFP under each Fe^{2+} concentration. Expression levels in biofilms grown in 50 μM Fe^{2+} were found to be significantly different than levels in biofilms grown in 0, 2 and 4 μM Fe^{2+}. P values ≤ 0.05 are represented in bold.

Biofilms (Fe^{2+} Conc.)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.367	0.413	**0.002**
2 μM	0.367	1	0.655	**0.017**
4 μM	0.413	0.655	1	**0.002**
50 μM	**0.002**	**0.017**	**0.002**	1

As illustrated by the data presented in Figures 16 and 22, we observe differences in the mean fluorescence levels detected for P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP under corresponding conditions of growth and iron concentrations. In planktonic cultures that lack additional iron in the media, we observe significantly stronger induction of P_{MSMEG_4515}-eGFP than of P_{hsp60}-eGFP. The fluorescence levels detected at 2 μM and 4 μM Fe^{2+} do not show significant differences between the two promoters. In planktonic cultures that grow in high iron, the level of P_{MSMEG_4515}-eGFP induction is significantly lower than that of P_{hsp60}-eGFP.

In four-day biofilms, once again, regardless of the trend in fluorescence levels observed for the same promoter at varying concentrations of iron, the levels of fluorescence differ when comparing across the two promoters. P_{MSMEG_4515}-eGFP exhibits significantly stronger mean fluorescence levels than P_{hsp60}-eGFP across all iron concentrations (Table 8). This observation is consistent with the statistically significant increase in expression of P_{hsp60}-eGFP in biofilms containing high amounts of iron (Table 7).

Table 8. Statistical significance for promoter expression in cultures and biofilms

Mean fluorescence values for P_{MSMEG_4515}-eGFP obtained under each growth condition were tested against the mean values obtained from P_{hsp60}-eGFP expression under the corresponding growth condition. Statistical significance in each case was determined by a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. Differences in eGFP expression levels were considered significant when P values ≤ 0.1, and are indicated in bold numbers.

Growth Condition	0 μM	2 μM	4 μM	50 μM
Planktonic Log Phase	**0.09**	0.40	0.23	**0.008**
Four-day Biofilm	**0.00005**	**0.005**	**0.0007**	**0.06**

2.3.1.4 Fluorescent Proteins and Biofilms

At the time that this study was initiated, the top green fluorescent proteins of choice were the GFP variants, enhanced GFP (eGFP) and Emerald Green (EmGFP) [90]. We chose eGFP due to prior success in laboratory use of this reporter, and successful green fluorescence detected from single-copy promoter fusions used in a colony assay (Figure 5B).

Our choice of a red fluorescent protein involved a few rounds of testing, before deciding on the mCherry construct that is reported in strains described in Table 3. Prior to using mCherry, we had tested DsRed2 as well as DsRed-Express for expression within biofilms; all three are derivatives of the original DsRed reporter. We obtained very weak signals using single-copy P_{hsp60}-DsRed2 and P_{hsp60}-DsRed-Express fusions, even when tested in planktonic cell cultures. As described earlier in this chapter, the mCherry fusion construct that we chose to use for our microscopy and flow cytometry-based experiments, carries a version of mCherry that is codon

optimized for use in the mycobacteria. Moreover, this cassette also contains the ribosome binding site from gp9 in phage TM4, which allows for stronger mCherry expression than from a cassette that lacks this modification (unpublished results from the Hatfull laboratory).

The complex structure of biofilms involves densely packed microcolonies encased in extracellular matrix material, and interstitial regions of voids that allow for flow of bulk solution and supply of nutrients and other metabolites across the film. These patterns of flow set up concentration gradients for nutrients across the biofilm, which in turn also determine the heterogeneity of cells within the biofilm [91]. Biofilm structure also sets up gradients of oxygen concentration across the film, with regions of high concentration as well as regions of low O_2 levels, or even anaerobic conditions [92]. The availability of oxygen is an important concern while studying gene expression using fluorescent proteins. FPs require molecular oxygen for optimal fluorophore maturation, and anaerobic conditions might pose a challenge to optimal functioning of these reporter proteins [90].

A study undertaken by Caroll *et al* to investigate the functional utility of mCherry in the mycobacteria, under conditions of oxygen depletion, or hypoxia, reported that the strength of fluorescence signal obtained from cultures grown under hypoxic conditions was comparable to signals obtained from aerobically grown liquid cultures [93]. The intensity of signal obtained from a P_{smyc}-mCherry fusion construct continued to increase even after three days, at which point the oxygen in the hypoxic culture condition was depleted. P_{smyc} represents the *M. smegmatis* rpsA (ribosomal protein-coding) promoter, with a *tetO* operator. The study concluded that mCherry can be used to reliably report promoter activity in *M. smegmatis* and *M. tuberculosis*, under hypoxic conditions [93].

Another study performed by Hansen *et al* to access GFP fluorescence in the low oxygen environment of biofilms of the Gram-positive oral bacterium *Streptococcus gordonii*¸ reported the functional use of this reporter under low oxygen tensions [94]. *S. gordonii* biofilms grown under aerobic conditions exhibit uniform fluorescence until a thickness of approximately 50 μm. When grown under anaerobic conditions, no fluorescence was observed in these films; however, when shifted to aerobic conditions, fluorescence was detected as early as four minutes post-shift. Fluorescence levels were restored to maximum intensity by twenty minutes post-shift. These observations were recorded by confocal microscopy. Overall, this study concludes that GFP can be used to effectively report up-regulation in promoter activity at levels of oxygen as low as 0.1 p.p.m. GFP produced at lower oxygen tensions can mature as fast as within 4-20 minutes, when exposed to even small amounts of oxygen prior to microscopy or other detection methods [94].

2.3.1.5 Heterogeneity in promoter induction within biofilms and in planktonic cells

As discussed in Chapter 1, one of the hallmarks of biofilms is the spatial difference in gene expression within the biofilm, with subsets of cells within a biofilm exhibiting heterogeneity in gene expression patterns.

Our confocal microscopy images of four-day biofilms grown in high iron show that a majority of cells express both eGFP and mCherry, indicating co-expression of P_{MSMEG_4515} and P_{hsp60} respectively. A subset of cells appear to exhibit only a single fluorescent signal - green or red (Figure 9C; inset panels). Flow cytometry-based analyses on biofilms grown under similar conditions enabled us to detect these subgroups and quantify their prevalence. Similar analyses were also performed on planktonic cultures.

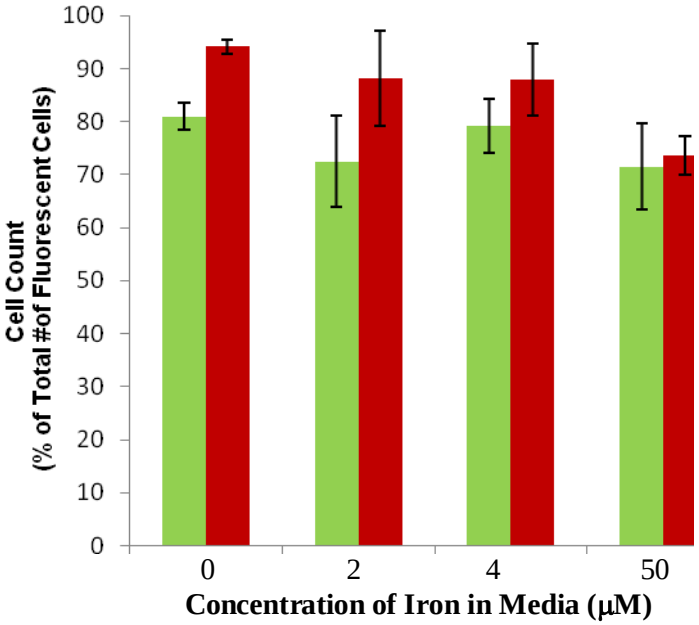
We observe that in planktonic cultures grown in the absence of iron, the number of cells expressing detectable eGFP (reporting P_{MSMEG_4515}) was similar to the number of cells expressing detectable mCherry (reporting P_{hsp60}). As observed in our microscopy studies, under high iron conditions, the overall number of cells expressing eGFP is significantly reduced (Figure 23B). In four-day biofilms, although we observe differences in the overall number of cells expressing eGFP and mCherry across iron concentrations, these differences are not found to be statistically significant for most conditions (Figure 23A).

When comparing changes in the number of cells that express eGFP, relative to increase in the amount of iron in the media, we find no significant change in four-day biofilms (Table 9A). In planktonic cells, we do find a significant reduction in the number of cells that express eGFP, when samples are grown in high iron (Table 10A). In the case of P_{hsp60}-mCherry, we find that planktonic cultures do not exhibit significant differences in the number of cells that express mCherry, when comparing samples grown in iron-deficient conditions with samples grown in high iron conditions Table 10B). We do observe a significant increase in the number of cells that express mCherry in four-day biofilms grown in high iron conditions, relative to the number of mCherry-expressing cells in lower iron (Table 9B).

Our analysis also enables us to investigate sub-populations of cells by fluorescent groups expressed (single- or dual-fluorophores expressed). In planktonic cultures grown in the absence of iron, we find that a significant number of cells express both mCherry and eGFP, confirming that most cells induce mycobactin biosynthesis under these conditions (Figure 24B). In high iron, a significant majority of the cells only express mCherry, confirming the iron-dependent regulation of P_{MSMEG_4515} (Figure 24B). In four-day biofilms, we find that the number of cells expressing both promoters is the predominant group in the absence of iron and in lower amounts

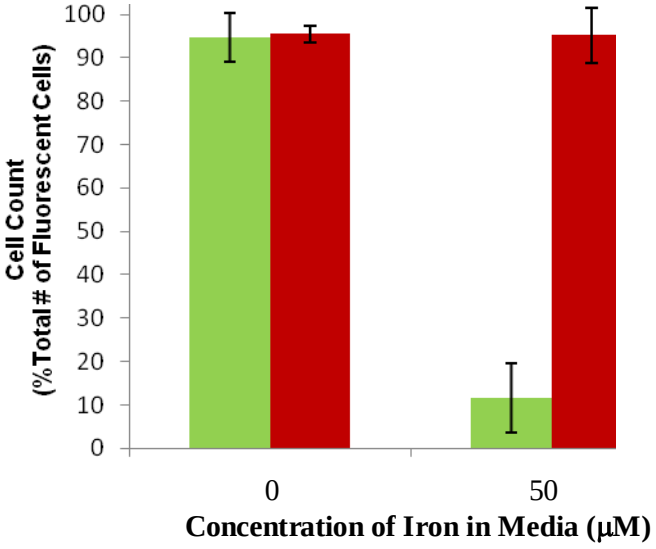
of iron; however, under high iron conditions the number of cells expressing only eGFP increases significantly (Figure 24A). Nevertheless, the number of cells that express only mCherry does not seem to change significantly in this condition. We observe that the population of cells in four-day biofilms that are grown in high iron consists of similar number cells that express *hsp60* alone, as in the lower iron conditions; however, fewer cells co-express *hsp60* and *MSMEG_4515*, and more cells express *MSMEG_4515* alone. Taken together, our data suggest heterogeneity in gene expression within sub-populations of cells in *M. smegmatis* biofilms, and that the expression of genes within the heterogeneous population changes when the growth conditions differ. Our overall flow cytometry-based data also suggest that P_{hsp60} is not a constitutively active promoter under biofilm conditions. In fact, we also observe that even under standard planktonic growth conditions, P_{hsp60} is not expressed at detectable levels in all cells, especially under high iron conditions of 50 μM Fe^{2+}.

A.



0 µM	2 µM	4 µM	50 µM
0.002	0.09	0.15	0.69

B.

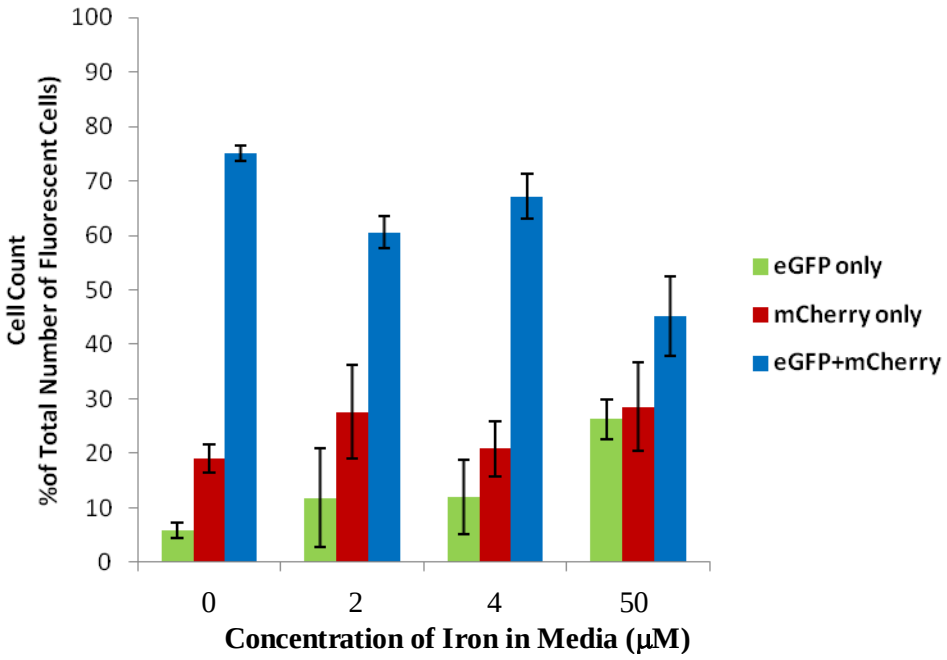


0 µM	50 µM
0.86	**0.007**

Figure 23. Total number of cells expressing candidate promoters in 4-day biofilms and planktonic cultures

Figure 23: Percentages of the total number of fluorescent cells that express green fluorescence (P_{MSMEG_4515}-eGFP) and red fluorescence (P_{hsp60}-mCherry) in four-day biofilms grown in 0, 2, 4 and 50 μM Fe^{2+} (**A**). Similar representation for planktonic cells in exponential phase, grown in 0 and 50 μM Fe^{2+} (**B**). Data are representative of three (A) and two (B) experimental replicates, and indicate standard deviation from the mean. Tables next to graphs provide P values for differences between the percentage of cells expressing green and red fluorescence at each concentration of iron, under each growth condition.

A.



B.

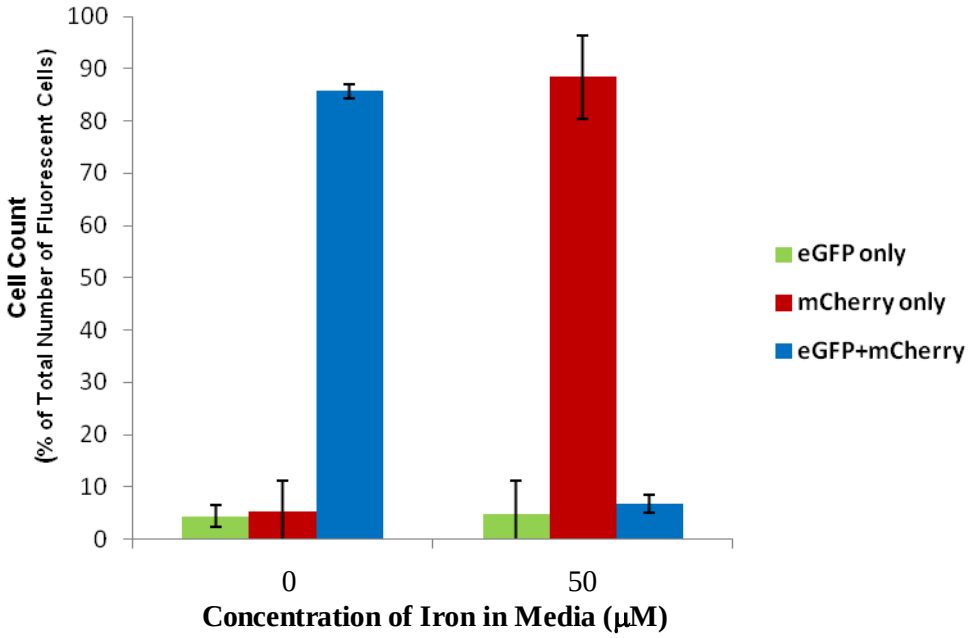


Figure 24. Fluorescent population profile in four-day biofilms and planktonic cultures

Figure 24: Percentages of the total number of fluorescent cells that express green fluorescence (P_{MSMEG_4515}-eGFP), red fluorescence (P_{hsp60}-mCherry) or dual-fluorescence, in four-day biofilms grown in 0, 2, 4 and 50 μM Fe^{2+} (**A**). Similar representation for planktonic cells in exponential phase, grown in 0 and 50 μM Fe^{2+} (**B**). Data are representative of three (A) and two (B) experimental replicates, and indicate standard deviation from the mean.

Table 9. Statistical significance for differences in fluorescent cell classes in four-day biofilms

Statistical significance in each case was determined by a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. No significant differences were observed for total number of green fluorescent cell counts in four-day biofilms (A). Differences in total number of red fluorescent cells in four-day biofilms were considered significant when P values ≤ 0.05, and are indicated in bold numbers for data (B).

A.

Green (P_{MSMEG_4515}-eGFP)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.17	0.63	0.13
2 μM	0.17	1	0.31	0.90
3 μM	0.63	0.31	1	0.24
50 μM	0.13	0.90	0.24	1

B.

Red (P_{hsp60}-eGFP)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.32	0.20	**0.0008**
2 μM	0.32	1	0.98	**0.06**
3 μM	0.17	0.98	1	**0.03**
50 μM	**0.0008**	**0.06**	**0.033**	1

Table 10. Statistical significance for fluorescent cell classes in planktonic cultures

Statistical significance in each case was determined by a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. Differences in total number of green and red fluorescent cells in four-day biofilms were considered significant when values ≤ 0.05, and are indicated in bold numbers for data (A and B, respectively).

A.

Green (P_{MSMEG_4515}-eGFP)	**0 μM**
0 μM	1
50 μM	**0.007**

B.

Red (P_{hsp60}-mCherry)	**0 μM**
0 μM	1
50 μM	0.95

Table 11. Statistical significance for cell classes in biofilms and planktonic cultures

Statistical significance in each case was determined by a Student's t-Test, with a two-tailed distribution, to obtain probability values for each pair. Population was assumed to be normally distributed, and each pair of samples was assumed to be of equal variance. Differences in total number of each class of fluorescent cells were considered significant when P values ≤ 0.05, and are indicated in bold numbers for data (A, B, C for four-day biofilms; D, E, F for planktonic cultures).

A.

Green only (P_{MSMEG_4515}-eGFP)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.32	0.20	**0.0008**
2 μM	0.32	1	0.98	**0.06**
4 μM	0.20	0.98	1	**0.03**
50 μM	**0.0008**	**0.06**	**0.03**	1

B.

Red only (P_{hsp60}-mCherry)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	0.18	0.63	0.13
2 μM	0.18	1	0.31	0.90
4 μM	0.63	0.31	1	0.24
50 μM	0.13	0.90	0.24	1

C.

Green + Red (both promoters)	**0 μM**	**2 μM**	**4 μM**	**50 μM**
0 μM	1	**0.001**	**0.03**	**0.002**
2 μM	**0.001**	1	0.08	**0.03**
4 μM	**0.03**	0.08	1	**0.01**
50 μM	**0.002**	**0.03**	**0.01**	1

D.

Green only	**0 μM**	**50 μM**
0 μM	1	0.93
50 μM	0.93	1

E.

Red only	**0 μM**	**50 μM**
0 μM	1	**0.007**
50 μM	**0.007**	1

F.

Gr + Rd	**0 μM**	**50 μM**
0 μM	1	**0.0004**
50 μM	**0.0004**	1

2.4 CONCLUSIONS AND FUTURE DIRECTIONS

This chapter presents work that was conceived based on observations made by Ojha and Hatfull in a study that provided the first instance of a comprehensive transcription profile on *M. smegmatis* biofilms [61]. Our work focuses on profiling and characterizing the expression of the mycobactin biosynthesis gene, *MSMEG_4515*, within *M. smegmatis* biofilms. Firstly, we have developed a stable set of fluorescent reporter vectors that are effective for use in imaging *M. smegmatis* biofilms by confocal microscopy. Secondly, using these reporter constructs, we have provided a detailed expression profile for *MSMEG_4515* in mature biofilms. We have also provided a comparative analysis of *MSMEG_4515* in planktonic cultures. These experiments have provided evidence for differential gene expression in biofilms and planktonic cells, even when cultured under identical nutrient conditions. Lastly, our results present evidence for the presence of subsets of cells within the biofilm, that exhibit heterogeneity in gene expression patterns. Although this aspect of biofilms has been well illustrated in other organisms, this is the first instance of a study that addresses heterogeneity in gene expression in *M. smegmatis* biofilm populations.

Our confocal microscopy and flow cytometry-based experiments confirm the iron-dependent regulation of *MSMEG_4515* in planktonic cultures. These studies also provide evidence for differential patterns of expression of *MSMEG_4515* in biofilms and planktonic cells. We find that *MSMEG_4515* continues to be expressed in biofilms even under high iron conditions that shut off expression in planktonic cells. The Ojha and Hatfull study did observe *MSMEG_4515* induction in four-day biofilms grown in 2 μM iron [61]. In quantifying *MSMEG_4515* gene expression in 50 μM iron by RT-PCR, the expression levels were

normalized to those observed in 2 μM iron [61], thus making their experimental analysis different from ours for comparison purposes. In using the *hsp60* promoter as a control for these experiments, we have also provided an expression profile for this gene within mature biofilms and in planktonic cultures. Our results indicate that *hsp60* expression in *M. smegmatis* is not entirely constitutive, and that the levels of expression differ between low and high iron environments.

M. smegmatis biofilms have been observed to exhibit variability in their robustness and exact timing of development. As a result of this biological variability, the absolute fluorescence distribution values can tend to vary across independent experimental repeats. Variability in biofilms can arise from differences in the starting population of cells that seeded the biofilm, variations arising during the course of biofilm development, differences within microcolonies that can form within the overall biofilm structure, or even sample bias introduced by the methodology employed for sample collection. Although these changes affect absolute values and distributions, the general pattern of induction levels observed as a compilation of several experimental repeats should provide a reliable estimate of the trend in gene expression.

