MYCOBACTERIAL DORMANCY AND PERSISTENCE: MOLECULAR MECHANISMS CONTROLLING STRESS RESPONSE, SURVIVAL AND ADAPTATION

CHIONH YOK HIAN
B.Sc. *Biological Science and Economics (1ˢᵗ Class Hons.)*
Nanyang Technology University

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Chionh Yok Hian
24 June 2015

ACKNOWLEDGEMENTS

different this thesis would had turn out if you had not came along, wanting to set up Wayne cultures. Last but not least, Wenwei, Michelle, Vanessa, Julia, Emily, Regina, Annabelle, Weixin, Jowin, Jian Hang, Lili, Li Ching, Zarina, Aakansha, Huimin, Eshele and Sze Wai for absorbing me into the SA lab through some kind of uptake mechanism that borders upon sorcery.

To my parents, I paraphrase P.G. Wodehouse, for without whose never-failing sympathy and encouragement this thesis would have been finished in half the time. I love you still. To my brother, Yok Teng, for without whose own excellent Ph.D. thesis as a reference, this thesis would have been finished in twice the time. I love you too.

You know how it is; you open this thesis, flip to the acknowledgements, and find that, once again, the author has dedicated it to a family member or loved one. Well, not this time. This thesis is dedicated to you, dear reader, for being the raison d'etre *for these words.*

Table of contents

Ph.D. Thesis

Mycobacterial dormancy and persistence: Molecular mechanisms controlling stress response, survival and adaptation

by

Yok Hian Chionh
Department of Microbiology
Yong Loo Lin School of Medicine
National University of Singapore

Summary

Tuberculosis is among the most prevalent infectious diseases on the planet. The causative pathogens from the *Mycobacterium tuberculosis* complex successfully counter host immunity and survive environmental stressors such as hypoxia and starvation to enter an antibiotic-tolerant persistent state. These "dormant" bacteria establish a latent disease that can relapse decades after the primary infection. Thus, elucidating the molecular mechanisms underlying persistence is critical to developing therapeutic interventions. To this end, systems-level analyses were performed to define the molecular responses to nutrient deprivation and hypoxia. For nutrient deprivation, transcriptional profiling was combined with flow cytometric measurements of microbe physiology and multiplexed metabolic phenotypic screens to deduce that starved mycobacteria enter a state of lipid metabolism-induced ketosis that results in formation of reactive oxygen species by upregulating cytochrome P450s. Further biochemical investigations established that targeted killing of persister populations could be accomplished using Fenton reactions that damage these enzymes. For hypoxic stress, data-driven analyses of the ribonucleome and proteome of hypoxic mycobacteria revealed chemical reprogramming of modified ribonucleosides in transfer RNAs, which caused selective translation of codon-biased mRNAs essential to the stress response. Disruption of this system by codon reengineering caused dormancy responses to be mistimed and this is detrimental to hypoxia survival. Together, these discoveries offer insights into the mechanisms underlying mycobacterial persistence, dormancy and drug tolerance, which provide new targets for drug development, platforms for drug screening, and biomarkers of disease state.

Thesis supervisors:
Peter C. Dedon
Underwood-Prescott Professor of Toxicology and Biological Engineering, Massachusetts Institute of Technology
Principle Investigator, Singapore-MIT Alliance for Research and Technology

Sylvie Alonso
Associate Professor, Department of Microbiology and LSI Immunology Programme, Yong Loo Lin School of Medicine, National University of Singapore
Co-Investigator, Singapore-MIT Alliance for Research and Technology

List of tables

List of figures

List of abbreviations

A	Adenosine
AA	Acetoacetate
ac^{4}C	N^4-Acetylcytidine
acp^{3}U	3-(3-amino-3-carboxypropyl)uridine
AIDS	Acquired immune deficiency syndrome
Ala	Alanine
Am	2'-O-Methyladenosine
AMK	Amikacin
Ar(p)	2'-O-Ribosyladenosine (phosphate)
Arg	Arginine
ARV	anti-retroviral
Asn	Asparagine
AsO4	Arsenite
Asp	Aspartic acid
ATP	Adenosine triphosphate
B	Nucleoside not A (G or C or T)
BCG	*Bacille de Calmette et Guérin*
C	Cytidine
CAP	Capreomycin
cDNA	Complementary DNA
CFU	Colony forming units
cFDA	Carboxyfluorescein diacetate,
Cm	2'-O-methylcytidine
Cm	2'-O-Methylcytidine
cmo^{5}U	uridine 5-oxyacetic acid
CoA	Coenzyme A
C$_T$	Threshold cycle
ct^{6}A	cyclic N^6-threonylcarbamoyladenosine
CYP	Cytochrome P450
Cys	Cysteine
D	Dihydrouridine
DM	Diabetes mellitus
DAPI	4',6-Diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dos	Dormancy survival
DOTS	Directly observed treatment short-course
DR	Drug resistant
DTT	Dithiothreitol
EMB	Ethambutol
ETH	Ethionamide
FAS	Fatty acid synthase
FITC	Fluorescein isothiocyanate

G	Guanosine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
Gm	2'-O-methylguanosine
Gm	2'-O-Methylguanosine
H	Nucleoside not G (A or C or T)
H_2O_2	Hydrogen peroxide
HCL	Hierarchical clustering analysis
His	Histidine
HIV	Human immunodeficiency virus
ho^5U	5-Hydroxyuridine
HPLC	High performance liquid chromatography
HSL	Hormone sensitive lipase
HSR	Head space ratio
I	Inosine
i^6A	N^6-Isopentenyladenosine
IAA	Iodoacetamide
ICL	Isocitrate lyase
IFN-γ	Interferon gamma
IL	Interleukin
Ile	Isoleucine
INH	Isoniazid
IP RPC	Ion-pair reversed phased chromatography
k^2C	lysidine
KAN	Kanamycin
LB	Luria-Bertani
LC	Liquid chromatography
LC-MS	Liquid chromatography tandem mass spectrometry
Leu	Leucine
LTBI	Latent tuberculosis infection
Lys	Lysine
m^2G	N^2-Methylguanosine
m/z	mass to charge ratio
m^1A	1-Methyladenosine
m^1G	1-Methylguanosine
m^1I	1-Methylinosine
m^2_2G	N^2,N^2-Dimethylguanosine
m^2A	1-methyladenosine
m^2G	$N2$-methylguanosine
m^3C	3-Methylcytidine
m^5C	5-Methylcytidine
m^5U	5-Methyluridine
m^6_2A	N^6,N^6-Dimethyladenosine

m^{6}A	*N^6*-methyladenosine
m^{6}t^{6}A	*N^6*-methyl-*N*6-threonylcarbamoyladenosine
m^{7}G	7-Methylguanosine
MBC	Minimal bactericidal concentration
mcm^{5}s^{2}U	5-methoxycarbonylmethyl-2-thiouridine
mcm^{5}U	5-Methoxycarbonyl-methyluridine
mcm^{5}U	5-Methoxycarbonylmethyluridine
mcmo^{5}U	Uridine 5-oxyacetic acid methyl ester
mcms^{2}U	5-Methoxycarbonylmethyl-2-thiouridine
MCS	Multiple cloning sites
MDR	Multi-drug resistant
Met	Methionine
MIC	Minimal inhibitory concentration
MIT	Massachusetts Institute of Technology
MMS	Methyl methanesulfonate
mnm^{5}s^{2}U	5-methylaminomethyl-2-thiouridine
mnm^{5}U	5-methylaminomethyluridine
mo^{5}U	5-Methoxyuridine
MRM	Multiple reactions monitoring
mRNA	Messenger RNA
MS	Mass spectrometry
ms^{2}i^{6}A	2-methylthio-*N6*-isopentenyladenosine
Mtb	*Mycobacterium tuberculosis*
Mw	Molecular weight
N	Unspecified or unknown nucleoside
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
ncm^{5}U	5-carbamoylmethyluridine
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NO	Nitric oxide
NRP	Non-replicating phase
NTM	Nontuberculous mycobacteria
NUS	National University of Singapore
OADC	Oleic Albumin Dextrose Catalase
OD	Optical density
OD$_{600nm}$	Optical density at 600 nm
oQ	epoxyqueuosine
ORF	Open reading frame
PA-824	Pretomanid
PAS	*p*-amino salicylic acid
PBS	Phosphate buffer saline
PBS	Phosphate-buffered saline
PC	Phosphatidylcholine

PCA	Principle component analysis
PDB	Protein data bank
PDIM	Phthiocerol dimycocerosate
Phe	Phenylalanine
Pro	Proline
PTM	Post-translational modification
PZA	Pyrazinamide
Q	Queuosine
qPCR	Quantitative PCR
QQQ	Triple quadrupole mass spectrometer
QTOF	Quadrupole time-of-flight mass spectrometer
R	Unspecified purine nucleoside
rBCG	Recombinant BCG
RE	Restriction enzymes
RIF	Rifampicin
RNA	Ribonucleic acid
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
RT-qPCR	Real time quantitative polymerase chain reaction
s^2C	2-thiocytidine
s^4U	4-thiouridine
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEM	Standard error of sample means
Ser	Serine
sRNAs	Small RNA
siRNA	Small interfering RNA
STM	Streptomycin
t^6A	N^6-threonylcarbamoyladenosine
TAG	Triacylglycerol
TB	Tuberculosis
TBHP	tert-Butyl hydroperoxide
TCA	Tricarboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
TDR	Totally-drug resistant
Thr	Threonine
TLC	Thin layer chromatography
TM207	Bedaquiline
TNF-α	Tumour necrosis factor alpha
tRNA	Transfer RNA
Trp	Tryptophan
Tyr	Tyrosine
U	Uridine

Um	2'-*O*-Methyluridine
UV	Ultra violet
V	nucleoside not T (A or G or C)
Val	Valine
VIO	Viomycin
vRNA	Viral RNA
WHO	World Health Organisation
WT	Wild type
x	Variable alkyl group
XDR	Extensively-drug resistant
Y	Unspecified pyrimidine nucleoside
yW	wybutosine
βHB	Beta-hydroxybutyrate
Ψ	Pseudouridine

Publications

1. Hia F, **Chionh YH**, Pang YL, DeMott MS, McBee ME, Dedon PC. *Mycobacterial RNA isolation optimized for non-coding RNA: high fidelity isolation of 5S rRNA from Mycobacterium bovis BCG reveals novel post-transcriptional processing and a complete spectrum of modified ribonucleosides.* Nucleic Acids Res. 2014 Dec 24. [Epub ahead of print]. (Featured in Chapter 2)

2. Su D, Chan CT, Gu C, Lim KS, **Chionh YH**, McBee ME, Russell BS, Babu IR, Begley TJ, Dedon PC. *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry.* Nat Protoc. 2014 Apr; 9(4):828-41. (Featured in Chapter 4)

3. Li Y, Chen Q, Zheng D, Yin L, **Chionh YH**, Wong LH, Tan SQ, Tan TC, Chan JK, Alonso S, Dedon PC, Lim B, Chen J. *Induction of functional human macrophages from bone marrow promonocytes by M-CSF in humanized mice.* J Immunol. 2013 Sep 15;191(6):3192-9.

4. **Chionh YH**, Ho CH, Pruksakorn D, Ramesh Babu I, Ng CS, Hia F, McBee ME, Su D, Pang YL, Gu C, Dong H, Prestwich EG, Shi PY, Preiser PR, Alonso S, Dedon PC. *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Res. 2013 Sep;41(17):e168. (Featured in Chapter 3)

5. Dong H, Chang DC, Hua MH, Lim SP, **Chionh YH**, Hia F, Lee YH, Kukkaro P, Lok SM, Dedon PC, Shi PY. *2'-O methylation of internal adenosine by flavivirus NS5 methyltransferase.* PLoS Pathog. 2012;8(4):e1002642.

