UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Molecular Medicine

Ph.D. COURSE IN: Molecular Medicine
CURRICULUM: Biomedicine
SERIES XXXV

Characterization of Rv0093c as a new anti-sigma factor of *Mycobacterium tuberculosis*

Coordinator: Prof. Riccardo Manganelli
Supervisor: Prof. Riccardo Manganelli

Ph.D. Student: Greta Segafreddo

TABLE OF CONTENTS

ABSTRACT **1**

1. INTRODUCTION **2**

1.1 Mycobacteria 2

1.1.1 *Mycobacterium tuberculosis* and tubercular disease 2

1.2 Copper homeostasis in *Mycobacterium tuberculosis* 5

1.2.1 Copper accumulation inside phagosomes 5

1.2.2 Copper detoxifying mechanisms 6

1.3 Sigma factors of *Mycobacterium tuberculosis* 7

1.3.1 ECF Sigma Factor C 9

1.3.2 SigC role in copper acquisition 10

1.3.3 Putative anti-sigma factor Rv0093c 11

1.4 General features of bacterial biofilms 13

1.4.1 Mycobacterial biofilms 14

2. AIM OF THE WORK **15**

3. MATERIALS AND METHODS **16**

3.1 Bacterial strains, media, and growth conditions 16

3.2 Biofilm growth and analyses 18

3.3 Plasmids and mutant strains construction 19

3.3.1 Construction of *M. tuberculosis* mutant TB543 and derived strains 22

3.3.2 Construction of *M. tuberculosis* mutant TB548 and derived strains 23

3.3.3 Construction of *M. smegmatis* mutant strains and M-PFC assay 25

3.4 Determination of copper susceptibility with a Resazurin microtiter assay 26

3.5 RNA analyses 26

3.5.1 RNA extraction for RNA sequencing 26

3.5.2 RNA extraction and retro-transcription for quantitative RT-PCR 27

4. RESULTS AND DISCUSSION **29**

4.1 *Mycobacterium tuberculosis* mutants lacking *sigC* and/or *rv0093c* show identical growth rates in complete medium 29

4.1.1 Copper deprivation does not impair growth of an *M. tuberculosis* mutant lacking *sigC* 30

4.2 *M. tuberculosis* mutants lacking *rv0093c* are more sensitive to copper stress 31

4.3 *rv0093c* deletion and copper overload loop activation of export mechanisms 33

4.4 *rv0093c* deletion affects transcription comparably to *sigC* induction 35

4.5 Rv0093c is a transmembrane protein with cytosolic terminal domains 38

4.6 SigC and Rv0093c interact *in vivo* 39

4.7 *sigC* deletion produces a color-shift phenotype in dehydrated cultures 40

5. CONCLUSIONS **44**

REFERENCES **46**

ABSTRACT

Mycobacterium tuberculosis genome encodes thirteen sigma factors, small subunits of the RNA polymerase that allow rapid adaptation of the bacterium transcriptional landscape, playing a fundamental role in stress-response and pathogenesis. Among them, alternative Sigma Factor C (SigC) is still poorly characterized: its function is seemingly linked to copper uptake and biofilm formation, but its regulon remains undefined. Furthermore, no anti-sigma has been associated to it yet, although the protein Rv0093c represents a suitable candidate.

In order to investigate SigC role in the physiology of *M. tuberculosis*, mutants lacking its structural gene, as well as *rv0093c*, were generated and analyzed in a number of contexts such as different culture conditions, SigC-related gene expression and biofilm development.

Although no significant differences were found between *M. tuberculosis* wildtype strain and its mutants when cultured in low-copper conditions, a marked sensitivity to high-copper concentrations was detected for the mutant lacking *rv0093c* compared to the wildtype strain. Moreover, genes that had been reported as part of SigC regulon were indeed upregulated in the *rv0093c* KO mutant, supporting the hypothesis of Rv0093c effectively being SigC-specific anti-sigma factor, and redefying SigC implication in copper metabolism.

Evidence of physical interaction between the two proteins was also obtained experimentally, by setting up a mycobacterial protein fragment complementation assay, executed in the non-pathogenic model species *Mycobacterium smegmatis*.

INTRODUCTION

1.1 Mycobacteria

The genus *Mycobacterium*, sole member of the family of Mycobacteriaceae, includes more than 100 species of mostly environmental bacilli, characterized by a GC-rich genome and a very thick, almost impenetrable cell wall, which provides resistance to acid-fast staining and sets the genus apart from Gram-positive and -negative bacteria[1].

Being for the most part innocuous saprophytes, the ability to cause disease in humans and animals is limited to a small number of opportunistic pathogen species, often referred to as Non-Tubercular Mycobacteria (NTM) or Mycobacteria Other-Than-Tuberculosis (MOTT)[2], and to an even smaller number of obligate pathogens. This last category contains the etiological agents of two of the most historically relevant afflictions of humankind, tuberculosis (TB) and leprosy. Tubercular mycobacteria, able to provoke typical chronic inflammatory lesions called granulomas or tubercules, are slow-growing bacilli generally grouped under the acronym of MTBC (*Mycobacterium tuberculosis* complex), which includes the primary causative agent of human tuberculosis, *M. tuberculosis*, together with *M. canetti* and *M. africanum*, *M. bovis*, which causes a bovine form of the pathology and is transmissible to humans, *M. microti*, *M. caprae* and *M. pinnipedii*[3,4]. Separate characteristics define instead *M. leprae*, the strictly intracellular pathogen responsible for leprosy, with its heavily restricted genome thought to contain only essential genes, and its extremely long replication time which makes it virtually impossible to manipulate in a laboratory setting[5].

1.1.1 *Mycobacterium tuberculosis* and tubercular disease

Mycobacterium tuberculosis, first isolated by Dr. Robert Koch in 1882, is the extremely successful class-three pathogen responsible for tuberculosis in humans. Among the several reasons that make the bacterium remarkably resistant to the host immune defenses and difficult to eradicate from the body, particularly relevant are the structure of its outer membrane, almost impermeable to most antibacterial drugs, as well as its characteristic replication time of 18-24 hours and penchant for dormancy, that heavily limit the number and availability of common antibiotic targets[1,6]. Together with the bacillus efficient adaptability to everchanging outside

stressful conditions, these features explain why the very ancient tubercular disease is still so deeply rooted in modern-day society, often in association with other pathologies, such as AIDS, diabetes, pulmonary chronic diseases, and malnutrition.

Indeed, in 2021 the World Health Organization estimated an incidence of 10.6 million new cases and registered approximately 1.6 million deaths worldwide (**Figure 1.1**), which define tuberculosis as the second deadliest single-pathogen disease in the world, after COVID_19. Notably, since the onset of the coronavirus pandemic, the previously positive trends in TB diagnosis and case-reporting have been heavily impacted, stalling and even reversing the progress made up to 2019 towards achieving the first milestones of the global commitment on disease eradication. Disruption of these essential services ultimately influenced the global TB burden by immediately increasing mortality and, later on, community transmission and disease development[7].

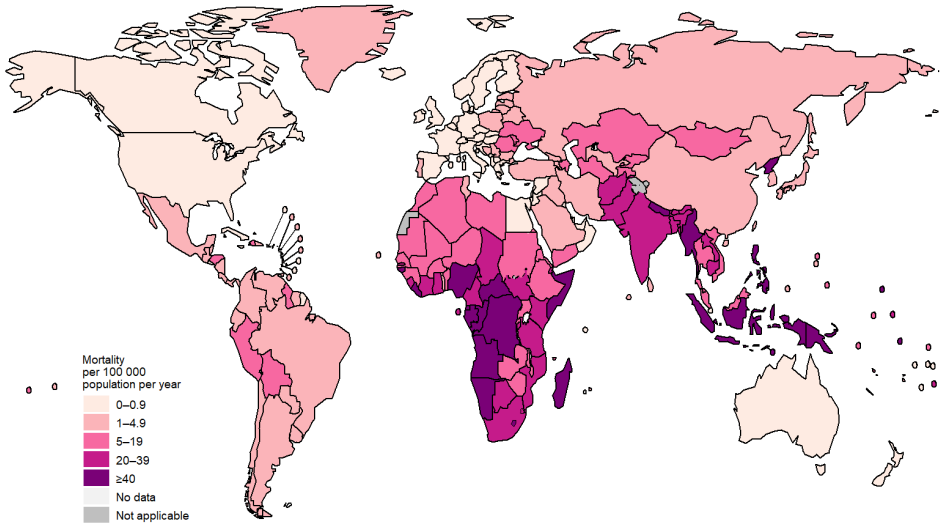

Figure 1.1 Estimated TB mortality rates at the end of 2021, excluding deaths among HIV-positive people. The total number of registered casualties in 2021 was approximately 1.6 M, of which an estimated 191,000 occurred due to MDR/RR-TB (adapted from "Global tuberculosis report 2022" [7]).

Infection with *M. tuberculosis* can remain latent and asymptomatic (LTBI) in immuno-competent individuals. About one-quarter of the world population is in fact estimated to be infected, but only about 5-15% of these people will become ill with active TB[7]. Present-day treatment requires a lengthy multidrug regimen, but prognosis is usually good if the infection remains localized to the lungs, and the causative strain is sensitive to first-line antibiotics

(Rifampicin, Isoniazid, Ethambutol and Pyrazinamide). Therefore, the emergence of multi drug-resistant (MDR), extensively drug-resistant (XDR) and totally drug-resistant (TDR) clinical strains, represents a major public health problem[7].

M. tuberculosis everlasting success is greatly due to its peculiar ability to escape degradation inside alveolar macrophages, by preventing the phago-lysosome fusion, and to tenaciously replicate in newly infected adjacent cells. When the active replication phase is eventually cushioned by the immune system, the infection foci rapidly structure into multi-cellular granulomas, aggregations of epithelioid and foamy macrophages, monocytes, and neutrophils, surrounded by a mantle of lymphocytes and a fibrous cuff produced by fibroblasts, which represent the veritable hallmark of tubercular disease[8] (**Figure 1.2**). Granuloma formation reportedly happens two to three weeks post-infection, when the acquired immunity effectors join their over-taxed innate colleagues, and hyper-activate them[9]. At even later stages, from the granuloma core outward, infected foamy macrophages necrosis produces a semi-solid, hypoxic, and acidic mass of debris, which is referred to as *caseum*[10]. Simultaneously confined and protected inside these pulmonary lesions, survivor bacilli can undergo a complex metabolic switch and persist in a dormant state for years, resulting in latency, to eventually reactivate due to a drop in the host immune defenses, and give rise to secondary TB[8–11].

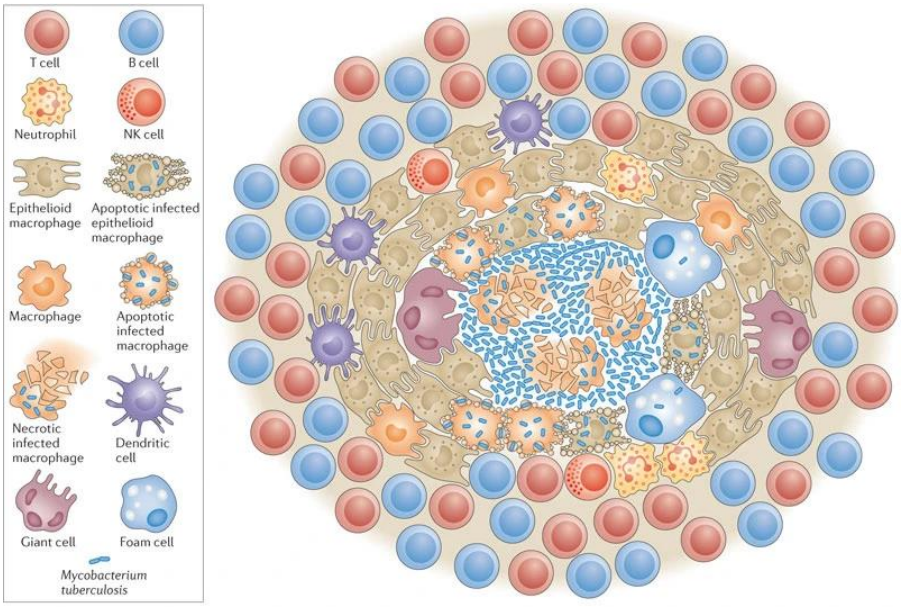

Figure 1.2 Schematic representation of a typical granuloma structure and cellular components (from "Ramakrishnan, *Nature Reviews Immunology*, 2012" [8]).

1.2 Copper homeostasis in *Mycobacterium tuberculosis*

Transition metals such as iron, zinc, copper, and manganese are essential for the growth and development of all organisms. Specifically, copper-dependent enzymes, or cuproproteins, are widely used for electron-transfer reactions in aerobic conditions[12]. As a result, one of the most accomplished ways to prevent and eventually counteract bacterial spreading is nutritional immunity, the host-mediated deprivation of metals[13]. On the contrary, the antimicrobial effects of copper have been known for thousands of years, and copper overload toxicity can be as lethal as its deprivation[14].

The indication that copper could also be exploited as a toxic agent comes from bacteria's own genomes, which encode very few, and almost exclusively extracellularly-active cuproproteins[15]. Little is known about *M. tuberculosis* dedicated mechanisms of copper import: aside from the P-type ATPase Cu^{+}-specific influx pump CtpB[16], no other importers or porins have been detected, and free transit of copper cations across both the mycomembrane and the inner membrane is nearly impossible due to the intrinsic impermeability of lipid layers. Nevertheless, at least two cuproproteins are essential during the pathogen's exponential growth and the replicative stages of infection, explaining the need for copper and thus the potential efficacy of nutritional immunity: SodC, a Cu/Zn superoxide dismutase, which is active against periplasmic and extracellular superoxide anions, and the most energy-efficient terminal oxidase, Cytochrome C Oxidase[17].

In spite of this specific evidence, the toxic effects attributable to copper are seemingly more related to overload rather than starvation in *M. tuberculosis*, and the pathogen's response to this specific stress has been under thorough examination over the last decade[18]. It is not to be excluded that mechanisms ultimately leading to metal shortage, especially characterized for iron[10,13], may in some cases apply to copper, and that disease progression or bacterial relocation to distinct anatomical compartments may vary the direction of the copper-related stress imposed on tubercular bacilli[19].

1.2.1 Copper accumulation inside phagosomes

Following phagocytosis by vanguard macrophages, *M. tuberculosis* begins its struggle against the immune system, stalling the phagosome acidification and its eventual fusion with lysosomes. Unable to efficiently kill the invading bacilli, macrophages recruit other effectors

of immunity to the infection site, such as IFN-γ-secreting $CD4^+$ T lymphocytes, and become activated in their respiratory burst. IFN-γ, and the hypoxic state that characterizes the inside of granulomas, induce the expression of the high-affinity copper importer Ctr1 and copper-chaperon Atox1 in macrophages. The increased copper intake is thus directed to ATP7A, that is partially relocated to the phagosome membrane, to actively pump copper ions directly into the phagosome[12,17]. Withstanding these mechanisms, the concentration of copper inside IFN-γ-activated macrophages infected with *M. tuberculosis* has been shown to range from a minimum of 25 to a maximum of 500 μM[20].

As a redox-active metal, copper engages in Fenton-like reactions with hydrogen peroxide (H_2O_2), actively produced during the respiratory burst. The two resulting toxic species, the hydroxyl radical (·OH) and hydroxyl anion (OH^-), irreversibly damage nucleic acids, lipids, and proteins, particularly those containing thiol groups. Additionally, excess of copper often leads to non-specific binding to proteins needing other metal cofactors, and their subsequent inactivation. Both scenarios concur to the bacterial cell death[20].

1.2.2 Copper detoxifying mechanisms

In case of excessive accumulation of copper, bacteria have evolved resistance mechanisms that entail both chelation and export of the damaging cations. A microarray analysis performed on *M. tuberculosis* in the presence of different concentrations of copper sulfate ($CuSO_4$) has highlighted a total of 30 differentially expressed genes compared to the non-treated controls. Considering that high Cu^+ levels are expected to also activate a general stress response, the 15 genes that were identified with the lowest concentration (50 μM) were deemed core genes, directly responding to copper influx, and 15 more that exhibited variations at the higher concentration (500 μM) were instead connected to a wider copper-related toxicity[20]. Notably, the genes *csoR* and *ricR*, that encode two paralogous copper-sensing repressors and have been later identified as the two main regulators in the copper toxicity response, showed induction at the two distinct conditions. Since copper fluctuations inside the phagosome, or during different stages of the infection have been detected, this difference suggests an underlying redundancy of the system, whose protagonists possess different binding affinities for the metal, and are seemingly activated in a progressive and concentration-dependent way[14,15].