2.4.1 Modular fluorescent reporter constructs as a tool to perform single-copy expression studies

The integrative fluorescent reporters that were generated for this study were optimized over several trials to achieve the current constructs. Our work with these vectors demonstrates that they can be used successfully to achieve detectable fluorescence read-outs from a single-copy fusion. We also observe that these constructs are most effective when reporting promoters that are strongly induced. In the case of the Tweety-based eGFP reporter construct, we obtained

significantly lower intensity of fluorescence from the characteristically strong P_{hsp60} promoter, when compared to the fluorescence intensity achieved by extrachromosomal replicative plasmids carrying the same promoter cassette. In light of this result, we conclude that this construct would be less than ideal for use with weak promoters or under very low levels of induction. Nevertheless, the modular Tweety and Giles-based fluorescent vectors provide a powerful tool for single-copy expression studies that utilize sensitive imaging techniques, and for experiments that require a stable, integrated reporter fusion. The integrative mCherry reporter construct optimized for our microscopy and flow cytometry experiments provide significantly stronger signals using the same *hsp60* promoter cassette; however, the expression pattern remains consistent with that obtained from the eGFP construct (Figure 25). This reporter construct can be successfully exploited for use under standard assay conditions that require single-copy expression.

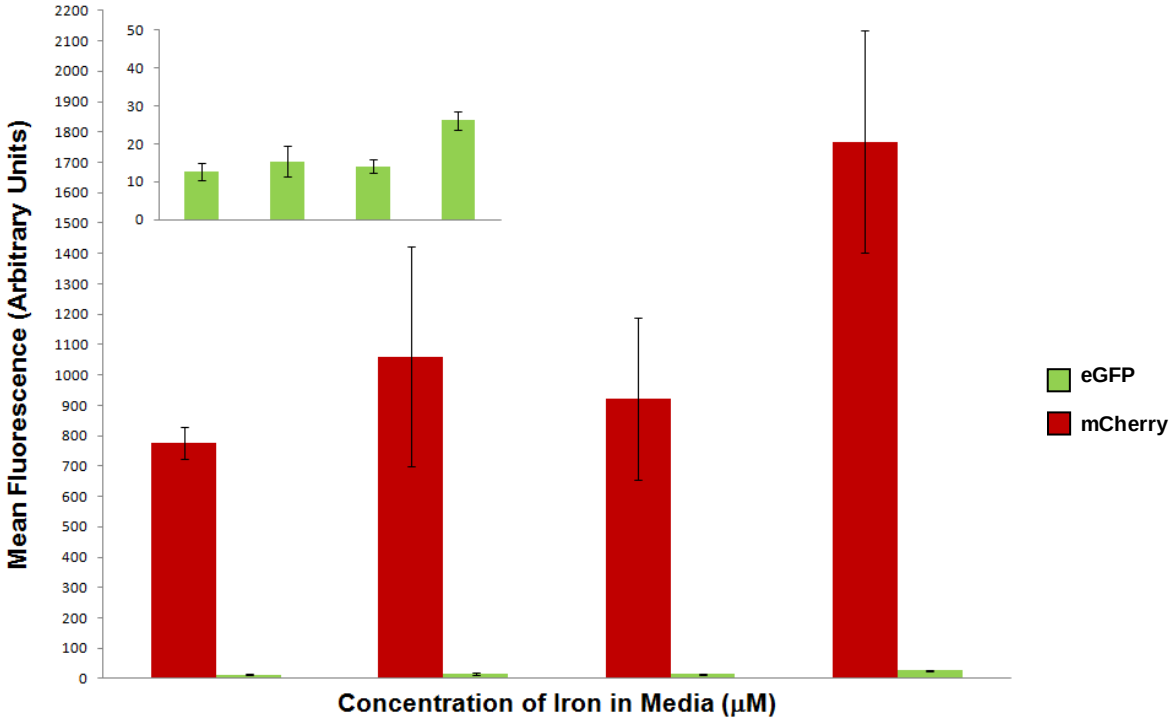


Figure 25. Comparative fluorescence intensities using eGFP and mCherry vector constructs

Figure 25: Mean fluorescence values for P_{hsp60}-mCherry and P_{hsp60}-eGFP in cells recovered from four-day biofilms grown in different concentrations of iron. Inset shows the plot for P_{hsp60}-eGFP, on a Y-axis range that allows for better resolution of these data points. Data represents three experimental replicates and denote standard deviation from the mean.

2.4.2 Mycobatin biosynthesis in mature biofilms

We present evidence that shows that unlike planktonic cultures where mycobactin biosynthesis is down-regulated in the presence of high amounts of iron, mature biofilms continue to induce mycobactin biosynthesis in the presence of similarly high amounts of iron. We also show that the percentage of cells expressing *MSMEG_4515* in the biofilm does not change

significantly with increase in iron concentration in the growth medium. These results suggest that cells within mature biofilms remain starved of the available iron in the media.

2.4.3 P_{hsp60} expression in mature four-day biofilms

In using P_{hsp60} as a control for the iron-responsive regulation of *MSMEG_4515*, we show that its expression is not completely constitutive in biofilms. Mature four-day biofilms contain a subset of cells that do not express P_{hsp60}-mCherry. We also observe an increase in P_{hsp60} expression in biofilms containing high amounts of iron, indicting possible stress-response induction at these conditions.

2.4.4 Future Directions

Our gene expression analyses on *M. smegmatis* biofilms relies on fluorescent proteins to report promoter induction under the growth conditions that were tested. Both eGFP and mCherry as fluorescent reporters exhibit slow protein turn-over, which could impact inferences derived from gene expression studies. This is an especially valid concern during prolonged studies such as one involving gene expression profiling over the course of biofilm development. Using fluorescent protein variants that have been engineered to have shorter turn-over times would be one possible way to address the above challenge. The development and use of a destabilized eGFP variant (dEGFP) with a half-life as short as two hours has been described successfully in mammalian cells [95].

Flow cytometry provides evidence for the presence of sub-populations of cells within the mature biofilm that exhibit differential gene expression patterns. The single fluorescent cell counts determined for this study provide a general snapshot of subpopulations of cells within the biofilm that exhibit different patterns and/or levels of gene expression. For a more convincing and accurate reflection of exact proportions of fluorescent classes, we will need to apply more stringent Ferdinand-based analyses for cell classification.

To study the spatial distribution of these subsets of cells, biofilms grown on solid surfaces (composed of biofilm media reinforced with a small amount of agarose) can be used for confocal microscopy following freezing and cryo-sectioning of the samples. A study published in 2008, on population heterogeneity within *B. subtilis* biofilms, reported the successful use of the aforementioned methods in tracking different subsets of cells within the biofilm [42].

The argument that cells in *M. smegmatis* biofilms grown in high iron media likely remain starved for iron can be tested by measuring the intracellular iron in these samples. A colorimetric ferrozine-based assay that was developed by Riemer *et al* to measure iron in cultured astrocytes, can be optimized for use with our mycobacterial samples. This ferrozine-based assay can be used to measure iron over a range of concentrations, from 0.2-30 nmol [96].

The presence of a lipid-rich extracellular matrix - a characteristic of *M. smegmatis* biofilms [61] - can be hypothesized to play a role in preventing cells from accessing the otherwise abundant iron in the media. The role of the extracellular matrix in restricting access to iron in the media can be partly addressed by using mutants that are defective in biofilm maturation, and lack a significant matrix.

3.0 EXPERIMENTAL AND BIOINFORMATIC ANALYSES OF *MSMEG_1240* TO INVESTIGATE ITS ROLE IN SLIDING MOTILITY IN *M. SMEGMATIS*

3.1 INTRODUCTION

Mycobacterial surface translocation was first reported in 1999 by the Kolter laboratory, with the demonstration that *M. smegmatis*, as well as the slow-growing opportunistic mycobacterium *M. avium*, exhibits a spreading phenotype on moist surfaces [70]. This surface spreading phenomenon does not depend on flagella, pili, or fimbrae, but instead involves passive translocation that is dependent on the expansive forces generated by growing cell fronts [70]. The next study on sliding motility involved a transposon mutagenesis screen for translocation-defective mutants in *M. smegmatis*. This screen identified twenty mutants carrying insertions that mapped to the *mps* gene. The mps gene product is involved in the biosynthesis of glycopeptidolipids (GPLs), which are amphiphilic glycosylated peptidolipids that are found in the cell envelope [72]. Mutants lacking GPLs were defective in sliding motility. These GPL-defective mutants were also found to be defective in forming biofilms on PVC, suggesting that GPLs were important for both processes to occur with full proficiency. Another translocation-defective mutant that was identified in the same screen carried the transposon insertion in its *tmptC* gene. TmptC is putatively involved in transporting GPLs to the cell envelope, and potentially also in GPL biosynthesis [72]. Another study by the Kolter lab identified a transposon

insertion in *atf1*, a gene that encodes an acetyltransferase that is involved in GPL biosynthesis. The GPLs in this mutant strain are non-acetylated, and the mutant shows partial defects in sliding motility and biofilm formation [55]. Figure 26 draws from the three aforementioned studies, and illustrates the proposed model for the role of GPLs in facilitating sliding motility on an agarose surface as well as biofilm formation on a PVC surface.

Further studies have indicated that a deletion or insertion in the *lsr2* gene, which encodes a small histone-like DNA-binding protein [97, 98], de-regulates the *mps* operon, causing hypermotility on agarose [99, 100]. This observation implicates a role for Lsr2 in GPL synthesis. Studies have also shown that a lack of inorganic polyphosphates (polyP) in the cell can affect fatty acid distribution on the cell wall, leading to differences in the ability to attach to or slide on a surface [101]. In 2007, *M. marinum* was also shown to be able to slide on a surface; although this bacterium does not synthesize GPLs, it generates antigenic glycolipids known as lipooligosaccharides (LOSs) which facilitate sliding [57].

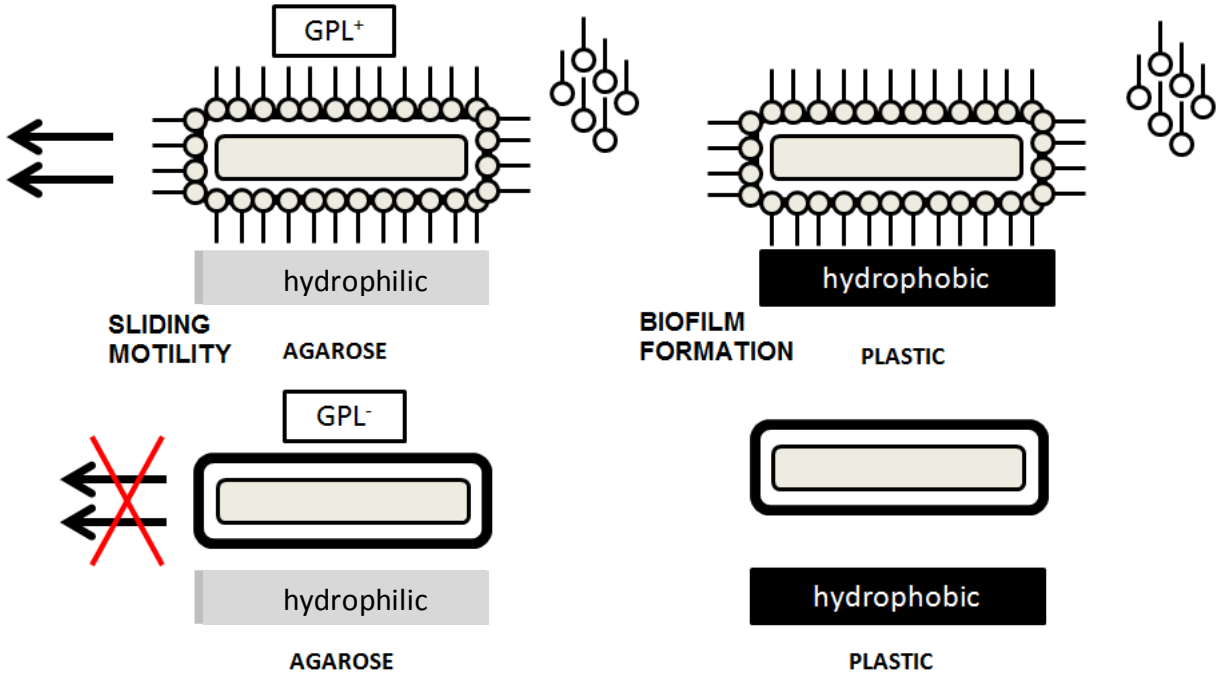


Figure 26. Role of GPLs in Sliding Motility and Biofilm Formation

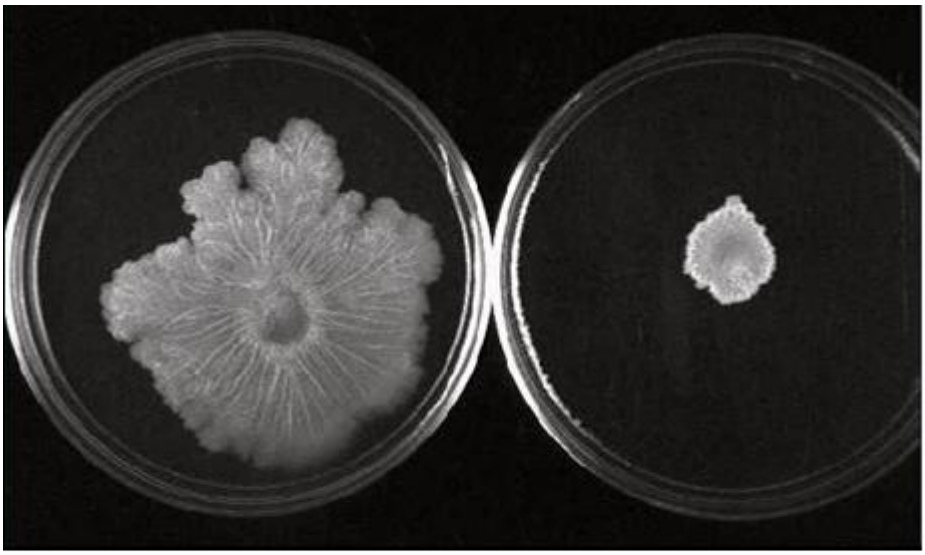
Figure 26: This schematic (adapted from Recht *et al*, 2000) [72] suggests that in the presence of GPLs on the cell envelope, the exposed hydrophobic fatty acid tails render the cell hydrophobic, and therefore unable to attach to a hydrophilic surface, such as agarose. As a consequence of the reduced friction between the bacterial cell and the agarose surface, the cell can slide over the surface. On the other hand, in the absence of GPLs, the hydrophilic capsular polysaccharides on the cell envelope are exposed, not allowing for interactions with the hydrophobic PVC surface, but instead with the hydrophilic agarose surface.

3.2 IDENTIFICATION OF A MUTANT DEFECTIVE IN SLIDING MOTILITY, YET PROFICIENT IN BIOFILM FORMATION

The sliding motility mutants screened and characterized since the original observation of surface translocation in *M. smegmatis*, are all also observed to lack the ability to form biofilms. Therefore, the ability to translocate on a surface and the ability to attach to surfaces to form biofilms, have been linked [102].

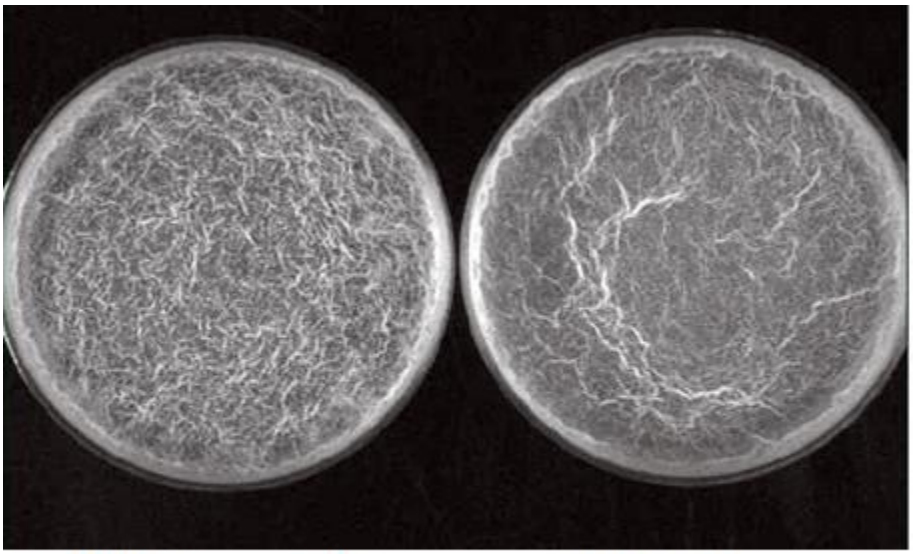
In considering sliding motility assays as a surrogate for biofilm assays, especially for ease of fluorescence microscopy, we decided to further investigate the link between the two surface phenomena. To do this, we performed a transposon mutagenesis screen, as described in Section 5.8 of Materials and Methods. Each transductant was screened for the ability to slide on a moist agarose-reinforced surface consisting of biofilm media that contained a limited carbon source. This screen was set up similar to the one performed by the Kolter lab, in the initial identification of hypomotile mutants [70]. Each mutant was also separately screened for the ability to attach to a surface (polystyrene petridishes) and form mature biofilms. Of 500 initial mutants screened, we identified one mutant that was severely defective in its ability to slide on agarose, as compared to the parent strain (Figure 27A). Although hypomotile, this mutant retained the ability to form mature biofilms (Figure 27B). The initial 500 mutants screened represented an approximately 7.2% coverage of the *M. smegmatis* genome. We similarly screened a few hundred additional transductants, and identified two hypermotile mutants; however, all further studies were performed only on the first hypomotile mutant, owing to the novel observation of distinct proficiencies in sliding motility and biofilm formation.

A.



Wild-type mc^{2}155 mc^{2}155 ::Tn*Mariner*

B.



Wild-type mc^{2}155 mc^{2}155 ::Tn*Mariner*

Figure 27. mc^{2}155::Tn*Mariner* is severely defective in sliding motility, but forms mature biofilms

Figure 27: mc^{2}155::Tn*Mariner* sliding motility and biofilm formation, when compared to wild type strain mc^{2}155. **A.** The mc^{2}155::Tn*Mariner* mutant is severely defective in sliding motility when compared to the wild type strain mc^{2}155. The growth observed is at 10 days post-inoculation, at 37ºC. Sliding motility is limited by the moisture content of the growth plates, and this can cause experimental variations in the extent of translocation. The qualitative comparison that is shown here compares the two strains when grown under similar conditions of temperature, humidity and media, in one representative experiment. **B.** The mc^{2}155::Tn*Mariner* mutant forms mature biofilms in standard biofilm media containing 4 μM iron. Growth observed is at 10 days post-inoculation, at 30ºC. mc^{2}155::Tn*Mariner* biofilms show pellicle-like structures, demonstrating proficiency in biofilm maturation. Wild type *M. smegmatis* biofilms show variability in the robustness of pellicles formed; nevertheless, the presence of pellicle-like structures indicates a mature biofilm.

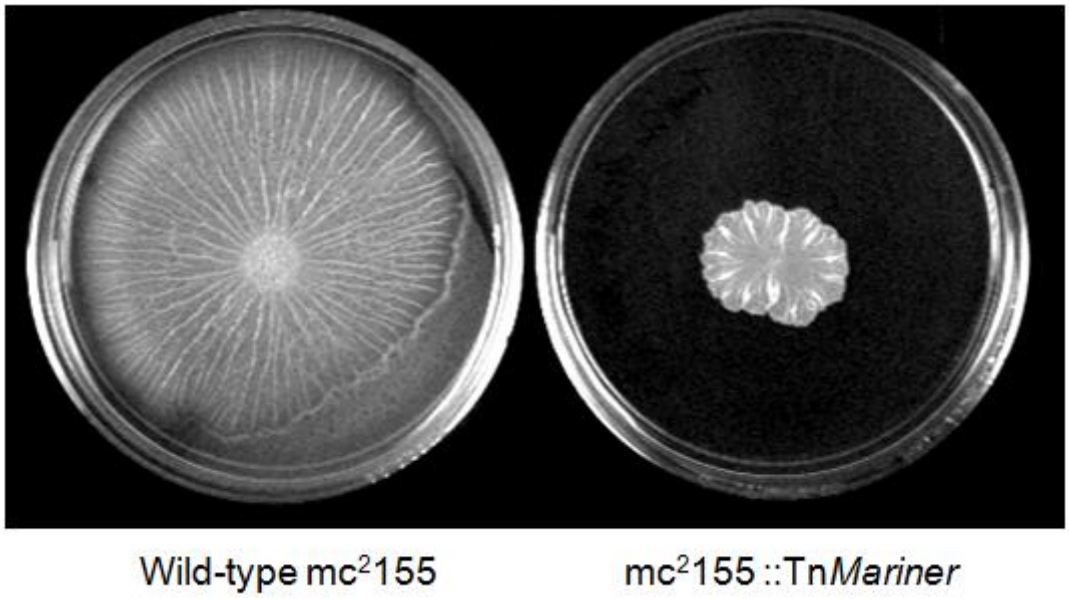


Figure 28. Supplemental iron does not rescue surface translocation defect

Figure 28: mc^{2}155::Tn*Mariner* and wild type strain mc^{2}155 sliding motility on supplemental iron. Plates contain 50 μM iron, and growth is observed 10 days post-inoculation, at 37ºC. Wild type shows robust motility; however, the mutant remains defective in surface translocation. These plates are representative of repeated observations.

Previous studies have indicated the requirement for iron in sliding motility [61]. Ojha and Hatfull have previously reported that wild-type mc^{2}155 is proficient in sliding motility on plates containing 2 μM iron; however, in the absence of additional iron in the media, this strain is significantly impaired in the extent of sliding observed over the same amount of time. They also showed that a mutant strain that carries a deletion in the iron uptake machinery was found to be defective in sliding even in the presence of 2 μM iron, thus reinforcing the importance of iron in facilitating sliding [61]. This mutant was also found to be defective in biofilm formation at 2 μM iron; however, the addition of 50 μM iron to the growth medium rescued this biofilm defect [61].