6. Chan CT, **Chionh YH**, Ho CH, Lim KS, Babu IR, Ang E, Wenwei L, Alonso S, Dedon PC. *Identification of N6,N6-dimethyladenosine in transfer RNA from Mycobacterium bovis Bacille Calmette-Guérin.* Molecules. 2011 Jun 21;16(6):5168-81. (Featured in Chapter 4)

Selected presentations

1. **Chionh YH,** Alonso S, Dedon, PC. *Decoding dormancy: Reprogrammed tRNAs read a code of codons to regulate the mycobacterial proteome.* American Society for Microbiology General Meeting 2014. Boston, USA. Selected speaker.
   *Young investigator presentation.

2. **Chionh YH,** Babu IR**,** Ng SW, Alonso S, Dedon, PC. *Decoding mycobacterial dormancy: Linking tRNA modifications with selective translation of genes.* 7th Asia-Pacific Organization For Cell Biology Congress 2014. Singapore. Poster presentation.

3. **Chionh YH,** Hia F**,** Alonso S, Dedon, PC. *Dynamic reprogramming of tRNA modifications is linked to Dos regulon activation in hypoxia-induced mycobacterial dormancy.* Boston Bacterial Meeting 2013. Poster Presentation.

4. **Chionh YH,** Hia F**,** Alonso S, Dedon, PC. *Dynamic Reprogramming of tRNA Modifications is Linked to Activation of the Dos Regulon in Hypoxia-induced Mycobacterial Dormancy.* American Society for Microbiology General Meeting 2013. Denver, USA. Poster presentation.
   *Awarded GM Outstanding Student Award and travel grant.

5. **Chionh YH,** Chan, TYC, Hia F, Alonso S, Dedon, PC. *Quantitative profiling of tRNA modification dynamics reveals key ribonucleoside signatures of non-replicative Mycobacterium bovis BCG.* Cold Spring Harbor Symposium for RNA Biology 2012. Suzhou, China. Selected speaker.

6. **Chionh YH,** Ravindran MS, Chan, TYC, Alonso S, Dedon, PC. *Ribonucleoside signatures of non-replicative Mycobacterium bovis BCG by multidimensional LC-MS/MS.* EMBO conference series: Tuberculosis 2012. Paris, France. Poster presentation.

7. **Chionh YH,** Chan, TYC, Ho C-H, Alonso S, Dedon, PC. *Translational Control in Microbial Pathogens:Defining the Spectrum of tRNA Modifications in Bacille Calmette-Guérin (BCG).* Joint American Society for Cancer Research and American Chemical Society meeting: Chemical in Cancer Research 2011. San Diego, USA. Poster presentation.

1. Background and Significance

1.1. Motivation and goal

This thesis is motivated by the convergence of three factors: the continued tuberculosis (TB) pandemic, the emergence of significant antibiotic resistance in its causative organism, *Mycobacterium tuberculosis (Mtb)*, and an evolving paradigm of translational control of cell response. Members of the Dedon research group at MIT recently described a new mechanism by which eukaryotic cells respond to stress, involving enzymatic reprogramming of chemically modified ribonucleotides in transfer ribonucleic acids (tRNAs) to control gene expression under stress. Given the conservation of tRNA modifications and translational machinery in all living organisms, I rationalized that this translational control mechanism would be operant in prokaryotes, including mycobacteria. Thus, the overarching goal of my research has been to understand how pathogens of the *Mtb* complex regulate their gene expression in response to nutrient deprivation and hypoxia – physiological stresses associated with the host's immune response and disease pathogenesis. This understanding would not only further our knowledge in microbiology in general and mycobacteriology specifically, but more importantly, it would open unexplored avenues in TB drug development and biomarker discovery.

1.2. Scope

The scope of the problem of tuberculosis (TB) is highlighted by both its long history and its prevalence. The historical context is set with the following milestones: 6000 years ago, mycobacteria evolved as human pathogens [1]; Hippocrates first described the disease 2400 years ago [2, 3]; Robert Koch identified the causative agent 133 years ago [4]; Albert Calmette and Camille

Guérin introduced attenuated *Mycobacterium bovis*, Bacille de Calmette et Guérin (BCG), as a vaccine against TB 94 years ago [5]; and 48 years ago, rifampicin was introduced to modern treatment regiments [6]. In spite of this long history of study, TB remains, save for human immunodeficiency virus infections and acquired immune deficiency syndrome (HIV/AIDS), as the greatest killer worldwide due to a single infectious agent [7], with an estimated one-third of humanity infected. The lack of progress in developing new antibiotics and vaccines for TB is apparent from the fact that today's TB drug regimen takes too long (6-24 months) to be effective and requires too many medications [8, 9]. Furthermore, current first- and second-line drugs are highly toxic [10-12], expensive and incompatible with common anti-retroviral (ARV) therapies used to treat HIV co-infections [13]. The most worrying problem, which motivates this thesis, is the recent advent of multi-drug resistant (MDR), extensively drug resistant (XDR) and even totally drug resistant (TDR) TB [7, 14]. Together, these developments highlight the urgent need for new therapeutics, active against current drug-resistant strains, with reduced side effects and increased potency. With these traits, new TB medications would shorten treatment times, improve patient adherence, and lessen the likelihood of bacterial strains developing drug resistance [15, 16]. Yet, disregarding serendipity, the discovery and development of TB drugs with novel mechanisms of action require a fundamental understanding of how pathogens of the *Mtb* complex evade the host innate and adaptive immunity, enter a persistent state in nutrient-limited, hypoxic granulomatous lesions, develop antibiotic resistance or tolerance (both genetic and phenotypic), and re-emerge years after the primary infection to the detriment of the host [17].

This thesis is aimed at gaining molecular insights in the fundamental mechanisms of mycobacterial-host interactions, which requires a critical

evaluation of reported experimental observations and occasionally a challenge to long-standing dogma. The scope of my research project is based on recent novel observations on the roles of metabolism and reactive oxygen species (ROS) in bacterial persistence [18, 19], and on the translational control of cell survival [20, 21]. My goals were to first to redefine these cellular and molecular systems in the context of mycobacterial survival of physiological stresses; second, to expand the original observations by quantifying how each component varies as a function of stress exposure; and third, to utilize this quantitative formulation of observed biological phenomena to build new models for mycobacterial-host interactions.

Throughout this process, quantitative biology allows us an "extra sense", as appreciated by Charles Darwin himself [22], to decide whether a given biological claim actually makes sense. This is illustrated in Chapter 2, in which we disproved earlier over-estimates on total RNA quantities mycobacteria and the length of mycobacterial 5S ribosomal RNA (rRNA) by providing its exact sequence and quantities within BCG. This recursive approach between hypothesis- and data-driven research enables us to gain a systems-level appreciation of mycobacterial stress responses, which is an appreciation sorely needed for the design and development of biomarkers of infection, diagnostic tools and new antibiotics.

In the following introductory chapter, we will travel down the scales of quantitative biology by first examining how TB epidemiology reflects *Mtb* pathophysiology, then narrowing the focus on *Mtb* cellular stress responses and how *Mtb* metabolism determines persistence. Next, we address gene expression by introducing an innovative new model of the translation control of stress response proteins in *Saccharomyces cerevisiae* and discussing its

implication for non-replicating mycobacteria. Finally, we conclude by providing an outline of the specific aims of each section of this thesis.

1.3. Tuberculosis: etiology, epidemiology and pathophysiology

TB is an "heirloom disease": one caused by mycobacteria that infected early humans and evolved with the species as people have spread around the world [23]. As the causative agent of TB, *Mtb* emerged relatively recently in evolutionary terms [1]. Although commonly defined as a latent, slowly debilitating disease, TB occasionally assumes an acute, rapidly progressive course. Initial tuberculous mycobacterial infection usually goes unnoticed and is a condition known as latent TB infection (LTBI). About 10% of immunocompetent adults with LTBI will eventually progress to active disease, and half of them will do so in the first two years following infection [8, 24]. The risk of progression to active disease is increased in immuno-compromised persons, such as in diabetics and patients co-infected with HIV, and children under 5 years of age [25-27]. The disease also affects practically all vertebrate species, which can serve as reservoirs for zoonotic infections. For instance, *M. bovis* infection of cattle is not only an economic threat on the scale of bovine spongiform encephalitis or hand, foot and mouth disease to the dairy industry, but it is still a significant zoonosis in non-industrialized countries of the world [28].

1.3.1. Disease burden of *Mtb* infections

Globally in 2013, 9 million people fell ill with TB, 11 million suffered from active TB, and 1.5 million died from the disease. Although gradual, the global incidence, prevalence and mortality declined 1.5%, 1.8% and 1.9%, respectively, on average yearly from 1990 - 2013 (**Fig 1.1**) [7]. Economically,

TB causes losses of US$12 billion in productivity annually [16], but this estimate does not consider the costs of vaccination, surveillance, prevention and eradication programs nor does it consider the losses caused by livestock TB [29]. These numbers, however, fail to capture the inequality with which the burden is shared. Most of the estimated number of cases of TB in 2013 occurred in Asia (56%) and the African Region (29%), with India and China alone accounting for 24% and 11% of global cases, respectively. The lowest epidemiological burden occurred predominantly in high-income countries including most countries in Western Europe, North America and Japan [7]. Singapore, interestingly, possesses the highest incidences rates for a developed country at ~40 new cases per 100,000 persons from 2000-2013 [30], though this is mostly attributed to the large, foreign-born, migrant workforce from developing nations [30, 31]. Indeed, TB and poverty are closely linked. Malnutrition, overcrowding, poor air circulation and sanitation-
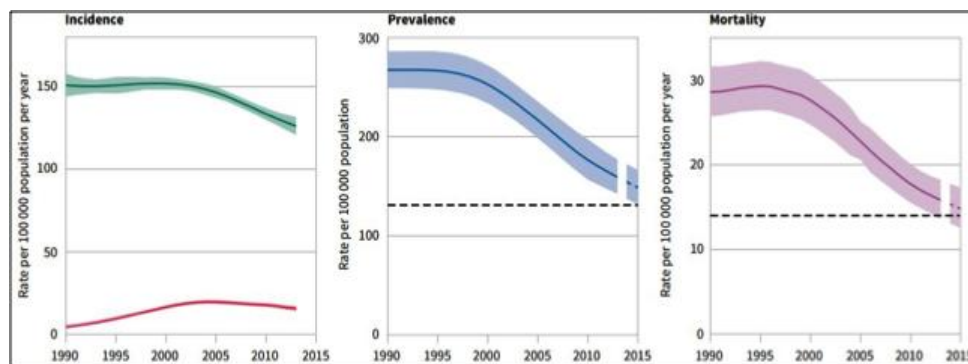
Figure 1.1. Estimated rates of TB incidence, prevalence and mortality (1990-2015). Left: Global incidence rate including HIV-positive TB (green) and estimated incidence rate of HIV-positive TB (red). Centre and right: TB prevalence and mortality rates 1990–2013 and forecast TB prevalence and mortality rates 2014–2015. The horizontal dashed lines represent the Stop TB Partnership targets of a 50% reduction in prevalence and mortality rates by 2015 compared with 1990. Shaded areas represent uncertainty bands. Mortality excludes TB deaths among HIV-positive people. Taken from [7], rights of noncommercial reprint for academic purposes obtained from http://www.who.int/licensing/reprints/en/.

factors associated with poverty all increase both the probability of becoming infected and the probability of developing clinical disease. Together, poverty and the tubercle bacillus create a vicious cycle: poor people go hungry and live in close, unhygienic quarters where TB flourishes; TB decreases people's capacity to work, and adds treatment expenses, exacerbating their poverty. Meanwhile, the poor receive inadequate health care, preventing even the diagnosis of their TB. Treatment, if received at all, is often erratic or simply incorrect. The poor are also less likely to seek and receive care from medical practitioners when ill, and are two- to three-times more likely than other income groups to self-medicate [16]. Self-medication encourages the emergence of drug-resistant TB strains [32], which might lead one to expect higher rates of drug resistant (DR) TB in developing countries. This assumption, however, is not fully supported by epidemiological evidence, which suggests that other factors are involved in the emergence and spread of DR TB.