CsoR binds DNA as a dimer of dimers in the absence of Cu^{+}, resulting in the repression of its downstream controlled genes: *csoR* itself, *rv0968*, *ctpV* and *rv0970*[21]. Metallation, or the binding of Cu^{+}, results in the de-repression of the operon, and activation of the leading resistance mechanism, efflux. This action is catered by CtpV, a putative copper-specific P-type ATPase, that pumps copper ions from the cytoplasm to the periplasmic space, possibly in concert with the later discovered porin MctB[12,17]. RicR also acts by directly repressing four other genes beside itself, in the absence of copper, encoding the multicopper oxidase MmcO, putative lipoprotein LpqS and putative permease Rv2963, the metallothionein MymT, and two predicted open reading frames named *socAB*. Activation of the regulon, and of its high chelating potential, is also due to copper-binding[22]. Of its two main studied effectors, MymT is a small cysteine-rich protein able to bind up to six copper ions, and thus able to act as a metal sink and possibly as a chaperon, delivering ions where needed. MmcO is instead a multicopper oxidase and, contrarily to most bacterial soluble MCOs, it is predicted to be lipidated, and membrane-bound, perhaps to oxidize copper extracellularly[14,15,17].

1.3 Sigma factors of *Mycobacterium tuberculosis*

As discussed, within macrophages and granulomas *M. tuberculosis* is exposed to a harsh environmental scenario, marked by acidic pH, depletion of nutrients, reactive oxygen and nitrogen species, and varied antimicrobial agents. Under these hostile circumstances, rapid adjustment of the bacterium transcriptional and translational landscapes is implemented by the engagement of a wide variety of responses. Among these, accessory sigma (σ) factors play a significant role in directing the transcriptional changes towards adaptation and survival.

Sigma factors are small interchangeable subunits of the bacterial RNA polymerase, required to lead the holoenzyme to specific gene promoters and initiate transcription, in concert with the alpha (α) subunit of the enzymatic core. *M. tuberculosis* genome encodes 13 sigma factors, the highest number among obligate pathogens in relation to genome length (ASFD, Alternative Sigma Factor Density). All 13 factors belong to the σ^{70} family, one of the two main families encoded by the bacterial pan-genome, which is divided into four structurally and functionally different subgroups[23–25]:

- Group 1 consists of primary factors, essential for the expression of housekeeping genes, and is represented by a single member in *M. tuberculosis*, SigA.
- Group 2 is composed by primary-like factors, not essential for in vitro growth and mostly related to adaptation to nutrient starvation and stationary phase-related stress. In *M. tuberculosis* it is represented by one component, SigB.
- Group 3 includes different sets of alternative factors, involved in general stress control, such as heat shock or sporulation. SigF of *M. tuberculosis* is part of this group.
- Group 4 gathers all remaining *M. tuberculosis* sigma factors (SigC, SigD, SigE, SigG, SigH, SigI, SigJ, SigK, SigL and SigM). Members of this group of accessory sigma factors, also known as Extra-Cytoplasmic Function (ECF) factors, mostly regulate extracellular functions, and are essential for virulence.

As summarized in **Figure 1.3**, four conserved regions and more sub-regions of the sigma factor protein structure can be identified, that are able to bind to promoter DNA: the N-terminal sub-region 1.2 contacts conserved guanines in the discriminator spacer between the -10 element (TATA box) and the transcription start site (TSS); sub-regions 2.3 and 2.4 recognize the TATA box when part of a double- or single-stranded DNA complex, respectively; sub-region 3.0 is involved in the recognition of conserved TG bases in the extended -10 promoter; and the C-terminal sub-region 4.2 recognizes the -35 element[23,25,26]. Notably, the protein structure of ECF sigma factors contains only regions 2 and 4.

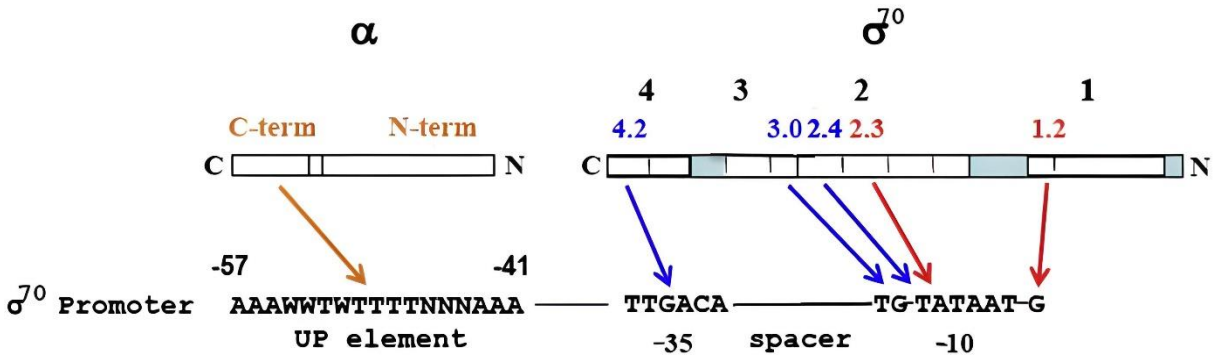

Figure 1.3 Promoter recognition by the two domains of the α subunit and the different sub-regions of the σ^{70} factor. Orange and blue arrows indicate recognition of promoter as double-stranded DNA, red arrows delineate a region of the non-template strand DNA recognized by the σ^{70} factor, following strand separation (from "Saecker *et al*, Journal of Medical Biology, 2011" [26]).

Induction of sigma factors in response to specific stimuli or stresses is crucial and finely regulated, as is the suspension of their activity, when it is no longer necessary. This modulation can occur at the transcriptional level, mediated by other sigma factors, by positive feedback loops, or by two-component systems; at the post-transcriptional level, with the inhibition given by antisense RNA molecules; or at the post-translational level, mediated by anti-sigma factors. Specifically, anti-sigma factors are soluble or membrane-linked proteins that physically restrain their counterparts, only to release them when their activity is required[27].

1.3.1 ECF Sigma Factor C

Alternative ECF Sigma Factor C (SigC) is a peculiar and still not very well characterized sigma factor of *Mycobacterium tuberculosis*. It is encoded by the gene *sigC* (*rv2069*), and its homologues can be found in most pathogenic mycobacteria. It is instead absent in non-pathogenic environmental mycobacteria, such as *M. smegmatis*, often used as a model organism due to safer manipulation and faster growth than its tubercular relative[23,24]. *sigC* was shown to be prevalently expressed during the exponential phase of bacterial growth, and switched off during stationary phase, when most alternative sigma factors are instead induced due to increasing stress conditions. This evidence seemingly indicates the factor involvement in the bacteria normal physiology, associating it to primary factor SigA[24].

Preliminary characterization of sigC upstream regulatory region identified two putative promoters, P1 and P2, that respectively resemble the consensus sequence of SigA–recognized promoters and the putative consensus sequence recognized by SigC itself, which still needs to be thoroughly characterized[28]. Studies conducted on *M. tuberculosis* hyper-virulent strain CDC1551 overall transcriptional profile demonstrated that more than 200 genes were downregulated by a defective *sigC* mutant, compared to the wildtype strain. Of these genes though, only 14-18, analyzed at different time points of the growth curve, were repressed more than two-fold. Among them, important virulence-related genes were found, such as *hspX*, encoding an α-crystallyn homologue, and *senX3*, which codes for a putative two-component sensor histidine kinase. Both genes resulted essential for survival inside macrophages. Other genes involved in the response to the host immune system were found to be down-regulated, such as the lipoprotein-coding gene *mpt83* and *fbpC*, which encodes a mycolyl-transferase. These results were consistent with reported mouse phenotypes, in which animals infected with the mutant strain displayed delayed lethality and reduced inflammatory lung infiltrate,

compared to their wildtype-infected counterparts[29]. In confirmation, experiments conducted on guinea pigs led to the formation of smaller and non-necrotic granulomas in animals infected with *M. tuberculosis* mutants lacking *sigC*, compared to those infected with the wildtype strain. When used to infect THP-1-derived macrophages, though, these mutants showed the same replication rate as the wildtype[30].

1.3.2 SigC role in copper acquisition

Recently published studies showed how artificial induction of *sigC* resulted in the upregulation of a panel of genes, including the *ppe-nrp* operon (*rv0096-rv0101*), and *ctpB*[31,32]. The *nrp* operon had initially been linked to biofilm maturation and was found selectively expressed at the later stages of pellicle formation, following the observation that a deletion mutant was unable to form these structures in comparison to the wildtype strain[33]. It was later determined that the operon encodes a biosynthetic machinery which produces a still uncharacterized diisonitrile lipopepdide (INLP). INLPs supposedly work as chalkophores, copper scavengers, and are possibly anchored to the mycomembrane. This study also showed how loss of the *nrp*, *fadD10* and *ppe1* genes of the operon causes sensitivity of the mutants to low-copper concentrations or copper chelation[34]. As mentioned, CtpB is a copper-specific pump localized at the inner membrane that works similarly to CtpV but is involved in influx. In a recently published paper, the growth defect of a *sigC* deletion mutant was reported after copper chelation. The study showed how single complementation with *ctpB* was not sufficient to recover growth, contrarily to *sigC* reinsertion. This result indicates that the action of CtpB alone is not sufficient to withstand copper influx, and fortifies the hypothesis of SigC-dependent production of INLPs, and their resulting involvement in copper uptake[16].

In a previous study conducted by the same authors, a similar growth defect had been reported for a *sigC*-deleted mutant after 3-4 culture passages in Sauton's minimal medium, which is copper-free. The growth phenotype could be complemented either with reinsertion of *sigC* or with addition of copper sulfate to the cultures, as showed in **Figure 1.4**. A transcriptomic analysis performed under these conditions confirmed overexpression of the *nrp* operon in the wildtype and, interestingly, showed the induction of copper-independent genes to mitigate respiratory distress in the mutant, such as the Alkyl hydroperoxide reductase-encoding

ahpCD as a probable substitute for SodC, and genes involved in hypoxic response or in general to anaerobic persistence (*cydABCD, hrp1*).

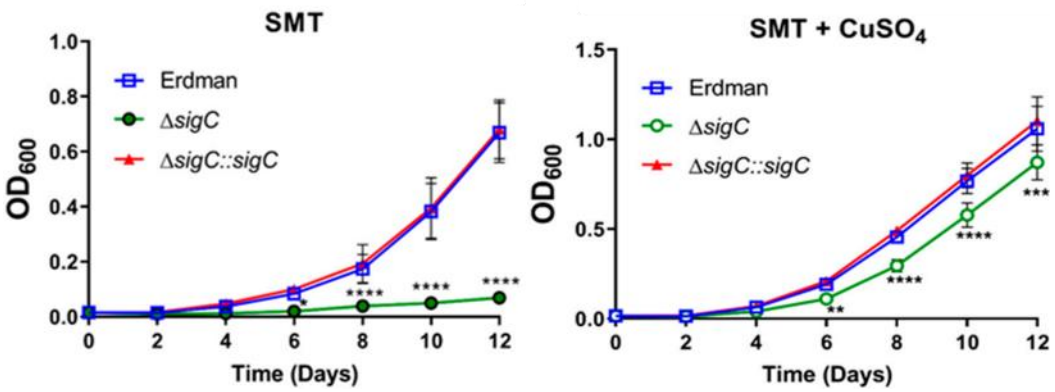

Figure 1.4 Wildtype strain Erdman (blue) and mutant strains Δ*sigC* (green) and Δ*sigC::sigC* (red) growth curves after several passages in Sauton's medium (left) and Sauton's medium + 6 µM $CuSO_4$ (right). Growth of the Δ*sigC* mutant is visibly hindered in copper-free conditions, while the phenotype is recovered by copper sulfate addition (adapted from "Grosse-Siestrup *et al*, Int J Mol sci, 2021" [32]).

Of note, evidence of copper toxicity genes upregulation was also detected in both transcriptomic experiments: extension of *sigC* artificial induction up to 48 hours led to the overexpression of *ricR* and five genes in a RicR-controlled genomic region (*rv0846-rv0850*)[31], and physiological induction of *sigC* after culture in Sauton's medium highlighted expression of the *csoR* operon[32]. This outcome may actually suggest a mechanism in which prolonged SigC activation causes excessive accumulation of copper inside *M. tuberculosis*, and consequently activates the export machinery.

1.3.3 Putative anti-sigma factor Rv0093c

As mentioned, sigma factors activity is controlled at the post-translational level by specific proteins called anti-sigma factors. Usually, sigma and related anti-sigma factors are encoded in the same operon, co-transcribed, and part of a tightly structured regulatory network. Similar conditions have not been found for *M. tuberculosis* SigC, for which the identity of a protein performing the function of anti-sigma has not been described[23,24].

A suitable candidate is represented by Rv0093c, a yet uncharacterized transmembrane protein. The sequence of its structural gene, *rv0093c*, was found to strongly resemble (85% identity) the 5'-terminal region of the *sigC* homologue in other mycobacterial species, such as the non-tubercular pathogens belonging to the *Mycobacterium avium complex* (MAC)[35], as shown in **Figure 1.5**. In these microorganisms, the resulting fusion protein is predicted to be activated by the proteolytic cleavage of the "anti-sigma" N-terminal portion. Furthermore, the protein sequence of Rv0093c is predicted to include a putative zinc-finger domain, zf-HC2, also found in other anti-sigma factors, such as the anti-sigP (YlaC) factor of *Bacillus subtilis*[24,36]. Regardless of this evidence, proof of Rv0093c and SigC protein interaction in *M. tuberculosis* is yet to be obtained experimentally.

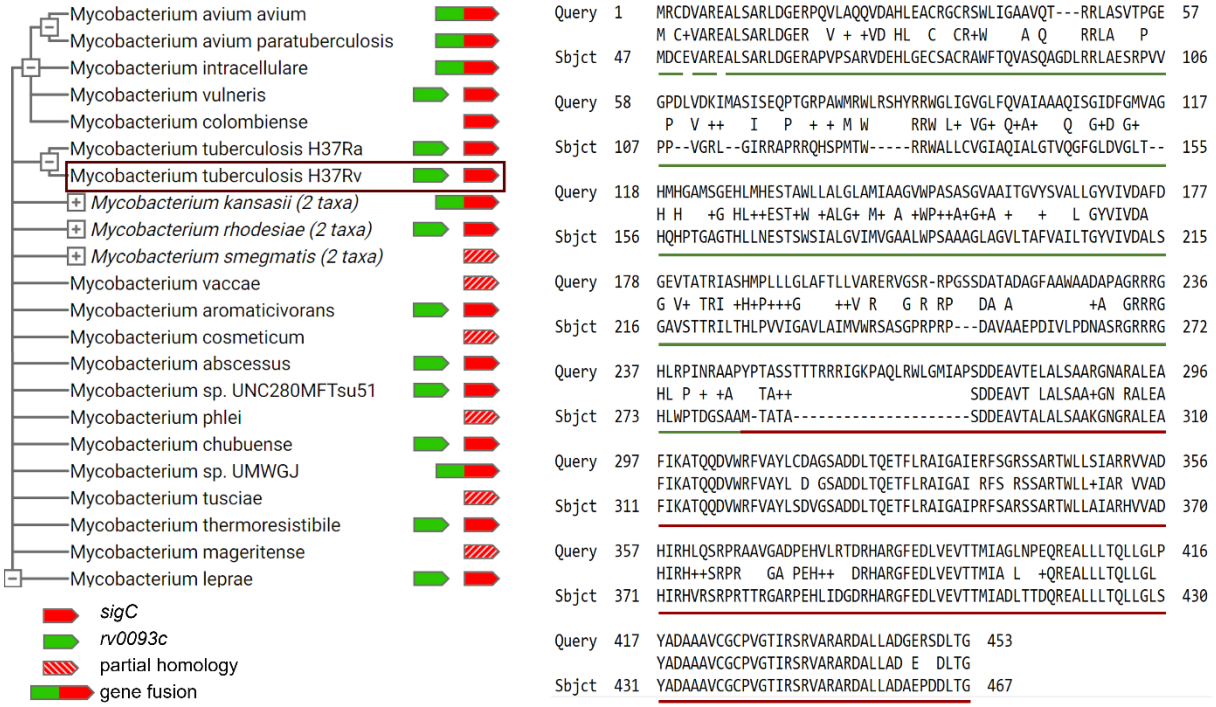

Figure 1.5 On the left, *sigC* (red) and *rv0093c* (green) fusion tree shows the two genes are fused in some mycobacterial species (eg. *Mycobacterium avium*) and are separated in others (eg. *M. tuberculosis* H37Rv). Of note, *rv0093c* is not present in all mycobacterial species analyzed, as well as *sigC*. Dashed arrows represent homology with other sigma factors (obtained on STRING Server[37], from ©STRING Consortium 2021). On the right, alignment of *M. avium* SigC protein sequence (query) with the sequences of *M. tuberculosis* Rv0093c (green) and SigC (red). The initial part of Rv0093c is not shown in the alignment (adapted from "Manganelli *et al, Microbiology spectrum,* 2014" [24]).