To rule out accessibility of iron as a reason for the defective surface translocation phenotype observed in the transposon mutant, the motility assay was repeated on media reinforced with 50 μM iron. The mutant remained impaired in sliding, therefore ruling out accessibility to iron as the cause for the observed translocation defect (Figure 28).

3.3 CHARACTERIZATION OF MC^{2}155:TN*MARINER*

3.3.1 Insertion site identification

Junction PCR, followed by sequencing as described in Section 5.8.7 of Materials and Methods, showed that the transposon insertion site lay within *MSMEG_1240*, a gene encoding a conserved hypothetical protein in *M. smegmatis*. A BLASTP search revealed only one homolog for the 528 amino acid sequence of MSMEG_1240 within the other mycobacteria: MCOL_00565 (612 aa) Figure 31, Table 12). MCOL_00565 is a putative MSMEG_1240 homolog found in *Mycobacterium colombiense*, which is a recently classified slow-growing, non-tuberculous member of the *M. avium* complex (MAC) [103]. MCOL_00565 is also classified as a protein of unknown function.

3.3.2 Deletion of *MSMEG_1240* does not cause hypomotility

To investigate the role of *MSMEG_1240* in surface translocation, we constructed an in-frame deletion mutant lacking the entire coding region of *MSMEG_1240*. We used mycobacterial recombineering to create this mutant [104]. Briefly, first, we constructed a deletion/targeting

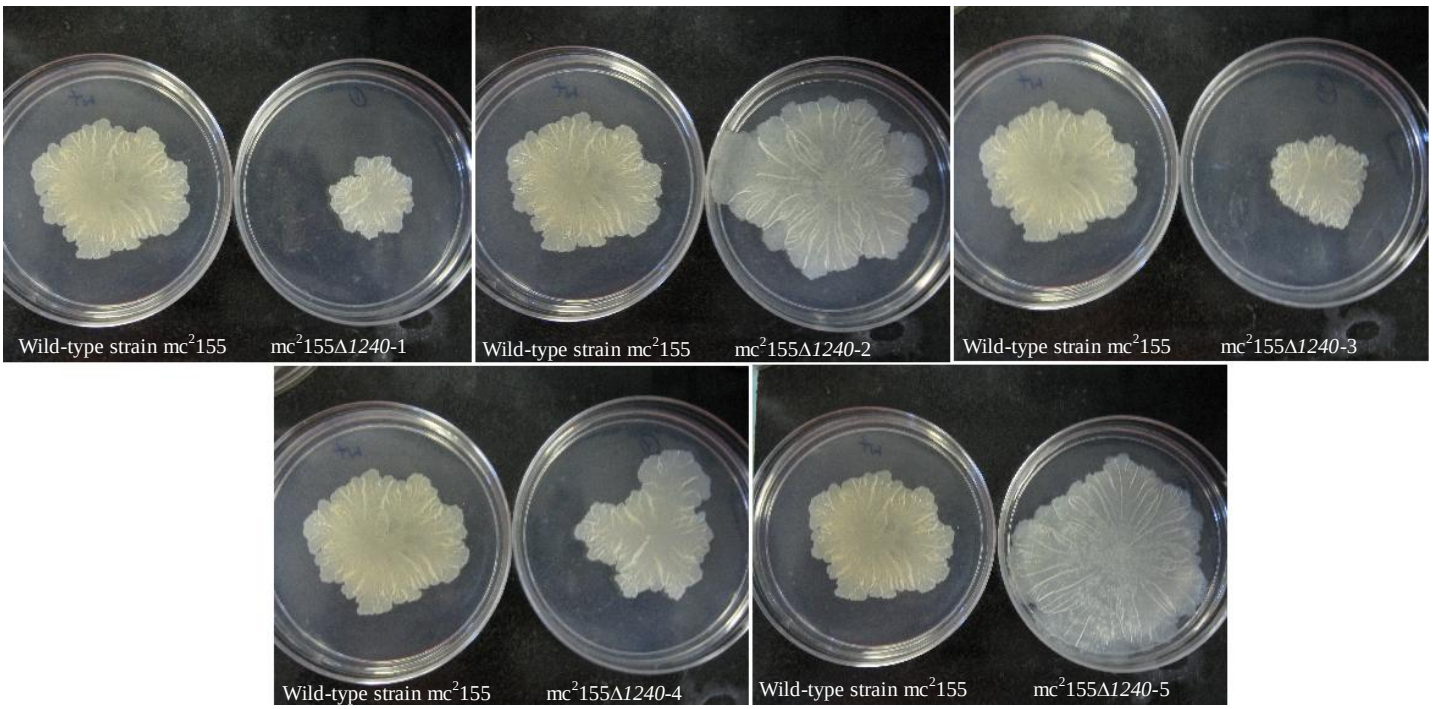
substrate that was designed to facilitate the replacement of the targeted region on the *M. smegmatis* chromosome, with a *hyg*-resistance cassette flanked by γσ resolvase sites. Primers were designed to amplify regions of the bacterial chromosome that were homologous to regions approximately 500 bp upstream and downstream of *MSMEG_1240,* including approximately 100 bp on the ends of the gene itself, as a caution against the deletion causing polar effects. Two separate PCRs were performed to generate the two homologous ends of the targeting substrate, using genomic DNA as template. Following PCR clean-up, these fragments were digested with restriction enzymes at sites engineered to match sites on the cloning vector used to generate the complete targeting substrate. The two PCR fragments were ligated in two separate steps into pYUB854, a vector carrying the hyg^R cassette flanked by resolvase sites that allow for ease of unmarking of the mutant if necessary. This vector, now carrying the targeting substrate for *MSMEG_1240,* was linearized, the reaction was cleaned up using the QIAGEN QIAquick PCR clean-up protocol, and quantified on a 0.8% agarose gel. 100 ng of this substrate was electroporated into *M. smegmatis* mc^2155 containing the recombineering plasmid, pJV53. pJV53 is an inducible vector that allows for the expression of proteins that promote homologous recombination between the bacterial chromosomal DNA and the targeting substrate [105]. The entire electroporation recovery mixture was plated on 7H10+ADC+Kan+Hyg plates, and five colonies were picked, cultured and screened for resistance to kanamycin and hygromycin. The putative mutants were cured of the pJV53 plasmid, re-tested for hygromycin resistance, and then purified and screened by PCR, using primers flanking the region targeted for deletion. Following confirmation of the deletion of *MSMEG_1240*, each of the five mutants was tested for proficiency in sliding motility.

Surprisingly, unlike the severely defective phenotype observed in mc^{2}155::Tn*Mariner* sliding motility, each of the five mc^{2}155Δ*MSMEG_1240* strains (strains 1-5) was comparable to the wild type strain in its ability to translocate on a 0.3% agarose surface (Figure 29A). The five mc^{2}155Δ*MSMEG_1240* strains shown in Figure 29A contained the pJV53 plasmid. The first two panels in Figure 29B show wild type strain mc^{2}155 alongside a representative mc^{2}155Δ*MSMEG_1240* strain (strain 3) after it had been cured of pJV53. Both strains look similar to one another and spread out much farther than the hypomotile mc^{2}155::Tn*Mariner* strain.

3.3.3 Complementation of strain mc^{2}155::Tn*Mariner*

In addition to constructing and testing the mc^{2}155Δ*MSMEG_1240* deletion strain as described above, we also tried to genetically complement mc^{2}155::Tn*Mariner* with *MSMEG_1240*, and tested for restoration of proficiency in surface translocation. To do this, we cloned the entire *MSMEG_1240* region into pMH94, a previously constructed L5-based integrative plasmid that was available in the lab [106], thereby introducing a single copy of this region into the bacterial chromosome at the L5 phage integration site. However, this method did not complement the defect in sliding motility observed in strain mc^{2}155::Tn*Mariner* (Figure 30).

A.



B.

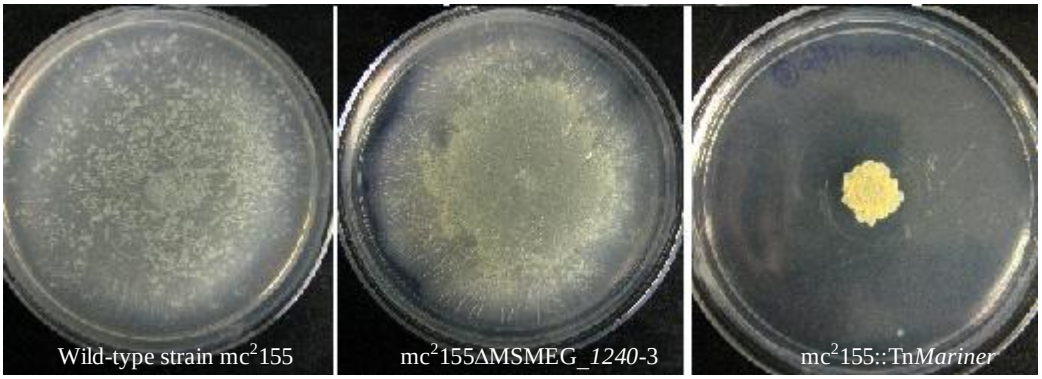


Figure 29. Strain mc²155Δ*MSMEG_1240* is not defective in surface translocation

Figure 29: Sliding motility in strain mc^{2}155Δ*MSMEG_1240* compared to wild type strain mc^{2}155.
A. Each panel shows the translocation radii of wild type mc^{2}155 as compared to the mc^{2}155Δ*MSMEG_1240* strain, at 5 days of growth at 37ºC. mc^{2}155Δ*MSMEG_1240*-1 through mc^{2}155Δ*MSMEG_1240*-5 represent five separate colonies screened for and confirmed to have the *MSMEG_1240* deletion. **B.** Wild type mc^{2}155, mc^{2}155Δ*MSMEG_1240*-3 and mc^{2}155::Tn*Mariner* grown at 37ºC, in a separate experiment than in Panel A; exhibits variability by experimental batch, introduced by experimental variation in inoculation, plate moisture and/or humidity conditions during incubation.

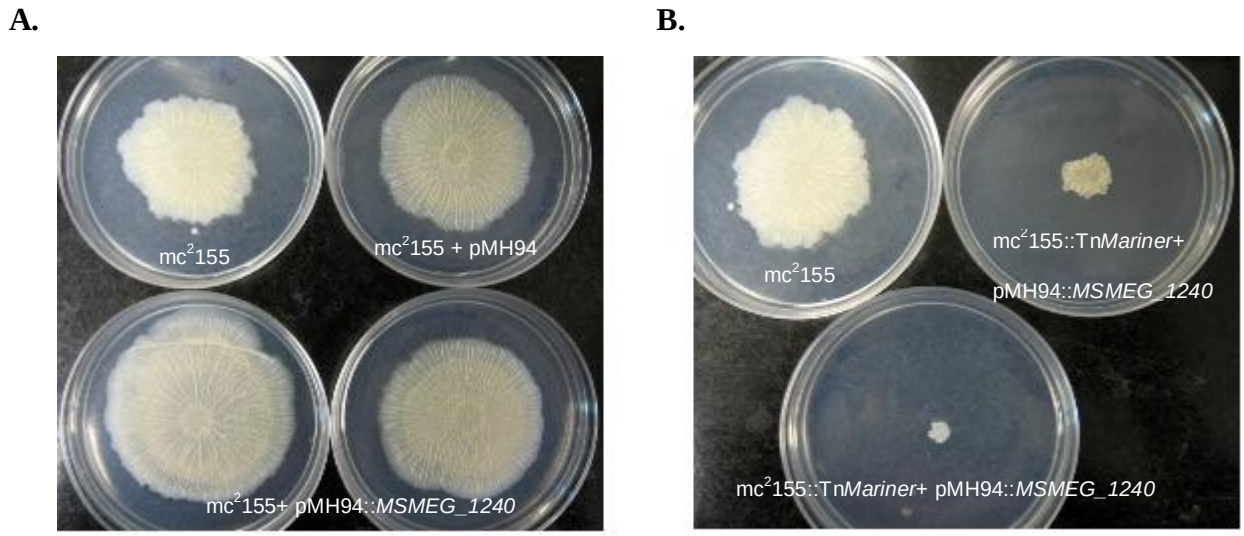


Figure 30. Complementation with *MSMEG_1240* failed to rescue the translocation defect

Figure 30: Sliding motility in strain mc²155::Tn*Mariner* complemented with a single copy of *MSMEG_1240*. **A.** Sliding motility in wild-type strain mc²155 carrying the integration-proficient vector pMH94 (empty vector) and the complementing plasmid pMH94::*MSMEG_1240* (two representative plates) **B.** Sliding motility in strain mc²155::Tn*Mariner* carrying the complementing plasmid pMH94::*MSMEG_1240* (two representative plates).

3.3.4 Difficulties in curing strain mc²155::Tn*Mariner*

Following the initial round of purification of the transposon insertion mutant strain mc²155::Tn*Mariner*, all other experimental manipulations, such as insertion mapping and complementation were performed using genomic DNA or cells obtained from this strain. Previous studies that have investigated phage infectivity on strains with altered surface translocation phenotypes have observed changes in the susceptibility of these strains to phage infection. These strains have typically been defective in a cell-wall or surface component, and the difference in phage infectivity includes changes such as resistance to certain phages which

otherwise infect the wild type parent strain [56, 99]. To test mc^{2}155::Tn*Mariner* for the infectivity of a representative series from our laboratory collection of mycobacteriophages, we grew the strain in liquid culture in the absence of Tween-80, at 37ºC. Our first indication of phage release in strain mc^{2}155::Tn*Mariner* cultures coincided with the culturing of this strain in Tween-free media. We also observed the incidence of plaques in the primary streak line of plates incubated at 37ºC. We tried curing this strain by passaging it several times in media containing Tween-80; however, these attempts proved to be largely unsuccessful in completely eliminating the phage. A PCR performed on this strain using TM4-specific primers revealed bands coincident with a control PCR performed on TM4 genomic DNA, indicating that the transposon delivery phage was the likely source of contamination/phage release.

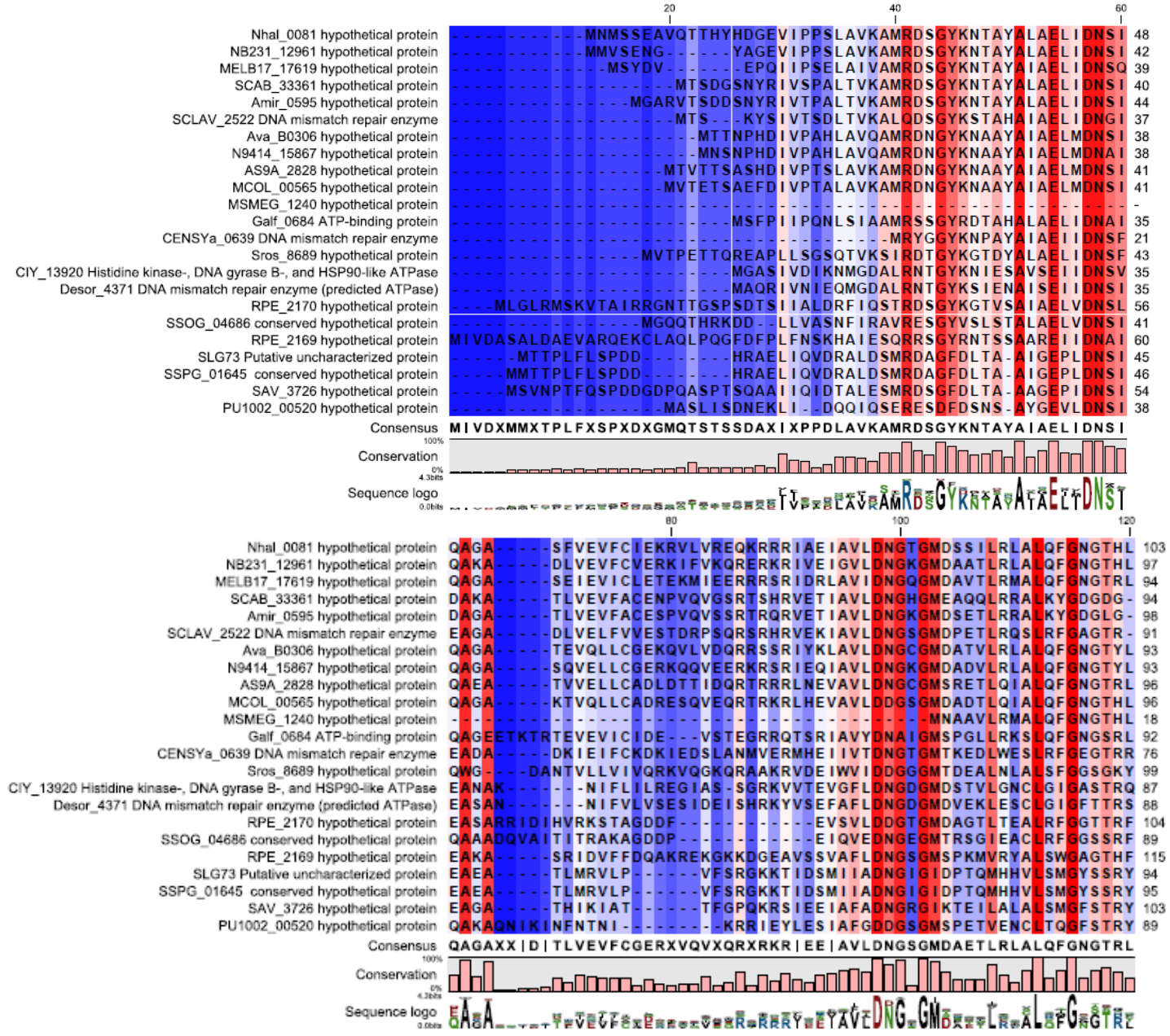
To rule out the possibility of a second-site insertion as the cause for the hypomotile phenotype, as well as to eliminate the problem of phage in the strain, we decided to tranduce the insertion in mc^{2}155::Tn*Mariner* into a wild type mc^{2}155 strain background. We used the generalized transduction methodology described by Lee *et al* [107]. The transduction protocol first called for propagation of the transducing phage, Bxz1, on the mutant strain. Strain mc^{2}155::Tn*Mariner* was grown in 7H9+ADC+$CaCl_2$ without Tween, and the culture was allowed to grow for a longer duration than usual to reach the desired cell density, due to clearing observed from the phage. The overgrown culture was passed through a 0.22 μM filter to break up clumps, and then used as described in Section 5.8.1 of Materials and Methods, to create top-agar lawns. Due to the continued presence of phage in the strain, we were unsuccessful in generating lawns suitable for phage propagation using this method, and therefore, unsuccessful in our attempt to move the insertion into a clean background to study the confirmed singular effect of this insertion.

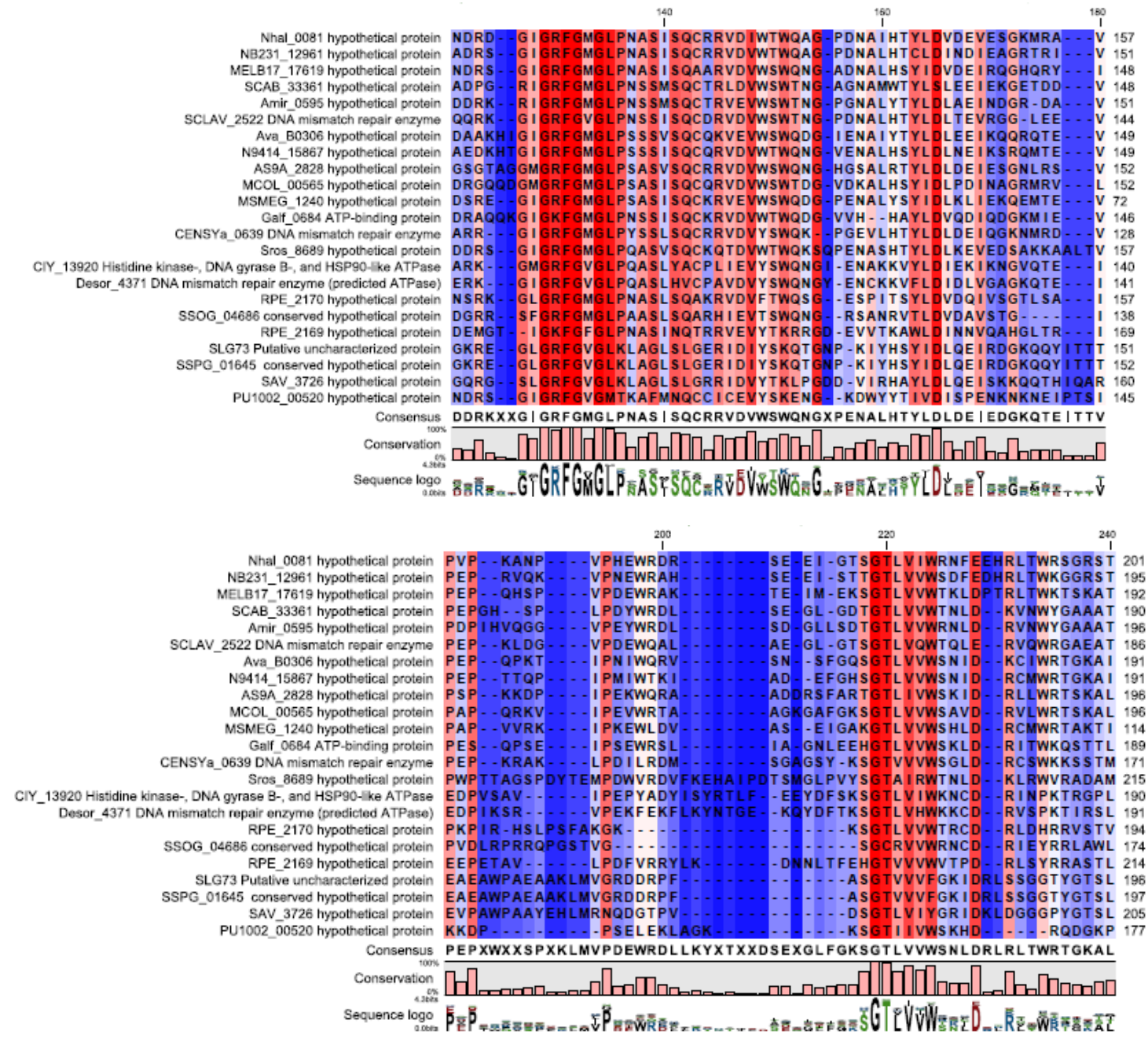
3.4 BIOINFORMATIC CHARACTERIZATION OF *MSMEG_1240*