1.3.2. Multidrug resistant tuberculosis is associated with disease relapse

The current standard 'directly observed treatment, short-course' (DOTS) treatment for pulmonary TB consists of isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), and ethambutol (EMB) for 2 months, then INH and RIF alone for a further 4-7 months depending on dosage [33]. Clinically, MDR-TB is defined as TB that is resistant to at least isoniazid (INH) and rifampin (RIF) [34]; strains further resistant to any member of the quinolone family and at least one of the second-line anti-TB injectable drugs (kanamycin, capreomycin, or amikacin) are classified as XDR-TB [7]. Obviously, TDR-TB strains are resistant to all tested first- and second-line drugs [14]. Globally, levels of MDR-TB among new TB cases had remained steady at 3.5% (95%

confidence interval, CI: 2.2–4.7%) for the last decade, while levels among previously treated cases, however, had been steadily increasing and reached 21% (95%CI: 14–28%) in 2013 [7]. This makes TB relapse, either re-infections or reactivation of latent infections, as the single largest risk factor for MDR-TB worldwide [35, 36]. Breaking this down regionally, we note that the African Region composed of mostly developing countries and possessing high incidences of TB, does not have higher rates of MDR-TB among either new or re-treated TB cases (**Table 1.1**). The surprising outlier is the European Region, in which 14% of new cases and 44% of previously treated cases had MDR-TB. This number is skewed by contributions from ex-Soviet bloc countries (e.g., Russian Federation- new: 19%, retreated: 49%; Belarus- new: 35%, retreated: 55%; Estonia- new: 17%, retreated: 48%; Kazakhstan- new: 25%, retreated: 55%) [7]. Since per capita gross domestic product (GDP) in 2013 is US$22,267 in the Eurasian Economic Union and US$2,320 in Africa, economics is not the sole determinant of MDR-TB[1] [37].

Table 1.1 **Estimated proportions of TB cases that have MDR-TB in WHO regions around the world in 2013**				
WHO region	New TB cases with MDR-TB (%)	95% confidence interval	Retreatment TB cases with MDR-TB (%)	95% confidence interval
Africa	2.4	0.2–5.0	13	0.02–27
Americas	2.2	1.3–3.0	13	4.9–22
Eastern Mediterranean	3.6	2.3–5.0	22	12–32
Europe	14	9.7–19	44	36–52
Southeast Asia	2.2	1.8–2.7	16	12–20
Western Pacific	4.4	2.6–6.3	22	18–26
Global	**3.5**	**2.2–4.7**	**21**	**14–28**

Source: World Health Organization Global TB Programme Information Resource Centre. **www.who.int/tb/data***. Retrieved 9th January 2015.*

---

[1] Further discussion on the role of national healthcare policies on limiting the development of DR TB is not within the purview of this review.

1.3.3. Development of antibiotic resistance in TB relapse reflects disease pathology

So why is TB relapse indicative of the development of MDR-TB? Our first clues come from the pathology of the disease itself. Inhalation of infected droplets expelled from the lungs is the usual route of TB infection, although ingestion, particularly via contaminated milk or water, also occurs [38]. Intrauterine, coital and intravenous routes of infection are less common [39, 40]. Inhaled bacilli are phagocytosed by alveolar macrophages that may either clear the infection or allow the mycobacteria to proliferate. In the latter instance, a primary focus may form, mediated by cytokines, notably tumour necrosis factor alpha (TNF-α), interferon gamma (IFN-γ) and interleukin 4 and 12 (IL-4/12) [41, 42], associated with a delayed hypersensitivity reaction that consists of dead and degenerate macrophages surrounded by epithelioid cells, granulocytes, lymphocytes, and later, multinucleated giant cells. The purulent to caseous, necrotic center may calcify, and the lesion may become surrounded by granulation tissue and a fibrous capsule to form the classic tubercle or granuloma. The primary focus, plus similar lesions formed in the regional lymph node, is known as the "primary complex." In alimentary forms of disease, the primary focus may be found in the pharynx or mesenteric lymph nodes or, less commonly, in the tonsils or intestines [43]. The cellular composition of and presence of acid-fast bacilli in tuberculous lesions differs between granulomatous lesions even within the same host [44, 45].

Established *Mtb* infections usually persist for the entire lifespan of the host and cannot be cleared without therapeutic intervention. Although a significant proportion of lesions remains unhealed in hosts, most infections (>90%) are latent [43]. However, in about 10% of individuals, the bacteria reactivates and rapidly multiplies and spreads to develop active TB [24]. Dissemination

through vascular and lymphatic channels may be generalized and rapidly fatal, as in acute miliary TB. Nodular lesions may form in many organs, including the pleura, peritoneum, liver, kidney, spleen, skeleton, mammary glands, reproductive tract, and central nervous system [46]. These potential life-threatening forms of reactivated disease accounts for 47–87% of active TB disease in low-prevalence regions, such as in the United States [47-49].

Throughout the course of the infection, *Mtb* is exposed to a range of microenvironments that induce compensatory metabolic pathways and physiological states. Prominent among these is the non-replicating persistent (NRP) or "dormant" state. Within the lung, regional differences exist in ventilation and perfusion, and in the degree of blood oxygenation. In a seminal study using resected lung tissue, lesions classified as 'open' (oxygen rich) were found to contain bacilli that grew readily on nutrient agar and a majority (~80%) produced cultures that were resistant to the drugs streptomycin (STM), 4-aminosalicylic acid (PAS) or INH, with which the patients had been treated. However, bacilli isolated from 'closed' (oxygen poor) lesions showed delayed growth (up to 4 months from 4-8 weeks) and were mostly drug sensitive upon resuscitation (83%) [50-52]. In agreement, recent studies in macaques demonstrate that granulomas whether non-necrotising, fibrocalcific or caseous, possessed varying bacterial numbers, even within the same individual. In these monkeys, the ability of the host immunity to sterilize infecting bacilli remains unchanged in both active and latent manifestations of the disease; however, killing efficiency is diminished in caseous granulomas and TB pneumonia resulting in residue bacterial growths in these regions [53]. Together, these results suggests that there are two aspects to drug resistance in *Mtb*, one acquired through genetics, dependent on the selection of mutant strains with increased fitness when

exposed to antibiotics and to the host immunity – as observed in the readily cultured bacilli found in 'open' lung lesions, and the second innate, determined by transient adaptations to the varied microenvironments of different granulomas – as observed in the dormant mycobacteria found in 'closed' lung lesions.

1.3.4. Acquired antibiotic resistance

Antibiotic resistance can emerge by spontaneous mutations in chromosomal genes or gained through the acquisition of genes by conjugation, transduction or transformation. Unlike other microorganisms, acquired drug resistance does not occur by horizontal gene transfer, since *Mtb* lacks plasmids and efficient genomic DNA transfer mechanisms – outside of mycobacterial phages – have not been described. Thus, acquired drug resistance most likely arises from spontaneous mutations in specific target genes rendering the bacteria resistant to a given drug [54].

These mutations could either occur within protein coding sequences, which would affect protein function or activity, or within promoter regions, which would affect gene expression. Recent studies on the mutation rates of *Mtb* clinical isolates in humans and non-human primates suggest that rates of spontaneous mutations are higher than expected; and though modulated by host environments these rates do not differ between active and latent infections [55-57]. Under selection pressure from antibiotics, transient mutator strains could readily gain the necessary mutations required to develop MDR-TB. Some of the most effective mutations (at least from the perspective of Mtb) are single-nucleotide polymorphism (SNPs). For instance, INH resistance in clinical MDR-TB strains is most frequently attributed to non-synonymous point mutations in *katG* catalase protein coding sequences,

which results in S315T mutants [58-60]. Another common function SNP involves 15C→T substitution in the promoter region for NADH-dependent enoyl-ACP reductase (*inhA*) [61-64]. INH is a prodrug and must be activated by KatG, which involves coupling INH with NADH to form isonicotinic acyl-NADH complexes. These complexes bind tightly to InhA, thereby blocking the natural enoyl-AcpM substrate and its fatty acid synthase activity. Hence, mycolic acid synthesis – an essential process for cell wall formation – is inhibited by INH through its activation by KatG [64]. S315T mutations decrease catalase-peroxidase activity of KatG, resulting in a deficiency in the formation of INH-NAD adducts [65], while15C→T mutations in *inhA* promoters cause InhA over-expression which helps the mutant overcome INH-NAD inhibition [64]. When exposed to non-sterlizing doses of INH, these mutants survive and proliferate, which spreads INH-resistant *Mtb* [63]. Resistance to RIF, PZA, STM, EMB and other second line drugs can arise through similar mechanisms (**Table 1.2**).

1.3.5. Innate antibiotic tolerance or phenotypic drug resistance

Not all cases of drug resistance in clinical isolates can be explained by the presence of stable gene mutations. The mutations in genes associated with INH and RIF mutants, for instance, are found in ~30-86% of INH-resistant and ~5-69% of RIF-resistant in *Mtb* clinical isolates [66, 83]. This suggests that other mechanisms of drug resistance exist. Intrinsic drug resistance in *Mtb* has been attributed to a combination of a highly impermeable mycolic acid containing cell wall and active drug efflux mechanisms [84-87]. ATP-binding cassettes (ABC), for example, make up ~2.5% of the *Mtb* genome [88]; several of which contribute to tolerance to β-lactams [89], novobiocins, pyrazolones, biarylpiperazines, bisanilinopyrimidines, pyrroles and pyridones

Table 1.2. Genes associated with acquired drug resistance

Drug[a]	Gene[b]	Function of gene product
INH	*katG*	Catalase/peroxidase
	inhA promoter	Enoyl-ACP reductase
	ahpC	Alkyl hydroperoxide reductase
RIF	*rpoB*	RNA polymerase β-subunit
PZA	*pncA*	PZase
STM	*rpsL*	Ribosomal protein S12
	rrs	16S rRNA
	gidB	RNA 7-methylguanosine methyltransferase
EMB	*embB*	Arabinosyl transferase
Fluoroquinolones	*gyrAB*	DNA gyrase
KAN,AMK	*rrs*	16S rRNA
	eis promoter	GCN5-related N-acetyltransferase
AMK,CAP,VIO	*tlyA*	rRNA methyltransferase
ETH	*inhA* promoter	Enoyl-ACP reductase
PAS	*thyA*	Thymidylate synthase A
PA-824	*ddn*	Deazaflavin-dependent nitroreductase
TM207	*atpE*	ATP synthase

[a] AMK, amikacin; CAP, capreomycin; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; PA-824, pretomanid; PAS*, p*-amino salicylic acid; RIF, rifampin; STM, streptomycin; TM207; bedaquiline; VIO, viomycin.
[b] Collated from literature searches [58, 63, 66-82] and the TB drug resistance database https://tbdreamdb.ki.se/Info/.

[90]. **Table 1.3** lists the known drug efflux pumps in *Mtb*. Together they form a formidable pharmacokinetic barrier preventing drug uptake and facilitating drug efflux.

Since caseous granulomas and inflammatory lung tissues in TB pneumonia are hypoxic, it should come as no surprise that *in vitro* studies show tha*t* under hypoxia *Mtb* becomes refractory to antibiotic killing [111, 112]. High O_2 tension exists in the upper lung, whereas the ventral lung experiences low O_2 tension. Consistent with anatomy and function, the partial O_2 pressure (pO_2) of atmospheric O_2 (150–160 mmHg) drops from ~150 to ~60 mm Hg from the trachea to the bronchioles [113, 114]. Across the alveoli, the diffusion distance is ~100–200 μm, resulting in a rapid diffusion to the blood vessels [114, 115]. Using redox-active dyes that are reduced at pO_2 lower than 10 mm Hg, studies in guinea pigs and non-human primates have shown that the granulomas are indeed hypoxic (~1. 9 mm Hg) [45, 116]. Thus, hypoxic exposures within inflammatory and granulatomous tissues is closely associated with an indifference to antibiotic killing.