1.4 General features of bacterial biofilms

Biofilms are multi-cellular bacterial communities that develop inside a self-deposed and highly complex three-dimensional matrix. They represent a widespread environmental circumstance, where nutrient shortage and hostile conditions stimulate cell aggregation, and the establishment of a close community acquires protective potential[38]. As for the colonization of biological surfaces, biofilms can be harmless/beneficial (eg. dental plaque caused by *Streptococcus mutans* and others) or pathological (eg. colonization of prosthetics and catheters, caused by *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus,* and others)[39].

Biofilms are formed through well-defined sequential steps, which involve an initial loose adhesion of individual cells to a substrate, achieved through the use of adhesive proteins, pili and fimbriae, which later becomes firmer thanks to the early deposition of a slimy matrix. This phase is followed by a more intense and coordinated production of matrix and formation of a three-dimensional architecture (**Figure 1.6**). As the structure expands, conquering more and more space, its interior is modelled to form individual colonies separated by water micro-channels, for nutrients transport[40,41]. The metabolic activity of the cells embedded in the biofilm remains lower than during free growth, and dormant, non-proliferative cells are a common occurrence, similarly to the stationary phase of the growth curve obtained experimentally[38]. Furthermore, the quenching activity of the matrix helps lowering antimicrobial substances to sub-inhibitory concentrations, contributing to the development of resistant bacterial forms[42,43].

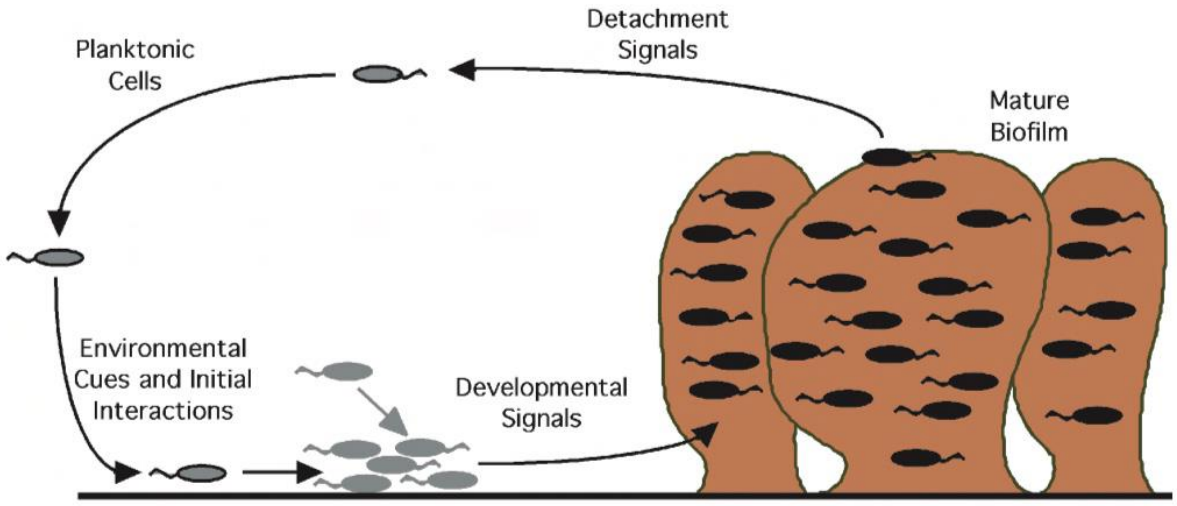

Figure 1.6 Schematic representation of the biofilm formation process. Following initial adhesion to the substrate, bacterial aggregation and matrix deposition eventually lead to the production of a complex three-dimensional structure (from O'Toole *et al, Annual review of Microbiology*, 2000" [41]).

1.4.1 Mycobacterial biofilms

Evidence of mycobacterial biofilms has been reported in the literature, describing the formation of pellicles and more complex structures at the air-liquid interface. The large quantities of mycolic acids and glycolipids released by the bacteria contribute to the impermeable nature of their membrane in forming the waxy, hydrophobic matrix that encapsulates mycobacterial cells and allows the pellicle development[44]. Several opportunistic mycobacterial pathologies were shown to be sustained by biofilm formation (eg. *M. avium* airway colonization, *M. abscessus* participation to cystic fibrosis, *M. fortuitum, M. chelonae* and *M. chimaera* post-surgery infections). In nosocomial environments, these bacterial species can spread and colonize prosthetics and catheters biomaterials or be transferred to the patients with use of infected surgical instruments. In all described cases, treatment of the infection results extraordinarily long and requires multidrug regimens[45].

Growth of *Mycobacterium tuberculosis* biofilm has been explored mostly in relation to antibiotic resistance. It was demonstrated that the frequencies of drug-tolerant bacteria in biofilms are higher than those in planktonic cultures, and that this difference may be related to the architectural structure of the pellicles. Cells within the biofilm are in fact subjected to prolonged exposure to sub-inhibitory concentrations of antibiotics, a mechanism that is known to foster the enrichment of resistants[46]. As mentioned, a recently published study identified a major role for biofilm maturation in isonitrile lipopeptides (INLPs), which were found to be selectively produced during biofilm advanced maturation (fourth week onwards), and never in planktonic cultures or on plated colonies. Expression of the *inlp* gene cluster was later linked to SigC[32].

2. AIM OF THE WORK

SigC is an alternative ECF sigma factor of *Mycobacterium tuberculosis*, whose function and regulation are not accurately defined. Even though some hypotheses on its activation context and related genes have been proposed, a consensus of its recognition sequence on mycobacterial promoters has not been identified, and thus its regulon and role remain uncharacterized. Moreover, the identity of its specific anti-sigma factor is still unknown, although the protein Rv0093c represents a good candidate, mainly due to sequence homology with SigC of other mycobacterial species.

Through the generation of mutants lacking *rv0093c* and *sigC*, this work mainly aims at exploring the connection between the two proteins, in order to define the role of the sigma factor in the complex physiology of *M. tuberculosis*, and confidently recognize its anti-sigma factor.

First and foremost, the hypothesis that *rv0093c* deletion from the bacterium's genome would work as an overexpression of *sigC* will be explored. Having been connected to copper uptake, SigC protracted action is expected to cause an overload of copper inside bacteria, initially leading to the activation of copper-efflux mechanisms and then, inevitably, to death. Thus, growth impairment and expression of copper-toxicity response genes will be evaluated after treatment with copper sulfate. Additionally, an RNA sequencing analysis will be performed in presence and absence of the putative anti-sigma factor, to further existent knowledge on SigC still undefined regulon.

Also, with a view to validating Rv0093c role as an anti-sigma, a mycobacterial protein-protein interaction assay based on a variation of the two-hybrid system will be performed on the non-pathogenic model species *Mycobacterium smegmatis*.

In conclusion, recent analyses on biofilm formation have highlighted that the *sigC*-deleted mutant is not impaired in pellicle maturation. Nevertheless, progressive dehydration of the cultures allows a mutant-specific color variation to arise, perhaps due to a different metabolic activity. This newfound phenotypic peculiarity, as well as its connection with the lack of the sigma factor, will be investigated further.

3. MATERIALS AND METHODS

3.1 Bacterial strains, media, and growth conditions

Planktonic cultures of *Mycobacterium tuberculosis* H37Rv, *Mycobacterium smegmatis* mc^2155 and derived strains were grown at 37 °C in 7H9 Broth (Difco™ Middlebrook, BD) supplemented with 0.2% glycerol (Sigma-Aldrich®) and 0.05% Tween® 80 (Sigma-Aldrich®), or in modified Sauton's medium[32], in either standing, rolling or orbital shaking conditions. When necessary, low-copper conditions were obtained through subsequent culture refreshings in copper-free Sauton's medium, while high-copper conditions were obtained with addition of copper sulfate to the final concentrations of 384 µM and 512 µM. For plate cultures, both 7H10 and 7H11 agarized media (Difco™ Middlebrook, BD) were used, supplemented with 0.5% glycerol and 0.05% Tween® 80. Tubercular strains growth in rich media (7H9 and 7H10) was enhanced by addition of 10% ADN growth supplement (5% BSA, 2% D-glucose, 0.85% NaCl, Sigma-Aldrich®). When necessary, Kanamycin (Km, Sigma-Aldrich®), Hygromycin B (Hyg, Invitrogen™), 5-Bromo-4-Cloro-Indolyl β-D-Galactopyranoside (X-Gal, Sigma-Aldrich®) and Trimethoprim (Trim, Sigma-Aldrich®) were added at the final concentrations of 20, 50, 50 and 25 µg/mL respectively. For gene induction purposes, Anhydrotetracycline (ATc, Sigma®) was added to liquid medium for a final concentration of 0.5 µg/mL. 7H10 agar was supplemented with 4-10% Sucrose (Sigma-Aldrich®) as necessary for selection processes. *M. tuberculosis* and derived strains were exclusively manipulated inside the Biosafety Level 3 (BSL-3) Laboratory.

For cloning purposes, *Escherichia coli* DH5α cells were grown at 37 °C in LB Broth (Sigma®) in shaking or on LB Agar plates. Antibiotics were added as follows: 50 µg/mL of Kanamycin and 150 µg/mL of Hygromycin B. As necessary, X-Gal was added at the final concentration of 40 µg/mL.

All strains used in this work are listed in **Table 3.1**:

Bacterial strain	Description	Reference
H37Rv	*M. tuberculosis* wildtype strain	Laboratory collection
TB543	H37Rv Δ*sigC*	This work
TB544	H37Rv Δ*sigC::sigC*	This work
TB545	H37Rv + pKM461	This work
TB547	TB543 + pKM461	This work
TB548	H37Rv Δ*rv0093c*	This work
TB549	H37Rv Δ*rv0093c::rv0093c*	This work
TB550	H37Rv Δ*sigC* Δ*rv0093c*	This work
TB551	H37Rv Δ*sigC::sigC* Δ*rv0093c::rv0093c*	This work
TB555	H37Rv Δ*sigC::sigC* Δ*rv0093c*	This work
TB556	H37Rv Δ*sigC* Δ*rv0093c::rv0093c*	This work
mc^2155	*M. smegmatis* wildtype strain	Laboratory collection
Ms166	mc^2155 $GCN4_{[DHFR\ F1,2]}/GCN4_{[DHFR\ F3]}$	Laboratory collection
Ms169	mc^2155 $hsp60_{[DHFR\ F1,2]}/hsp60_{[DHFR\ F3]}$	Laboratory collection
Ms327	mc^2155 $SigC_{[DHFR\ F1,2]}/Rv0093c_{[DHFR\ F3]}$	This work
Ms328	mc^2155 $_{[DHFR\ F1,2]}SigC/_{[DHFR\ F3]}Rv0093c$	This work
Ms329	mc^2155 $SigC_{[DHFR\ F1,2]}/_{[DHFR\ F3]}Rv0093c$	This work
Ms330	mc^2155 $_{[DHFR\ F1,2]}SigC/Rv0093c_{[DHFR\ F3]}$	This work
EC DH5α	*E. coli* cloning strain	Laboratory collection

Table 3.1 Bacterial strains used or generated throughout this work.

3.2 Biofilm growth and analyses

Bacteria were pre-grown in complete Sauton's medium, in rolling conditions, until late exponential phase (OD_{540} from 0.8 to 1.2). The cultures were then diluted to OD_{540} = 0.01 in detergent-free Sauton's medium and dispensed into 12-well plates (ThermoFisher™), which were then firmly wrapped with Parafilm (Bemis®) and incubated at 37 °C in standing for up to 5 weeks (*M. tuberculosis*). Biofilm formation was documented weekly. Experiments were performed in parallel in controlled (5% CO_2, H_2O saturation) and uncontrolled atmosphere. For viability testing, suspended pellicles and submerged materials were collected with centrifugation, and serial dilutions of the pellets were plated. Before subsequent analyses, the samples incubated in uncontrolled atmosphere were resuspended in 1 mL distilled water, filtered on 0.22 µm pore-sized membrane filters (Millipore™) and sterility tested by plating.

Following analyses performed on recovered samples were carried out in collaboration with Prof. A. Roveri, Prof. G. Miotto, Prof L. Zennaro and Dr. M. Rossetto (Dept. Molecular Medicine, University of Padova). Absorption spectra were acquired on a 280-700 nm wavelength range, with a Cary UV-Vis Multicell peltier spectrophotometer (Agilent). The samples where then concentrated on a DiaFlo (Amicon) equipped with a YM3000 ultrafiltration membrane. Chromatographic analyses were performed on a 1260 Infinity II UHPLC system (Agilent Technologies) with a 166 UV-VIS detector (Beckman Coulter©) at a set wavelength of 275 nm. Whole samples were diluted 1:5 in H_2O and resolved on an Ultrasphere HPLC column (ODS, 80 Å, 5 µm, 10 x 250 mm, Beckman Coulter©) at 30°C, using H_2O and acetonitrile as elution solvents, with a 3 mL/min flow and linear gradient from 5 to 40% acetonitrile in 7.5 min, raised to 95% in 4 min and kept for 1 min, and a following column re-equilibration of 14 min at 5% of acetonitrile. For fraction collection, about 200 µL of the initial samples were processed in 4 subsequent 50 µL-injections. In preparation for mass spectrometry evaluation, the fractions were dried and then reconstituted with 30 µL of 5 mM ammonium formate with 0.1 mM formic acid in H_2O and MeOH 60:40 (v/v).

For the acquisition of NMR spectra, 370 µl of ultra-filtered samples were mixed with 55 µl of 40 mM TSP (3-(Trimethylsilyl)propionic-2,2,3,3-d 4 acid sodium salt) in D_2O, buffered with 125 µl of 1.5 M potassium phosphate at pH 7.4, and transferred into an NMR 535-PP tube (Wilmad®). All ^{1}H-NMR spectra were acquired using a 300 MHz Bruker Avance III NMR spectrometer (Bruker© Biospin, Rheinstetten, Germany) equipped with a 5 mm Z-gradient probe at the constant temperature of 300 ± 0.1 K. For the analysis of each sample, 1D-Noesy spin echo sequence was used applying the following acquisition parameters: 512 scans,

64 K data points, 20 ppm spectral width, and 4'' as relaxation delay. A 0.3 Hz line broadening was applied to all the spectra before performing the Fourier Transform. Spectra acquisition and processing were performed using Topspin 3.6.1 software (Bruker©).

For the evaluation of minerals and heavy metals content by Inductively Coupled Plasma – Optical Emission Spectroscopy (ICP-OES) 0.8 mL of ultra-filtered samples were diluted with 2.75 mL of 5% HNO_3 prepared in distilled water. An ICP-OES 5110 (Agilent) spectrometer, operating with Argon plasma, was used to investigate the content of Al, As, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, Se, Sr, Zn. For the calibration curves, certified single-element standards (Agilent) were used to prepare six standard solutions at different concentrations (0, 0.050, 0.200, 0.400, 1.000, 2.000 and 4.000 ppm for Na; 0, 0.010, 0.050, 0.100, 0.500, 1.000 and 2.000 ppm for Ca and Mg; 0, 0.050, 0.100, 0.400, 1.000, 2.000, 5.000 and 15.000 ppm for K; 0, 0.005, 0.010, 0.040, 0.100, 0.200, 0.500 and 1.500 ppm for the other elements). To quantify the elements at higher concentrations (Ca, K, Mg, Na, P) the samples were further diluted 1: 250 (v/v) with a 5% aqueous solution of HNO_3. The operating parameters were RF power 1.30 (kW), plasma gas flow rate 12.0 (L min-1), auxiliary gas flow 1.0 (L min-1), nebulizer flow 0.70 (L min-1), nebulizer pump speed 12 (rpm), read time 30(s), reading replicates 3.