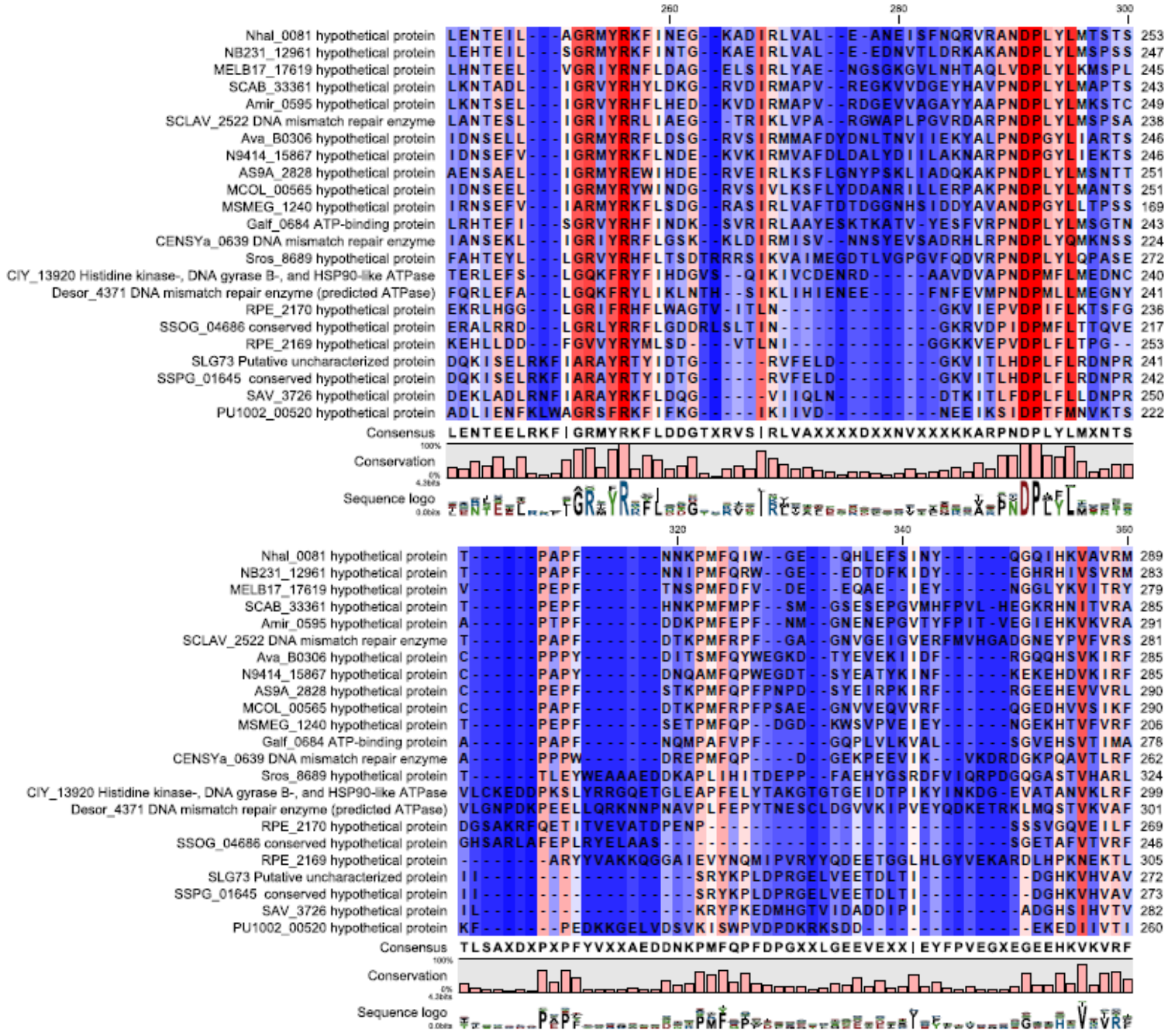
3.4.1 Genomic Arrangement and Homologs

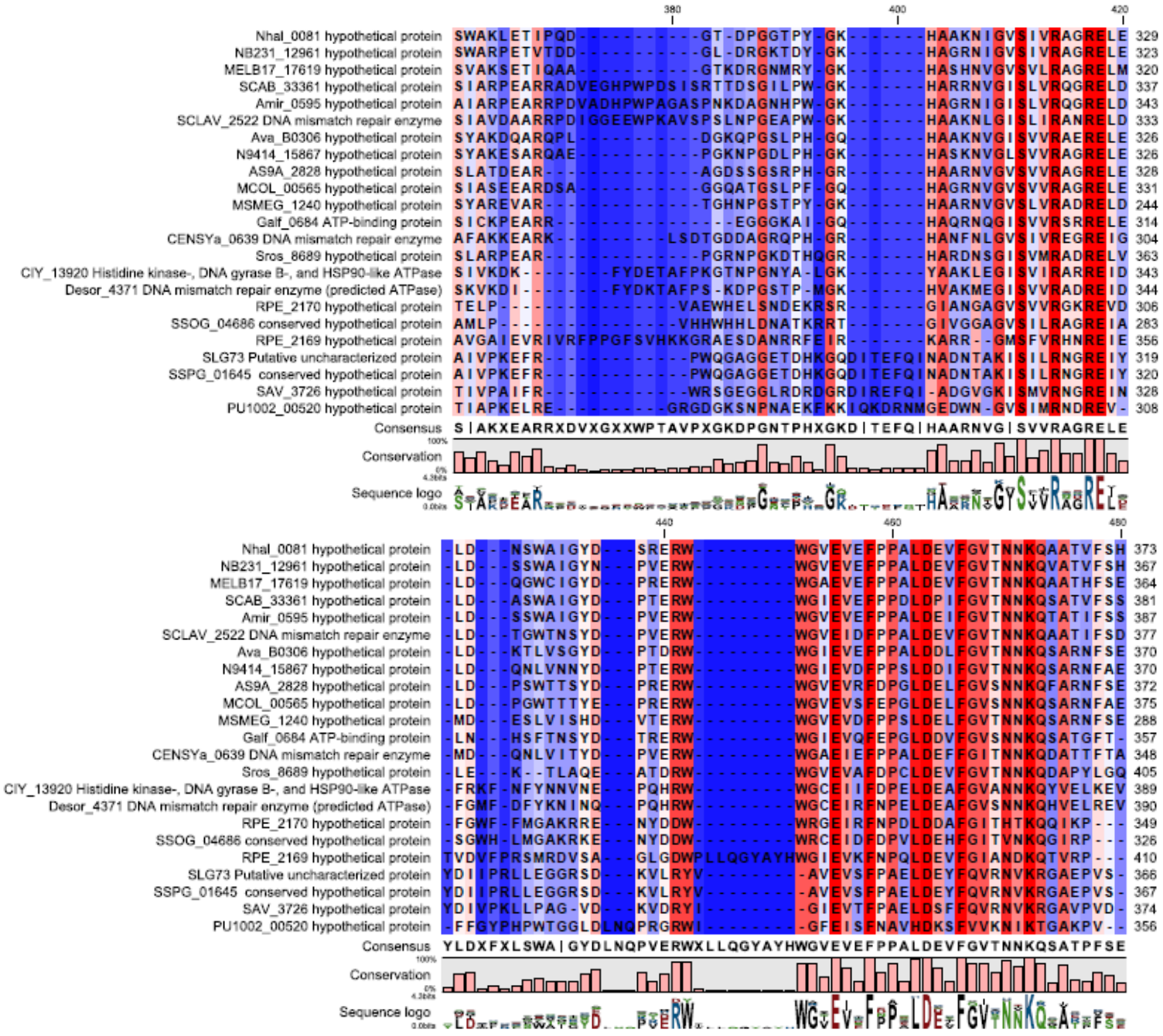
MSMEG_1240 lies immediately upstream of *MSMEG_1241*, another uncharacterized gene. The two genes are predicted to function as an operon, based on conservation patterns in other organisms, and as predicted by an OperonDB search [108]. Homologs of the two genes appear in a cluster in three other bacterial genomes - *Anabaena variabilis* strain ATCC 29413, *Actinosynnema mirum* strain DSM 43827, and *Streptomyces scabiei* strain 87.22. Moreover, the genomic arrangement of the region of the chromosome is similar to the arrangement of these genes in *M. smegmatis*, as determined through the JCVI Comprehensive Microbial Resource Database and the Kyoto Encylcopedia of Genes and Genomes (KEGG) [109, 110]. While *Actinosynnema mirum* and *Steptomyces scabiei* also belong to the Actinobacteria, *Anabaena variabilis* is a cyanobacterium. Furthermore, a search for conserved gene clusters on the KEGG SSDB (Sequence Similarity Database) indicated eight other genomes that carry putative orthologs of *MSMEG_1240*, *MSMEG_1241*, as well as the neighboring genes in the chromosomal region, in a conserved positional context of the chromosome. Figure 31 provides a ClustalW-based multiple sequence alignment of the MSMEG_1240 amino acid sequence and its eight predicted orthologs. It is noteworthy to mention that MSMEG_1240 shows the highest similarity to Ava_B0306 in *Anabaena variabilis*; the two proteins are 51% identical and 68% similar at the amino acid level. A BLASTP search against the database reference proteins revealed fourteen additional proteins that share significant similarity to MSMEG_1240 (Table 12); most notable of these is the hypothetical protein encoded by *MCOL_00565* in

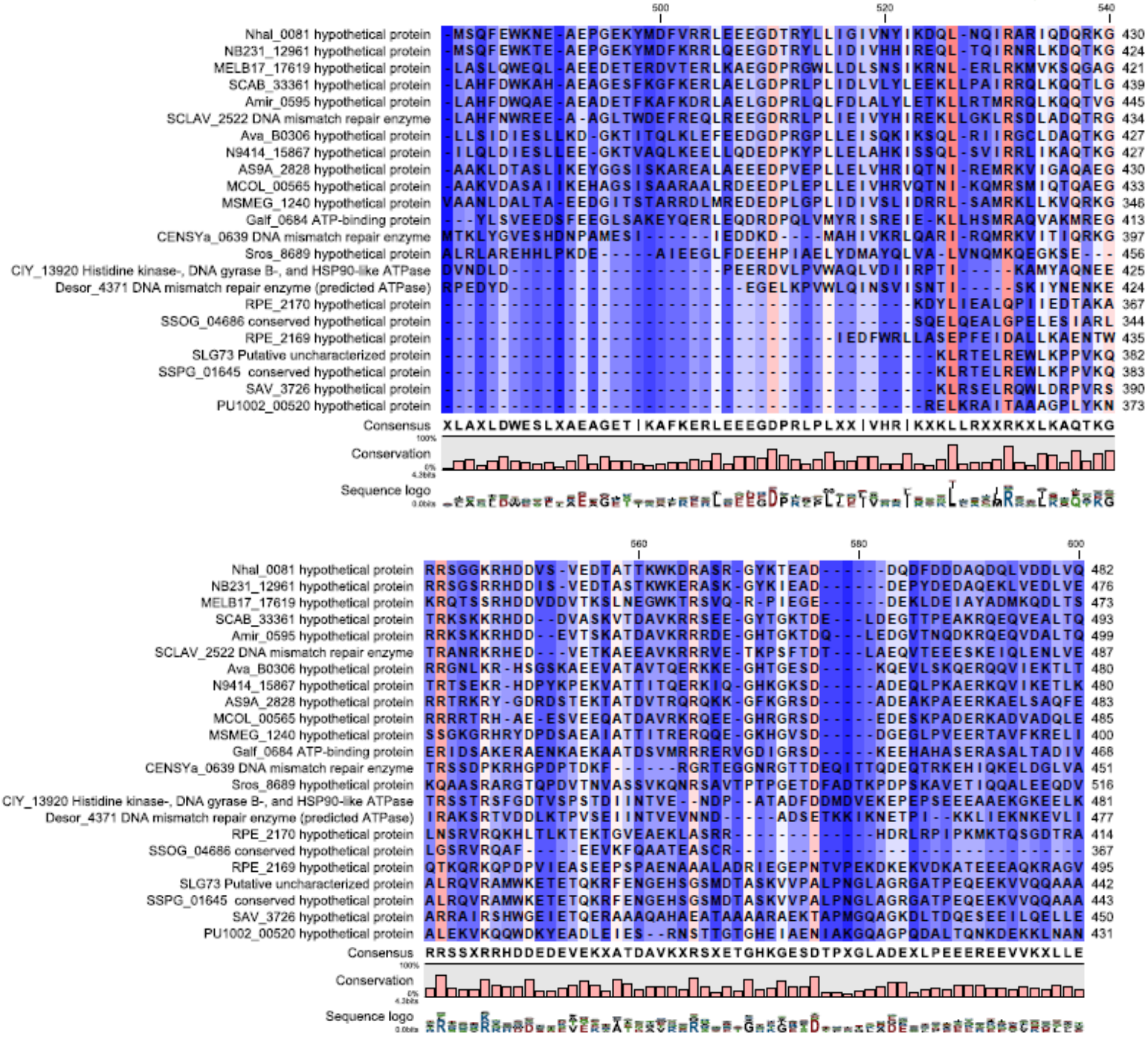
Mycobacterium colombiense. Of the twenty-two putative orthologs considered for this analysis, ten occur in the Actinobacteria, seven in the Proteobacteria, two in the Cyanobacteria, two in the Firmicutes, and one in the Archaea.

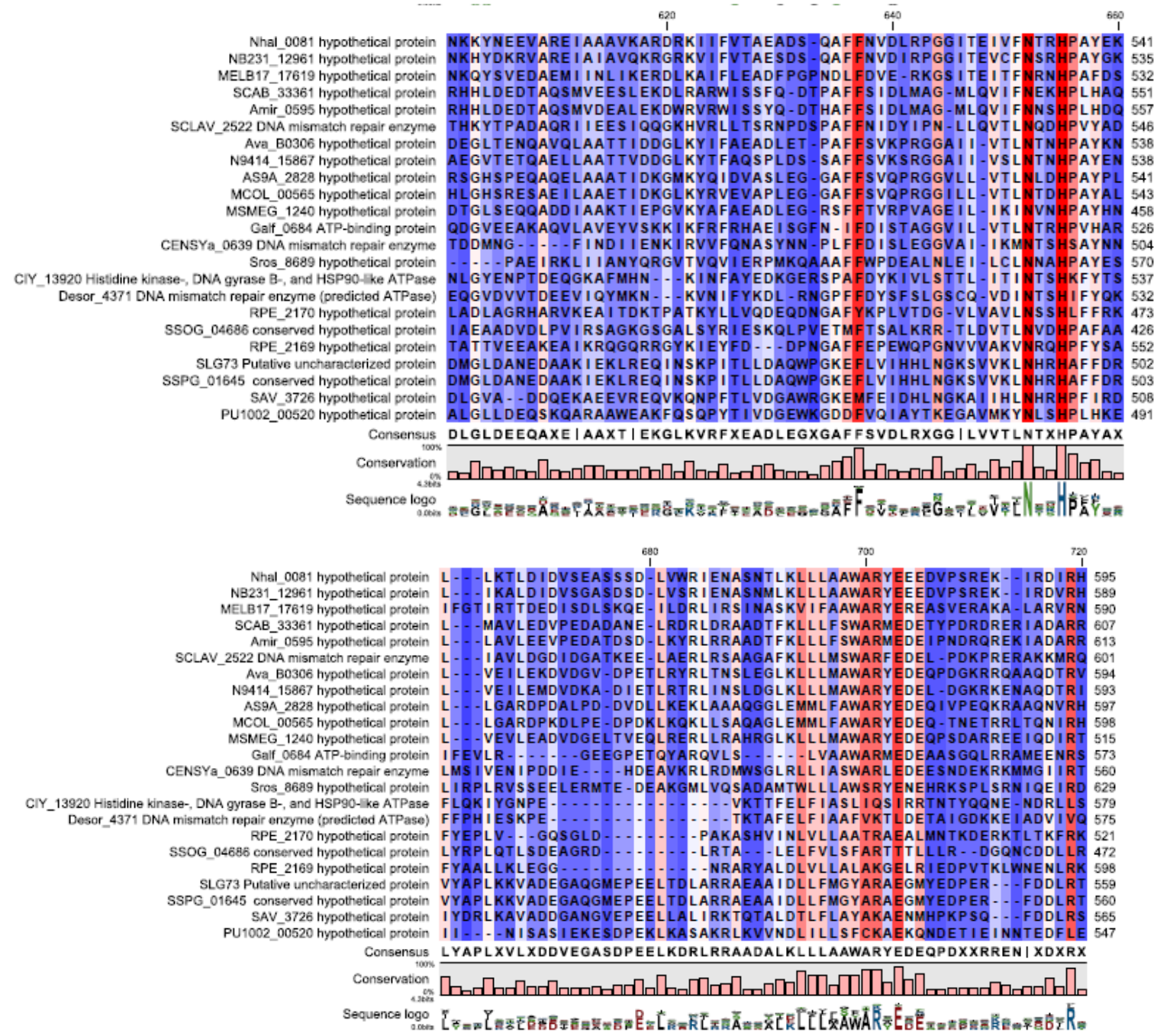

Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo


140
160
180
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo
200
220
240
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo


260
280
300
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo
320
340
360
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo


380
400
420
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
S | AKXEARRXDVXGXXWPTAVPXGKDPGNTPHXGKD | TEFQ | HAARNVG | SVVRAGRELE
100%
Conservation
0%
4.3bits
Sequence logo
0.0bits
440
460
480
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
YLDXFXLSWA | GYDLNQPVERWXLLQGYAYHWGVEVEFPPALDEVFGVTNNKQSATPFSE
100%
Conservation
0%
4.3bits
Sequence logo
0.0bits


500
520
540
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo
560
580
600
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo


620
640
660
Nhal_0081 hypothetical protein
NB231_12961 hypothetical protein
MELB17_17619 hypothetical protein
SCAB_33361 hypothetical protein
Amir_0595 hypothetical protein
SCLAV_2522 DNA mismatch repair enzyme
Ava_B0306 hypothetical protein
N9414_15867 hypothetical protein
AS9A_2828 hypothetical protein
MCOL_00565 hypothetical protein
MSMEG_1240 hypothetical protein
Galf_0684 ATP-binding protein
CENSYa_0639 DNA mismatch repair enzyme
Sros_8689 hypothetical protein
CIY_13920 Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
Desor_4371 DNA mismatch repair enzyme (predicted ATPase)
RPE_2170 hypothetical protein
SSOG_04686 conserved hypothetical protein
RPE_2169 hypothetical protein
SLG73 Putative uncharacterized protein
SSPG_01645 conserved hypothetical protein
SAV_3726 hypothetical protein
PU1002_00520 hypothetical protein
Consensus
Conservation
Sequence logo
680
700
720

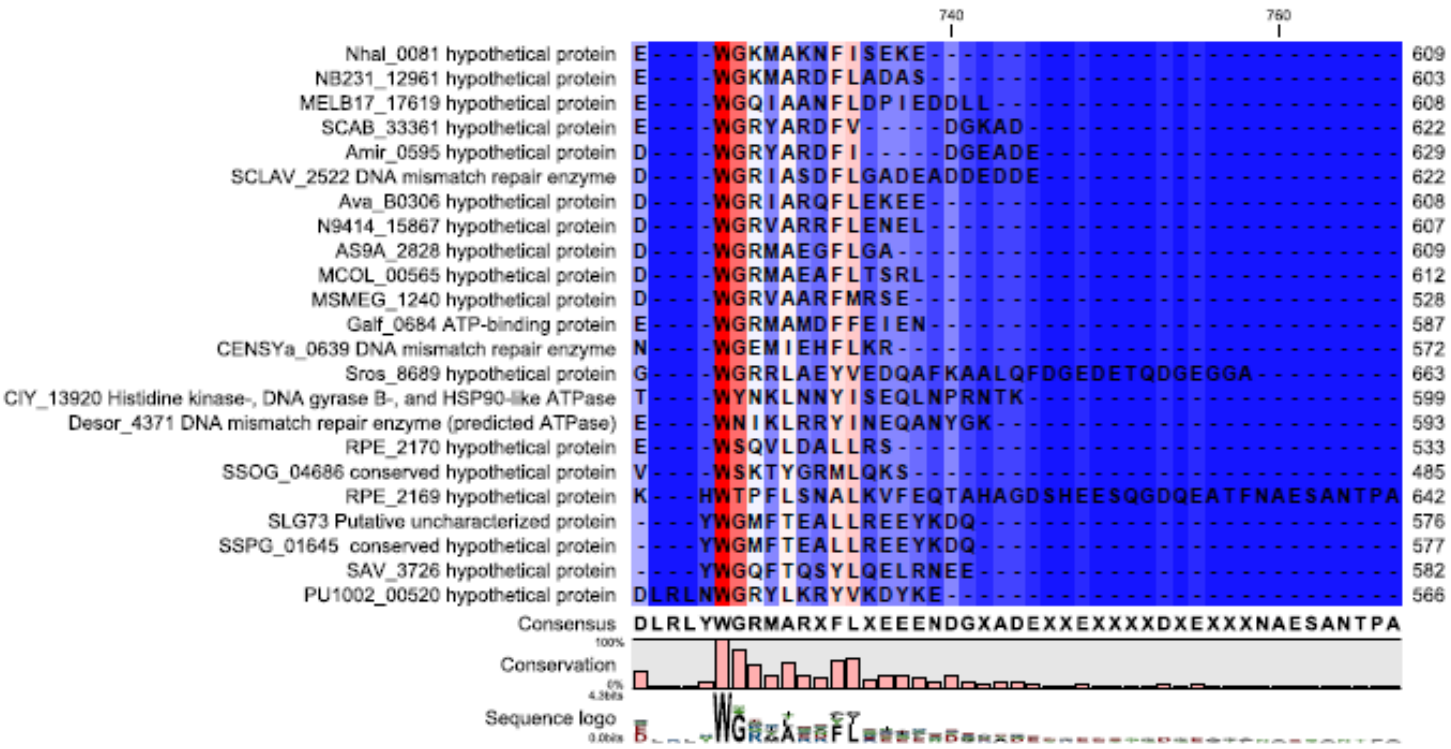


Figure 31. MSMEG_1240 amino acid sequence alignment with 22 putative homologs

Figure 31: Clustal-based alignment of MSMEG_1240 and 22 putative homologs. Alignments were assigned a free end gap; residue conservation follows a gradient from red to dark blue, with red representing the highest conservation.