The hypoxic stress response of *Mtb* is well defined. Part of this response involves a dynamic shift in cellular metabolism. A consensus from studies of gene expression of *Mtb* in infected primary human macrophages, mouse bone marrow derived macrophages, rodents models and *in vitro* culture systems leads to a model involving an upregulation of genes involved in the glyoxylate cycle, gluconeogenesis and fatty acid metabolism, including isocitrate lyases (*icl*, *aceA1*), citrate synthase (*gltA*), PEP carboxykinase (*pckA*), pyruvate phosphate dikinase (*ppdK*), fructose-bisphosphate aldolase (*fba*), acyl CoA dehydrogenases (*fadE*) and enoyl CoA hydratase (*echA*), as

Table 1.3. Known drug efflux pumps in *Mtb*.

Genes[a]	Drug extruded[b]	Transporter family[c]
Rv0194	STM	ABC
IniABC (Rv0341-Rv0343)	INH, EMB	Membrane transporter
MmpL5-MmpS5 (Rv0676c-Rv0677c)	Azoles, TET	RND
EmrB (Rv0783)	RIF	MFS
Rv0849	β-Lactams	MFS
PstB (Rv0933)	INH, RIF, EMB, CIP	ABC
Rv1218c	β-Lactams	ABC
Rv1258c	INH, RIF, EMB, OFL, β-lactams	MFS
P55 (Rv1410c)	STM,INH, RIF	MFS
Rv1634	Fluoroquinolones	MFS
Rv1747	INH	ABC
BacA (Rv1819c)	Bleomycin	ABC
Rv1877	TET, KAN, erythromycin	MFS
Rv2333c	TET, spectinomycin	MFS
Rv2459	INH, EMB	MFS
Rv2686c-Rv2888c	CIP	ABC
DrrABC (Rv2936-Rv2938)	TET, STM, EMB, RIF	ABC
EfpA (Rv2846c)	INH, ETH	MFS
MmpL7 (Rv2942)	INH	RND
Mmr (Rv3065)	Erythromycin, β-lactams	SMR

[a] Collated from literature searches [89-101] and genomic annotation comparisons from http://tuberculist.epfl.ch/; http://www.tbdb.org/ and http://www.genome.jp/kegg/.
[b] CIP, ciprofloxacin; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; OFL, ofloxacin; RIF, rifampin; STM, streptomycin; TET, tetracycline.
[c] ABC, ATP-binding cassette; MFS, major facilitator superfamily; SMR, small multidrug resistance family; RND, resistance/nodulation/cell division family.

well as the genes for uptake of glycerol-3-phosphate (*ugp* and *glpD*) and cholesterol (*mce4* operon). There is also down-regulation of key cytochrome complexes in the electron transport chain (ETC) [117-121]. The net effect: a shift towards lipotropy and a five-fold reduction in steady-state ATP levels in dormant *Mtb* compared to their replicating counterparts [107, 122]. A unique feature of this lipolytic shift is the conservation of carbon through the cycling of acetyl-CoA, from even-chain fatty acid β-oxidation, to glyoxylate and succinate by the glyoxylate shunt [123, 124]. The former is condensed with a molecule of acetyl CoA to form oxaloacetate that can be further converted to phosphoenolpyruvate, which allows further reductive carboxylation to take place [125]. The latter is coupled to ATP synthesis and membrane potential maintenance, respectively, through the interaction of succinate dehydrogenase (SDH) with the ETC and by it accumulation within hypoxic cells [124]. Isocitrate lyase, in turn, is required for *Mtb* persistence in mouse models [126, 127], highlighting its importance in dormant mycobacteria. Though the precise mechanism is unclear, these studies suggest that altered lipid utilization is closely associated with the antibiotic tolerance of dormant mycobacteria. Further characterization of the close links between lipid catabolism and antibiotic resistance is presented in Chapter 5 of this thesis, wherein we highlight the consequences of this catabolic shift and its effects on phenotypic drug resistance in mycobacterial persisters.

1.3.6. Regulation of stress responses in *Mtb*

Mtb contains a plethora of regulatory systems controlling growth, physiology and metabolism during stress. These include serine/threonine protein kinases (STPKs – 11 identified) that are coupled to sensors of environmental signals and mediate host-pathogen interactions, cell division and developmental changes [128, 129]; WhiB proteins (7 identified) that bind to iron-sulfur

clusters and DNA, thus coupling transcription to intracellular redox potential [130, 131]; guanosine pentaphosphate alarmone signaling (Rel_{Mtb})and toxin-antitoxin systems (~80 pairs identified) best known their role in bacterial stringent responses [132, 133] (further discussed in Chapter 5); sigma factors (13 identified) regulating gene expression under stress [134-136]; and two-component systems (11 identified) that use phosphorelay to "sense" the environment and induce the expression of gene regulons necessary for survival within host cells [137-141].

The coordination of stress responses is a complex process, involving multiple dynamic interactions between regulatory systems, though the full extent of these interactions, their triggers and effectors are still unknown. For hypoxic stress, it is generally accepted that the DosRS/T two-component system plays a prominent role mediating adaptations to O_2-limited environments by influencing the reversible shifts between aerobic and anaerobic respiration [117, 120, 142-146]. Recent work on Ser/Thr Protein Kinase B (PknB), however, details its DosR-independent activity in resuming growth after hypoxic exposures [147]. Nonetheless, the dormancy survival (Dos) regulon, a set of 48 genes controlled by DosRS/T [117, 144, 148, 149], remains the best characterized regulatory network in mycobacteria.

First identified in a screen for genes differentially expressed (DevSR) in the virulent lab adapted strain of *Mtb* (H37Rv) compared to the avirulent H37Ra strain [150], these proteins were found to be important in dormancy survival (Dos) and alternatively named DosR and DosS [151, 152]. In this system, DosS and DosT are heme-containing, redox- and hypoxia-sensing histidine kinases that activate DosR, a transcription factor, which induces transcription of Dos regulon genes and small non-coding RNAs (ncRNA), as well as activating STPK [153-156]. In addition to hypoxia, nitric oxide (NO) produced

by activated macrophages [120], and carbon monoxide (CO) from pro-inflammatory responses in lung lesions [157, 158], have also been demonstrated to activate the Dos regulon. Extended exposure to hypoxia further induces a larger enduring hypoxia response leading to dormancy [145]. During hypoxia, one of the most highly upregulated genes in the regulon is *hspX* which encodes an alpha-crystallin (Acr) heat shock chaperon protein that is strongly associated with growth arrest *in vitro* [144, 159, 160] (**Fig. 1.2**). The significance of this regulon for survival within hosts has been demonstrated in rabbit and guinea pig models of TB – both displaying a human-like histopathology and hypoxic lesions – and in humans wherein Dos regulon proteins are found to be expressed in latent infections [161-164].

1.4. Models for the study of *Mtb* dormancy and persistence *in vitro*

Much of our knowledge about dormant *Mtb* comes from the analysis of differentially expressed transcripts upon stress exposures [165-168]. Variation in transcript levels, however, only shows 30-40% correlation with protein abundance, and thus, fails to fully explain complex biological phenomena such as Mtb dormancy and persistence [169-171]. Since variation in underlying transcript levels cannot account for the majority of variation observed in the corresponding protein levels our knowledge of the control of mycobacterial stress responses is incomplete without a thorough understanding of post-transcriptional process that regulate gene expression. For instance, it is assumed, based on transcript levels, that DosR plays a prominent role in the switch towards, but not the maintenance of, non-replicating persistence [145]. However, in Chapter 6 we demonstrate that while *dosR* transcript levels fall after the establishment of the NRP state, DosR protein levels remain elevated until resuscitation (by re-aeration). This discovery and the others detailed in the latter sections of this thesis were

facilitated through the use of deterministic models that mimic physiological

stresses *in vitro*.
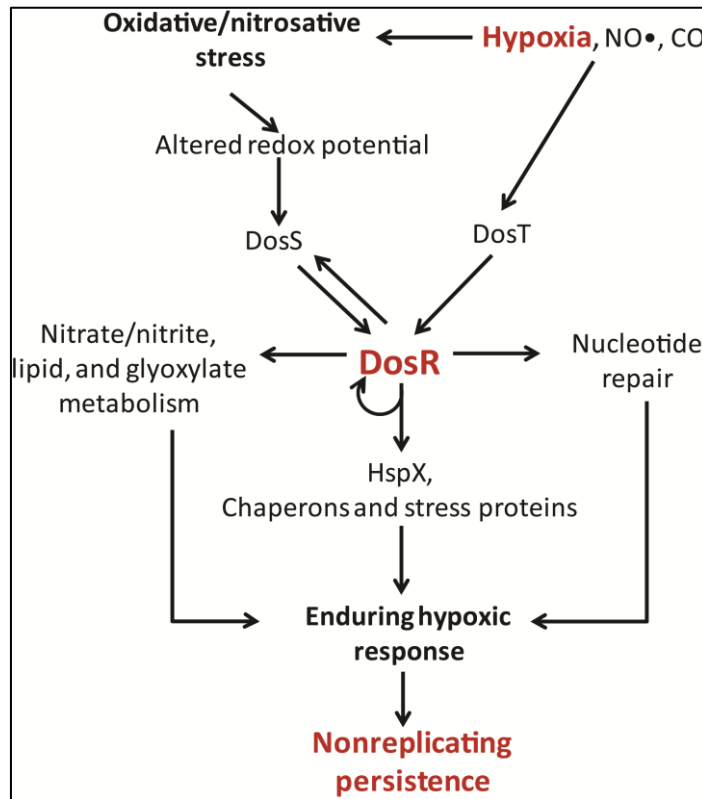
Figure 1.2. Regulation of hypoxia-induced nonreplicating persistence by DosR. Under hypoxic conditions, DosS and DosT are sensor histidine kinases that activate DosR. Induction of DosR upregulates the dos regulon (which dosR and dosS are members of), leading to changes in cellular metabolism, protein homeostasis and nucleotide repair. A subsequent enduring hypoxic response is induced, enabling the cell to enter nonreplicating persistence.

1.4.1. The Wayne model for hypoxia-induced dormancy

An ideal model for *Mtb* dormancy should be tractable and biologically

relevant. The Wayne model and its variants seek to imitate the hypoxic

conditions faced by *Mtb* in closed, necrotic lesions by inducing gradual

depletion of oxygen within a sealed, stirred culture tube of specific dimensions

[172]. The model starts with aerobic cultures in balanced growth. Over time,

oxygen is consumed and, when dissolved oxygen falls below 1% in culture,

the bacilli enters the microaerophilic non-replicating phase (NRP1). This phase is defined by a slow increase in turbidity without a corresponding increase in cell numbers (determined by colony forming units, CFU) or cellular DNA. The bacilli remain viable, steadily producing ATP while regenerating NAD through the activity of glycine dehydrogenase [173, 174]. As time progresses, dissolved oxygen drops below 0.06%, and now, the bacilli enters a second non-replicating phase phase (NRP2). Under these anaerobic conditions, steady-state ATP production glycine dehydrogenase activity decreases while culture turbidity and viability remain unchanged. Evidence for the clinical and therapeutic relevance of these Wayne bacilli comes from their selective susceptibility to metronidazole, a drug that damages bacterial DNA only under anaerobiosis [175-177]. Aerobically growing bacilli are not affected by the drug [172, 178]. Studies in rabbits – a TB animal model presenting hypoxic lesions – showed that metronidazole significantly reduces the number of bacilli in the lung [45]. These results are consistent with metronidazole studies in macques and human patients, with the latter showing increased efficacy when used against INH- and RIF-resistant *Mtb* [179, 180]. This is the basis for the development of nitroimidazoles – a new class of TB antibiotics of which nitrodihydroimidazooxazole (Delamanid, OPC-67683; Otsuka Pharmaceutical Co., Ltd.) and nitroimidazoloxazine (Pretomanid, PA-824; TB Alliance) are currently in phase III and II trials, respectively [181].