3.3 Plasmids and mutant strains construction

Throughout this work, the laboratory strains *Mycobacterium tuberculosis* H37Rv and *M. smegmatis* mc^2155 were used as parental wildtype strains for genetic manipulations. Deletion of target gene *sigC*, to obtain TB543 strain, was performed using a two-step homologous recombination approach, as described[47]. Deletion of *rv0093c*, to obtain TB548 and TB543-derived TB550, was performed using a recently developed tool for genetic engineering of mycobacterial chromosomes, called ORBIT (*Oligonucleotide-mediated Recombineering followed by Bxb1 Integrase Targeting*)[48].

Plasmids and primers used for mutant strains construction are listed in **Tables 3.2** and **3.3:**

Plasmid name	Description	Reference
pTOPO	pCR®-Blunt II-TOPO®, *KmR*	Invitrogen™
pTOPO_U	*sigC*_upstream in pTOPO	This work
pTOPO_D	*sigC*_downstream in pTOPO	This work
pTOPO_U+D	*sigC*_upstream + *sigC*_downstream in pTOPO	This work
pTOPO_sigC	P_{sigC}-*sigC* in pTOPO	This work
p1NIL	Gene manipulation vector	[47]
p1NIL_U+D	*sigC*_upstream + *sigC*_downstream in p1NIL	This work
pGOAL19	*Hyg* P_{ag85}-*lacZ* P_{hsp60}-*sacB* PacI cassette vector	[47]
pGRE1	*sigC*_upstream + *sigC*_downstream + pGOAL PacI cassette in p1NIL	This work
pKM461	P_{Tet}-*Che9c RecT-Bxb1-Int*, *SacRB*, *TerR, KmR*	[48]
pKM464	Bxb1 attB, *HygR*	[48]
pMV306	*attP*, *Int, KanR*	[49]
pGRE2	P_{sigC}-*sigC* in pMV306	This work
pGRE4	$P_{rv0093c}$-*rv0093c* in pMV306	This work
pGRE10	P_{sigC}-*sigC* + $P_{rv0093c}$-*rv0093c* in pMV306	This work
pUAB100	P_{hsp60}-*gcn4-mDHFR*$_{[F1,2]}$	[50]
pUAB200	P_{hsp60}-*gcn4-mDHFR*$_{[F3]}$	[50]
pUAB300	P_{hsp60}-*mDHFR*$_{[F1,2]}$	[50]
pUAB400	P_{hsp60}-*mDHFR*$_{[F3]}$	[50]
pGRE6	*sigC-mDHFR*$_{[F1,2]}$ in pUAB100	This work
pGRE7	*rv0093c-mDHFR*$_{[F3]}$ in pUAB200	This work
pGRE8	*mDHFR*$_{[F1,2]}$-*sigC* in pUAB300	This work
pGRE9	*mDHFR*$_{[F3]}$-*rv0093c* in pUAB400	This work

Table 3.2 Plasmids used and generated throughout this work.

Primer name	Sequence	Description
RP2075	5' – **aagctt**gatcggacatgacggccacca – 3'	Fw *sigC*_upstream
RP2076	5' – **aggcct**cgtcgcggtcatgggcataaa – 3'	Rv *sigC*_upstream
RP2077	5' – **aggcct**gacctcaccggctaggcagac – 3'	Fw *sigC*_downstream
RP2078	5' – **ctcgag**ctgtccacacccataccgccc – 3'	Rv *sigC*_downstream
RP2079	5' – **gatatc**tcgattctgtccaccgggcc – 3'	Fw P_{sigC}-*sigC*
RP2080	5' – gagacaccgcaacgcgtgcgcatc – 3'	Rv P_{sigC}-*sigC*
RP2086	5' – gatatctcaccaccaccctgaccgac – 3'	Fw $P_{rv0093c}$-*rv0093c*
RP2087	5' – gatatccgacgttgtacccttgcgc – 3'	Rv $P_{rv0093c}$-*rv0093c*
RP2108	5' – gagacaccgcaacgcgtgcgcatc – 3'	Fw external to RP2075
RP2109	5' – cgcaacattgggttcaatgccgcgg – 3'	Rv external to RP2078
RP2149	5' – cgagccccacggcgctcacgcaccacccgggccaccgtgctcgcac aagccacaacggccgggagc**ggtttgtctggtcaaccaccgcggtctcag tggtgtacggtacaaaacc**gcatcacgcggacgacggcgcggtcatttgtgg cccaccgacggttccgcagcctagcgggtatcgccc – 3'	Target specific oligo for *rv0093c*
RP2159	5' – **caattg**gtgctcgcacaagccaca – 3'	Fw *rv0093c*-MfeI
RP2160	5' – **atcgat**ctaggctgcggaaccgtc – 3'	Rv *rv0093c*-ClaI
RP2165	5' – cctggtatctttatagtcctgtcg – 3'	Specific primer for 5' junction[48]
RP2166	5' – tgcacgggaccaacaccttcgtgg – 3'	Specific primer for 3' junction[48]
RP2194	5' – gccaagacaattgcgatgaccgcgacggcaagc – 3'	Fw *sigC* in pUAB100 for NEBuilder
RP2195	5' – gccaccgccaccatcctagccggtgaggtcgtc – 3'	Rv *sigC* in pUAB100 for NEBuilder
RP2196	5' – ggaggtggtgggtccggaatgaccgcgacggcaagc – 3'	Fw *sigC* in pUAB300 for NEBuilder
RP2197	5' – agtcgatcgtacgctagtctagccggtgaggtcgtc – 3'	Rv *sigC* in pUAB300 for NEBuilder

Table 3.3 Primers used throughout this work for mutant strains construction. Bold stretches represent recognition sites for restriction enzymes HindIII (aagctt), StuI (aggcct), XhoI (ctcgag), EcoRV (gatatc), MfeI (caattg) and ClaI (atcgat), and the attP site sequence in RP2149.

3.3.1 Construction of *M. tuberculosis* mutant TB543 and derived strains

Approximately 1000 bp upstream and downstream of *sigC* were amplified through PCR, using primers specifically designed to contain different restriction enzymes target sequences, as shown in **Table 3.3**. *sigC* upstream (U) and downstream (D) regions were designed in order to contain the very first and very last codons of the gene, and to maintain the reading frame. The obtained amplicons were cloned inside the commercial vector pTOPO (Zero Blunt® TOPO® PCR Cloning Kit, Invitrogen™), obtaining plasmids pTOPO_U and pTOPO_D. The downstream fragment was then extracted from pTOPO_D using restriction enzymes StuI and XhoI (NEB®) and inserted right after the upper fragment sequence in pTOPO_U, previously linearized with StuI, and resulting in plasmid pTOPO_U+D. The newly formed U+D fragment was later extracted from pTOPO_U+D, using enzymes HindIII (NEB®) and XhoI, and inserted into the vector p1NIL[47], previously linearized with ScaI (NEB®). The resulting plasmid, p1NIL_U+D, was then completed with the insertion of the PacI-cassette extracted from plasmid pGOAL19[47], containing the selection markers hyg^R, *sacB* and *lacZ*. The final plasmid, named pGRE1, was eventually transformed into electro-competent *M. tuberculosis* H37Rv cells: the wildtype strain was grown up to OD_{540} = 0.4, washed twice with a 10% glycerol solution and resuspended in 1 mL of 10% glycerol solution. 50 µL aliquots of fresh competent cells were electroporated with 1 µg of pGRE1. After a 24 hrs recovery period at 37 °C in 800 µL 7H9 ADN, cells were plated on 7H10 ADN Km/Hyg/X-Gal plates and incubated at 37 °C for 3-4 weeks. Blu colonies appearing after this time period represented cells in which pGRE1 had successfully integrated into *M. tuberculosis* genome through homologous recombination (first cross-over event). These colonies were PCR-tested with primer couples RP2108/RP2080 and RP2079/RP2109, to verify the correct insertion of the plasmid, and to investigate its positioning in relation to the wildtype copy of *sigC*. This last aspect depends exclusively on which homology region is exploited by the DNA polymerase of the bacterium to "transfer" to the plasmid, before "returning" to the original genomic template, during genomic replication. Both resolutions contain the entire sequence of pGRE1.

In order to stimulate the expulsion of pGRE1 through a second cross-over event, validated colonies were initially coated on drug-free 7H10 ADN plates. Dilutions of the resulting patinas were plated on 7H10 ADN X-Gal plates supplemented with 4% Sucrose. The *sacB* gene contained within pGRE1 generates a lethal sucrose metabolite, thus forcing mutant cells to expel the plasmid to avoid death. *lacZ* represents an additional screening method, since the blue phenotype is maintained only by those cells which do not manage to entirely expel

pGRE1 backbone. Non-blue colonies were further tested by subsequent streaking on Km, Hyg and drug-free 7H10 ADN plates. Only colonies surfacing on drug-free conditions, but not on Kanamycin- and Hygromycin-supplemented plates, were PCR-tested. The previously designed primers RP2108 and RP2109 were used to investigate whether the second event of homologous recombination eventually resulted in the desired genetic deletion (3879 bp), of if it restored the wildtype condition (4437 bp). Genetic deletion was ultimately confirmed via Sanger sequencing (BMR Genomics), and the resulting unmarked mutant strain was called TB543 (*M. tuberculosis* H37Rv Δ*sigC*).

For complementation, the genomic region containing *sigC* and its putative promoter was amplified via PCR, using EcoRV-tailed primers RP2079 and RP2080 as shown in **Table 3.3**. The amplicon was cloned into the commercial vector pTOPO, amplified, and ultimately re-extracted using EcoRV (NEB®) restriction enzyme. The fragment was then cloned into the integrative plasmid pMV306[49], to obtain pGRE2. Electro-competent cells were prepared as described above, starting from the newly obtained mutant TB543, and electroporated with 1 µg of plasmid. Following a 24 hrs recovery period, cells were plated on 7H10 ADN Km plates and incubated at 37 °C for at least 3 weeks. Transformation of TB543 with pGRE2 led eventually to the generation of the mutant strain TB544 (*M. tuberculosis* H37Rv Δ*sigC::sigC*). The effective integration of the plasmid, and presence of the inserted gene, was confirmed via PCR.

3.3.2 Construction of *M. tuberculosis* mutant TB548 and derived strains

The target specific oligonucleotide RP2149 was designed as follows: considering the lagging strand of the genome, roughly 60 bases were selected across the start and stop codons of *rv0093c*, and the Bxb1 attP site sequence was inserted in-between, as indicated[48]. The wildtype strain of *Mycobacterium tuberculosis* H37Rv, and mutant strain TB543, lacking the *sigC* gene, were initially transformed with the plasmid pKM461, to obtain the acceptor strains TB545 and TB547. Both strains were grown up to OD_{540} = 0.8, treated with ATc for 8 h to induce expression of Che9c RecT recombinase and Bxb1 integrase from a TetR inducible promoter, and then with 2 M glycine for 16 h. Afterwards, the cultures were pelleted at 4000 rpm for 10', washed twice in a 10% glycerol solution and finally resuspended in 2 mL of 10% glycerol solution. 380 µL aliquots of fresh competent cells were co-electroporated with 1 µg of RP2149 and 200 ng of payload plasmid pKM464. After a 24 hrs recovery period at 37 °C in 2 mL 7H9 ADN, cells were plated on 7H10 ADN Hyg plates and incubated at 37 °C for 3 weeks.

According to protocol, during the first replication cycle after transformation, RP2149 would pair to *rv0093c* genomic region due to sequence homology, and Bxb1 would trigger integration of the pKM464 plasmid into RP2149 sequence, after Che9c RecT-driven recombination between the attP and attB sites. Successful substitution of the original genomic sequence with the RP2149-pKM464 combination would then result in the generation of mutant daughter cells. Recombinant colonies resulting from both transformations were processed for rapid DNA extraction: genome-plasmid junctions were PCR tested, using primers that anneal externally to the oligonucleotide arms, and primers specifically designed on pKM464, as illustrated in **Table 3.3**. As a final control, the entire genomic region was Sanger-sequenced to verify replacement of the target gene by correct insertion of pKM464. Serial dilutions of the resulting deletion mutants were then plated on 7H10 ADN plates, supplemented with 10% Sucrose to drive expulsion of pKM461. After this treatment, surfacing cells were tested on 7H10 ADN Km plates, and one sensitive clone for each transformation was selected. The final mutant strains were named TB548 (*M. tuberculosis* H37Rv Δ*rv0093c*) and TB550 (*M. tuberculosis* H37Rv Δ*sigC* Δ*rv0093c*).

For complementation, the genomic region containing *rv0093c* and its putative promoter was amplified via PCR, using EcoRV-tailed primers RP2086 and RP2087, as shown in **Table 3.3**. The amplicon was cloned into the commercial vector pTOPO, amplified, and ultimately re-extracted using EcoRV restriction enzyme. The fragment was then cloned into the integrative plasmids pMV306[49] and pGRE2, to obtain pGRE4 and pGRE10 respectively. Electro-competed cells of the mutants TB548 and TB550 were prepared as described, and electroporated with 1 µg of plasmid, as follows: TB548 was transformed with pGRE4 only, TB550 was transformed with pGRE2, pGRE4 and pGRE10. Following the 24 hrs recovery period, cells were plated on 7H10 ADN Km/Hyg plates and incubated at 37 °C for at least 3 weeks. Transformation of TB548 with pGRE4 led eventually to the generation of TB549 (*M. tuberculosis* H37Rv Δ*rv0093::rv0093c*). Transformation of TB550 with plasmids pGRE2, pGRE4 and pGRE10 led to the generation of mutants TB555, TB556 and TB551, respectively. The effective integration of the plasmids, and presence of the inserted genes, were later confirmed via PCR.

3.3.3 Construction of *M. smegmatis* mutant strains and M-PFC assay

The gene *sigC* was directly cloned inside plasmids pUAB100 and pUAB300[50] using the NEBuilder® HiFi DNA Assembly Cloning Kit (NEB®). Both plasmids were initially restricted with the enzymes BamHI and ClaI. According to the manufacturer instructions, the forward primers RP2194 and RP2196 were designed to anneal with their 5' half on the BamHI end of the plasmids, and with their 3' half on the beginning of *sigC*. Accordingly, the reverse primers RP2195 and RP2197 were designed to anneal with their 5' half on the end of *sigC* and their 3' half on the ClaI end on the plasmids. A PCR reaction was carried out to amplify *sigC* with the described primers, and a second enzymatic reaction allowed to fuse the obtained amplicon with the two linearized vectors. The final plasmids were named pGRE6 and pGRE8. Following a distinct cloning strategy, the gene *rv0093c* was PCR amplified with the MfeI-tailed forward primer RP2159 and ClaI-tailed reverse primer RP2160, and cloned in the commercial plasmid pTOPO. The fragments were then excised from pTOPO and cloned inside the plasmids pUAB200 and pUAB400, previously restricted with the same enzymes. The final plasmids were named pGRE7 and pGRE9. Plasmid couples pGRE6+pGRE7, pGRE8+pGRE9, pGRE6+pGRE9 and pGRE7+pGRE8 were co-electroporated into fresh *M. smegmatis* competent cells, prepared as described above, considering 1 µg of DNA for integrative plasmids (pGRE6 and pGRE8) and 0.5 µg for episomes (pGRE7 and pGRE9). After 3 hrs of incubation in 800 µL of 7H9 medium, in orbital shaking at 37 °C, the samples were plated on 7H11 Km/Hyg plates and incubated for 3-4 days at 37 °C. Recombinant colonies were tested by streaking onto new 7H11 Km, 7H11 Hyg and 7H11 Km/Hyg plates, and then one clone for each transformation was selected and stored at -80°C. The mutant strains thus generated were named Ms327, Ms328, Ms329 and Ms330.

In order to carry out the mycobacterial protein fragment complementation assay (M-PFC[50]), the *M. smegmatis* mutant strains described above, and the control strains Ms166 and Ms169, were streaked on 7H11 Trim plates. After 3-4 days of incubation at 37 °C, the resistant colonies were re-streaked on 7H11 Km, 7H11 Hyg, 7H11 Km/Hyg and 7H11 Trim plates to confirm the observed phenotype.