Table 12. List of Putative MSMEG_1240 Homologs

Proteins that show significant homology to MSMEG_1240, as determined by a BLASTP search (19/22 proteins listed). All listed proteins, except for CENSYa_0639, SCLAV_2522 and Galf_0684 are annotated as hypothetical proteins. *CENSYa_0639 and SCLAV_2522 are annotated as DNA mismatch repair enzymes. Galf_0684 is annotated as an ATP-binding protein. Of the three unlisted proteins, Desor_4371 and CIY_13920 are annotated as predicted ATPases (DNA mismatch repair enzyme and HATPase, respectively).

Organism	Motility	Protein	Amino Acid Identity	Amino Acid Similarity
Anabaena variabilis	+	Ava_B0306	51%	68%
Nodularia spumigena	?	N9414_15867	51%	67%
Amycolicicoccus subflavus	?	AS9A_2828	45%	61%
Mycobacterium colombiense	?	MCOL_00565	45%	61%
Nitrosococcus halophilus	+	NhaI_0081	40%	59%
Nitrococcus mobilis	+	NB231_12961	41%	58%
Streptomyces clavuligerus	?	SCLAV_2522*	37%	55%
Streptomyces scabiei	?	SCAB_33361	37%	53%
Gallionella capsiferriformans	?	Galf_0684*	34%	54%
Cenarchaeum symbiosum A	?	CENSYa_0639*	34%	55%
Actinosynnema mirum	-	Amir_0595	36%	53%
Marinobacter sp.	?	MELB17_17619	37%	53%
Streptosporangium roseum	?	Sros_8689	30%	45%
Rhodopseudomonas palustris	+	RPE_2170	39%	60%
Streptomyces hygroscopicus	-	SSOG_04686	30%	44%
Candidatus Pelagibacter ubique	+	PU1002_00520	21%	40%
Rhodopseudomonas palustris	+	RPE_2169	34%	54%
Streptomyces lividans	?	SSPG_01645	23%	38%
Streptomyces avermitilis	?	SAV_3726	25%	41%

3.4.2 MSMEG_1240 contains a conserved N-terminal HATPase motif

Pfam and HHpred analyses detect a Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase (HATPase) motif in MSMEG_1240 [111, 112]. Pfam also predicts HATPase motifs in the twenty-two proteins that were considered as putative homologs of MSMEG_1240. A MEME analysis [113] of all twenty-three protein sequences revealed fourteen motifs that were considered relevant based on conservation in greater than ten of the twenty-two queried proteins. MSMEG_1240 contains thirteen of these motifs, as illustrated in Figure 32. Motifs 1 and 2 are conserved in all twenty-three proteins tested; motif 1 is the likely to be the HATPase domain, based on Pfam predictions of the location of this motif in MSMEG_1240.

Pfam analysis also shows a PLDc or Phospholipase D active site motif in MSMEG_1241. The MSMEG_1241 amino acid sequence shows homology to proteins at the matching position with respect to MSMEG_1240, in eight of the species noted in Table 12.

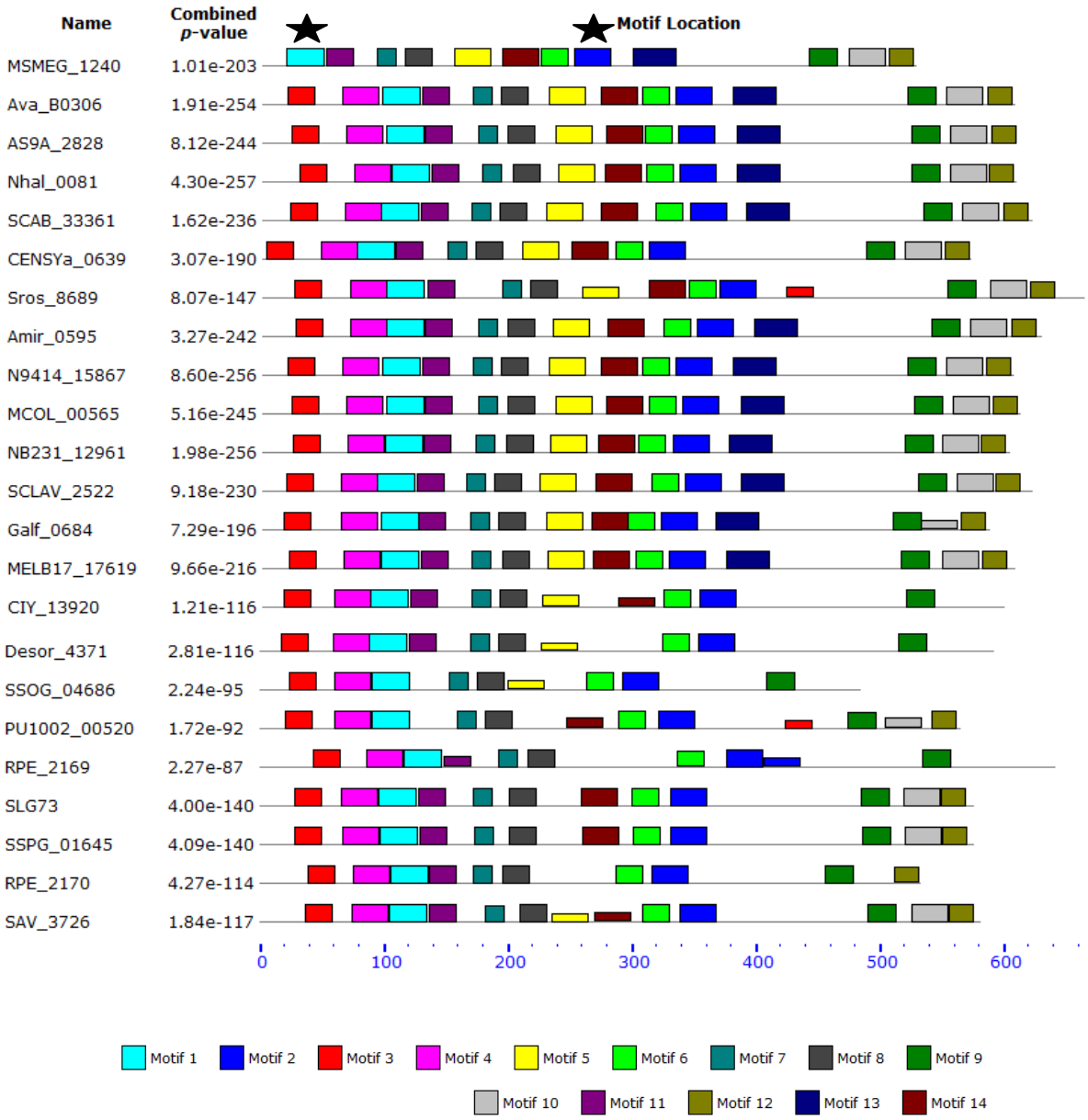


Figure 32. Conserved Motifs in MSMEG_1240 and 22 Putative Homologs

Figure 32: A combined block diagram illustrating the location of 14 motifs conserved between MSMEG_1240 and 22 putative homologs. Block height is proportional to the combined p-value, and only shown when p-value > 0.0001. Stars denote Motifs 1 and 2, which are conserved in all 23 protein sequences. (Diagram identified and graphic generated by MEME at http://meme.sdsc.edu/meme/cgi-bin/meme.cgi).

3.5 CONCLUSIONS AND FUTURE DIRECTIONS

Amongst the mycobacteria, surface translocation has been previously reported in *M. smegmatis*, *M. avium* and more recently, in *M. marinum*. In *M. smegmatis*, sliding motility has been shown to require the biosynthesis and export of GPLs; mutants that fail to synthesize, export or process GPLs, also fail in or are impaired in surface translocation. The work presented in this chapter began with the need for further assessment of the link between sliding motility and biofilm formation, to decide on the possibility of using the former assay as a surrogate for biofilms that form at the liquid-air interface. Our transposon mutagenesis and screen was designed to test mutants for their ability to slide on a surface as well as the ability to attach to a polystyrene plate and form mature biofilms at the liquid-air interface. This independent screen led to the identification of a transposon mutant that was severely defective in surface translocation, but proficient in biofilm formation. The observation of a hypomotile strain that retains the ability to form biofilms is novel in *M. smegmatis*.

Our efforts at genetically characterizing this mutant have led us to find that although an insertion in *M. smegmatis* gene *MSMEG_1240* - which encodes an uncharacterized protein - creates this separation in processes, the mutant cannot be complemented with a single-copy integrative copy of *MSMEG_1240*. We also find that a complete deletion of *MSMEG_1240* does not recapitulate hypomotility. One possible explanation for our inability to complement the mc^{2}155::Tn*Mariner* with *MSMEG_1240*, and the lack of a hypomotile phenotype in a mc^{2}155Δ*MSMEG_1240* strain is that the insertion into *MSMEG_1240* causes polarity on *MSMEG_1241*. Preliminary RT-PCR performed on wild-type mc^{2}155 and the mc^{2}155Δ*MSMEG_1240* strain does not reveal differences in the expression of *MSMEG_1241*

between the two. RNA samples used for this experiment were obtained from cultures that were grown planktonically in liquid culture. To re-test for the possibility of differences arising specifically during growth on a surface, we could perform RT-PCR using RNA samples obtained from a surface-grown pellicle. Another immediate effort to test for the possible role of *MSMEG_1241* in contributing to the motility defect, would involve constructing an in-frame deletion of *MSMEG_1241*, and also a double deletion of *MSMEG_1240* and *MSMEG_1241*.

To rule out the possibility of a secondary insertion elsewhere in the genome, we had planned to perform Southern blot analysis on genomic DNA extracted from the mc^2155::Tn*Mariner* strain. However, further work with this strain has proved challenging; the strain was found to harbor the TM4-based delivery phage. Repeated attempts at curing the strain have proved unsuccessful. An alternate strategy to overcome the problem of phage contamination would be to sequence the genome, or at least parts of the genome to identify any possible secondary insertions or modifications to the genome.

Bioinformatic analyses of MSMEG_1240 did not reveal any homologs in the mycobacteria until late 2011, when a BLASTP search revealed a significant match with MCOL_00565, a protein of unknown function in *Mycobacterium colombiense*. *M. colombiense* is a member of the *M. avium* complex [103]. The whole genome sequence for the type strain CECT 3035 was first published in October of 2011 [114]; this explains the absence of MCOL_00565 from results obtained during previous searches. The closest MSMEG_1240 homolog is found in the cyanobacterium *Anabaena variabilis*; MSMEG_1240 and Ava_B0306 share 51% amino acid identity and 68% similarity. Moreover, we find that similar to the case of *MSMEG_1240* and *MSMEG_1241*, *Ava_B0306* also occurs in a putative operon with *Ava_B0305*. The organization of these genes as well as the surrounding region of the

chromosome appears to be similar, although the orientation differs. Moreover, *A. variabilis* also exhibits surface translocation by means of twitching/gliding motility [115, 116].

HHPred, Pfam and MEME-based searches suggest the presence of a Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase (HATPase) motif in the N-terminal region of MSMEG_1240 and its twenty-two putative homologs, including Ava_B0306 and MCOL_00565. Of the twenty-two, four proteins have been annotated as either DNA mismatch repair enzymes or ATPases/ATP-binding proteins. *M. avium* is the only other member of the mycobacteria that is reported to synthesize GPLs and exhibit GPL-based sliding motility. Although it may be a coincidence that the only homolog to MSMEG_1240 within the mycobacteria occurs in the MAC complex, the high degree of identity at the amino acid level is noteworthy.

Overall, the work described in this chapter indicates that a Tn*Mariner* insertion into a putative ATPase compromises sliding motility in *M. smegmatis,* without affecting biofilm formation. In the absence of the complete genetic characterization of the insertion mutant, it is likely that there exists another plausible explanation for this observation, as discussed above. Even so, the separation between the ability to translocate on a surface and the ability to form biofilms is a novel observation in *M. smegmatis,* and warrants further characterization.

4.0 DISCUSSION

4.1 GENE EXPRESSION IN *M. SMEGMATIS* PLANKTONIC CULTURES AND BIOFILMS

Bacterial biofilms consist of organized communities of cells encased in extracellular matrix material, leading to a distinct biofilm-specific cellular architecture [8, 117]. Biofilm formation is reminiscent of a development pathway, involving attachment to a solid surface or a liquid-air interface, colonization of the surface, maturation, and production of an extracellular matrix [17]. The complex biofilm structural architecture gives rise to gradients of nutrients and oxygen availability, therefore creating distinct micro-environments within the same biofilm [15]. Bacterial cells that form the biofilm respond to these different environmental niches by changing their gene expression patterns [43]. A significant volume of studies on biofilms of a number of bacterial species, including *E. coli, P. aeruginosa*, and *B. subtilis* amongst others, suggest that different subsets of cells within the same biofilm exhibit different patterns of gene expression [33-35, 41, 42] . These studies also show that gene expression patterns change over the course of biofilm development [34].

The first comparative transcriptome analysis of *M. smegmatis* biofilms relative to planktonic cultures was published by Ojha and Hatfull in 2007 [61]. This study showed that

approximately one hundred genes in *M. smegmatis* are up-regulated in a biofilm-specific pattern, and that a significant percentage of these genes are involved in iron acquisition [61].

The importance of iron for the growth and metabolism of the mycobacteria in general, and in host-associated virulence of *M. tuberculosis* has been extensively studied and well characterized [73, 118]. In addition to *in vitro* studies, studies in mouse models of infection [119] and on patient isolates [73] demonstrate that excess iron availability in the host enhance tuberculosis infections. On the other hand, anemic patients have better outcomes from human tuberculosis [73]. Administration of iron supplements to anemic patients with active TB infections is in fact detrimental to the prognosis of the infection [120]. Taken together with current research on *M. tuberculosis* biofilms [58], and their likely role in drug tolerance and infections, the study of iron response in mycobacterial biofilms gathers increasing importance.

4.1.1 P_{MSMEG_4515} Expression in Biofilms and Planktonic Cells

As early as 1971, Ratledge and Hull had determined the induction of the mycobactin siderophore in *M. smegmatis* grown in standing cultures in trace metal-free media. They determined that the concentration of mycobactin produced increased with growth, and that mycobactin biosynthesis was induced under deficient conditions of Fe^{2+}, Zn^{2+}, Mn^{2+}, and Mg^{2+}. The greatest degree of mycobactin biosynthesis was observed under iron deficient conditions, at 1.8 μM iron. Addition of 9-90 μM Fe^{2+} repressed mycobactin biosynthesis to 50% of the maximally induced concentration [121]. The above studies provide important justification for our selection of the mycobactin biosynthesis gene as the candidate of choice to profile gene expression in biofilms. The comparative profiling of mycobactin biosynthesis gene expression in

mature biofilms and planktonic cultures, and the additional profiling of expression across a range of iron levels, provides a detailed report on the iron-related siderophore induction levels experienced by mycobacterial biofilms.

Apart from the Ojha and Hatfull transcriptome study [61], very little is known about gene expression patterns within *M. smegmatis* biofilms. Our fluorescent reporter-based studies with P_{MSMEG_4515}-eGFP and P_{hsp60}-eGFP/ P_{hsp60}-mCherry are the first examples of detailed cellular-level analyses of gene expression in *M. smegmatis* biofilms. Our fluorescence confocal microscopy as well as flow cytometry-based analyses show that P_{MSMEG_4515}-eGFP is differentially induced during planktonic growth and within biofilms. As expected for an iron-regulated siderophore biosynthesis gene, expression levels are significantly induced during planktonic growth in iron-deficient media, and significantly down-regulated in iron-supplemented and iron-rich media. Within mature four-day biofilms, however, P_{MSMEG_4515} shows an iron-independent pattern of gene expression, relative to the amount of iron present in the growth medium. The differential expression patterns for P_{MSMEG_4515} induction is consistent with observations of differential gene expression in biofilms as compared to planktonic cultures. The apparent iron-independent induction of P_{MSMEG_4515} suggests the possibility of occlusion of cells from the available iron, perhaps due to the thick matrix in a mature biofilm.

To test the possibility of the mature biofilm structure playing a role in preventing access to iron even under iron-rich conditions, we can repeat rigorous analyses on biofilm maturation-defective mutants. We have performed one set of identical flow cytometry-based experiments as listed above, in two mutant strains that have been previously reported to be defective in biofilm maturation. Strain $mc^2155\Delta$*groEL1* is defective in biofilm maturation [49], and strain

$mc^2155\Delta$*lsr2* is defective in overall biofilm formation [100]. Although this strain does not form a pellicle-like biofilm, it does form a slimy layer when grown unperturbed biofilm medium.

The *M. smegmatis mbtB* gene, which is currently annotated as *MSMEG_4515*, is part of the mycobactin biosynthesis cluster that is responsible for assembling the core scaffold of the mycobactin siderophore molecule [122]. In *M. smegmatis* as well as in *M. tuberculosis*, this gene is under the control of the iron dependent repressor, or IdeR [73]. In *M. smegmatis*, a study involving mutational analyses of the *mbt* gene cluster has conclusively proved *mbtB* to be essential for mycobactin biosynthesis, along with *mbt*A,C,D,E, and F [123]. The P_{MSMEG_4515} promoter sequence used in my studies was not mutationally altered to test its iron-dependent regulation. Nevertheless, in *M. tuberculosis*, *in vitro* binding assays performed by Gold *et al* show that purified IdeR binds to a region proximal to the -10 position of the *mbtA-mbtB* promoter. IdeR therefore acts as a transcriptional repressor by preventing RNA polymerase from binding; *mbtB* is repressed by IdeR under iron-rich conditions [76]. One way to test the iron-dependent IdeR-mediated regulation of P_{MSMEG_4515} would be to perform microscopy and flow cytometry on an *ideR* mutant strain grown planktonically. As quantitated in Figure 16A, P_{MSMEG_4515} expression is down-regulated in iron-rich media. In an *ideR* mutant strain background, P_{MSMEG_4515} expression should remain unchanged even in the presence of iron-rich conditions, if this promoter sequence were indeed iron responsive.

To test for temporal changes in gene expression patterns and expression levels over the course of biofilm development, I have performed flow cytometry on biofilm samples harvested at three-day and five-day stages of development. Following experimental repeats and analyses using Ferdinand, these data will provide a temporal pattern of *MSMEG_4515* gene expression in *M. smegmatis* biofilms.

The reporter constructs generated for our study have a potential for broader applications than fluorescence-based studies. The P_{MSMEG_4515}-eGFP reporter construct can be developed into an inducible vector for protein purification, especially for proteins that require folding within the mycobacterial cell. Our flow cytometry data confirm the iron-responsive regulation of *MSMEG_4515* in planktonic cultures. Replacement of the *eGFP* gene in the fusion construct with the gene of choice, growth of the resultant strain in standard mycobacterial growth media and addition of a chelating agent as an inducer, can allow for significant protein production. Our studies have already demonstrated that this promoter is very highly induced under iron-deficient conditions, even when present as a single copy. Cloning the promoter fusion cassette into an extrachromosomal vector has the potential to induce the production of massive amounts of protein.

4.1.2 P_{hsp60} Expression in Biofilms and Planktonic Cells

Although P_{hsp60} was originally conceived as a constitutively active control for our gene expression studies, our flow cytometry-based analyses show that this promoter does not maintain a standard induction level across iron conditions. As quantified in Figures 16 and 22 for planktonic growth and biofilms, respectively, P_{hsp60}-eGFP is induced to a higher level in iron-rich conditions than in iron-deficient and low-iron conditions. These observed differences in mean fluorescence levels from P_{hsp60}-eGFP are statistically significant, and hold true in planktonic cultures as well as in biofilms.

Iron is essential for bacterial growth; however, an excess of iron can be toxic to the cell. When the concentration of iron is higher than what the bacterial cell can sequester, it leads to

oxidative stress caused by the generation of reactive oxygen species via the Fenten reaction [124]. The Ojha and Hatfull transcriptome analysis was performed in biofilms grown at 2 μM iron, and even so, stress response genes were observed to be up-regulated [61]. It is likely that the higher level of P_{hsp60} induction in 50 μM iron is a response to oxidative damage, although further experimentation will be required to confirm this hypothesis.

Ojha and Hatfull had previously reported the lack of robust biofilm formation in iron-deficient media [61]. As illustrated in Figure 33A, biofilms grown in iron-deficient media do not form mature pellicle-like structures at the four-day stage. Similarly, we consistently observe that biofilms formed in 50 μM Fe^{2+} also do not form a pellicle-like film, and instead produce a slime-like, viscous monolayer.

The Ojha and Hatfull study reported a biofilm maturation defect exhibited by an iron uptake-deficient mutant, suggesting that iron is required for the early transitioning into the biofilm stage. Iron has also been shown to be associated with the production of fatty acids that make up the mycobacterial matrix material [61]. Therefore, under iron-deficient conditions, it is not surprising to observe a defect in biofilm maturation.

The failure to form robust biofilms in iron-rich conditions suggests the likelihood that high iron levels could be toxic to *M. smegmatis* biofilms. A preliminary test for planktonic growth in varying iron concentrations shows similar growth rates in all four iron concentrations, when observed over a period of twenty hours until late exponential stage (Figure 33B).