1.4.2. Nutrient deprivation models for *Mtb* persistence

Hypoxia induces a reversal in the susceptibility of *Mtb* to INH and metronidazole as the bacilli switch between replicating and non-replicating states. This is similar to type I *Escherichia coli* persisters formed in response to external triggers such as starvation (rather than "type II", which are produced by spontaneous switching without environmental triggers and

stressors) [182]. First proposed by Imlay and later verified through single-cell observations [183-185], this theory suggests that bacterial populations are heterogeneous (containing both persisters and non-persisters), with physiological stresses in the environment pushing a greater part of the population to become dormant persisters [186]. These dormant bacteria are thought to be the basis for phenotypic antibiotic tolerance [18].

A shift towards lipid catabolism is detected in *Mtb* recovered from granulomas [187-189]. This suggests that the availability of lipids for nutrition could be an environmental cue for phenotype switching. Lipids required for survival early during infections are likely obtained from foamy macrophages found within granulomas [190, 191]. Direct observation of the interior of *Mtb*-containing granulomatous lesions showed that bacilli could be found in close proximity to cholesterol and triacylglycerol rich intracellular lipid droplets within foam cells [187]. Further studies also showed that phagocytosed bacteria preferentially utilize host lipids as carbon sources [151, 190, 192, 193]. But what happens when this nutrient source runs out?

Indeed, there is evidence to suggest that non-replicating mycobacteria recovered in resected lung lesions are type I persisters induced by nutrient starvation. *Mtb* isolated from lung lesions were found to display altered morphology and staining properties that are remarkably similar to those observed in mycobacteria left to survive in distilled water [194, 195]. When starved in phosphate-saline buffers, *Mtb* stopped growing, gained tolerance to several antibiotics, and retained their virulence [112, 189, 196]. Restricted mycobacterial growth was also observed under phosphate-limiting conditions and the persisters are phenotypically resistance to INH [197]. Nutrient starvation also induces large shifts in transcriptional activities that share many common features with the transcriptional reprogramming observed in Wayne

bacilli [192]. In Chapter 5, we revisited the pioneering work of Loebel [198, 199], developed a nutrient starvation model to characterize starved mycobacteria at the molecular (transcriptomic and metabolic) and cellular (single-cell and population) levels, and found that mycobacterial persisters possess a selective but reversible vulnerability to hydrogen peroxide (H_2O_2) – a phenomenon that is reminiscent of the hypoxia-induced susceptibility to metronidazole in dormant *Mtb*.

1.5. Control of gene expression in response to stress

Adaptation of gene expression through the regulation of transcription and translation are key mechanisms in microbial responses to fluctuating environmental conditions. For pathogens of the *Mtb* complex, stresses in the microenvironment induced by host immune responses cause the activation of a variety of regulatory mechanisms corresponding to their inducers – oxidative, nitrosative and acid pH stresses and macro- and micro-nutrient depletion [19, 120, 199-202]. In parallel, this evokes cellular survival and evasion responses to cell-mediated immunity and resistance to reactive oxygen species (ROS), reactive nitrogen species (RNS) and therapeutic agents [117, 196]. Though mRNA copy number is often used as a proxy for gene expression [203], transcription, translation and degradation processes (for both RNA and protein) can be extensively coupled and may frequently regulate each other through feedback loops [204]. This is highlighted in several time-course studies of bacterial stress responses wherein large differences between protein and mRNA abundance changes were detected [205, 206]. Clearly, our understanding of perturbed systems is still incomplete.

Alternate transcriptional start sites, alternative splicing, RNA editing, RNA interference, stop codon read-through, protein degradation and modulation of

enzyme function and activities are just some of the post-transcriptional and post-translational processes that impact gene expression and cellular homeostasis [207-211]. Yet, these processes only explain cell phenotypes in a piecemeal manner. Recently, however, a new paradigm for the translational control of cellular stress responses was developed. Based on the response of *S. cerevisiae* to cytotoxins, this paradigm unifies longstanding, but unexplained, observations in tRNA modifications and codon usage bias to redefine translational control as a function of tRNA reprogramming [20, 21].

1.5.1. RNA modifications – a well characterized but poorly understood aspect of molecular biology

All canonical ribonucleosides (adenosine, A; uridine, U; guanosine, G; and cytosine, C) can be modified enzymatically *in vivo*. These modifications range from simple alkylation, isomerizations, sulfonation, oxidation, and deamination to complex linkages to glycosides and amino acids [212-215]. Though biosynthesis pathways and enzyme mechanisms for many of these modifications are unknown, several of these modifications are catalyzed by processes that involve multiple enzymes and reactions [215-217]. Currently, there are more than 120 known post-transcriptional structural modifications in all RNA species across all living organisms and, while most commonly found in transfer RNA (tRNA), a significant number of modifications are also observed in ribosomal RNA (rRNA) and even mRNA [218-220]. Not only are modified nucleosides contained in tRNA from all three phylogenetic domains -- Archea, Bacteria and Eukarya [221] – eight of them (pseudouridine, Ψ; dihydropseudouridine, D; N^4-acetylcytidine, ac^{4}C, (cyclic)N^6-threonylcarbamoyladenosine, (c)t^{6}A; 5-methyluridine, m^{5}U; 1-methylguanosine, m^{1}G; 1-methyladenosine, m^{1}A; and 2'-*O*-methylcytidine, Cm. **Fig 1.3**) are present in the same position and in the same subpopulation

of tRNAs in organisms from all domains [222-224]. Furthermore, recent theoretical and experimental studies have endeavored to derive the minimal set of genes that are necessary and sufficient to sustain a functioning cell, with the repeated finding that several tRNA modifying enzymes are essential for life [225-228]. Surprisingly, despite their conservation and profusion in biological systems, the precise function of these modifications has eluded understanding, a problem that is compounded by a paucity of knowledge about the identities of RNA modifications in important pathogens such as *Mtb*.
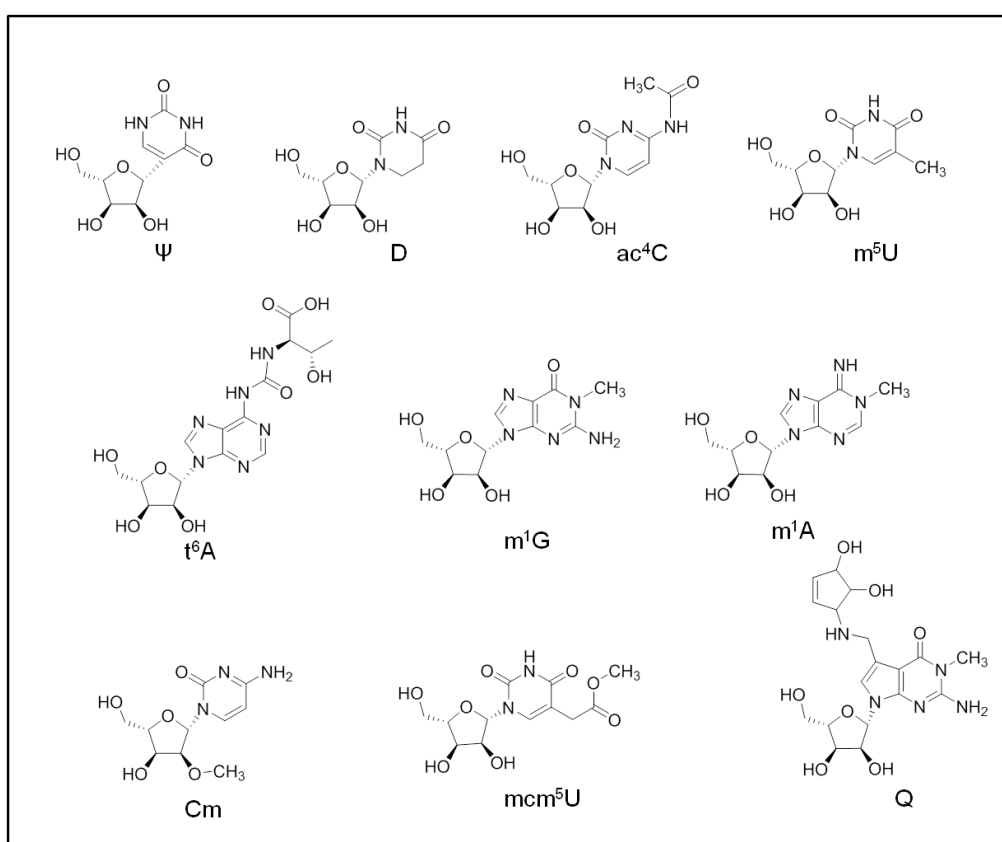
Figure 1.3. Chemical structures of conserved tRNA modifications. Modifications derived from all 4 canonical ribonucleosides are shown. Pseudouridine, Ψ; dihydropseudouridine, D; *N4*-acetylcytidine, ac⁴C; (cyclic)*N6*-threonylcarbamoyladenosine, (c)t⁶A; 5-methyluridine, m⁵U; 1-methylguanosine, m¹G; 1-methyladenosine, m¹A; and 2'-O-methylcytidine, Cm are common to all three kingdoms of life. 5-methoxycarbonyl-methyluridine (mcm⁵U) and queuosine (Q) are wobble modifications thought to be conserved in eukaryotes and most prokaryotes respectively; their functions are discussed in text.

To date, there has been no complete definition of the spectrum of RNA modifications in any microorganism, with the possible exception of *E. coli* and *S.cerevisiae* (**Table 1.4**).

1.5.2. Known functions of tRNA modifications

Modified ribonucleosides are found in virtually all coding and non-coding RNA species [229-231], but they are most abundant in tRNA, the adaptor biomolecule serving as the physical link between genetic information coded in mRNA transcripts and amino acids. tRNAs are typically 75 to 90 nucleotides (nt) long and on average, each contain 8 modifications [232]. These modifications can occur at virtually any position, but are most consistently found at positions 34 – the wobble position – and position 37 (**Fig. 1.4**). Early observations show that these modifications can modulate base pairing and pi-stacking, thus, stabilizing secondary and tertiary RNA structures [234, 235]. Their presence also serves as a tRNA maturation marker [236, 237]. For instance, wybutosine (yW) only replaces m^{1}G at position 37 of tRNA^{Phe} in *S. cerevisae* after intron splicing taken place [238]. Hypomodified tRNA would represent misfolded tRNA or aberrant splicing events, allowing the cell to target them for degradation [239]. *In vitro* experiments with unmodified tRNAs caused translational pausing and increased translational errors suggesting that modifications also help maintain reading frames at the ribosomal A site and reduce the rates of translational frameshift mutations that might occur through tRNA slippage [240, 241]. Most importantly, tRNA modifications are essential for accurately decoding the genome, through both the expansion and restriction of genetic decoding by codon recognition [242, 243]. tRNA modifications a play a key role in the wobble hypothesis by expanding possible codon-anticodon pairing [242, 244]. These wobble base pairings allows organisms to use less and 61 codons to translate mRNAs. For

Table 1.4. Modified ribonucleosides in total tRNA*

Ribonucleoside	Symbol	E. coli	S. cerevisiae
Pseudouridiine	Ψ	■	■
dihydrouridine	D	■	■
5-methylcytidine	m^5C	■	■
2'-O-methylcytidine	Cm	■	■
3-methylcytidine	m^3C		■
5-methyluridine	m^5U	■	■
2-thiocytidine	s^2C	■	
4-thiouridine	s^4U	■	■
inosine	I	■	■
1-methyladenosine	m^2A	■	
N6-methyladenosine	m^6A	■	■
1-methyladenosine	m^1A		■
2'-O-methyladenosine	Am	■	■
1-methylinosine	m^1I		■
N4-acetylcytidine	ac^4C	■	■
5-methylaminomethyluridine	mnm^5U	■	
7-methylguanosine	m^7G	■	■
1-methylguanosine	m^1G	■	■
2'-O-methylguanosine	Gm	■	■
N2-methylguanosine	m^2G		■
5-carbamoylmethyluridine	ncm^5U		■
5-methylaminomethyl-2-thiouridine	mnm^5s^2U	■	
N2,N2-dimethylguanosine	m^2_2G		■
5-methoxycarbonylmethyluridine	mcm^5U		■
uridine 5-oxyacetic acid	cmo^5U	■	
5-methoxycarbonylmethyl-2-thiouridine	mcm^5s^2U		■
uridine 5-oxyacetic acid methyl ester	$mcmo^5U$	■	
N6-isopentenyladenosine	i^6A	■	■
3-(3-amino-3-carboxypropyl)uridine	acp^3U	■	■
lysidine	k^2C	■	
2-methylthio-N6-isopentenyladenosine	$ms^{2}i^6A$	■	■
cyclic N6-threonylcarbamoyladenosine	ct^6A	■	
queuosine	Q	■	■
N6-threonylcarbamoyladenosine	t^6A	■	■
epoxyqueuosine	oQ	■	
N6-methyl-N6-threonylcarbamoyladenosine	m^6t^6A	■	■
wybutosine	yW		■
Total		27	24

* Shaded (grey) and unshaded cells represent presence or absence of the modified ribonucleoside respectively. Based on tRNA modification studies of Suzuki and coworkers [217, 233]. And cross validated with tRNA sequences catalogued on Modomics (http://modomics.genesilico.pl/).

example, wobble 5-oxyacetyl-uridine (cmo^{5}U) allows for the decoding of G-ending codons [245, 246]. This property, as we demonstrate Chapter 6, allows for the selective translation of hypoxia-induced transcripts in mycobacteria entering dormancy.