3.4 Determination of copper susceptibility with a Resazurin microtiter assay

Bacterial susceptibility to copper sulfate ($CuSO_4$, Sigma-Aldrich®) was measured using a well-established Resazurin microtiter assay (REMA[51]), generally employed for drug MIC determination. Briefly, bacteria were cultured in rolling at 37 °C up to the exponential growth phase (OD_{540} = 0.4) and subsequently resuspended in detergent-free Sauton's medium at OD_{540} = 0.01 (10^6 CFU/mL). Serial dilutions of the copper salt were performed in black microtiter plates (Thermofisher Scientific™ Nunc™) in the same medium, for a final volume of 100 µL/well. Specifically, susceptibility was tested for concentrations ranging from 12 µM to 3 mM $CuSO_4$. 100 µL of the previously prepared bacterial suspension were then added to each well, thus obtaining a two-fold dilution of both bacteria and copper salt. Growth controls, represented by the appropriately diluted bacterial cultures, were added for each technical replicate, and negative controls were added for each tested concentration. The plates were then incubated for seven days at 37 °C in a water-saturated incubator. At the end of the incubation, 20 µL of the resazurin-containing reagent alamarBlue (alamarBlue™ Cell Viability Reagent, Invitrogen™) were added to each well. Fluorescence of the samples was measured after 24 hrs on a high speed GloMax® Explorer Detection System, with the software GloMax® Discover (v. 3.2.2, Promega), setting the excitation wavelength at 520 nm and the emission on a range between 580 and 640 nm.

3.5 RNA analyses

3.5.1 RNA extraction for RNA sequencing

RNA extraction for RNA sequencing was performed starting from 35 mL volume cultures grown approximately to OD_{540} = 0.4. The samples were pelleted at 3000 rpm for 5', resuspended in 1 mL TRIzol® Reagent (Ambion®) and transferred to 2-mL screw cap tubes containing 500 µL Zirconia/Silica beads (0.1 mm diam, BioSpec Products). Cells were mechanically disrupted with three rounds of 30-seconds pulses in a Mini Bead Beater (BioSpec Products), alternating with 30-seconds incubations on ice. The samples were incubated 5' at RT and periodically inverted, and finally centrifuged at maximum speed for 1'. The lysates were transferred to new centrifuge tubes over a heavy phase lock gel (Eppendorf©), and thoroughly mixed with 300 µL of chloroform-isoamyl alcohol mixture 24:1 (Sigma-Aldrich®)

for 2'. Following a 10' centrifugation at maximum speed, the aqueous phase was recovered, mixed with 0.5 volumes of isopropanol (Sigma-Aldrich®) and 0.5 volumes of high salt solution (0.8 M sodium citrate, 1.2 M sodium chloride) and precipitated overnight at 4 °C. The day after, RNA samples were pelleted for 15' at 4 °C, washed in 75% ethanol, air-dried, and resuspended in 90 µL RNase-free water. Finally, they were treated 1 hour with DNaseI (Ambion™) and purified with the QIAgen RNeasy Mini Kit before storing at -80 °C.

RNA sequencing was executed at Cribi NGS facility (Complesso Interdipartimentale "A. Vallisneri", Padova). Briefly, RNA was quantified using a Qubit fluorimeter (Thermofisher™) and quality evaluation was performed on an automated electrophoresis system TapeStation (Agilent). 2,5 µg RNA samples were processed for subsequent ribodepletion, library generation and 100PE sequencing on an Illumina Novaseq platform. Transcriptomic analysis was performed in collaboration with Prof. S. Campanaro, Dr. L. Treu and Dr. V. Biruel, from the Dpt. of Biology (Unipd). Exploratory data clustering was performed with Cluster and PCA analysis, and final log2FC data was obtained with DESeq2 (Bioconductor).

3.5.2 RNA extraction and retro-transcription for quantitative RT-PCR

RNA extraction from small volume tubercular samples was performed following a three-day long protocol, which includes two phase-separation steps. Starting from log phase coltures (OD_{540} = 0.4-0.5), 2 mL samples were collected, pelleted at maximum speed for 5', resuspended in 1 mL TRIzol® Reagent and transferred to 2-mL screw cap tubes containing 800 µL Zirconia/Silica beads. Cell disruption was performed mechanically with three rounds of 45-seconds pulses in a Mini Bead Beater, alternating with 30-seconds incubations on ice. In order to obtain the aqueous-organic phase separation, 100 µL of chloroform-isoamyl alcohol mixture 24:1 were thoroughly mixed into the samples. Following a 30' centrifugation at maximum speed, the aqueous phase was recovered, and mixed with 500 µL isopropanol and 2 µL glycogen (Invitrogen™) for an overnight precipitation at -20 °C. Everything but the cell disruption step was repeated during the second day. After the second overnight precipitation, RNA pellets were washed twice in 75% ethanol, air-dried, resuspended in 20 µL RNase-free water and conserved at -80 °C. 500 ng of the RNA samples were then retro-transcribed to first strand cDNA with M-MLV Reverse Transcriptase (Invitrogen™) following manufacturer instructions, and conserved at -20 °C.

Quantitative Real-Time PCR (RT-qPCR) was performed on a 7000 Sequence Detection System (Applied Biosystems) using PowerUP™ SYBR™ Green Master Mix (Applied Biosystems). Non-retrotranscribed samples were added to the analysis in order to quantify genomic contamination, and CT raw data were corrected with standardization curves specifically calculated for all primer couples used throughout the work (**Table 3.4**). Sample mRNA copy numbers were obtained after normalization on the housekeeping gene *sigA,* and fold-change levels were calculated as copy number ratios between different strains or tested conditions.

All primers used for RT-qPCR analyses are listed in **Table 3.4**:

Primer name	Sequence	Description
RP1401	5' – caggggtcgatcccaagaat – 3'	*rv0097* fw
RP1402	5' – gatctccgggtggtcttcgt – 3'	*rv0097* rv
RP1516	5' – ccatcccgaaaaggaagacc – 3'	*sigA* fw
RP1517	5' – aggtctggttcagcgtcgag – 3'	*sigA* rv
RP2128	5' – cacatccgccacgtccgatc – 3'	*sigC* fw
RP2129	5' – cggtggttaggtcggcgatc – 3'	*sigC* rev
RP2161	5' – caccgatgacctggcgacgt – 3'	*rv0093c* fw
RP2162	5' – ccgcgcccaccatgatcaca – 3'	*rv0093c* rev
RP2182	5' – gtttcaatcagctcgcctac – 3'	*fcoT* fw
RP2183	5' – cagatctcgacacaggaggc – 3'	*fcoT* rv
RP2184	5' – acggttcggggccatcttga – 3'	*csoR* fw
RP2185	5' – ccgcttgcccatgaccatcc – 3'	*csoR* rv
RP2218	5' – acgtccgtaggtagcgaatt – 3'	*lpqS* fw
RP2219	5' – gtaagcgcggctgtcat – 3'	*lpqS* rv

Table 3.4 Primers used throughout this work for RT-qPCR analyses.

4. RESULTS AND DISCUSSION

Mycobacterium tuberculosis ECF Sigma Factor C (SigC) is peculiar under several aspects. It is an alternative sigma factor but, as shown, its expression pattern more closely resembles that of constitutive sigma factor SigA rather than that of other alternative factors[24]. Homologues of its structural gene are found in most pathogenic mycobacteria, such as *Mycobacterium leprae*, whose genome is thought to contain essential genes only, but *sigC* is not essential for *M. tuberculosis* survival. Or, more accurately, no condition has been found yet in which SigC is thoroughly, definitely essential. On this note, the observation that low copper concentrations impair the growth of an Erdman mutant lacking a functional form of the gene has come closest to the definition of a specific SigC-related phenotype[32]. As a confirmation, genes involved in copper uptake, and upregulated under copper-limiting conditions, were also induced by the artificial overexpression of *sigC*[31,32].

In this work, a preliminary characterization of SigC role and regulation has been carried out, and the validation of the still uncharacterized membrane protein Rv0093c as SigC-specific anti-sigma partner has been specifically approached with the generation of a deletion mutant. As theorized, in fact, the analysis of a mutant lacking the putative anti-sigma factor would produce similar results to the overexpression of *sigC*.

4.1 *Mycobacterium tuberculosis* mutants lacking *sigC* and/or *rv0093c* show identical growth rates in complete medium

The wildtype laboratory strain of *M. tuberculosis* H37Rv was chosen as host for the generation of a panel of mutants that consists in single and double deletions of *sigC* and the putative anti-sigma-encoding gene *rv0093c*, and their single and double complemented strains. Daily optical density readings were implemented as a way to measure growth rates in liquid cultures, and detected no significant differences between the strains (**Figure 4.1**), in all tested growth conditions. Obtained data establishes that further variations highlighted among these strains are not attributable to growth differences.

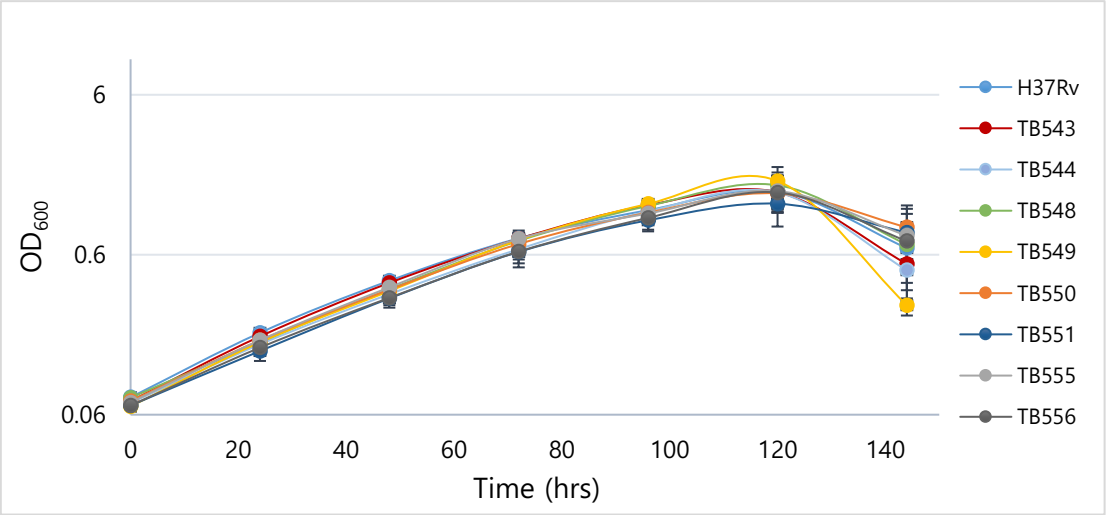

Figure 4.1 Growth curves obtained for wildtype strain H37Rv, and mutant strains TB543 (Δ*sigC*), TB544 (Δ*sigC::sigC*), TB548 (Δ*rv0093c*), TB549 (Δ*rv00930c::rv0093c*), TB550 (Δ*sigC* Δ*rv0093c*), TB551 (Δ*sigC::sigC* Δ*rv00930c::rv0093c*), TB555 (Δ*sigC::sigC* Δ*rv00930c*) and TB556 (Δ*sigC* Δ*rv00930c::rv0093c*), cultured in 7H9-ADN, in rolling. Represented data is the average result of at least three independent biological replicates and error bars indicate standard deviation.

4.1.1 Copper deprivation does not impair growth of an *M. tuberculosis* mutant lacking *sigC*

Several culture passages in copper-free Sauton's minimal medium were performed with the objective of treating *M. tuberculosis* strains to copper deprivation, and fully exhaust their reservoirs. Contrarily to previously described results[32], no growth defects could be observed between the wildtype strain and the *sigC*-deleted mutant, TB543. The strains were grown both in rolling and in standing conditions, and comparisons were made in different media, for up to six culture refreshings. The inability to replicate published data may be due to intrinsic dissimilarities between different laboratory host strains (Erdman and H37Rv), and the possible copper contamination of homemade media. The employment of known copper chelators might be of use, should the measurement of residual copper content validate this last hypothesis.

Nevertheless, genes reported under SigC control did result upregulated in mutant TB543, indicating that the achieved level of copper-deprivation is sufficient to elicit a targeted

transcriptional response, even though it has no visible effect on growth impairment. **Figure 4.2** recapitulates discussed results.

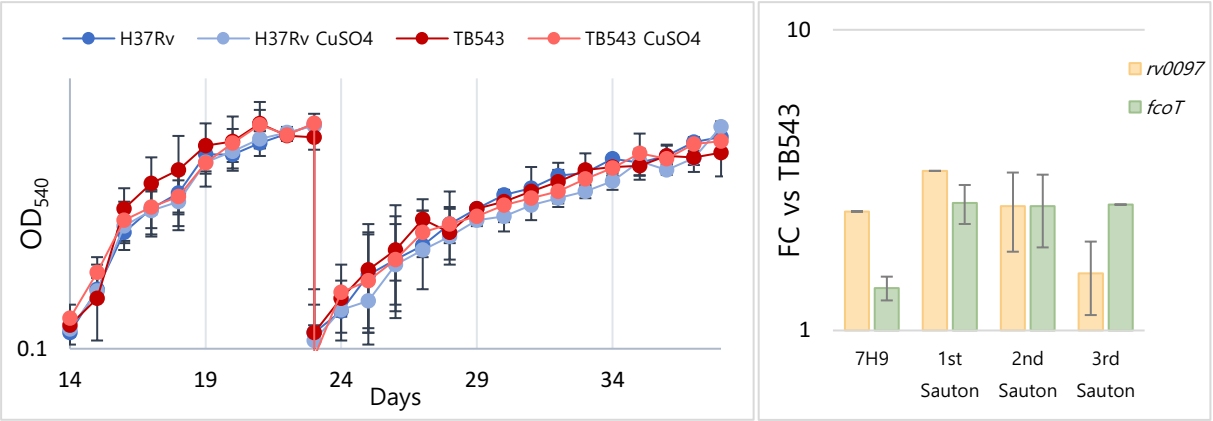

Figure 4.2 On the left, fifth and sixth culture passages of H37Rv and TB543, in Sauton's minimal medium, with or without 6 µM copper addition. On the right, *rv0097* and *fcoT* fold change levels, expressed as ratio between wildtype strain H37Rv and Δ*sigC* mutant TB543 copy numbers, when cultured in 7H9 or along 3 subsequent passages in Sauton's medium. Represented data is the average result of at least three independent biological replicates and error bars indicate standard deviation.

4.2 *M. tuberculosis* mutants lacking *rv0093c* are more sensitive to copper stress

The resazurin microtiter assay (REMA) was exploited to examine the mutants' susceptibility to different copper concentrations. Once inside bacteria, resazurin is reduced to highly fluorescent resorufin by metabolically active bacteria, and fluorescence measuring allows to quantify the effect of added copper sulfate on bacterial growth[51]. The nature of the copper salt and range of tested concentrations were based on the 6 µM $CuSO_4$-containing 7H9 medium, representing standard laboratory conditions. As visible in **Figure 4.3**, growth of *M. tuberculosis* mutant TB548, lacking *rv0093c*, was significantly reduced in comparison to both the wildtype (P = 0.0056) and TB543 mutant (P = 0.0014), starting from a concentration of 384 µM of $CuSO_4$. This evidence is supported by preliminary data obtained with the complemented strains. Of note, mutant TB555 carries both the double deletion of *sigC* and *rv0093c,* and *sigC* complementation. As expected, this mutant shows the same pattern of heightened sensitivity to

copper sulfate that is visible for the genotypically identical TB548. Mutant TB550, lacking the two genes, shows instead a similar pattern to both H37Rv and the double complemented mutant TB551, implying that *sigC* deletion superimposes that of *rv0093c*, and supporting the correlation between the two encoded proteins. In fact, considering the hypothesis of Rv0093c being SigC-specific anti-sigma factor, *rv0093c* deletion would work as a *sigC* induction, producing a persistent hyper-activation of copper import mechanisms and consequently leading to the opposite effect of copper overload toxicity, at lower concentrations to those inevitably affecting the wildtype. Taken together, these results confirm both SigC connection to copper uptake, that could not be replicated with the growth experiments, and Rv0093c possible role as an anti-sigma factor.