Studies by the Buckling lab on the genetically linked traits of siderophore production and biofilm formation in *P. aeruginosa*, show that these traits are also socially linked [125]. As in the case of *M. smegmatis*, *P. aeruginosa* also requires iron for biofilm formation. Mutants that are defective in siderophore production are also found to be defective in their ability to form

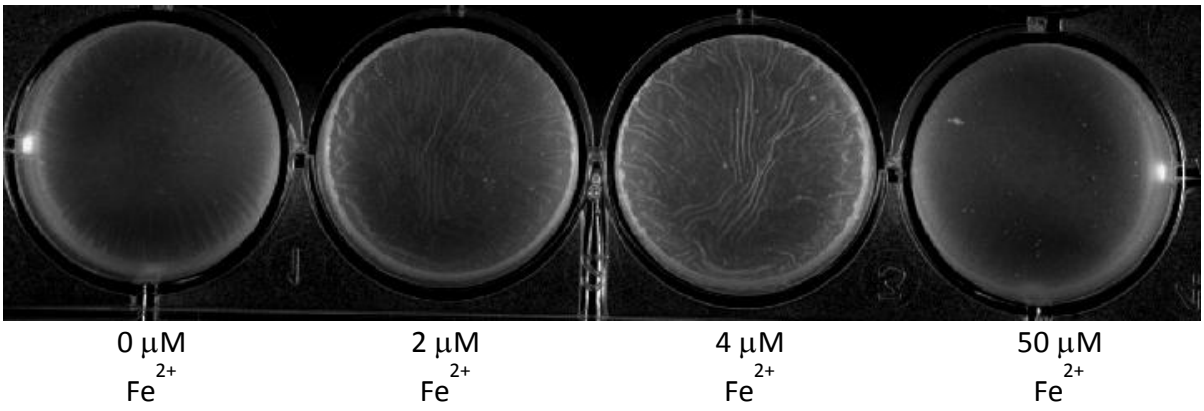
biofilms. In mixed biofilms of siderophore-producing cells and non-producing clones, or cheaters, the cheaters do not fare better than they do in planktonic mixed populations. Therefore, although cheats may arise in the population for one trait (in this case, for siderophore dependence), they are not likely to also spread and thrive within a biofilm structure. Since siderophore production is important for biofilm formation, cheaters sweeping across a biofilm without contributing to the biomass would lead to a weak biofilm system. The study finds that cheaters for siderophore production do not have a higher fitness advantage in biofilms than they do in planktonic cultures. The matrix material produced by the wild type population does not confer an extra advantage to the cheaters [125].

Ojha and Hatfull found that *M. smegmatis* biofilm formation requires the availability of iron. The same study showed that strains that lack the ability to produce the exochelin class of siderophores are defective in biofilm formation; however, mutants lacking the mycobactin biosynthesis genes form robust biofilms [61]. Our current data are not sufficient to adopt or refute the Buckling lab's model in the case of *M. smegmatis* biofilms. To test for the social link between siderophore production and biofilm formation in *M. smegmatis*, it would be worthwhile to consider the interactions between the wild type strain and a strain lacking the ability to produce exochelin. This study could also be addressed by flow cytometry using two different, compatible fluorescent proteins to tag each strain. The relative abundance of each cell type can then be quantitated after growing them as mixed biofilms.

Our studies involving P_{MSMEG_4515} and P_{hsp60} are one of the first examples of a detailed cellular-level investigation into gene expression in *M. smegmatis* biofilms. Comparative profiling of P_{MSMEG_4515} in biofilms and planktonic cultures provide detailed information on the differential levels of promoter induction. The scope of this study can be extended further to include a greater

number of promoters, to study spatial patterns of gene expression, and to fine tune the level of accuracy on data obtained for population categories.

A.



B.

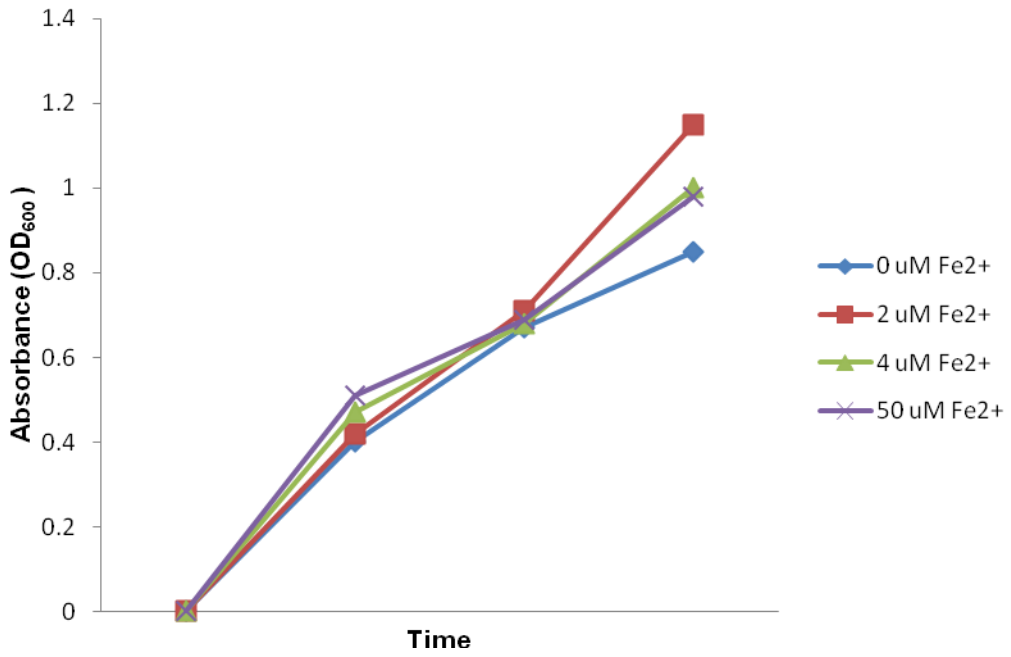


Figure 33. Biofilm Formation and Planktonic Growth in Varying Concentrations of Iron

Figure 33: Comparison of growth in **A. m**ature four-day biofilms and **B.** planktonic cultures. Each assay is performed in each concentration of iron (0 μM, 2 μM, 4 μM, 50 μM Fe^{2+}). Panels A and B show representative data from repeatable experiments. Data in this figure are for strain mc^2155 with P_{MSMEG_4515}-eGFP + P_{hsp60}-mCherry. For planktonic growth, the X-axis represents time in hours. The four data points represent: inoculation time t = 0 hours, t = 15 hrs post-inoculation (o/n growth), t = 17.5 hrs post-inoculation, t = 20 hrs post-inoculation. Planktonic cultures were grown in the same media as the biofilms, with the addition of Tween-80 to prevent clumping.

4.2 SEPARATION OF SURFACE TRANSLOCATION AND BIOFILM FORMATION

As described in section 3.1, all previous studies on *M. smegmatis* sliding motility have linked defects in sliding motility to defects in biofilm formation. All hypomotile mutants that have been characterized so far are found to be defective in GPL biosynthesis, transport, or processing, thereby also affecting biofilm formation. Our study describes the first instance of a transposon mutant strain, mc^{2}155::Tn*Mariner* that shows separate outcomes for sliding motility and the ability to form mature biofilms. The separation of these phenotypes suggests that sliding motility is not a necessary requisite for biofilm formation.

Chapter 3 describes the challenges faced while planning and performing experiments aimed at complete characterization of the transposon mutant strain, and also illustrates some alternate methods that can be employed to address some of these challenges.

Since mc^{2}155::Tn*Mariner* could form mature biofilms when tested, it is unlikely that the mutation affects GPL biosynthesis or processing, since GPLs are required for biofilm formation. We already knew that the reverse is not the case, that is, defects in biofilm formation do not indicate defects in sliding motility. It should also be mentioned that a different mating strain of M. *smegmatis*, strain Jucho, although found to be severely defective in sliding motility, formed robust biofilms. The genomic sequence for *M. smegmatis* strain Jucho is not yet available as a resource to investigate the presence of a *MSMEG_1240* homolog. Further experimentation to elucidate the exact genetic nature of this mutant will enable us to further understand the requisites for sliding motility.

Also, this study has led us to investigate MSMEG_1240 as a putative ATPase; close homology to other proteins of similar predicted function suggest a likely functional role for this

gene. Biochemical assays can be used test whether MSMEG_1240 is a functional ATPase. The finding that MSMEG_1240 has a homolog in another member of the mycobacteria - one that belongs to a group that is also known to exhibit sliding motility - lends further justification for the characterization of this novel mutant.

Section 3.5 in Chapter 3 provides a detailed account of future experiments that need to be performed to characterize mc^{2}155::Tn*Mariner* further, and to confirm the role of *MSMEG_1240* in sliding motility. As it stands at the moment, a number of possible explanations can account for the phenotype that we observe for mc^{2}155::Tn*Mariner*. First, it is possible that the transposon insertion into *MSMEG_1240* disrupted *MSMEG_1241*, giving rise to the observed phenotype. A deletion of *MSMEG_1241* and a double deletion of the putative *MSMEG_1240-MSMEG_1241* operon can help address this possibility. Second, the phenotype that we observe could be the result of a secondary transposon insertion elsewhere in the bacterial genome. This possibility can be addressed by performing a Southern Blot on genomic DNA extracted from strain mc^{2}155::Tn*Mariner*. Before performing the Southern, we will have to first revisit the technical details for curing the currently phage-contaminated mc^{2}155::Tn*Mariner* strain. Third, the phenotype that we observe could be caused by the presence of phage contamination in strain mc^{2}155::Tn*Mariner*. This possibility can be addressed by curing the strain and then re-testing it for the hypomotile phenotype. Since curing the strain of the phage has proved unsuccessful following repeated attempts, another possible method that we considered for testing the transposon insertion was to transduce the insertion sequence into a wild type strain background. This method requires propagation of the transducing phage on lawns generated from mc^{2}155::Tn*Mariner*; however, the presence of the phage in the strain makes it very challenging to generate a successful bacterial lawn. Yet another possibility for testing the transposon

insertion mutant would be to re-engineer the insertion in a wild type strain into the exact region within *MSMEG_1240* that the transposon maps to, and to test the resulting mutant for a phenotype. This method can also address the possibility that the transposon insertion causes a dominant negative phenotype which cannot be complemented with a copy of *MSMEG_1240*.

5.0 MATERIALS AND METHODS

5.1 BACTERIAL STRAINS AND MEDIA

5.1.1 *Mycobacterium smegmatis* Strains

Almost all experiments and manipulations on *M. smegmatis* utilized strain mc^2155 as the parental wild-type strain. This strain is amenable to molecular manipulations owing to the mutations that confer an efficient plasmid transformation (Ept) phenotype [126], allowing for greater efficiency of plasmid transformation than in non-Ept strains.

M. smegmatis mating strains Jucho, Nishi and Rabinowitchi (Rab) [127] were used for experiments to test their proficiency in biofilm formation and sliding motility, as compared to strain mc^2155.

5.1.2 *E. coli* Strains

Ligation reactions for standard molecular cloning protocols were transformed into commercially obtained GC5 (Gene Choice) or NEB5alpha competent cells (New England Biolabs) for high efficiency transformations. Purified, tested plasmid stocks were propagated in laboratory-prepared DH5α competent cells [128]. HB101 cells (Promega) were used to passage plasmids to

be used for mycobacterial recombineering [104]. DH5α λpir cells were used to test clones for transposon insertions.

5.1.3 Media for *M. smegmatis*

All strains of *M. smegmatis* were grown in liquid culture in Middlebrook 7H9 broth (Difco), supplemented with 10% Albumin Dextrose Complex (ADC), and 0.2% glycerol. Planktonic cultures of strain mc^2155 were grown in the presence of 0.05% Tween-80, while strains Jucho and Rabinowitchi were grown in the presence of 0.20% Tween-80, to prevent clumping. Strain Nishi was grown in Luria-Bertani broth (LB broth), in the presence of 0.20% Tween-80.

For growth on solid media, *M. smegmatis* was plated on 7H10 agar (Difco) supplemented with 10% ADC and 0.5% glycerol. For most growth, all media were also supplemented with carbenicillin (Cb, Difco, 50 μg/mL) and cycloheximide (Chx, Sigma, 10 μg/mL). In the presence of plasmids that conferred antibiotic resistance, the selection media was also supplemented with the corresponding antibiotics - kanamycin (Kan, Sigma, 20 μg/mL), hygromycin B (Hyg, Sigma, 150 μg/mL) or tetracycline (Tet, Sigma, 5 μg/mL).

For growth as biofilms, strains mc^2155, Rab and Jucho were inoculated in modified M63 media (54), without any detergents or antibiotics, and incubated without disturbance at 30°C for 3-7 days of development. For each 1 mL of reconstituted biofilm media, 1 μL of a saturated culture was used as inoculum. Strain Nishi biofilms were grown in LB broth, without detergents or antibiotics. For planktonic growth in minimal media, strains mc^2155, Rab and Jucho were inoculated in the same M63 growth media, supplemented with 0.05% or 0.2% Tween-80,

respectively. Strain Nishi was again inoculated in LB broth, supplemented with 0.2% Tween-80. For each 1 mL of reconstituted biofilm media, 4 μL of a saturated culture was used as inoculum.

5.1.4 Media for *E. coli*

All *E.coli* strains were grown in LB broth for liquid cultures, or on LB-agar plates containing appropriate antibiotics - Kan (20 μg/mL), Hyg (150 μg/mL), or Cb (50 μg/mL).

5.2 DNA MANIPULATIONS

5.2.1 Standard Molecular Cloning Methodology

The generation of plasmids used for this study involved the insertion of DNA sequences of interest into vector DNA. Vector DNA was prepared by restriction digestion to cut and/or excise and eliminate specific regions/cassettes from available laboratory stocks of plasmids. Insert DNA was generated by PCR amplification, using either plasmid DNA or *M. smegmatis* genomic DNA (gDNA) as template. All PCR products were cleaned up using the QIAquick PCR Purification Kit (QIAGEN), and the product DNA was concentrated and eluted in 50 μL elution buffer (Buffer EB). Restriction digests were performed in a reaction volume of 20 μL-50 μL. Restriction enzymes were supplied by New England Biolabs (NEB), and all reactions were performed in accordance with the manufacturer's instructions. For parent plasmids that were used as vectors for cloning, restriction digestion was typically followed by treatment with calf alkaline

phosphatase (CIP, NEB). In some cases, when necessary, the DNA was treated with Klenow (NEB) or T4 polynucleotide kinase (Roche). Vector DNA when treated with CIP, which cannot be heat inactivated, was always purified and concentrated by gel extraction, using the QIAquick Gel Extraction Kit (QIAGEN). Vector DNA was eluted from the spin column in 30 μL of dH_20. Vector and insert DNA fragments were quantified on 0.8-0.9% agarose gels, using the Gene Choice DNA Ladder I (Gene Choice). In some cases, the DNA was also quantified on the Nanodrop before setting up ligation reactions. All ligation reactions were set up following the manufacturer's instructions, at room temperature, in 15-20 μL final volume, using the Fast-Link DNA Ligation Kit (Epicentre). Ligation reactions were allowed to continue for 1-2 hours, followed by heat inactivation of the ligase enzyme at 75ºC for 15 minutes.

5.2.2 PCR Conditions and Parameters

5.2.2.1 Standard PCR

All PCR amplification reactions for standard molecular cloning purposes were performed using the following reagents : Pfu polymerase (Stratagene), 1mM deoxyribonucleotides (dNTPs) - 0.25 mM each of dATP, dTTP, dCTP, dGTP, 5% dimethyl sulfoxide (DMSO, 0-5%) as an additive for enhanced amplification of GC-rich mycobacterial template DNA, primers at 0.25 μM final concentration.

Thermocycler conditions were set as follows: 95°C denaturation for 5 min; 25 cycles of 95°C for 30 seconds; annealing temperature based on primers used (2°C lower than the lowest melting

temperature) for 30 seconds; extension at 72°C, depending on length of the product (approximately 1 min/1 kbp product); final extension for 7 minutes; 4°C cooling/storage.

5.2.2.2 Colony PCR

In some instances, especially for diagnostic PCR amplification of a target sequence, instead of using purified plasmid DNA or purified gDNA as template, cells were scraped off a bacterial colony on a petridish to use as the source of template DNA for PCR. Prior to setting up the reaction as described in section 5.3.2.1, the cells were resuspended in 1 μL of dH_20 and boiled at 95ºC for 5 minutes. 5 μL of this mixture was used for PCR.

5.2.3 Primer Design

Primers were designed using DNA Strider or A plasmid Editor (ApE). Primers used for cloning purposes were designed to be 18-30 nucleotides in length, with almost equal number of nucleotide homology flanking either side of the PCR target sequence. Restriction endonuclease cut sites were engineered into the primer target region, as required, with modest changes in the native sequence. Primers were verified/virtually tested using Amplify 3 to eliminate the probability for primer dimers. Primers used for sequencing were designed to be 25-30 nucleotides in length, with complete homology to the target sequence. All primers and other oligonucleotides were manufactured by Integrated DNA Technologies (IDT), Inc. Upon receipt, lyophilized primer DNA was resuspended in TE buffer or dH_20 to 100 μM/μL (parental stock

concentration), and diluted in dH_2O to 10 μM/μL (working stock). Primer parental stocks and working stocks were stored at -20ºC.

5.2.4 Transformation and Electroporation of DNA into Bacterial Cells

Purified plasmid DNA was always propagated in *E.coli* cells to generate stocks for storage, and for electroporation into mc^2155 cells. For plasmids generated by standard cloning methods, the ligation mix was transformed into *E. coli*, and the recovery mix was plated on selection media. Colonies retrieved were then mini-prepped, and screened by restriction digestion. Verified plasmids were then electroporated into mc^2155 cells.

Purified plasmids were transformed into $CaCl_2$ competent cells [128] of *E. coli* strain DH5α. Plasmids generated by ligation reactions were transformed into high-efficiency, commercially-obtained *E. coli* strain GC5 or NEB5α cells. pYUB854-based plasmids used for mycobacterial recombineering, were transformed into *E.coli* strain HB101 cells [105] by heat shock transformation methods, as previously described [129]. Following recovery in LB media (without antibiotics), at 37ºC, the cells were plated on LB agar that contained the appropriate antibiotics as selection agents.

Purified plasmids were electroporated into *M smegmatis* strain mc^2155 electrocompetent cells, prepared as previously described [104]. 50-100 ng plasmid DNA was introduced into electrocompetent cells using a Bio-Rad Gene Pulser Electroporator System. For mycobacterial recombineering, pYUB854-based plasmids that carried the deletion substrate were electroporated into mc^2155 cells that contained pJV53, for recombination efficiency [104]. In some cases,

when protocols did not require very high transformation efficiencies, purified plasmids were electroporated just as described above, into quick-prepped electrocompetetent cells (Gregory Broussard, unpublished). 1.5 mL of a saturated culture of mc^2155 cells was pelleted by centrifugation at 13,000 rpm for 1 minute. The pellet was resuspended in 1 mL of ice cold 10% glycerol, washed by pipetting up and down a few times, and then pelleted as described above. Following a total of three such washes, the pellet was resuspended in 100 μL ice cold 10% glycerol, and immediately used for electroporation with 1 μL of a 50-100 ng/μL plasmid DNA stock.

5.2.5 Sequencing

Plasmids generated by standard molecular cloning methods described above, were sequenced through submission to GeneWiz, Inc., following their specified guidelines. 500-800 ng of purified plasmid DNA was mixed with 8 pmol of sequencing primer. Sequencing primers were constructed to read from 100-200 bp upstream or downstream of the region targeted for sequencing.

5.3 BIOFILM SET-UP

M. smegmatis strains mc^2155, Jucho and Rabinowitchi, and mc^2155 strains carrying pAB reporter constructs, pNIT, and pNIT-eGFP, and mutants $mc^2155\Delta$*groEL1* and $mc^2155\Delta$*lsr2*, were set up to form biofilms as previously described [49]. Biofilms were set up in modified M63

[55] base media, supplemented with 2% glucose, Casamino Acids, $MgSO_4$, and $CaCl_2$ (hereafter referred to as biofilm media). The final concentration of iron in standard biofilm media was maintained at approximately 4 μM. Initial biofilm assays performed during this study, maintained iron at a final concentration of 2 μM; however, this standard concentration was modified to 4 μM, due to repeated observations of improvement in the robustness of the biofilms that form in 4 μM iron. Unless otherwise specified, biofilm assays were typically set up in 10 mL of standard biofilm media, in 60 x 15 mm polystyrene petridishes, and inoculated with 10 μL of a saturated culture of the bacterial strain used. Plates were incubated at 30ºC, without disturbance, for 3-6 days.

For confocal microscopy, biofilms were set up in 6 mL of biofilm media held in 50 mL conical tubes. Cover-slips were placed in each tube, at an angle (approximately 45 degrees) that allowed for partial submergence in the media, so that biofilms would form with one end of attachment at the liquid-air interface on the cover-slip. Each tube was inoculated with 6 μL of a saturated bacterial culture.

For flow cytometry, biofilms were set up in 48-well assay plates, each well containing 1 mL of biofilm media, and inoculated with 1 μL of saturated bacterial culture.

5.4 MICROSCOPY

5.4.1 Fluorescence Microscopy

Almost all fluorescence microscopy was performed using a Zeiss Axiostar-Plus microscope (Carl Zeiss), and images were captured on an AxioPlus MRc 5 digital camera and AxioVision Rel 4.2 image processing software (Carl Zeiss). Images were processed using Adobe Photoshop (Adobe Systems Incorporated), using minimal adjustment for brightness and contrast, and maintaining identical settings across controls and experimental slides. For planktonic cells, slides were prepared by placing 2-5 μL of liquid culture on glass slides (VWR International, LLC., or Thermo Fisher Scientific, Inc.), angling a glass cover-slip over the liquid, and pressing down to release and wipe off any excess liquid. Slides were then sealed with VALAP (mixture of Vaseline, lanolin, paraffin wax), to allow for better imaging.

5.4.2 Confocal Microscopy

All confocal microscopy was performed using a Confocal Scanning Laser Microscope (Nikon, Bio-Rad system). Biofilms grown on cover-slips were carefully pulled out of the conical flasks using forceps, and gently placed on a glass microscope slide, with the biofilm facing upwards. A fresh cover-slip was gently placed over the biofilm, before placing the slide under the 40x and 60x objective lenses.