1.5.3. Translation control of stress responses by tRNA modifications

tRNA modifications are not static. There is solid evidence to indicate that their abundance changes dynamically in response to stress. For example, the levels of pseudouridines in tRNA are induced by nutrient deprivation while 2-thiouridine (s2U) and 2-thiouridine (s4U) content decreases upon cyanogen bromide and near-UV exposures respectively [247-249]. The virulence of fungi pathogens has been shown to be linked to 7-methylguanosine (m^{7}G46), a modification found on position 46 in the hypervariable loop of tRNA in eurkaryotes and prokaryotes (**Fig. 1.4**). Mutants lacking this crucial modification grow as well as their wild-type counterparts in culture, but fail to invade their natural plants host. Loss of virulence was due to a lack of adaptive stress responses to oxidative and osmotic stresses [250]. Later, it was determined that m^{7}G46 modifications support the introduction of wobble position modifications [251].

Wobble modifications, in particular, play an important role in cellular responses to stress. For example, queuosine (Q) (**Fig. 1.3**), a cyclopentadiol derivative of 7-cyano-7-deazaguanosine (pre-Q$_0$) synthesized in bacteria by 4 enzymes QueC-F, is post-transcriptionally exchanged with wobble guanine by tRNA-guanine transglycosylase (TGTase) [252-254]. Wobble Q has been implicated in a number of physiological processes such as cell proliferation [255, 256], oxidative metabolism [257], tyrosine biosynthesis [258], and bacterial virulence [259, 260]. While non-essential (Q-deficient HeLa cell
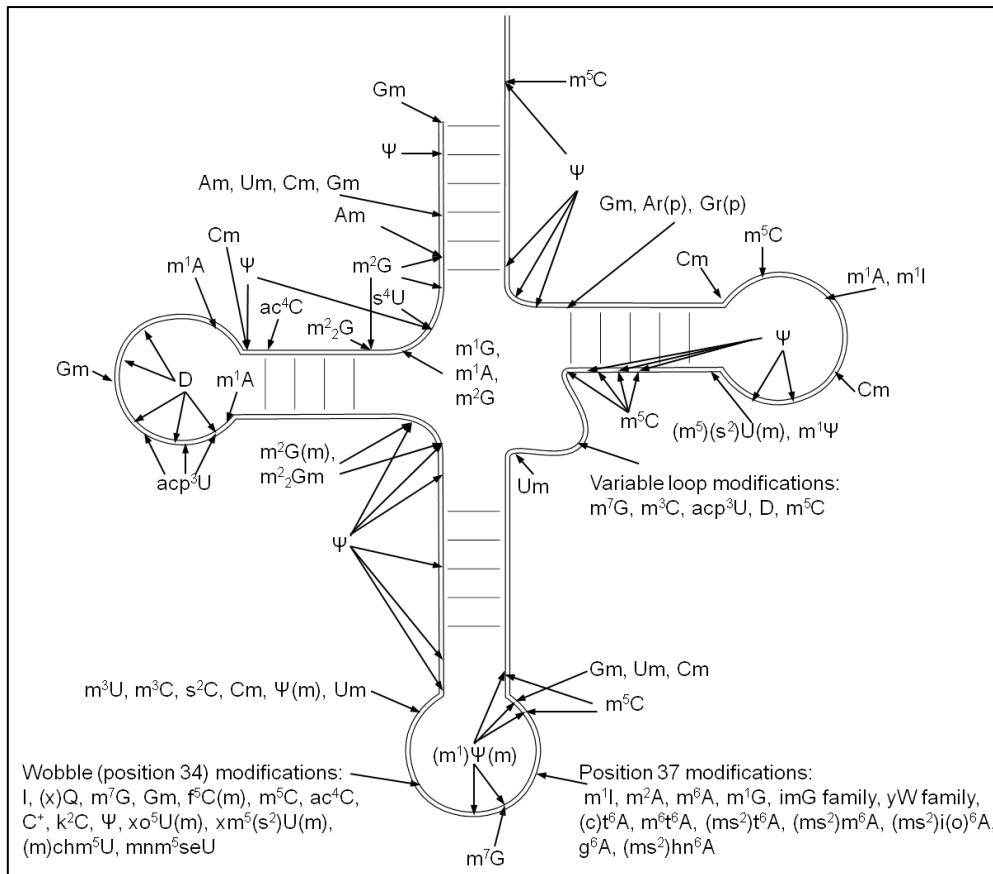
Figure 1.4. Distribution of modified nucleosides in tRNA. Representation of a tRNA showing the positions at which modifications in archea, bacteria and eukaryotes may be found. A variety of modifications had been identified at virtually all positions in tRNA. Positions 34 and 37 are most frequently hypermodified. Figure adapted from references [222, 224].

grows well under aerobic and hypoxic conditions), proliferation of HeLa cells improved under aerobic conditions but declined under hypoxia in the presence of queuine; suggesting a role for Q in the cellular response to hypoxia [256, 261]. The biological activity of Q-base is likely due to its influence on the coding capacity of tRNA [241]. tRNA[His] with guanidine at the wobble base (G34) clearly prefers the histidine codon CAC to the CAU, whereas tRNA[His] (Q34) has little preference for the codon CAC [262]. Furthermore, queuosine-modified tRNAs have been shown to suppress stop codons as tRNA[Tyr](G34) allows stop codon readthrough, whereas tRNA[Tyr](Q34) prevents it [263].

How does the decoding property of a modified ribonucleoside regulate cellular responses to stress? An emerging model, in which modification-dependent tRNA decoding activities are keyed to codon use in genetic programs, thereby allowing for the switching between genetic programs when cells are exposed to environmental stresses, provides insights into the translational control of stress responses. In *S. cerevisiae*, deletion of tRNA anticodon methyltransferase Trm9 increase sensitivity to DNA damage induced by alkylating stress. Trm9 is necessary for formation of wobble 5-methylcarbonylmethyluridine (mcm^{5}U34, **Fig. 1.3**), which is necessary to decode AGA and GAA codons. When treated with the alkylating cytotoxin methylmethansulfonate (MMS), levels of mcm^{5}U34 increase and, interestingly, the stress-induced transcripts that were required to rescue the cell had a large over-abundance of the codons AGA and GAA – the very same codons preferentially read by mcm^{5}U34 containing tRNAs [264]. Subsequent experiments performed at the systems-level show that, while every stress has its own characteristic tRNA modification signature, the general model remains unchanged [264]. Codon bias – the overabundance of specific codons and the deficiency of others – is innate in all organisms and is 'keyed' to the levels of modifications in a tRNA pool [265].

Though it is clear that limits to the amount of protein produced from mRNAs are determined by levels of the copy numbers cognate tRNAs [266, 267], this model predicts that tRNA levels could be regulated to match codon use by specific mRNAs as part of a genetic program. In this model, the tRNA pool is reprogrammed during stress, either through increased activity of tRNA modifying enzymes or through changes in the number of copies of a tRNA [20, 21, 237]. This altered tRNA pool results in the selective translation of mRNA transcripts enriched in a specific codon that are selectively decoded by

the reprogrammed tRNA [242]. This model also provides a biological role for codon usage bias, long shown to be non-random and evolutionarily driven by selection pressures [268], in the cellular responses to stress (**Fig. 1.5**).
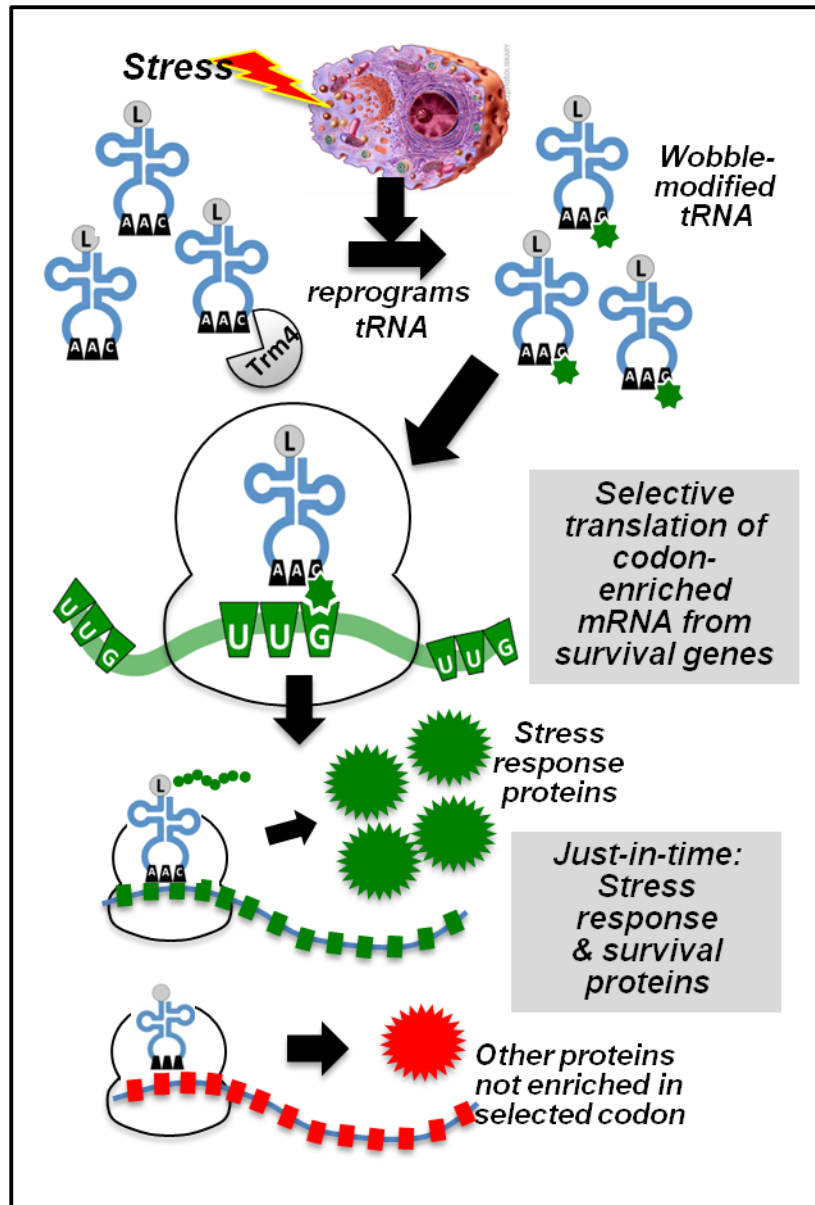
Figure 1.5. General model for the translational control of stress responses by tRNA modifications. Following stress, cells reprogram their tRNA pools enriching for tRNA species with specific wobble modifications. These wobble modified tRNAs bind to particular codons more efficiently than others driving the selective translation of transcripts enriched these codons. Here, for illustration, hydrogen peroxide (H_2O_2) exposures induces Trm4 activity in yeast which modifies tRNA$^{Leu(\underline{C}AA)}$ to tRNA$^{Leu(m5\underline{C}AA)}$. tRNA$^{Leu(m5\underline{C}AA)}$ preferentially decodes transcripts overusing Leu^{UUG} codons. As a corollary, mRNA transcripts lacking these selected codons are translated less efficiently. The result is a rapid, coordinated and just-in-time stress response allowing cells to survive otherwise cytotoxic stresses. Adapted from [21, 269].