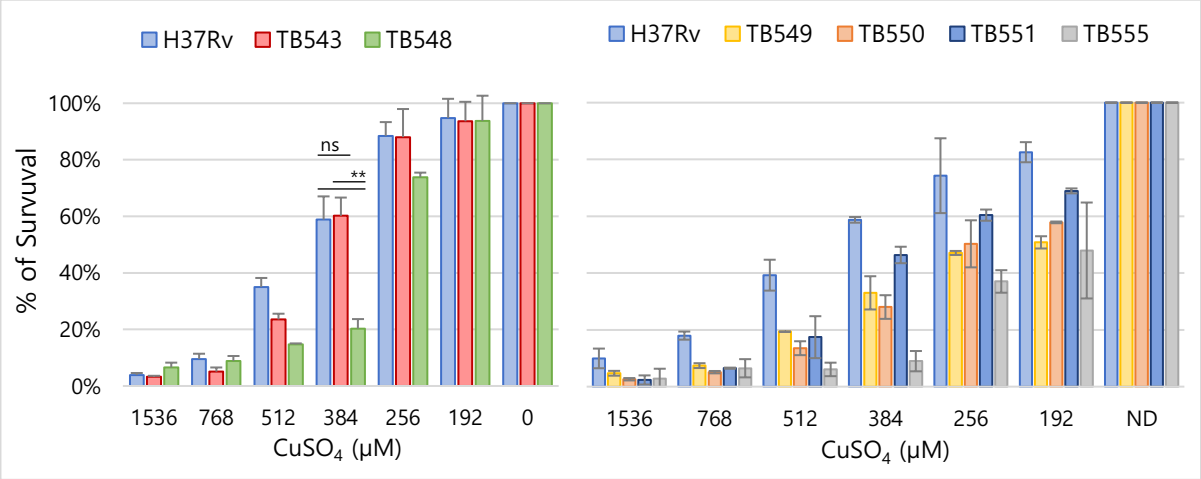

Figure 4.3 Percentage of survival of *M. tuberculosis* wildtype strain H37Rv and mutant strains TB543, TB548, TB549, TB550, TB551 and TB555 after treatment with $CuSO_4$. Tested concentrations represent multiples of the standard 7H9 concentration of 6µM $CuSO_4$. Represented data is the average result of at least two independent biological replicates and error bars indicate standard deviation. Statistical analysis was carried out with a two-sample T test assuming equal variances when the number of replicates was higher than three. Significance levels are indicated as follows: extremely significant: $P < 0.001$ (***), very significant: $0.001 < P < 0.01$ (**), significant: $0.01 < P < 0.05$ (*), non-significant: $P > 0.05$ (ns).

4.3 *rv0093c* deletion and copper overload loop activation of export mechanisms

In order to evaluate SigC indirect connection to copper export mechanisms, and validate Rv0093c activity as an anti-sigma factor, the regulation of *csoR* and RicR reporter gene *lpqS* was assessed in different media conditions and across different strains.

CsoR is the negative regulatory element of the *cso* operon, which encodes a copper efflux system, and LpqS is a putative lipoprotein with a still unknown role in the response to copper overload toxicity. Binding of copper ions due to increasing cytoplasmic burden causes CsoR and RicR to detach from their own promoters, resulting in the de-repression of all downstream genes, including *csoR* itself and *lpqS*[21,22]. Interestingly, expression of the two genes has been shown to increase as a secondary effect of *sigC* overexpression, suggesting that rising copper concentrations due to a SigC-driven induction of influx mechanisms are later evened out with the activation of the efflux machinery[32].

As shown in **Figure 4.4**, the two genes show a marked overexpression in H37Rv after treatment with high copper concentrations. Specifically, induction of *csoR* seems to increase in a time-dependent manner over the span of 90 minutes, while *lpqS* seemingly reaches a plateau before the initial recording time. Of note, while *csoR* is downregulated in the wildtype strain in copper-free Sauton's medium, when compared to 7H9, *lpqS* levels are kept steady in the two different media, suggesting a stricter binding of CsoR to its operon the lower the copper concentration gets.

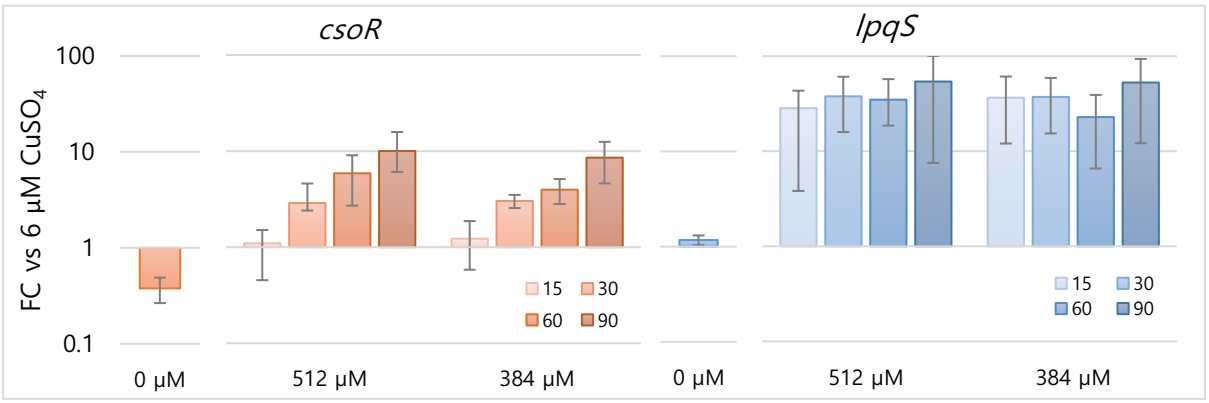

Figure 4.4 *csoR* and *lpqS* fold change levels in the wildtype strain H37Rv, expressed as ratio between absence (single bar) or excess (grouped bars) copper sulfate, and physiological copper concentration (6 µM). Data is representative of at least two independent biological replicates and error bars indicate standard deviation.

In the mutant lacking *sigC* (TB543), *csoR* and *lpqS* expression is decreased or equal, at highest, compared to the wildtype. This effect is expected due to inactive copper import mechanisms, which are reportedly under SigC control, and the added virtual absence of copper itself in Sauton's minimal medium. In the case of the *rv0093c*-deleted TB548 mutant, the protracted action of SigC, and resulting increased activity of the copper influx mechanisms, would eventually result in overexpression of the two genes, according to the copper concentration control loop that has been proposed[32]. As can be seen in **Figure 4.5**, the obtained results are in agreement with what is reported in the literature. Specifically, the indirect overexpression of *sigC* in 7H9 does not trigger an increase in *csoR* levels, but only in the RicR-related gene *lpqS.* Conversely, both genes are equally induced when the mutant is cultured in copper-free medium. The fact that, as explained, Sauton's medium cannot be considered completely deplete of copper, and that the conditions for CsoR and RicR metallation and detachment from their promoters may not be fully respected or maintained in these experiments, remain to be taken into consideration, as they might be important to properly define these regulators binding and unbinding thresholds.

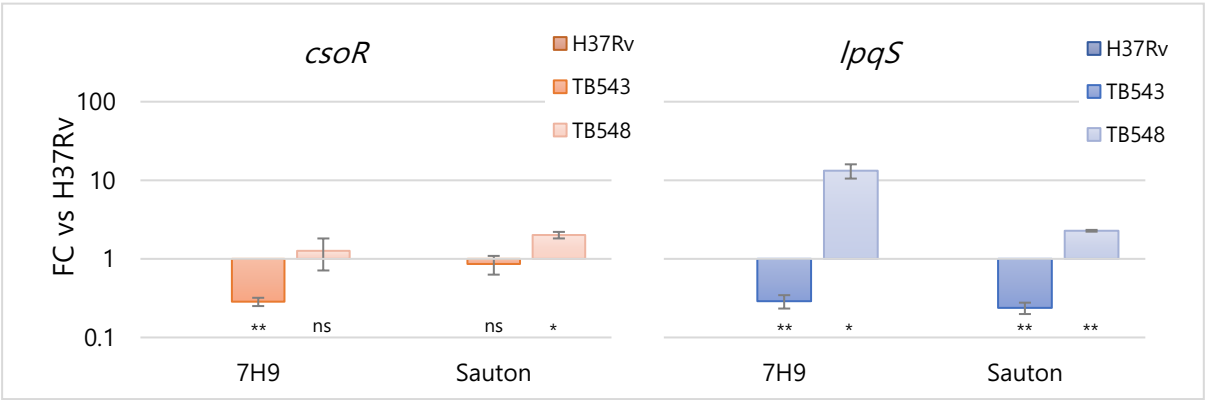

Figure 4.5 *csoR* and *lpqS* fold change levels in the wildtype strain H37Rv, expressed as ratio between mutant strains TB543 and TB548, and wildtype strain H37Rv copy numbers, measured in 7H9 (6 µM copper supfate) and copper-free Sauton's medium. Data is representative of at least three independent biological replicates and error bars indicate standard deviation. Statistical analysis was carried out comparing samples to H37Rv with a two-sample T test assuming equal variances. Significance levels are indicated as follows: extremely significant: $P < 0.001$ (***), very significant: $0.001 < P < 0.01$ (**), significant: $0.01 < P < 0.05$ (*), non-significant: $P > 0.05$ (ns).

4.4 *rv0093c* deletion affects transcription comparably to *sigC* induction

fcoT is a central gene in the *ppe1-nrp* operon, which is upregulated both in the context of artificial *sigC* overexpression[31], and under copper-limiting conditions[32], which physiologically induce *sigC*. It encodes a long-chain fatty acyl-CoA transferase (FcoT), involved in the production of isonitrile lipopeptides, which have been shown to strictly correlate to biofilm formation[33], and to have a possible role in copper direct uptake and transportation[34].

The gene expression was investigated in physiological concentrations of copper sulfate across the entire mutant panel, at two different growth stages. As shown in **Figure 4.6**, the gene results downregulated in the mutants carrying a complete *sigC* deletion (TB543, TB550, TB556) during exponential growth, and is instead induced in mutants lacking *rv0093c*, as would happen after a *sigC* induction. Again, this evidence supports the hypothesis of Rv0093c being SigC-specific anti-sigma factor. Furthermore, expression of the gene during the stationary (log) phase of the growth curve tends to uniform to wildtype baseline for mutants lacking *sigC*, consistently with reported evidence that *sigC* levels physiologically decrease as the bacterial population approaches stationary growth. On the contrary, the effect of sigC overexpression by deletion of *rv0093c* is intensified on log phase.

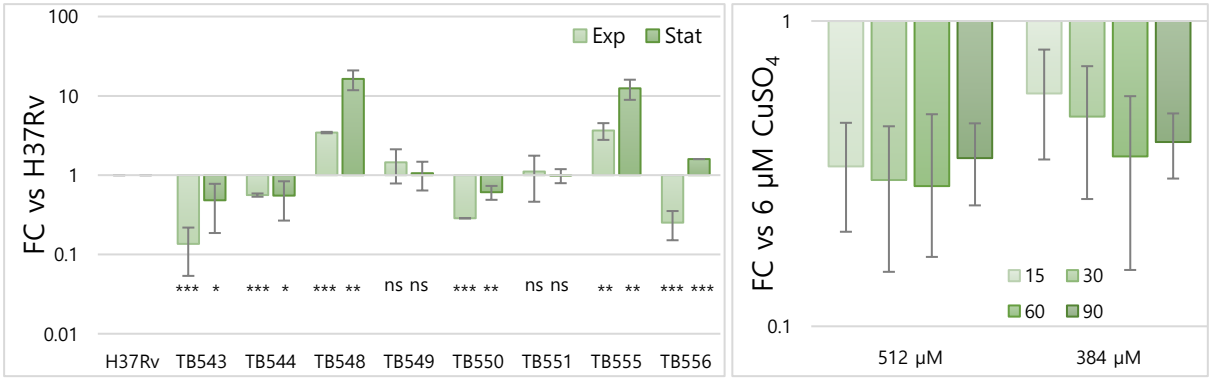

Figure 4.6 On the left, *fcoT* fold change levels, expressed as ratio between mutant strains and wildtype strain H37Rv copy numbers. Statistical analysis was carried out comparing samples to H37Rv with a two-sample T test assuming equal variances. Significance levels are indicated as follows: extremely significant: $P < 0.001$ (***), very significant: $0.001 < P < 0.01$ (**), significant: $0.01 < P < 0.05$ (*), non-significant: $P > 0.05$ (ns). On the right, *fcoT* fold change levels in the wildtype strain H37Rv, expressed as ratio between non-treated controls and samples collected 15-, 30-, 60- and 90-minutes post-treatment with $CuSO_4$. Data is representative of two independent biological replicates and error bars indicate standard deviation.

As visible in **Figure 4.6**, *fcoT* downregulation was detected after copper overload treatment of the wildtype strain H37Rv, when compared to the untreated controls. While a time-dependent decreasing trend seems to be visible for the lowest concentration of copper sulfate (384 µM) over the span of 90 minutes, *fcoT* levels seemingly reach a plateau before the first post-treatment recorded time at the higher concentration of copper sulfate (512 µM).

In order to properly characterize the entire panel of genes influenced by *rv0093c* deletion, RNAseq experiments were performed comparing the mutant strain TB548 to the wildtype H37Rv, in collaboration with Pr. Campanaro, Dr. Treu and Dr. Biruel, from the Department of Biology (Unipd). Fold change analysis allowed to identify a total of 22 differentially expressed genes, 19 of witch were upregulated in the mutant (log2FC > 1,5). As expected, these genes correspond almost exactly to published transcriptomic results obtained after transposon-mediated overexpression of *sigC*[31], further confirming Rv0093c role as SigC-specific anti-sigma factor. **Table 4.1** summarizes the described results. Of note, the *csoR* gene, *nrp* and *lpqS* operons, as well as *ricR* regulon, result induced in both experiments. The function and SigC-dependent or -indepentent regulation of all other significantly induced genes will be further studied, and subsequent experiments will be designed in order to define whether a consensus sequence can be identified for the entire SigC-related gene panel. Furthermore, the role and connection with SigC of the two underrepresented genes in the TB548 mutant, the aspartate oxidase-encoding *nadB* (log2FC = -1,58) and probable nicotinate-nucleotide pyrophosphatase-encoding *nadC* (log2FC = -1,69) will also be clarified (not shown).

Gene	Δ*rv0093c*	OE *sigC* [31]	Description
rv0092	1,47		Cation transporter P-type ATPase A (CtpA)
rv0095c	6,65		Conserved hypothetical protein
rv0096	5,47	5,94	PPE family protein (PPE1)
rv0097	4,13	4,66	Possible oxidoreductase
rv0098	4,53	4,6	Probable fatty acyl CoA thioesterase type III (FcoT)
rv0099	4,51	4,63	Possible fatty-acid-CoA ligase (FadD10)
rv0100	2,75	4,18	Conserved hypothetical protein

rv0101	3,73	4,41	Probable peptide synthetase (Nrp)
rv0103c	2,75	2,2	Probable cation-transporter P-type ATPase B (CtpB)
rv0104		1,23	Conserved hypothetical protein
rv0106		1,28	Conserved hypothetical protein
rv0186	1,51		Probable beta-glucosidase (BglS)
rv0186A	2,88		Metallothionein (MymT)
rv0190	1,61		Conserved protein (RicR)
rv0846c	2,90	2,61	Probable oxidase (MmcO)
rv0847	4,64	2,92	Probable lipoprotein (LpqS)
rv0848	3,73	3,66	Possible cysteine synthase a (CysK2)
rv0849	3,52	3,58	Probable conserved integral membrane transport protein
rv0850	3,76	3,67	Putative transposase (fragment)
rv0851c	2,19		Probable short-chain type dehydrogenase/reductase
rv0967	2,01	1,38	Copper-sensitive operon repressor (CsoR)
rv0968		1,41	Conserved protein
rv0969		1,38	Probable metal cation transporter P-type ATPase (CtpV)
rv0970		0,94	Probable conserved integral membrane protein
rv1813c		2,01	Conserved hypothetical protein
rv2963	2,12	1,76	Probable integral membrane protein
rv2964		0,65	Probable formyltetrahydrofolate deformylase (PurU)

Table 4.1 Log2FC data obtained from RNAseq experiments. Genes upregulated in both experiments are included in black boxes. Operons and regulons are indicated with concordantly coloured genes (*nrp*: green; *csoR*: red; *ricR* regulon and *lpqS* operon: blue). Data is representative of at least three independent biological replicates for the Δ*rv0093c* (TB548) mutant, and refers to "Rustad *et al, Genome Biology*, 2014" [31] for the OE *sigC* mutant.