5.5 FLOW CYTOMETRY ASSAYS

5.5.1 Flow Cytometry with Biofilm Samples

To perform flow cytometry on bacterial samples grown as biofilms, cells were inoculated in 48-well plates, as described in Section 5.4, in biofilm media containing no additional iron (0 μM), 2 μM, 4 μM or 50 μM Fe^{2+}. Each bacterial strain was inoculated in triplicate samples in each concentration of iron. Biofilms were harvested by mechanically breaking each film with a micropipette tip to add 0.06% Tween-80 to disperse the film, followed by pipetting up and down several times with a standard P-1000 pipette tip. Four-day biofilms were then incubated at room temperature for 1-2 hours, followed by further pipetting to break up clumps, before passage through 35 μm strainer-cap tubes. These strained samples were immediately used for flow cytometry. When unavoidable, samples were placed at 4ºC prior to processing at the cell analyzer. All flow cytometry assays were performed on a BD LSRFortessa cell analyzer (BD Biosciences), and raw data was acquired and saved using BD FACSDiva Software Version 6.1.3 (BD Biosciences). All further analysis of this data was performed using Ferdinand, a flow cytometry data analysis software that was developed by Dr. Jeffrey Lawrence, Department of Biological Sciences, University of Pittsburgh. Data analysis methodology is described in Section 5.6.3.

5.5.2 Flow Cytometry with Planktonic Samples

To perform flow cytometry on planktonic samples, for comparison with the same strains grown as biofilms, cells were inoculated in 125 mL flasks of reconstituted biofilm media containing 0.05% Tween-80. Cells were inoculated to an initial optical density ($OD_{600\ nm}$) of 0.001, and grown at 37ºC on a shaker set at 200 r.p.m. Samples were harvested at early log, late log, stationary and late stationary phases of growth, and passaged through a strainer-cap, if necessary. All flow cytometry readings with planktonically grown bacterial samples were acquired as described in Section 5.6.1, and and analyzed as described in Section 5.6.3.

5.5.3 Flow cytometry data analysis using Ferdinand

All data analyses following cytometry runs with samples processed from planktonic cultures and biofilms were performed using Ferdinand, a flow cytometry data analysis software developed at the University of Pittsburgh.

5.5.3.1 Data Import and Initial Filtering of Events by Forward and Side Scatter Parameters

All flow cytometry data, encoded in flow cytometry standard (FCS) files that were originally generated and saved on the BD FACSDiva platform, were re-opened and read using Ferdinand for any further data analysis. The initial plot of positive events were derived after application of pre-defined thresholds on the forward scatter and side scatter detectors, for scoring positive events that represent single bacterial cells. These pre-defined values were described using a

sample set of data files, and set to accommodate events over a normal distribution curve. Table 13 lists global gate parameters that were applied to all flow cytometry files, for biofilms as well as planktonic sample runs.

Table 13. Detector Threshold Gates and Filter Parameters

List of threshold values used for scoring events as positive for global gates.
"Z" indicates Z-values for scoring positives.

Gate Name	Threshold Gating (Absolute Values)		Filter Parameters	
	Min. Gate	Max. Gate	Min. Z	Max. Z
FSC-A	1	10000	-3	3
FSC-H	1	10000	-3	3
SSC-A	1	100000	-3	3
SSC-H	1	100000	-3	3
FSC-W	10	1000	-3	3
SSC-W	10	800	-3	3
GFP	10	2000	-1.5	1.5
mCherry	10	2000	-1.5	1.5

5.5.3.2 Channel Data Parameters and Compensation Matrices for Fluorescent Channels

Following the filtering of events past the threshold values listed in Table 13, a raw fluorescence signal distribution plot could be generated for each of the two fluorescent channels - eGFP and mCherry. At this stage, however, the fluorescence signal distribution plot obtained for each channel was not corrected for spillover of fluorescence across channels. To compensate for the detection of eGFP or mCherry fluorescence across detectors, standard curves obtained from

the single fluorescent protein-tagged strains listed in Table 3, were used to derive raw compensations on Ferdinand.

Spillover compensations were derived afresh for each experimental replicate. Single fluorescent protein-tagged strains grown under different concentrations of iron were used to derive raw compensations for data from experimental samples grown in the corresponding concentration of iron. Compensation matrices were also derived separately for samples grown planktonically or as biofilms. Following the derivation and application of compensation matrices, the resulting adjustment in fluorescence spillover was tested by reloading the FCS files for each single fluorescent strain. The mean fluorescence values for the theoretically absent fluorescence signal that is recorded from the single fluorescent protein-tagged strains, should be close to zero. Tables 14 and 15 show the signal compensation values applied to each fluorescence channel while processing data obtained from four-day biofilms and planktonic cultures grown in 0 μM, 2 μM, 4 μM and 50 μM Fe^{2+}. As an example for the adjustment that these compensations apply to the raw data, Table 16 shows raw mean fluorescence signals and compensated mean fluorescence values obtained for one experimental run of single fluorescent protein-tagged strains grown as biofilms or planktonic cultures. As an illustration of the change in overall fluorescence signals recorded upon transformation of raw data using the compensation matrices, Figures 10 and 14 in Chapter 2 show histograms of raw (Panel A,C) and compensated (Panel B,D) fluorescence signal distributions for single fluorescent protein-tagged strains grown planktonically in 0 μM and 50 μM Fe^{2+} conditions, respectively, for one representative experimental set of data. Similarly, Figure 17 shows raw (Panel A,C) and compensated (Panel B,D) fluorescence signal distributions for single fluorescent protein-tagged strains grown to four-day mature biofilms in 0 μM Fe^{2+}.

Promoter expression was expressed as mean fluorescence values. For data obtained from biofilm samples, the average of mean fluorescence values determined from three experiments were plotted graphically to compare relative levels of fluorescence as a function of iron availability (Figures 16A and 22A). For planktonic data, the average of mean fluorescence values determined from two experiments, were plotted in Figure 16B and 22B.

To determine the subsets of fluorescent cells that make up the overall sample, simple threshold counts obtained on Ferdinand for the eGFP and mCherry detectors were recorded following signal compensation of flow cytometry data, as described above. These thresholds classified cells as positive for green and/or red fluorescence based on the threshold parameters listed in Table 13, for the eGFP and mCherry detectors. This method provides a crude read-out of the overall abundance of each fluorescent tag in the sample, when used simply based upon detector threshold counts.

Table 14. Compensation Matrices for Four-Day Biofilms

Signal compensation values derived using single fluorescent protein-tagged strains grown as biofilms, under each concentration of Fe^{2+}. These values were applied to convert raw fluorescence signals to compensated fluorescence signals in dual fluorescent protein-tagged strains.

4-day Biofilms; 0 μM Fe^{2+}	***Detectors: eGFP***		***mCherry***	
	eGFP	**mCherry**	**eGFP**	**mCherry**
Compensation Mean	100	1.21	55.49	100
Compensation Variable	0	1.47	38.21	0
4-day Biofilms; 2 μM Fe^{2+}	**Detectors: eGFP**		**mCherry**	
	eGFP	**mCherry**	**eGFP**	**mCherry**
Compensation Mean	100	1.23	48.99	100
Compensation Variable	0	1.64	36.95	0
4-day Biofilms; 4 μM Fe^{2+}	**Detectors: eGFP**		**mCherry**	
	eGFP	**mCherry**	**eGFP**	**mCherry**
Compensation Mean	100	1.19	34.47	100
Compensation Variable	0	1.79	21.07	0
4-day Biofilms; 50 μM Fe^{2+}	**Detectors: eGFP**		**mCherry**	
	eGFP	**mCherry**	**eGFP**	**mCherry**
Compensation Mean	100	1.19	35.29	100
Compensation Variable	0	1.99	25.35	0

Table 15. Compensation Matrices for Planktonic Cultures

Compensation matrices derived using single fluorescent protein-tagged strains grown as planktonic cultures, under each concentration of Fe^{2+}. These matrices were applied to convert raw fluorescence signals to compensated fluorescence signals in dual fluorescent protein-tagged strains.

0 µM Fe^{2+}	Detectors: eGFP		mCherry	
	eGFP	mCherry	eGFP	mCherry
Compensation Mean	100	0.49	18.57	100
Compensation Variable	0	1.09	16.83	0

2 µM Fe^{2+}	Detectors: eGFP		mCherry	
	eGFP	mCherry	eGFP	mCherry
Compensation Mean	100	0.5	19.67	100
Compensation Variable	0	0.87	15.27	0

4 µM Fe^{2+}	Detectors: eGFP		mCherry	
	eGFP	mCherry	eGFP	mCherry
Compensation Mean	100	0.41	19.67	100
Compensation Variable	0	0.63	15.27	0

50 µM Fe^{2+}	Detectors: eGFP		mCherry	
	eGFP	mCherry	eGFP	mCherry
Compensation Mean	100	0.55	21.76	100
Compensation Variable	0	1.3	16.34	0

Table 16. Sample Data Representing Raw and Compensated Fluorescence Values

Raw and compensated mean fluorescence values obtained upon application of compensation matrices such as those in Tables 14 and 15. These data represent values derived for single fluorescent protein-tagged strains, for one experimental run at 0 μM or 50 μM Fe^{2+}.

Planktonic Cultures	**0 μM Fe^{2+}**				**50 μM Fe^{2+}**			
Strain	**Mean Channel Values**				**Mean Channel Values**			
	Raw		**Compensated**		**Raw**		**Compensated**	
	eGFP	**mCherry**	**eGFP**	**mCherry**	**eGFP**	**mCherry**	**eGFP**	**mCherry**
mc²155 with P_{hsp60}-eGFP	17.3	3.8	17.3	1.6	17.3	4.5	17.3	1.6
mc²155 with P_{hsp60}-mCherry	2.7	837.5	0	837.7	3.4	990.8	-0.6	991.3
Four-Day Biofilms	**0 μM Fe^{2+}**				**50 μM Fe^{2+}**			
Strain	**Mean Channel Values**				**Mean Channel Values**			
	Raw		**Compensated**		**Raw**		**Compensated**	
	eGFP	**mCherry**	**eGFP**	**mCherry**	**eGFP**	**mCherry**	**eGFP**	**mCherry**
mc2155 with P_{hsp60}-eGFP	25.2	13.5	25.2	1.1	30.3	11.5	30.3	1.8
mc2155 with P_{hsp60}-mCherry	6.3	730.7	-1.5	732.1	9.1	1529.7	-7.5	1533.0

5.6 SURFACE TRANSLOCATION ASSAYS

Unless otherwise specified, all surface translocation assays were performed on standard biofilm media, with the only changes being that the amount of glucose added while reconstituting the biofilm base, was reduced from 2% to 0.2% to facilitate sliding. The medium was solidified with 0.3% agarose. The iron concentration in the standard media used for surface translocation assays was typically maintained at 4 μM, similar to standard biofilm media used for biofilm assays. This media was poured fresh, and the plates were allowed to solidify and dry at room temperature for one day, prior to inoculation for the assay. 1 μL of a saturated bacterial culture was gently placed (to avoid piercing the agarose) at the center of the plate, and allowed to dry. Plates were incubated inverted at 37ºC in sealed bags containing damp tissue paper, to allow for their moisture content to be retained. The tissue paper was typically maintained moist, over the course of the assay. When performing assays with multiple spots on the same plate, samples were placed equidistant from the center of the plate.

5.7 SPECIALIZED TRANSDUCTION AND SCREENING OF MUTANTS

5.7.1 Generation of High Titer Lysates of Specialized Transducing Phage MycoMariner

High titer lysates of the specialized tranducing phage MycoMariner, obtained from the Bill Jacobs laboratory, were generated as described by Sarkis and Hatfull [87], by propagating the phage on *M. smegmatis* lawns at the permissive temperature of 30ºC. *M. smegmatis* was grown

in 7H9+ADC+$CaCl_2$ without Tween, to an OD_{600} of approximately 1. Phage dilutions of 10^0-10^{-9} in phage buffer containing 1 mM $CaCl_2$, were each used to infect 300 μL of this bacterial culture. Following 30 minutes of adsorption, each infection was mixed with 7H9, 1mM $CaCl_2$, and 0.35% MBTA, and plated as top agar lawns on 7H10+CB+CHX plates. Plates were incubated at the permissive temperature of 30ºC overnight, or until plaques became visible. The nearly cleared plate was flooded with phage buffer, incubated at room temperature for 4-5 hours, and the lysate was then siphoned off the plate and filtered by passage through a 0.22 μm filter, before storage at 4ºC.

5.7.2 Specialized Transduction using Tn MycoMariner

The methodology for specialized transduction developed by Bardarov, *et. al.,* was utilized to generate transposon mutants of *M. smegmatis* strain mc^2155 [130]. Bacterial cells were grown in 7H9 supplemented with ADC and $CaCl_2$, without Tween, to an OD_{600} of 0.8-1. 10 mL of this culture were spun down by centrifugation, and resuspended in 1 mL of pre-warmed 7H9+ADC maintained at 37ºC. The appropriate volume of MycoMariner specialized transducing delivery phage lysate, which was generated as described in Section 5.8.1, was added to this culture to allow for a multiplicity of infection (m.o.i.) of 10. This mixture was incubated at the non-permissive temperature of 37ºC for 30 minutes, to allow for phage adsorption, followed by outgrowth in 50 mL pre-warmed 7H9+ADC+Tween at 37ºC for 30 minutes. The culture was then centrifuged to pellet the cells, and resuspended in 1 mL of PBS+Tween. This entire volume was split into several microliter aliquots and plated on 7H10+ADC+Hyg plates. Following incubation at 37ºC for 3 days, all tranductants were inoculated into 1 mL of 7H9+ADC+Tween,

grown to an OD_{600} of approximately 1, and then separately assayed for proficiency in biofilm formation and surface translocation, as compared to the parent strain.

5.7.3 Extraction of Genomic DNA from *M. smegmatis*

To map the transposon insertion site in the mutant we chose to characterize, the first step taken was to extract genomic DNA, which was used for sequencing reactions. Genomic DNA was extracted following a similar protocol as described by van Kessel and Hatfull [104]. A 50 mL culture of *M. smegmatis* was grown to late log phase, to an OD_{600} of 0.8-1. Cells were harvested by centrifugation at 5000 r.p.m. for 10 minutes, and the pellet was resuspended in 5 mL of GTE solution (50 mM glucose; 25 mM Tris-HCl, pH8; 10 mM EDTA) containing 10 mg/mL of lysozyme. This mixture was incubated without disturbance at 37ºC for 18-20 hours, followed by treatment with Proteinase K (0.1 mg/mL), RNase A (10 μg/ml), 15 mM EDTA, and SDS to a final concentration of 1 percent. This mixture was incubated at 37ºC for 4-5 hours, followed by incubation at 60ºC for 30 minutes. Once the mixture was cooled to room temperature, it was treated with an equal volume of 25:24:1 of phenol:chloroform:isoamyl alcohol, gently mixed, and then centrifuged at 12,000 r.p.m. for 5 minutes. This step was repeated with the aqueous phase retrieved from the previous spin. The aqueous phase retrieved after the second spin was mixed with an equal volume of 24:1 of chloroform:isoamyl alcohol, gently mixed, and centrifuged at 12,000 r.p.m. for 5 minutes. The aqueous phase was mixed with a tenth of its volume of 3M sodium acetate at pH 5.2, and two volumes of ethanol were slowly added along the sides of the tube, to form a layer over the aqueous phase. The DNA that precipitated out was spooled with a glass loop, rinsed with 70% ethanol, and then dissolved in TE for storage.

Working stocks of this DNA preparation were generated by diluting to $1/10^{th}$ of the stock concentration, in dH_20.

5.7.4 Mapping of the Transposon Insertion Site

The transposon insertion site was determined by performing PCR on genomic DNA, using a junction primer that was specific to the hygromycin cassette delivered by the transducing phage, and a random primer as the second primer. 1 μL of the PCR product obtained from this reaction was then used as template for a round of two-step PCR, using a second junction primer that annealed to the product amplified by the first junction primer, and a T7 nested primer that annealed to the product amplified by the random primer used in the first PCR. The thermocycler parameters for this PCR were set as follows: 95ºC for 5 minutes, followed by 3 cycles of 95ºC for 30 seconds, 45ºC for 30 seconds, 68ºC for 3 minutes, followed by 27 cycles of 95ºC for 30 seconds, 60ºC for 30 seconds, 68ºC for 3 minutes, followed by 68ºC for 7 minutes, and 4ºC to end. PCR products were then run on a 0.9% agaorse gel, excised and purified, before being sent for sequencing. The sequencing primer was designed to match the hygromycin cassette introduced by the transposon. The sequence retrieved from GeneWiz was analyzed using a BLASTN search against the nucleotide collection, with the program selection relaxed for somewhat similar sequences.

5.8 PLASMIDS USED IN THIS STUDY

Table 17. Plasmids constructed by others

Plasmid	Description	Features	Ab^R
pJL37	Extrachromosomal; carries P_{hsp60} from *M. bovis* BCG	Cloning region, *ori*E, *ori*M	Kan^R
pJL37-eGFP	Extrachromosomal; derived from pJL37, with *eGFP* fused downstream of P_{hsp60}	eGFP, *ori*E, *ori*M	Kan^R
pTTP1A	Integration proficient vector carrying mycobacteriophage Tweety attP-Integrase cassette; reverse orientation	*ori*E, Tweety attP-Int cassette [88]	Kan^R
pTTP1B	Integration proficient vector carrying mycobacteriophage Tweety attP-Integrase cassette in forward orientation	*ori*E, Tweety attP-Int cassette [88]	Kan^R
pGH1000A	Integration proficient vector carrying mycobacteriophage Giles attP-Integrase cassette in forward orientation	*ori*E, Giles attP-Int cassette [89]	Hyg^R
pGH1000B	Integration proficient vector carrying mycobacteriophage Giles attP-Integrase cassette in reverse orientation	*ori*E, Giles attP-Int cassette [89]	Hyg^R
phsp60-Cherry	Replicative vector backbone obtained from the Eric Rubin	*ori*E, *ori*M, mCherry	Kan^R

pMH94	Integration proficient vector carrying mycobacteriophage L5 attP-Integrase cassette	*ori*E,L5attP-Int cassette	Kan^R
pMS0605eGFP	L5 integration vector carrying P_{MSMEG_0615}-*eGFP*	*ori*E, L5 attP-Int cassette	Kan^R
pYUB854	Used to generate targeting substratefor mycobacterial recombineering	Hyg^R cassette flanked by MCS and γσ resolvase sites	Hyg^R
pGH542	Constitutively expresses γσ resolvase; used to unmark gene knockouts	*ori*E, *ori*M	Tet^R
pJV53	Carries Che9c genes 60-61, under the regulation of the acetamidase promoter	*ori*E, *ori*M [104]	Kan^R

Table 18. Plasmids constructed by AB

Plasmid	Description	Features	Ab^R
pAB03	pJL37 backbone; P_{hsp60} replaced by P_{MSMEG_6758}	*ori*E, *ori*M	Kan^R
pTweetyhsp60-eGFP	P_{hsp60}-eGFP cloned into pTTP1B backbone	eGFP, *ori*E	Kan^R
pTweetyhsp60-eYFP	P_{hsp60}-eYFP cloned into pTTP1B backbone	eYFP, *ori*E	Kan^R
pTweetyhsp60-eCFP	P_{hsp60}-eCFP cloned into pTTP1B backbone	eYFP, *ori*E	Kan^R
pTweetyhsp60-DsRed2	P_{hsp60}-DsRed2 cloned into pTTP1B backbone	*ori*E, DsRed2	Kan^R
pTweetyhsp60-DsRedexpress	P_{hsp60}-DsRedExpress cloned into pTTP1B backbone	*ori*E, DsRedExpress	Kan^R
pGH1000hsp60-eGFP	P_{hsp60}-eGFP cloned into pGH1000A backbone	*ori*E, eGFP	Hyg^R
pGH1000hsp60-mCherry	RBSgp9-hsp60-mCherryBomb cassette cloned into pGH1000A	*ori*E,mCherry	Hyg^R
pTweety4515-eGFP	pTweetyhsp60-eGFP; P_{hsp60} replaced by P_{MSMEG_4515}	*ori*E,eGFP	Kan^R

pTweety0011-eGFP	pTweetyhsp60-eGFP, with P_{hsp60} replaced by P_{MSMEG_0011} between *Nde*I and *Kpn*I	*ori*E,eGFP	Kan^R
pTweety0911-eGFP	pTweetyhsp60-eGFP, with P_{hsp60} replaced by P_{MSMEG_0911} between *Nde*I and *Kpn*I	*ori*E,eGFP	Kan^R
pJL374515-eGFPdir	P_{hsp60} of pJL37 replaced with P_{MSMEG_4515}-EGFP cassette	*ori*E, *ori*M	Kan^R
pJL374515-eGFPrev	P_{hsp60} of pJL37 replaced with P_{MSMEG_4515}-EGFP cassette in reverse orientation	*ori*E, *ori*M	Kan^R
pMS_1240c, MS_1240-41c	pMH94 with *MSMEG_1240* genomic region and MSMEG_1240-1240 to test for complementation of MSMEG_1240::Tn*Mariner*.	*ori*E, Kan^R	

5.9 BIOINFORMATIC ANALYSES

5.9.1 Programs Used

5.9.1.1 Sequence Alignments and Comparisons

BLASTN: nucleotide sequence alignment retrieval. Settings used: database - nucleotide collection; program optimization - somewhat similar sequences [131].

ClustalW: Multiple sequence alignments for proteins and DNA. Settings used: Default; free end gap [132].

5.9.1.2 Protein Analysis

BLASTP: protein database search using query sequence; Reference Sequences were used as the search database [131]

HHpred: protein homology detection and protein structure prediction using hidden Markov models. Settings used: Default [112].

MEME: motif based sequence analysis tool used to identify conserved motifs in MSMEG_1240 and it homologs [113].

CLC Main Workbench 6: software used for nucleotide and amino acid sequence analysis and generation of alignments and graphics.

5.9.2 Databases

5.9.2.1 Bacterial Genes and Proteins

KEGG: information on genomes, proteins, and enzymatic pathways distributed over a collection of databases, established by the Japanese Human Genome Programme [110].

JCVI Comprehensive Microbial Resources: comprehensive information on publically available prokaryotic genomes [109].

Pfam (Protein family database): database of protein families based on sequence homology [111].