Subsequent experiments performed at the systems-level, show that toxicants with distinct molecular mechanisms induce has its own characteristic tRNA modification signature, unique in a manner that enables the bias translation of specific survival proteins [265, 269]. The general model was further corroborated by numerous experiments showing that loss of specific tRNA modifying enzymes sensitizes yeast to the associated stress esposures [270, 271]. While these systems have hitherto been described in eukaryotes, the ubiquitous nature of its components led us to hypothesis that they are present in all organisms. In particular, these molecular mechanisms may play an important role in the events associated with TB pathogenesis; modulating key processes such as *Mtb* dormancy.

1.6. Thesis overview and research aims

My doctoral thesis research is split into two parts. The first is an exploratory characterization of the stress responses of pathogenic mycobacteria. The second is to test the hypothesis that the spectrum of tRNA modifications in mycobacteria changes as a function of cellular environment, and that these pathogens utilize them to control their translation to survive physiological stresses associated with disease pathogenesis. I pursued this research in three integrated stages, each with its own specific aims; starting with the development of a sensitive analytical platform to quantify stress responses at the population, cellular and molecular levels; applying it to *M. bovis BCG* as a tractable surrogate for *Mtb*, and then translating the these findings into molecular mechanisms which could be applied in the development of disease-state biomarkers, antibiotic targets and drug screening platforms.

Aim 1: Develop a platform to characterize and quantify tRNA secondary modifications

Objectives: Develop and validate bioanalytical tools for identification and quantification of the cellular stress response with a focus on RNA and their spectrum of modified ribonucleosides.

Experimental strategy: Overall this involves developing methods for the unbiased and efficient isolation of total RNA (Chapter 2), purification of each non-coding RNA species by high performance liquid chromatography (HPLC) (Chapter 3), identification, structural characterization and quantification of every modified ribonucleoside in the enzymatic digests from purified tRNA by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) (Chapter 4).

Aim 2: Characterize mycobacterial stress responses

Objectives: Establish and define *in vitro* models for mycobacterial dormancy and persistence.

Experimental strategy: We revisited the batch culture models of Loebel and Wayne [172, 198]. In the former, we established a nutrient starvation model that reproducibly generated antibiotic tolerant persisters in a well-defined culture media and used flow cytometry, phenotypic microarrays, RNA-seq and targeted biochemical measurements of cellular metabolism to characterize the persistent state of mycobacteria (Chapter 5). In the latter, we scaled-up the original hypoxic batch culture protocol 20 folds and used the bioanalytical platform developed in aim 1 to determine the full spectrum of tRNA modifications in BCG. tRNA reprogramming during the transistion into and out of a hypoxia-induced dormancy was tracked by the quantification of tRNA modifications and copy numbers while dynamic changes in the proteome was determined using isobaric tags for relative and absolute quantitation (iTRAQ) proteomics (Chapter 6).

Aim 3: Integrate findings at the systems level to identify vulnerabilities that can be exploited as antibiotic targets

Experimental strategy: New metabolic pathways and molecular mechanisms of translational control were identified through multivariate data analysis. Precise predictions on population phenotypes under stress were made and validated using whole-cell chemical susceptibility screens or codon re-engineering approaches. Specific models of mycobacterial adaptation to nutrient starvation (Chapter 5) and hypoxic (Chapter 6) were made and general biological principles applied to design drug discovery platforms (Chapter 7).

Together, the studies present in this thesis advance our understanding of mycobacterial dormancy and persistence: cellular states that contribute to clinical presentations of latent and relapse TB. We discovered metabolic pathways, tRNA modifications and translation control mechanisms that give us further insights into bacterial physiology and disease pathogenesis. These findings increase our scientific resources against TB – one of humanity's oldest and greatest foes.

1.7. References

1.   Bos, K.I., et al., *Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis.* Nature, 2014. **514**(7523): p. 494-7.
2.   Hippocrates. *Aphorisms.* 400 B.C.E; Available from: http://classics.mit.edu/Hippocrates/aphorisms.5.v.html.
3.   Daniel, T.M., *The history of tuberculosis.* Respir Med, 2006. **100**(11): p. 1862-70.
4.   Gradmann, C., *Robert Koch and the pressures of scientific research: tuberculosis and tuberculin.* Med Hist, 2001. **45**(1): p. 1-32.
5.   Colditz, G.A., et al., *Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature.* JAMA, 1994. **271**(9): p. 698-702.
6.   Sharma, S.K., et al., *Rifamycins (rifampicin, rifabutin and rifapentine) compared to isoniazid for preventing tuberculosis in HIV-negative*

people at risk of active TB. Cochrane Database Syst Rev, 2013. **7**: p. CD007545.

7. WHO, *Global Tuberculosis Report 2014.* 2014, World Health Organization: Geneva. p. 147.

8. Persing, D.H., *Latent tuberculosis: interferon and beyond?* J Mol Diagn, 2015. **17**(1): p. 2-3.

9. Keshavjee, S. and P.E. Farmer, *Tuberculosis, drug resistance, and the history of modern medicine.* N Engl J Med, 2012. **367**(10): p. 931-6.

10. Lee, K.K., et al., *Isoniazid-induced cell death is precipitated by underlying mitochondrial complex I dysfunction in mouse hepatocytes.* Free Radic Biol Med, 2013. **65**: p. 584-94.

11. Olufunsho, A. and A. Alade, *Investigation of lipid peroxidation as probable mechanism of rifampicin toxicity in vivo.* Ann Neurosci, 2012. **19**(2): p. 68-70.

12. Shih, T.Y., et al., *A novel mechanism underlies the hepatotoxicity of pyrazinamide.* Antimicrob Agents Chemother, 2013. **57**(4): p. 1685-90.

13. Syed, J., *TB Drug and Vaccine Pipeline 2006.* What's in the Pipeline: New HIV Drugs, Vaccines, Microbicides, HCV and TB Therapies in Clinical Trials, 2006: p. 66.

14. Udwadia, Z.F., et al., *Totally drug-resistant tuberculosis in India.* Clin Infect Dis, 2012. **54**(4): p. 579-81.

15. Ginsberg, A.M. and M. Spigelman, *Challenges in tuberculosis drug research and development.* Nat Med, 2007. **13**(3): p. 290-4.

16. WHO, *Global tuberculosis report 2013.* 2013: World Health Organization.

17. Russell, D.G., C.E. Barry, 3rd, and J.L. Flynn, *Tuberculosis: what we don't know can, and does, hurt us.* Science, 2010. **328**(5980): p. 852-6.

18. Balaban, N.Q., et al., *A problem of persistence: still more questions than answers?* Nat Rev Microbiol, 2013. **11**(8): p. 587-91.

19. Kumar, A., et al., *Redox homeostasis in mycobacteria: the key to tuberculosis control?* Expert Rev Mol Med, 2011. **13**: p. e39.

20. Dedon, P.C. and T.J. Begley, *A system of RNA modifications and biased codon use controls cellular stress response at the level of translation.* Chem Res Toxicol, 2014. **27**(3): p. 330-7.

21. Gu, C., T.J. Begley, and P.C. Dedon, *tRNA modifications regulate translation during cellular stress.* FEBS Lett, 2014. **588**(23): p. 4287-4296.

22. Cohen, J.E., *Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better.* PLoS biology, 2004. **2**(12): p. e439.

23. Wirth, T., et al., *Origin, spread and demography of the Mycobacterium tuberculosis complex.* PLoS Pathog, 2008. **4**(9): p. e1000160.

24. Lin, P.L. and J.L. Flynn, *Understanding latent tuberculosis: a moving target.* J Immunol, 2010. **185**(1): p. 15-22.

25. The Lancet, D., amp, and Endocrinology, *Diabetes and tuberculosis— a wake-up call.* The Lancet Diabetes & Endocrinology. **2**(9): p. 677.

26. Daley, C.L., et al., *An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms.* N Engl J Med, 1992. **326**(4): p. 231-5.

27. Nelson, L.J. and C.D. Wells, *Global epidemiology of childhood tuberculosis.* Int J Tuberc Lung Dis, 2004. **8**(5): p. 636-47.

28. Muller, B., et al., *Zoonotic Mycobacterium bovis-induced tuberculosis in humans.* Emerg Infect Dis, 2013. **19**(6): p. 899-908.
29. Bloom, D.E., et al., *The global economic burden of noncommunicable diseases.* 2012, Program on the Global Demography of Aging.
30. Wah, W., et al., *Time series analysis of demographic and temporal trends of tuberculosis in Singapore.* BMC Public Health, 2014. **14**: p. 1121.
31. Lim, K.L., et al., *Molecular epidemiology of norovirus in Singapore, 2004-2011.* J Med Virol, 2013. **85**(10): p. 1842-51.
32. Pasipanodya, J.G. and T. Gumbo, *A meta-analysis of self-administered vs directly observed therapy effect on microbiologic failure, relapse, and acquired drug resistance in tuberculosis patients.* Clin Infect Dis, 2013. **57**(1): p. 21-31.
33. Moonan, P.K., et al., *Does directly observed therapy (DOT) reduce drug resistant tuberculosis?* BMC Public Health, 2011. **11**: p. 19.
34. Dalton, T., et al., *Prevalence of and risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in eight countries: a prospective cohort study.* Lancet, 2012. **380**(9851): p. 1406-17.
35. Ormerod, L.P., *Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment.* Br Med Bull, 2005. **73-74**: p. 17-24.
36. Faustini, A., A.J. Hall, and C.A. Perucci, *Risk factors for multidrug resistant tuberculosis in Europe: a systematic review.* Thorax, 2006. **61**(2): p. 158-63.
37. IMF, *World Economic Outlook, October 2014*, I.M. Fund, Editor. 2014, International Monetary Fund: Washington, D.C.
38. Gagneux, S., *Host-pathogen coevolution in human tuberculosis.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1590): p. 850-9.
39. Molina, R.L., K. Diouf, and N.M. Nour, *Tuberculosis and the obstetrician-gynecologist: a global perspective.* Rev Obstet Gynecol, 2013. **6**(3-4): p. 174-81.
40. Huang, D. and H. Yin, *Primary inoculation tuberculosis after an accidental scalpel injury.* Infection, 2013. **41**(4): p. 841-844.
41. Cooper, A.M., et al., *Interleukin 12 (IL-12) is crucial to the development of protective immunity in mice intravenously infected with mycobacterium tuberculosis.* J Exp Med, 1997. **186**(1): p. 39-45.
42. Fenhalls, G., et al., *Distribution of IFN-gamma, IL-4 and TNF-alpha protein and CD8 T cells producing IL-12p40 mRNA in human lung tuberculous granulomas.* Immunology, 2002. **105**(3): p. 325-35.
43. Zumla, A., et al., *Tuberculosis.* N Engl J Med, 2013. **368**(8): p. 745-55.
44. Barry, C.E., 3rd, et al., *The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.* Nat Rev Microbiol, 2009. **7**(12): p. 845-55.
45. Via, L.E., et al., *Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates.* Infect Immun, 2008. **76**(6): p. 2333-40.
46. Garcia-Rodriguez, J.F., et al., *Extrapulmonary tuberculosis: epidemiology and risk factors.* Enferm Infecc Microbiol Clin, 2011. **29**(7): p. 502-9.
47. Mack, U., et al., *LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement.* Eur Respir J, 2009. **33**(5): p. 956-73.