4.5 Rv0093c is a transmembrane protein with cytosolic terminal domains

According to the literature, Rv0093c is a multi-pass transmembrane protein of *Mycobacterium tuberculosis*, which contains a putative zf-HC2 domain found in other anti-sigma factors[24], such as the anti-sigP (YlaC) factor of *Bacillus subtilis*[36]. More accurately, the position of the zinc-finger domain was confirmed between residues 49 and 83, part of the cytosolic N-terminal domain of the protein (InterPro[52] and CDD/SPARKLE[53]).

In silico predictions of Rv0093c cross-membrane topology with at least three published models (DeepTMHMM[54], HMMTOP[55], Phobius[56]) were concordant in indicating the localization of the membrane spanning regions of the protein, as well as the orientation of its N- and C-terminal domains. 3D rendition of the protein secondary structure was also obtained with the AlphaFold database[57]. Described results are represented in **Figure 4.7**.

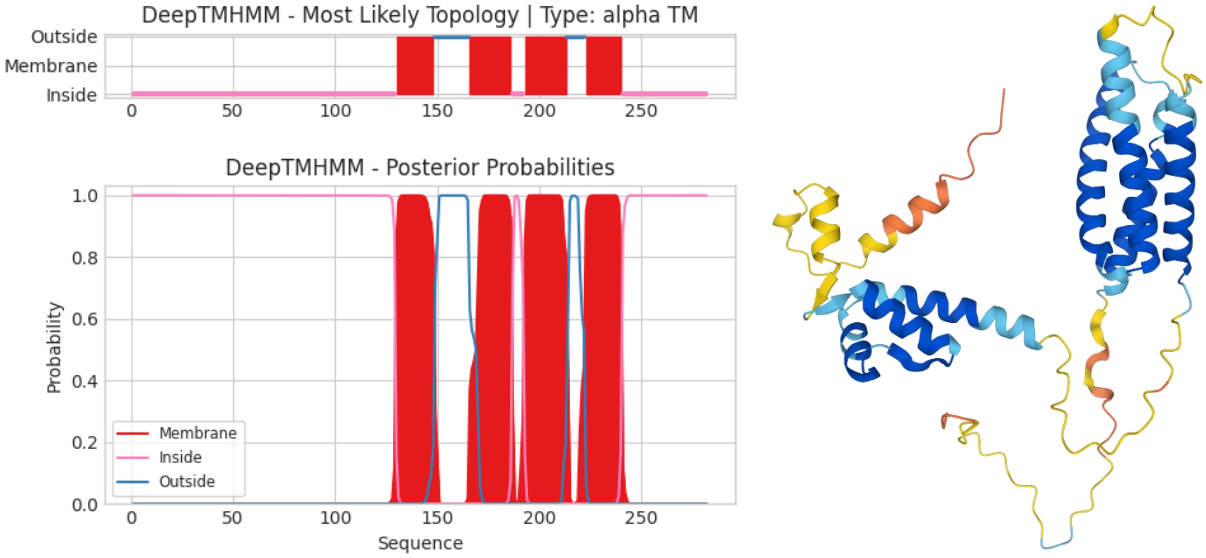

Figure 4.7 On the left, DeepTMHMM[54] most likely topology and posterior probabilities outputs for Rv0093c indicate the presence of cytosolic protein termini and four transmembrane α-helices connected by three short segments. On the right, 3D rendering prediction of Rv0093c secondary structure obtained from AlphaFold Protein Structure Database[57] (Deep Mind and EMBL-EBI), with confidence of prediction ranging from blue (high) to red (low).

4.6 SigC and Rv0093c interact *in vivo*

The mycobacterial protein fragment complementation (M-PFC) assay[50] was exploited to investigate SigC and putative anti-sigma factor Rv0093c physical interaction *in vivo*, in the environmental and rapidly growing mycobacterial species *M. smegmatis*. Bait and prey proteins were generated as fusions between the two target proteins and the two separate domains of the murine dihydrofolate reductase enzyme (mDHFR). As per the assay conditions, SigC-Rv0093c interaction would determine the functional reconstitution of mDHFR, conferring resistance to treatment with Trimethoprim, to which endogenous DHFR is lethally sensitive.

Figure 4.8 is representative of the output obtained with the experiment. Survival on the Trimethoprim-containing medium defines the two proteins as interacting partners and validates Rv0093c status as SigC-specific anti-sigma factor. Notably, the possibility of mDHFR domains steric hindrance on protein binding, which could be indicative of direct interaction between the protein termini, was evaluated by producing different fusion combinations. All produced strains, Ms327, Ms328, Ms329 and Ms330 showed ability to grow in Trimethoprim-containing medium, seemingly indicating that the sigma-anti-sigma interaction involves internal residues for both proteins, and neither the very N-terminal nor the C-terminal amino acids are strictly needed or implicated in the binding.

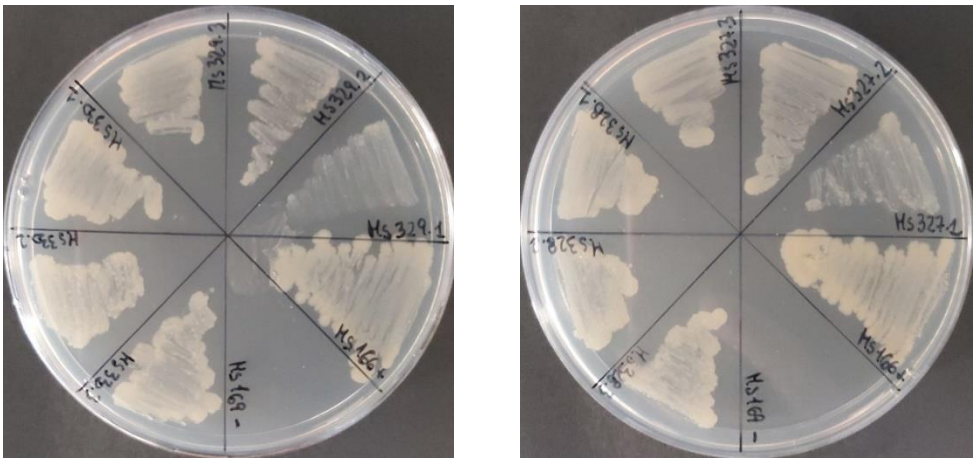

Figure 4.8 Survival of Ms327 (SigC-mDHFR[F1,2] and Rv0093c-mDHFR[F3]), Ms328 (mDHFR[F1,2]-SigC and mDHFR[F3]-Rv0093c), Ms329 (SigC-mDHFR[F1,2] and mDHFR[F3]-Rv0093c) and Ms330 (mDHFR[F1,2]-SigC and Rv0093c-mDHFR[F3]) on a Trim-containing 7H11 plate. Growth is shown also for the assay positive control (strain Ms166) and expected absence of growth is noticeable for the negative control Ms169.

Although in support of the standing hypothesis, the result of this assay adds little to the nature of the interaction between SigC and Rv0093c. Furthermore, the possibility of directly tagging the two target proteins in their natural host has been discarded also due to the assumption that the physiological expression levels of these regulators in the bacterial cell would not be enough to produce a detectable signal. Thus, the use of a reporter enzyme customized for protein interaction studies and optimized for detecting interactions under native conditions, such as the NanoBit® (Promega©) luciferase, will be implemented. These studies will be fundamental to better define the time-course dynamics and possible copper-dependency of the binding, and to quantify its intensity and duration.

4.7 *sigC* deletion produces a color-shift phenotype in dehydrated cultures

As described[46], *M. tuberculosis* is able to develop biofilms as pellicles at the air-liquid interface. Maturation of these structures has been reported to strictly correlate with activation of the *nrp* operon[33], whose expression was later shown to be dependent on SigC[32].

The initial hypothesis that the mutant lacking *sigC* (TB543) would not be able to form biofilms, as suggested by these premises, was disprooved by obtained results. As depicted in **Figure 4.9**, *M. tuberculosis* wildtype and mutant strains form a very convoluted pellicle that covers the entire well surface and walls, but no macroscopic differences can be observed in terms of temporal and spatial expansion of the membranes. Interestingly, the outcome varied substantially when the biofilm plate was incubated without humidification: the culture medium underwent massive evaporation (about 95% after five weeks), and no pellicles or other morphologies could be observed in either strain. In these conditions, an evident color-shift phenotype could be recognized for mutant TB543, which was restored to wildtype conditions in the *sigC*-complemented strain TB544. The phenotype-causing molecule appeared homogenously distributed in the whole water-based medium volume, indicating its probable hydrophylic nature, and the color was not altered after test treatment with either NaOH or HCl, allowing to exclude pH variation between the samples. This evidence was later confirmed with direct pH measurements, that reported non significant differences between the wildtype (pH = 6.45) and TB543 mutant (pH = 6.58).

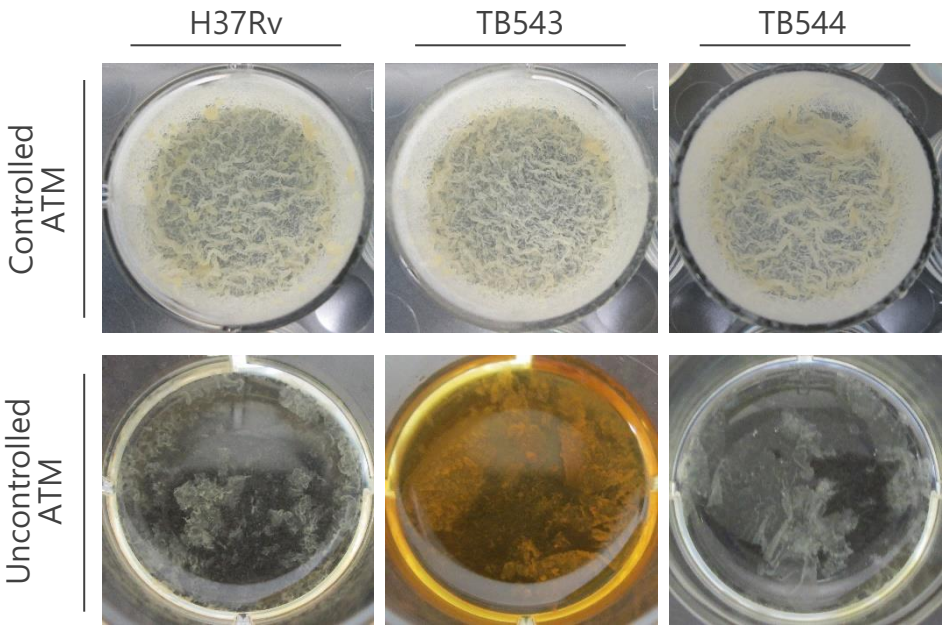

Figure 4.9 H37Rv and mutants TB543 and TB544 low-CFU inoculi, cultured in standing at 37 °C in detergent-free Sauton's medium. Biofilm formation is successful when incubation occurs in controlled atmosphere (5% CO_2, water-saturated) and aborted in uncontrolled atmosphere. Medium inoculated with TB543 shows a color shift from light yellow to vivid orange. Photographs were taken at the 5th week post inoculum.

In order to properly identify the target molecules direcly or indirectly causing this phenotype, further analyses have been performed in collaboration with Professors Roveri, Miotto and Zennaro, and Dr. Rossetto, from the Department of Molecular Medicine (Unipd). Initial evaluation of absorption spectra was unhelpful in detecting majour differences between the wildtype and TB543 mutant samples. An overall higher signal intensity was recorded for the mutant sample, possibly indicating that accumulation of one or more metabolites could be responsible for the highly visible color phenotype. Ultraconcentration did not separate the colored substance for both strains, possibly indicating complexation with larger molecules. Comparable spectrometry profiles were obtained after a ^{1}H-NMR analysis, performed on doubly ultra-filtrated samples. The overall metabolic profile of the NMR spectra resulted almost

entirely superimposable for the entire spectral amplitude, without it being possible to identify compounds selectively present in one of the two samples (**Figure 4.10**).

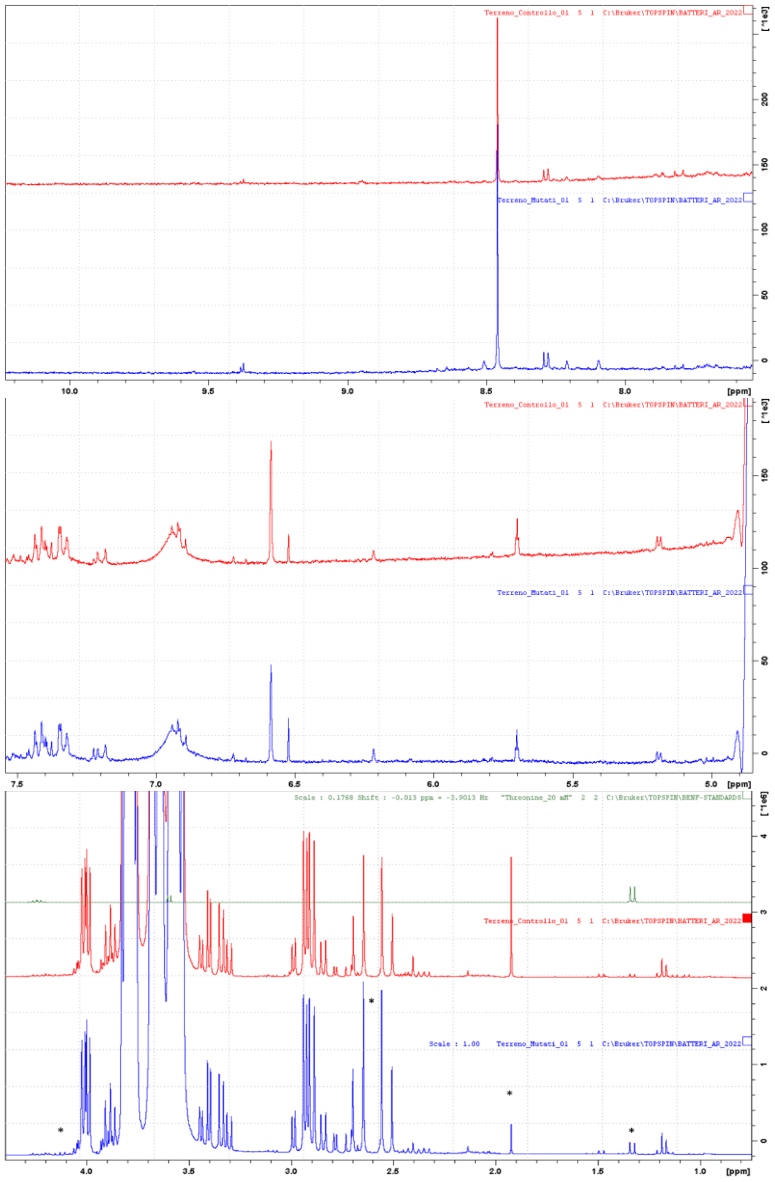

Figure 4.10 ^{1}H-NMR spectra aquired for H37Rv (red) and mutant TB543 (blue) show overall superimposable peaks with differential intensities at 1.32-1.34 ppm (Threonine), 1.93 ppm (Acetate) and 8.46 ppm (Formate). High glycerol content masks a region between 3.5 and 3.85 ppm, forcing acquisition at a lower gain and resulting in a loss of sensitivity. Curtesy of Prof. Lucio Zennaro and Dr. Monica Rossetto.

Differences in the intensity of peaks attributable to Threonine (1.32-1.34 ppm), Acetate (1.93 ppm) and Formate (8.46 ppm) have been identified, but the color of the samples is

unlikely to be due to these molecules, which result colorless. Minerals and heavy metals composition of Sauton's medium, H37Rv and TB543 preparations has been investigated with ICP-OES, but resulted in no significant differences between the samples for all studied analytes. Interestingly, copper residual contamination resulted very close to the detection limit of 30 µM for all three samples.

In a similar manner, an almost complete pattern overlay was also obtained with preliminary HPLC separations, as shown in **Figure 4.11**. Nevertheless, substantial peak-intensity differences were displayed between the TB543 mutant and the wildtype strain, mainly concerning two areas, respectively at retention times of 3.3 and 7.6 minutes post-injection. In order to identify the chemical composition of the samples, both these fractions were collected and will be further investigated with mass spectrometry. Of note, only the first fraction (3.3 min) maintained the original "yellow-orange" phenotype after recovery.