BIBLIOGRAPHY

1. WHO, W.H.O., *Global Tuberculosis Control Report*, 2011.
2. Lawn, S.D. and A.I. Zumla, *Tuberculosis.* The Lancet. **378**(9785): p. 57-72.
3. Gandhi, N.R., et al., *Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis.* The Lancet, 2010. **Volume 375**(Issue 9728): p. 1830-1843.
4. Barry, C.E., et al., *The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.* Nat Rev Micro, 2009. **7**(12): p. 845-855.
5. Jindani, A., C.J. Doré, and D.A. Mitchison, *Bactericidal and Sterilizing Activities of Antituberculosis Drugs during the First 14 Days.* American Journal of Respiratory and Critical Care Medicine, 2003. **167**(10): p. 1348-1354.
6. Kaufmann, S.H.E., *How can immunology contribute to the control of tuberculosis?* Nat Rev Immunol, 2001. **1**(1): p. 20-30.
7. Stewart, G.R., B.D. Robertson, and D.B. Young, *Tuberculosis: a problem with persistence.* Nat Rev Micro, 2003. **1**(2): p. 97-105.
8. Costerton, J.W., P.S. Stewart, and E.P. Greenberg, *Bacterial Biofilms: A Common Cause of Persistent Infections.* Science, 1999. **284**(5418): p. 1318-1322.
9. Lewis, K., *Persister Cells.* Annual Review of Microbiology, 2010. **64**(1): p. 357-372.
10. Darzins, E. and G. Fahr, *Cord-Forming Property, Lethality and Pathogenicity of Mycobacteria.* Diseases of the Chest, 1956. **30**(6): p. 642-648.
11. Ojha, A.K., et al., *Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria.* Molecular Microbiology, 2008. **69**(1): p. 164-174.
12. Lenaerts, A.J., et al., *Location of persisting mycobacteria in a Guinea pig model of tuberculosis revealed by r207910.* Antimicrob Agents Chemother, 2007. **51**(9): p. 3338-45.
13. Pang, J.M., et al., *The Polyketide Pks1 Contributes to Biofilm Formation in Mycobacterium tuberculosis.* J Bacteriol, 2012. **194**(3): p. 715-721.
14. Costerton, J.W., G.G. Geesey, and K.J. Cheng, *How bacteria stick.* Sci Am, 1978. **238**(1): p. 86-95.
15. de Beer, D., P. Stoodley, and Z. Lewandowski, *Liquid flow in heterogeneous biofilms.* Biotechnology and Bioengineering, 1994. **44**(5): p. 636-641.
16. Lawrence, J.R., et al., *Optical sectioning of microbial biofilms.* J Bacteriol, 1991. **173**(20): p. 6558-6567.

17. O'Toole, G., H.B. Kaplan, and R. Kolter, *BIOFILM FORMATION AS MICROBIAL DEVELOPMENT.* Annual Review of Microbiology, 2000. **54**(1): p. 49-79.
18. Pratt, L.A. and R. Kolter, *Genetic analysis of Escherichia coli biofilm formation: roles of flagella, motility, chemotaxis and type I pili.* Molecular Microbiology, 1998. **30**(2): p. 285-293.
19. Dewanti, R. and A.C.L. Wong, *Influence of culture conditions on biofilm formation by Escherichia coli O157:H7.* International Journal of Food Microbiology, 1995. **26**(2): p. 147-164.
20. O'Toole, G.A. and R. Kolter, *Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development.* Molecular Microbiology, 1998. **30**(2): p. 295-304.
21. Heilmann, C., et al., *Characterization of Tn917 insertion mutants of Staphylococcus epidermidis affected in biofilm formation.* Infection and Immunity, 1996. **64**(1): p. 277-82.
22. Muller, E., et al., *Isolation and characterization of transposon mutants of Staphylococcus epidermidis deficient in capsular polysaccharide/adhesin and slime.* Infection and Immunity, 1993. **61**(2): p. 551-558.
23. Mack, D., et al., *Characterization of transposon mutants of biofilm-producing Staphylococcus epidermidis impaired in the accumulative phase of biofilm production: genetic identification of a hexosamine-containing polysaccharide intercellular adhesin.* Infection and Immunity, 1994. **62**(8): p. 3244-3253.
24. Mack, D., et al., *The intercellular adhesin involved in biofilm accumulation of Staphylococcus epidermidis is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis.* J Bacteriol, 1996. **178**(1): p. 175-83.
25. McKenney, D., et al., *The ica Locus of Staphylococcus epidermidis Encodes Production of the Capsular Polysaccharide/Adhesin.* Infection and Immunity, 1998. **66**(10): p. 4711-4720.
26. Cramton, S.E., et al., *The Intercellular Adhesion (ica) Locus Is Present in Staphylococcus aureus and Is Required for Biofilm Formation.* Infection and Immunity, 1999. **67**(10): p. 5427-5433.
27. Deighton, M. and R. Borland, *Regulation of slime production in Staphylococcus epidermidis by iron limitation.* Infection and Immunity, 1993. **61**(10): p. 4473-4479.
28. O'Gara, J.P., *ica and beyond: biofilm mechanisms and regulation in Staphylococcus epidermidis and Staphylococcus aureus.* FEMS Microbiology Letters, 2007. **270**(2): p. 179-188.
29. Flemming, H.C., T.R. Neu, and D.J. Wozniak, *The EPS matrix: the "house of biofilm cells".* J Bacteriol, 2007. **189**(22): p. 7945-7.
30. Branda, S.S., et al., *A major protein component of the Bacillus subtilis biofilm matrix.* Molecular Microbiology, 2006. **59**(4): p. 1229-1238.
31. Romero, D., et al., *Amyloid fibers provide structural integrity to Bacillus subtilis biofilms.* Proceedings of the National Academy of Sciences, 2010. **107**(5): p. 2230-2234.
32. Parsek, M.R. and P.K. Singh, *BACTERIAL BIOFILMS: An Emerging Link to Disease Pathogenesis.* Annual Review of Microbiology, 2003. **57**(1): p. 677-701.
33. Whiteley, M., et al., *Gene expression in Pseudomonas aeruginosa biofilms.* Nature, 2001. **413**(6858): p. 860-864.

34. Waite, R.D., et al., *Transcriptome Analysis of Pseudomonas aeruginosa Growth: Comparison of Gene Expression in Planktonic Cultures and Developing and Mature Biofilms.* J Bacteriol, 2005. **187**(18): p. 6571-6576.
35. Waite, R.D., et al., *Clustering of Pseudomonas aeruginosa transcriptomes from planktonic cultures, developing and mature biofilms reveals distinct expression profiles.* BMC Genomics, 2006. **7**: p. 162.
36. Schembri, M.A., K. Kjærgaard, and P. Klemm, *Global gene expression in Escherichia coli biofilms.* Molecular Microbiology, 2003. **48**(1): p. 253-267.
37. Beloin, C., et al., *Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression.* Molecular Microbiology, 2004. **51**(3): p. 659-674.
38. Ren, D., et al., *Gene expression in <i>Escherichia coli biofilms.* Applied Microbiology and Biotechnology, 2004. **64**(4): p. 515-524.
39. Stanley, N.R., et al., *Identification of Catabolite Repression as a Physiological Regulator of Biofilm Formation by Bacillus subtilis by Use of DNA Microarrays.* J Bacteriol, 2003. **185**(6): p. 1951-1957.
40. Beenken, K.E., et al., *Global Gene Expression in Staphylococcus aureus Biofilms.* J Bacteriol, 2004. **186**(14): p. 4665-4684.
41. Domka, J., et al., *Temporal gene-expression in Escherichia coli K-12 biofilms.* Environmental Microbiology, 2007. **9**(2): p. 332-346.
42. Vlamakis, H., et al., *Control of cell fate by the formation of an architecturally complex bacterial community.* Genes Dev, 2008. **22**(7): p. 945-53.
43. Lopez, D., H. Vlamakis, and R. Kolter, *Biofilms.* Cold Spring Harb Perspect Biol, 2010. **2**(7): p. a000398.
44. Teng, R. and T. Dick, *Isoniazid resistance of exponentially growing Mycobacterium smegmatis biofilm culture.* FEMS Microbiol Lett, 2003. **227**(2): p. 171-4.
45. Carter, G., et al., *Characterization of biofilm formation by clinical isolates of Mycobacterium avium.* J Med Microbiol, 2003. **52**(Pt 9): p. 747-52.
46. Primm, T.P., C.A. Lucero, and J.O. Falkinham, 3rd, *Health impacts of environmental mycobacteria.* Clin Microbiol Rev, 2004. **17**(1): p. 98-106.
47. Hall-Stoodley, L., et al., *Mycobacterium marinum biofilm formation reveals cording morphology.* FEMS Microbiol Lett, 2006. **257**(1): p. 43-9.
48. Hall-Stoodley, L. and H. Lappin-Scott, *Biofilm formation by the rapidly growing mycobacterial species Mycobacterium fortuitum.* FEMS Microbiol Lett, 1998. **168**(1): p. 77-84.
49. Ojha, A., et al., *GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria.* Cell, 2005. **123**(5): p. 861-73.
50. Ojha, A.K., et al., *Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria.* Mol Microbiol, 2008. **69**(1): p. 164-74.
51. Schulze-Robbecke, R., B. Janning, and R. Fischeder, *Occurrence of mycobacteria in biofilm samples.* Tuber Lung Dis, 1992. **73**(3): p. 141-4.
52. Kressel, A.B. and F. Kidd, *Pseudo-outbreak of Mycobacterium chelonae and Methylobacterium mesophilicum caused by contamination of an automated endoscopy washer.* Infect Control Hosp Epidemiol, 2001. **22**(7): p. 414-8.

53. Marsollier, L., et al., *Aquatic snails, passive hosts of Mycobacterium ulcerans.* Appl Environ Microbiol, 2004. **70**(10): p. 6296-8.
54. Marsollier, L., et al., *Aquatic Plants Stimulate the Growth of and Biofilm Formation by Mycobacterium ulcerans in Axenic Culture and Harbor These Bacteria in the Environment.* Appl Environ Microbiol, 2004. **70**(2): p. 1097-1103.
55. Recht, J. and R. Kolter, *Glycopeptidolipid acetylation affects sliding motility and biofilm formation in Mycobacterium smegmatis.* J Bacteriol, 2001. **183**(19): p. 5718-24.
56. Chen, J.M., et al., *Roles of Lsr2 in colony morphology and biofilm formation of Mycobacterium smegmatis.* J Bacteriol, 2006. **188**(2): p. 633-41.
57. Ren, H., et al., *Identification of the lipooligosaccharide biosynthetic gene cluster from Mycobacterium marinum.* Mol Microbiol, 2007. **63**(5): p. 1345-59.
58. Ojha, A., et al., *Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria*. Molecular Microbiology, 2008. **69**(1): p. 164-74.
59. Ojha, A.K., et al., *Enzymatic hydrolysis of trehalose dimycolate releases free mycolic acids during mycobacterial growth in biofilms*. J Biol Chem, 2010. **285**(23): p. 17380-9.
60. Zambrano, M.M. and R. Kolter, *Mycobacterial biofilms: a greasy way to hold it together*. Cell, 2005. **123**(5): p. 762-4.
61. Ojha, A. and G.F. Hatfull, *The role of iron in Mycobacterium smegmatis biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonic growth.* Molecular Microbiology, 2007. **66**(2): p. 468-483.
62. Höner Zu Bentrup, K., et al., *Characterization of Activity and Expression of Isocitrate Lyase in Mycobacterium avium andMycobacterium tuberculosis.* J Bacteriol, 1999. **181**(23): p. 7161-7167.
63. Wayne, L.G. and K.Y. Lin, *Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions.* Infection and Immunity, 1982. **37**(3): p. 1042-1049.
64. Henrichsen, J., *Bacterial surface translocation: a survey and a classification.* Bacteriol Rev, 1972. **36**(4): p. 478-503.
65. Harshey, R.M., *Bacterial motility on a surface: many ways to a common goal.* Annu Rev Microbiol, 2003. **57**: p. 249-73.
66. Harshey, R.M., *Bees aren't the only ones: swarming in gram-negative bacteria*. Mol Microbiol, 1994. **13**(3): p. 389-94.
67. Claret, L. and C. Hughes, *Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during Proteus swarming.* J Bacteriol, 2000. **182**(3): p. 833-6.
68. Mattick, J.S., *Type IV pili and twitching motility.* Annu Rev Microbiol, 2002. **56**: p. 289-314.
69. McBride, M.J., *Bacterial gliding motility: multiple mechanisms for cell movement over surfaces.* Annu Rev Microbiol, 2001. **55**: p. 49-75.
70. Martinez, A., S. Torello, and R. Kolter, *Sliding motility in mycobacteria.* J Bacteriol, 1999. **181**(23): p. 7331-8.

71. Matsuyama, T., A. Bhasin, and R.M. Harshey, *Mutational analysis of flagellum-independent surface spreading of Serratia marcescens 274 on a low-agar medium.* J Bacteriol, 1995. **177**(4): p. 987-91.
72. Recht, J., et al., *Genetic analysis of sliding motility in Mycobacterium smegmatis.* J Bacteriol, 2000. **182**(15): p. 4348-51.
73. De Voss, J.J., et al., *Iron Acquisition and Metabolism by Mycobacteria.* J Bacteriol, 1999. **181**(15): p. 4443-4451.
74. Rodriguez, G.M., et al., *ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response.* Infect Immun, 2002. **70**(7): p. 3371-81.
75. Wisedchaisri, G., et al., *Crystal structures, metal activation, and DNA-binding properties of two-domain IdeR from Mycobacterium tuberculosis.* Biochemistry, 2007. **46**(2): p. 436-47.
76. Gold, B., et al., *The Mycobacterium tuberculosis IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages.* Molecular Microbiology, 2001. **42**(3): p. 851-865.
77. Dussurget, O., M. Rodriguez, and I. Smith, *An ideR mutant of Mycobacterium smegmatis has derepressed siderophore production and an altered oxidative-stress response.* Molecular Microbiology, 1996. **22**(3): p. 535-544.
78. Ratledge, C. and L.G. Dover, *IRON METABOLISM IN PATHOGENIC BACTERIA.* Annual Review of Microbiology, 2000. **54**(1): p. 881-941.
79. Banin, E., M.L. Vasil, and E.P. Greenberg, *Iron and Pseudomonas aeruginosa biofilm formation.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 11076-81.
80. Singh, P.K., et al., *A component of innate immunity prevents bacterial biofilm development.* Nature, 2002. **417**(6888): p. 552-5.
81. Singh, P.K., *Iron sequestration by human lactoferrin stimulates P. aeruginosa surface motility and blocks biofilm formation.* Biometals, 2004. **17**(3): p. 267-70.
82. Hancock, V., M. Dahl, and P. Klemm, *Abolition of Biofilm Formation in Urinary Tract Escherichia coli and Klebsiella Isolates by Metal Interference through Competition for Fur.* Appl Environ Microbiol, 2010. **76**(12): p. 3836-3841.
83. De Voss, J.J., et al., *Iron acquisition and metabolism by mycobacteria.* J Bacteriol, 1999. **181**(15): p. 4443-51.
84. Barclay, R. and C. Ratledge, *Mycobactins and exochelins of Mycobacterium tuberculosis, M. bovis, M. africanum and other related species.* J Gen Microbiol, 1988. **134**(3): p. 771-6.
85. Fiss, E.H., S. Yu, and W.R. Jacobs, Jr., *Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways.* Mol Microbiol, 1994. **14**(3): p. 557-69.
86. De Voss, J.J., et al., *The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1252-7.
87. Sarkis, G.J. and G.F. Hatfull, *Mycobacteriophages*
Mycobacteria Protocols, T. Parish and N.G. Stoker, Editors. 1998, Humana Press. p. 145-173.

88. Pham, T.T., et al., *Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria.* Microbiology, 2007. **153**(Pt 8): p. 2711-23.
89. Morris, P., et al., *Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination.* J Bacteriol, 2008. **190**(6): p. 2172-82.
90. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins.* Nat Methods, 2005. **2**(12): p. 905-9.
91. de Beer, D., P. Stoodley, and Z. Lewandowski, *Liquid flow in heterogeneous biofilms.* Biotechnol Bioeng, 1994. **44**(5): p. 636-41.
92. de Beer, D., et al., *Effects of biofilm structures on oxygen distribution and mass transport.* Biotechnology and Bioengineering, 1994. **43**(11): p. 1131-1138.
93. Carroll, P., et al., *Sensitive Detection of Gene Expression in Mycobacteria under Replicating and Non-Replicating Conditions Using Optimized Far-Red Reporters.* PLoS ONE, 2010. **5**(3): p. e9823.
94. Hansen, M.C., et al., *Assessment of GFP fluorescence in cells of Streptococcus gordonii under conditions of low pH and low oxygen concentration.* Microbiology, 2001. **147**(Pt 5): p. 1383-91.
95. Li, X., et al., *Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter.* Journal of Biological Chemistry, 1998. **273**(52): p. 34970-34975.
96. Riemer, J., et al., *Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells.* Analytical Biochemistry, 2004. **331**(2): p. 370-375.
97. Chen, J.M., et al., *Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein.* Nucleic Acids Research, 2008. **36**(7): p. 2123-2135.
98. Gordon, B.R.G., et al., *Lsr2 of Mycobacterium Represents a Novel Class of H-NS-Like Proteins.* J Bacteriol, 2008. **190**(21): p. 7052-7059.
99. Arora, K., et al., *Inactivation of lsr2 results in a hypermotile phenotype in Mycobacterium smegmatis.* J Bacteriol, 2008. **190**(12): p. 4291-300.
100. Kocincova, D., et al., *Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and affects surface properties in Mycobacterium smegmatis.* Tuberculosis (Edinb), 2008. **88**(5): p. 390-8.
101. Shi, T., T. Fu, and J. Xie, *Polyphosphate deficiency affects the sliding motility and biofilm formation of Mycobacterium smegmatis.* Curr Microbiol, 2011. **63**(5): p. 470-6.
102. Schorey, J.S. and L. Sweet, *The mycobacterial glycopeptidolipids: structure, function, and their role in pathogenesis.* Glycobiology, 2008. **18**(11): p. 832-841.
103. Murcia, M.I., et al., *Mycobacterium colombiense sp. nov., a novel member of the Mycobacterium avium complex and description of MAC-X as a new ITS genetic variant.* International Journal of Systematic and Evolutionary Microbiology, 2006. **56**(9): p. 2049-2054.
104. van Kessel, J.C. and G.F. Hatfull, *Mycobacterial recombineering.* Methods Mol Biol, 2008. **435**: p. 203-15.
105. van Kessel, J.C., L.J. Marinelli, and G.F. Hatfull, *Recombineering mycobacteria and their phages.* Nat Rev Micro, 2008. **6**(11): p. 851-857.
106. Lee, M.H., et al., *Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and*

bacille Calmette-Guérin. Proceedings of the National Academy of Sciences, 1991. **88**(8): p. 3111-3115.
107. Lee, S., et al., *Bxz1, a new generalized transducing phage for mycobacteria*. FEMS Microbiol Lett, 2004. **241**(2): p. 271-6.
108. Pertea, M., et al., *OperonDB: a comprehensive database of predicted operons in microbial genomes.* Nucleic Acids Res, 2009. **37**(Database issue): p. D479-82.
109. Davidsen, T., et al., *The comprehensive microbial resource.* Nucleic Acids Res, 2010. **38**(Database issue): p. D340-5.
110. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
111. Finn, R.D., et al., *The Pfam protein families database.* Nucleic Acids Res, 2010. **38**(Database issue): p. D211-22.
112. Soding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W244-8.
113. Bailey, T.L., et al., *MEME Suite: tools for motif discovery and searching.* Nucleic Acids Research, 2009. **37**(suppl 2): p. W202-W208.
114. Gonzalez-Perez, M., et al., *Genome sequence of the Mycobacterium colombiense type strain, CECT 3035.* J Bacteriol, 2011. **193**(20): p. 5866-7.
115. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets.* Nucleic Acids Res, 2012. **40**(Database issue): p. D109-14.
116. Hoiczyk, E., *Gliding motility in cyanobacteria: observations and possible explanations.* Archives of Microbiology, 2000. **174**(1): p. 11-17.
117. Hall-Stoodley, L., J.W. Costerton, and P. Stoodley, *Bacterial biofilms: from the natural environment to infectious diseases.* Nat Rev Microbiol, 2004. **2**(2): p. 95-108.
118. Ratledge, C., *Iron, mycobacteria and tuberculosis.* Tuberculosis (Edinb), 2004. **84**(1-2): p. 110-30.
119. Lounis, N., et al., *Iron and Mycobacterium tuberculosis infection.* J Clin Virol, 2001. **20**(3): p. 123-6.
120. Murray, M.J., et al., *The adverse effect of iron repletion on the course of certain infections.* Br Med J, 1978. **2**(6145): p. 1113-5.
121. Ratledge, C. and M.J. Hall, *Influence of metal ions on the formation of mycobactin and salicylic acid in Mycobacterium smegmatis grown in static culture.* J Bacteriol, 1971. **108**(1): p. 314-9.
122. Quadri, L.E., et al., *Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin.* Chem Biol, 1998. **5**(11): p. 631-45.
123. Chavadi, S.S., et al., *Mutational and Phylogenetic Analyses of the Mycobacterial mbt Gene Cluster.* J Bacteriol, 2011. **193**(21): p. 5905-5913.
124. Touati, D., *Iron and Oxidative Stress in Bacteria.* Archives of Biochemistry and Biophysics, 2000. **373**(1): p. 1-6.
125. Harrison, F. and A. Buckling, *Siderophore production and biofilm formation as linked social traits.* ISME J, 2009. **3**(5): p. 632-634.
126. Snapper, S.B., et al., *Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis.* Mol Microbiol, 1990. **4**(11): p. 1911-9.

127. Mizuguchi, Y., K. Suga, and T. Tokunaga, *Multiple mating types of Mycobacterium smegmatis.* Jpn J Microbiol, 1976. **20**(5): p. 435-43.
128. Sambrook, J. and D.W. Russell, *Molecular cloning: a laboratory manual*. 2001: Cold Spring Harbor Laboratory Press.
129. Bergmans, H.E., I.M. van Die, and W.P. Hoekstra, *Transformation in Escherichia coli: stages in the process.* J Bacteriol, 1981. **146**(2): p. 564-70.
130. Bardarov, S., et al., *Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis.* Microbiology, 2002. **148**(Pt 10): p. 3007-17.
131. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
132. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.