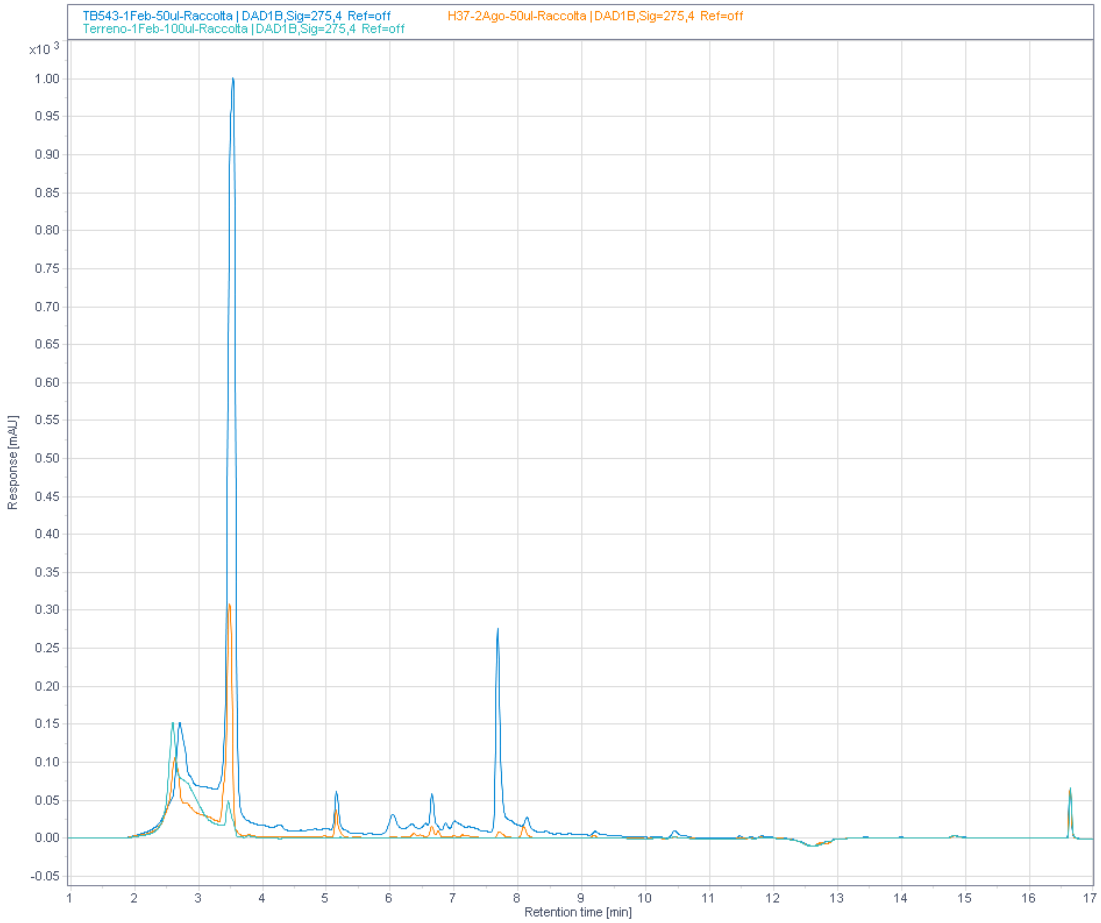

Figure 4.11 HPLC chromatogram aquired for H37Rv (orange) and mutant TB543 (blue) show overall superimposable peaks with differential intensities at retention times of 3.3 min and 7.6 min. Curtesy of Prof. Giovanni Miotto.

5. CONCLUSIONS

Sigma Factor C of *Mycobacterium tuberculosis* (SigC) has been reported to direct copper uptake in starvation, both with the demonstration that a mutant lacking its coding gene is not able to grow in a medium lacking copper, and with transcriptomic analyses that returned upregulation of copper-related genes when *sigC* was artificially or physiologically overexpressed. Based on the hypothesis that *rv0093c* may encode SigC-specific anti-sigma factor, the transcriptomic analysis of a mutant lacking the gene produces similar results to the upregulation of *sigC*, as expected.

Accordingly, a *rv0093c* KO mutant is more sensitive to high copper concentrations than its parental strain and upregulates genes that are both directly and indirectly connected with SigC. In particular, *fcoT* expression was shown to increase both in copper-lacking media and upon *rv0093c* elimination, and symmetrically to decrease in the absence of the sigma factor coding gene, or after treatment with high copper concentrations.

Evaluation of *rv0093c* deletion on main copper export effectors was performed in a similar way, by investigating the expression of *csoR* and RicR reporter gene *lpqS.* As hypothesized, in fact, excessive copper burden derived from SigC-directed intake results in the activation of the efflux machinery. The output is a seeming, but secondary, SigC-looped induction of both genes. *csoR* and *lpqS* indeed resulted overexpressed after copper treatment and in the absence of the putative anti-sigma factor encoding gene, *rv0093c.*

Lastly, Rv0093c configuration is interesting in itself. The protein was predicted to be inserted in the inner cytoplasmic membrane with four alpha helices, and to have a prominent cytosolic N-terminal domain, which could possibly be related to the interaction with SigC. Even though the nature and possible copper-dependency of this binding have not been clarified, evidence of *in vivo* interaction between the two proteins has been obtained, with a protein fragment complementation assay carried out in the model species *M. smegmatis*. Further study will allow to exactly define SigC-Rv0093c binding dynamics along the growth curve and under copper overload or deprivation.

In light of these comprehensive evidences, it would be appropriate to confidently appoint the protein Rv0093c as Regulator of Sigma factor C (RscA), a new anti-sigma factor of *Mycobacterium tuberculosis*.

Additionally, the complete characterization of SigC regulon will be performed, including new genes highlighted by transcriptomic studies, to obtain a more detailed analysis on SigC-controlled genes upstream regulatory regions, and possibly identify a consensus sequence specifically recognized by the sigma factor.

This study will further clarify SigC hypothesized connection with copper, with biofilm maturation and, possibly, with metabolic shifts occurring at low-density inoculi, another aspect of this research that will be properly examined.

REFERENCES

1. Cook, G. M. *et al.* Physiology of Mycobacteria. *Advances in Microbial Physiology* vol. 55 Preprint at https://doi.org/10.1016/S0065-2911(09)05502-7 (2009).

2. Johansen, M. D., Herrmann, J. L. & Kremer, L. Non-tuberculous mycobacteria and the rise of Mycobacterium abscessus. *Nature Reviews Microbiology* vol. 18 392–407 Preprint at https://doi.org/10.1038/s41579-020-0331-1 (2020).

3. Percival, S. L. & Williams, D. W. Mycobacterium. *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks: Second Edition* 177–207 (2014) doi:10.1016/B978-0-12-415846-7.00009-3.

4. Kanabalan, R. D. *et al.* Human tuberculosis and Mycobacterium tuberculosis complex: A review on genetic diversity, pathogenesis and omics approaches in host biomarkers discovery. *Microbiological Research* vol. 246 Preprint at https://doi.org/10.1016/j.micres.2020.126674 (2021).

5. Chavarro-Portillo, B., Soto, C. Y. & Guerrero, M. I. Mycobacterium leprae's evolution and environmental adaptation. *Acta Trop* **197**, 105041 (2019).

6. Ehrt, S., Schnappinger, D. & Rhee, K. Y. Metabolic principles of persistence and pathogenicity in Mycobacterium tuberculosis. *Nature Reviews Microbiology* vol. 16 496–507 Preprint at https://doi.org/10.1038/s41579-018-0013-4 (2018).

7. *Global Tuberculosis reporT 2022*. http://apps.who.int/bookorders. (2022).

8. Ramakrishnan, L. Revisiting the role of the granuloma in tuberculosis. *Nature Reviews Immunology* vol. 12 352–366 Preprint at https://doi.org/10.1038/nri3211 (2012).

9. Russell, D. G., Barry, C. E. & Flynn, J. L. Tuberculosis: What we don't know can, and does, hurt us. *Science* vol. 328 852–856 Preprint at https://doi.org/10.1126/science.1184784 (2010).

10. Sarathy, J. P. & Dartois, V. Caseum: a Niche for Mycobacterium tuberculosis Drug-Tolerant Persisters. (2020) doi:10.1128/CMR.

11. Ehrt, S. & Schnappinger, D. Mycobacterial survival strategies in the phagosome: Defence against host stresses. *Cellular Microbiology* vol. 11 1170–1178 Preprint at https://doi.org/10.1111/j.1462-5822.2009.01335.x (2009).

12. Wolschendorf, F. *et al.* Copper resistance is essential for virulence of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **108**, 1621–1626 (2011).

13. Hood, M. I. & Skaar, E. P. Nutritional immunity: Transition metals at the pathogen-host interface. *Nature Reviews Microbiology* vol. 10 525–537 Preprint at https://doi.org/10.1038/nrmicro2836 (2012).

14. Darwin, K. H. Mycobacterium tuberculosis and copper: A newly appreciated defense against an old foe? *Journal of Biological Chemistry* **290**, 18962–18966 (2015).

15. Samanovic, M. I., Ding, C., Thiele, D. J. & Darwin, K. H. Copper in microbial pathogenesis: Meddling with the metal. *Cell Host and Microbe* vol. 11 106–115 Preprint at https://doi.org/10.1016/j.chom.2012.01.009 (2012).

16. Shey-Njila, O. *et al.* CtpB Facilitates Mycobacterium tuberculosis Growth in Copper-Limited Niches. *Int J Mol Sci* **23**, (2022).

17. Rowland, J. L. & Niederweis, M. Resistance mechanisms of Mycobacterium tuberculosis against phagosomal copper overload. *Tuberculosis* vol. 92 202–210 Preprint at https://doi.org/10.1016/j.tube.2011.12.006 (2012).

18. Neyrolles, O., Wolschendorf, F., Mitra, A. & Niederweis, M. Mycobacteria, metals, and the macrophage. *Immunol Rev* **264**, 249–263 (2015).

19. Buglino, J. A. *et al.* Diisonitrile Lipopeptides Mediate Resistance to Copper Starvation in Pathogenic Mycobacteria. *mBio* **13**, e0251322 (2022).

20. Ward, S. K., Hoye, E. A. & Talaat, A. M. The global responses of Mycobacterium tuberculosis to physiological levels of copper. *J Bacteriol* **190**, 2939–2946 (2008).

21. Marcus, S. A., Sidiropoulos, S. W., Steinberg, H. & Talaat, A. M. CsoR Is Essential for Maintaining Copper Homeostasis in Mycobacterium tuberculosis. *PLoS One* **11**, e0151816 (2016).

22. Festa, R. A. *et al.* A novel copper-responsive regulon in Mycobacterium tuberculosis. *Mol Microbiol* **79**, 133–148 (2011).

23. Rodrigue, S., Provvedi, R., Jacques, P. É., Gaudreau, L. & Manganelli, R. The σ factors of Mycobacterium tuberculosis. *FEMS Microbiol Rev* **30**, (2006).

24. Manganelli, R. Sigma Factors: Key Molecules in Mycobacterium tuberculosis Physiology and Virulence. *Microbiol Spectr* **2**, (2014).

25. Paget, M. S. Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules* vol. 5 Preprint at https://doi.org/10.3390/biom5031245 (2015).

26. Saecker, R. M., Record, M. T. & Dehaseth, P. L. Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of Molecular Biology* vol. 412 754–771 Preprint at https://doi.org/10.1016/j.jmb.2011.01.018 (2011).

27. Österberg, S., Peso-Santos, T. del & Shingler, V. Regulation of alternative sigma factor use. *Annual Review of Microbiology* vol. 65 Preprint at https://doi.org/10.1146/annurev.micro.112408.134219 (2011).

28. Chang, A., Smollett, K. L., Gopaul, K. K., Chan, B. H. Y. & Davis, E. O. Mycobacterium tuberculosis H37Rv sigC is expressed from two promoters but is not auto-regulatory. *Tuberculosis* **92**, (2012).

29. Sun, R. *et al.* Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol Microbiol* **52**, (2004).

30. Karls, R. K., Guarner, J., McMurray, D. N., Birkness, K. A. & Quinn, F. D. Examination of Mycobacterium tuberculosis sigma factor mutants using low-dose aerosol infection of guinea pigs suggests a role for SigC in pathogenesis. *Microbiology (N Y)* **152**, (2006).

31. Rustad, T. R. *et al.* Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biol* **15**, 502 (2014).

32. Grosse-Siestrup, B. T. *et al.* A role for mycobacterium tuberculosis sigma factor c in copper nutritional immunity. *Int J Mol Sci* **22**, 1–18 (2021).

33. Richards, J. P. *et al. Adaptation of Mycobacterium tuberculosis to Biofilm Growth Is Genetically Linked to Drug Tolerance*. http://aac.asm.org/ (2019).

34. Harris, N. C. *et al.* Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria. *Proc Natl Acad Sci U S A* **114**, 7025–7030 (2017).

35. Sechi, L. A. *et al.* Genome and transcriptome scale portrait of sigma factors in Mycobacterium avium subsp. paratuberculosis. *Infection, Genetics and Evolution* **7**, (2007).

36. Borriss, R. *et al.* Bacillus subtilis, the model Gram-positive bacterium: 20 years of annotation refinement. *Microb Biotechnol* **11**, 3–17 (2018).

37. von Mering, C. *et al.* STRING: A database of predicted functional associations between proteins. *Nucleic Acids Research* vol. 31 258–261 Preprint at https://doi.org/10.1093/nar/gkg034 (2003).

38. Yang, Y. *et al.* Defining a temporal order of genetic requirements for development of mycobacterial biofilms. *Mol Microbiol* **105**, 794–809 (2017).

39. Fux, C. A., Costerton, J. W., Stewart, P. S. & Stoodley, P. Survival strategies of infectious biofilms. *Trends in Microbiology* vol. 13 34–40 Preprint at https://doi.org/10.1016/j.tim.2004.11.010 (2005).

40. Davies, D. Understanding biofilm resistance to antibacterial agents. *Nature Reviews Drug Discovery* vol. 2 114–122 Preprint at https://doi.org/10.1038/nrd1008 (2003).

41. O'toole, G., Kaplan, H. B. & Kolter, R. *BIOFILM FORMATION AS MICROBIAL DEVELOPMENT*. www.annualreviews.org (2000).

42. Flemming, H. C. *et al.* Biofilms: An emergent form of bacterial life. *Nature Reviews Microbiology* vol. 14 563–575 Preprint at https://doi.org/10.1038/nrmicro.2016.94 (2016).

43. López, D., Vlamakis, H. & Kolter, R. Biofilms. *Cold Spring Harbor perspectives in biology* vol. 2 Preprint at https://doi.org/10.1101/cshperspect.a000398 (2010).

44. Esteban, J. & García-Coca, M. Mycobacterium biofilms. *Frontiers in Microbiology* vol. 8 Preprint at https://doi.org/10.3389/fmicb.2017.02651 (2018).

45. Chakraborty, P. & Kumar, A. The extracellular matrix of mycobacterial biofilms: Could we shorten the treatment of mycobacterial infections? *Microbial Cell* vol. 6 105–122 Preprint at https://doi.org/10.15698/mic2019.02.667 (2019).

46. Ojha, A. K. *et al.* Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* **69**, 164–174 (2008).

47. Parish, T. & Stoker, N. G. *Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement*. *Microbiology* vol. 146 (2000).

48. Murphy, K. C. *et al.* Orbit: A new paradigm for genetic engineering of mycobacterial chromosomes. *mBio* **9**, 1–20 (2018).

49. Stover, C. K. *et al. 47. Braithwaite, A. W. eta*. *N. R. eta/. Proc. natn. Acad Sci. US.A* vol. 13 (1989).

50. Singh, A., Mai, D., Kumar, A. & Steyn, A. J. C. *Dissecting virulence pathways of Mycobacterium tuberculosis through protein-protein association*. www.pnas.orgcgidoi10.1073pnas.0602817103 (2006).

51. Palomino, J. C. *et al.* Resazurin microtiter assay plate: Simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **46**, 2720–2722 (2002).

52. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res* (2022) doi:10.1093/nar/gkac993.

53. Lu, S. *et al.* CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res* **48**, D265–D268 (2020).

54. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. doi:10.1101/2022.04.08.487609.

55. Tusnády, E., Tusnády, T., Istv´, I. & Simon, I. *The HMMTOP transmembrane topology prediction server*. *BIOINFORMATICS APPLICATIONS NOTE* vol. 17 http://www.cbs.dtu.dk/ (2001).

56. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* **50**, W276–W279 (2022).

57. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).