48. Horsburgh, C.R., Jr. and E.J. Rubin, *Clinical practice. Latent tuberculosis infection in the United States.* N Engl J Med, 2011. **364**(15): p. 1441-8.
49. Linas, B.P., et al., *Priorities for screening and treatment of latent tuberculosis infection in the United States.* Am J Respir Crit Care Med, 2011. **184**(5): p. 590-601.
50. Willis, H.S., et al., *The Death and Resurrection of the Tubercle Bacillus.* Transactions of the American Clinical and Climatological Association, 1956. **67**: p. 132-138.
51. Loring, W.E. and H.M. Vandiviere, *The treated pulmonary lesion and its tubercle bacillus. I. Pathology and pathogenesis.* Am J Med Sci, 1956. **232**(1): p. 20-9.
52. Vandiviere, H.M., et al., *The treated pulmonary lesion and its tubercle bacillus. II. The death and resurrection.* Am J Med Sci, 1956. **232**(1): p. 30-7; passim.
53. Lin, P.L., et al., *Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing.* Nat Med, 2014. **20**(1): p. 75-9.
54. Ramaswamy, S. and J.M. Musser, *Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update.* Tuber Lung Dis, 1998. **79**(1): p. 3-29.
55. Ford, C.B., et al., *Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection.* Nat Genet, 2011. **43**(5): p. 482-6.
56. Ford, C.B., et al., *Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis.* Nat Genet, 2013. **45**(7): p. 784-90.
57. McGrath, M., et al., *Mutation rate and the emergence of drug resistance in Mycobacterium tuberculosis.* J Antimicrob Chemother, 2014. **69**(2): p. 292-302.
58. Zhang, Y., et al., *The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis.* Nature, 1992. **358**(6387): p. 591-3.
59. van Doorn, H.R., et al., *Public health impact of isoniazid-resistant Mycobacterium tuberculosis strains with a mutation at amino-acid position 315 of katG: a decade of experience in The Netherlands.* Clin Microbiol Infect, 2006. **12**(8): p. 769-75.
60. van Soolingen, D., et al., *Mutations at amino acid position 315 of the katG gene are associated with high-level resistance to isoniazid, other drug resistance, and successful transmission of Mycobacterium tuberculosis in the Netherlands.* J Infect Dis, 2000. **182**(6): p. 1788-90.
61. Rawat, R., A. Whitty, and P.J. Tonge, *The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: adduct affinity and drug resistance.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 13881-6.
62. Ramaswamy, S.V., et al., *Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2003. **47**(4): p. 1241-50.
63. Hazbon, M.H., et al., *Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2006. **50**(8): p. 2640-9.
64. Leung, E.T., et al., *Molecular characterization of isoniazid resistance in Mycobacterium tuberculosis: identification of a novel mutation in inhA.* Antimicrob Agents Chemother, 2006. **50**(3): p. 1075-8.

65. Vilcheze, C. and W.R. Jacobs, Jr., *The mechanism of isoniazid killing: clarity through the scope of genetics.* Annu Rev Microbiol, 2007. **61**: p. 35-50.

66. Rodwell, T.C., et al., *Predicting extensively drug-resistant Mycobacterium tuberculosis phenotypes with genetic mutations.* J Clin Microbiol, 2014. **52**(3): p. 781-9.

67. Banerjee, A., et al., *inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis.* Science, 1994. **263**(5144): p. 227-30.

68. Sherman, D.R., et al., *Compensatory ahpC gene expression in isoniazid-resistant Mycobacterium tuberculosis.* Science, 1996. **272**(5268): p. 1641-3.

69. Georghiou, S.B., et al., *Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review.* PLoS One, 2012. **7**(3): p. e33275.

70. Sandgren, A., et al., *Tuberculosis drug resistance mutation database.* PLoS Med, 2009. **6**(2): p. e2.

71. Maruri, F., et al., *A systematic review of gyrase mutations associated with fluoroquinolone-resistant Mycobacterium tuberculosis and a proposed gyrase numbering system.* J Antimicrob Chemother, 2012. **67**(4): p. 819-31.

72. Telenti, A., et al., *Direct, automated detection of rifampin-resistant Mycobacterium tuberculosis by polymerase chain reaction and single-strand conformation polymorphism analysis.* Antimicrob Agents Chemother, 1993. **37**(10): p. 2054-8.

73. Scorpio, A. and Y. Zhang, *Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus.* Nat Med, 1996. **2**(6): p. 662-7.

74. Gillespie, S.H., *Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective.* Antimicrob Agents Chemother, 2002. **46**(2): p. 267-74.

75. Okamoto, S., et al., *Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria.* Mol Microbiol, 2007. **63**(4): p. 1096-106.

76. Telenti, A., et al., *The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol.* Nat Med, 1997. **3**(5): p. 567-70.

77. Takiff, H.E., et al., *Cloning and nucleotide sequence of Mycobacterium tuberculosis gyrA and gyrB genes and detection of quinolone resistance mutations.* Antimicrob Agents Chemother, 1994. **38**(4): p. 773-80.

78. Alangaden, G.J., et al., *Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 1998. **42**(5): p. 1295-7.

79. Johansen, S.K., et al., *Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs.* Mol Cell, 2006. **23**(2): p. 173-82.

80. Rengarajan, J., et al., *The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria.* Mol Microbiol, 2004. **53**(1): p. 275-82.

81. Manjunatha, U.H., et al., *Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2006. **103**(2): p. 431-6.

82. Andries, K., et al., *A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis.* Science, 2005. **307**(5707): p. 223-7.

83. Louw, G.E., et al., *A balancing act: efflux/influx in mycobacterial drug resistance.* Antimicrob Agents Chemother, 2009. **53**(8): p. 3181-9.

84. Jarlier, V. and H. Nikaido, *Mycobacterial cell wall: structure and role in natural resistance to antibiotics.* FEMS Microbiol Lett, 1994. **123**(1-2): p. 11-8.

85. De Rossi, E., J.A. Ainsa, and G. Riccardi, *Role of mycobacterial efflux transporters in drug resistance: an unresolved question.* FEMS Microbiol Rev, 2006. **30**(1): p. 36-52.

86. Escribano, I., et al., *Importance of the efflux pump systems in the resistance of Mycobacterium tuberculosis to fluoroquinolones and linezolid.* Chemotherapy, 2007. **53**(6): p. 397-401.

87. Spies, F.S., et al., *Identification of mutations related to streptomycin resistance in clinical isolates of Mycobacterium tuberculosis and possible involvement of efflux mechanism.* Antimicrob Agents Chemother, 2008. **52**(8): p. 2947-9.

88. Braibant, M., P. Gilot, and J. Content, *The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis.* FEMS Microbiol Rev, 2000. **24**(4): p. 449-67.

89. Danilchanka, O., C. Mailaender, and M. Niederweis, *Identification of a novel multidrug efflux pump of Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2008. **52**(7): p. 2503-11.

90. Balganesh, M., et al., *Rv1218c, an ABC transporter of Mycobacterium tuberculosis with implications in drug discovery.* Antimicrob Agents Chemother, 2010. **54**(12): p. 5167-72.

91. Black, P.A., et al., *Energy metabolism and drug efflux in Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2014. **58**(5): p. 2491-503.

92. da Silva, P.E., et al., *Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis.* FEMS Immunol Med Microbiol, 2011. **63**(1): p. 1-9.

93. Ainsa, J.A., et al., *Molecular cloning and characterization of Tap, a putative multidrug efflux pump present in Mycobacterium fortuitum and Mycobacterium tuberculosis.* J Bacteriol, 1998. **180**(22): p. 5836-43.

94. Jiang, X., et al., *Assessment of efflux pump gene expression in a clinical isolate Mycobacterium tuberculosis by real-time reverse transcription PCR.* Microb Drug Resist, 2008. **14**(1): p. 7-11.

95. Silva, P.E., et al., *Characterization of P55, a multidrug efflux pump in Mycobacterium bovis and Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2001. **45**(3): p. 800-4.

96. De Rossi, E., et al., *The multidrug transporters belonging to major facilitator superfamily in Mycobacterium tuberculosis.* Mol Med, 2002. **8**(11): p. 714-24.

97. Gupta, A.K., et al., *Microarray analysis of efflux pump genes in multidrug-resistant Mycobacterium tuberculosis during stress induced by common anti-tuberculous drugs.* Microb Drug Resist, 2010. **16**(1): p. 21-8.

98. Wilson, M., et al., *Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12833-8.

99. Choudhuri, B.S., et al., *Overexpression and functional characterization of an ABC (ATP-binding cassette) transporter encoded by the genes drrA and drrB of Mycobacterium tuberculosis.* Biochem J, 2002. **367**(Pt 1): p. 279-85.

100. Pasca, M.R., et al., *Rv2686c-Rv2687c-Rv2688c, an ABC fluoroquinolone efflux pump in Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2004. **48**(8): p. 3175-8.

101. Pasca, M.R., et al., *mmpL7 gene of Mycobacterium tuberculosis is responsible for isoniazid efflux in Mycobacterium smegmatis.* Antimicrob Agents Chemother, 2005. **49**(11): p. 4775-7.

102. Palomino, J.C., D.F. Ramos, and P.A. da Silva, *New anti-tuberculosis drugs: strategies, sources and new molecules.* Curr Med Chem, 2009. **16**(15): p. 1898-904.

103. Tommerup, I.C., *Spore dormancy in vesticular-arbuscular mycorrhizal fingi.* Trans Br Mycol Soc, 1983. **81**: p. 37-45.

104. Segev, E., Y. Smith, and S. Ben-Yehuda, *RNA dynamics in aging bacterial spores.* Cell, 2012. **148**(1-2): p. 139-49.

105. Xu, M., et al., *Timing of transcriptional quiescence during gametogenesis is controlled by global histone H3K4 demethylation.* Dev Cell, 2012. **23**(5): p. 1059-71.

106. Cutting, S.M. and E. Ricca, *Bacterial spore-formers: friends and foes.* FEMS Microbiol Lett, 2014. **358**(2): p. 107-9.

107. Rao, S.P., et al., *The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11945-50.

108. Shi, L., et al., *Changes in energy metabolism of Mycobacterium tuberculosis in mouse lung and under in vitro conditions affecting aerobic respiration.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15629-34.

109. Boshoff, H.I. and C.E. Barry, 3rd, *Tuberculosis - metabolism and respiration in the absence of growth.* Nat Rev Microbiol, 2005. **3**(1): p. 70-80.

110. Chakraborty, S., et al., *Para-aminosalicylic acid acts as an alternative substrate of folate metabolism in Mycobacterium tuberculosis.* Science, 2013. **339**(6115): p. 88-91.

111. Warner, D.F. and V. Mizrahi, *Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy.* Clin Microbiol Rev, 2006. **19**(3): p. 558-70.

112. Gengenbacher, M., et al., *Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability.* Microbiology, 2010. **156**(Pt 1): p. 81-7.

113. Rasmussen, K.N., *The apical localization of pulmonary tuberculosis.* Acta Tuberc Scand, 1957. **34**(3-4): p. 245-59.

114. Dirken, M.N. and H. Heemstra, *Alveolar oxygen tension and lung circulation.* Q J Exp Physiol, 1948. **34**(3-4): p. 193-211.

115. Dirken, M.N. and H. Heemstra, *The alveolar-arterial difference in oxygen-tension.* Arch Neerl Physiol Homme Anim, 1948. **28**(3-4): p. 501-9.

116. Aly, S., et al., *Oxygen status of lung granulomas in Mycobacterium tuberculosis-infected mice.* J Pathol, 2006. **210**(3): p. 298-305.

117. Schnappinger, D., et al., *Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment.* J Exp Med, 2003. **198**(5): p. 693-704.

118. Tailleux, L., et al., *Probing host pathogen cross-talk by transcriptional profiling of both Mycobacterium tuberculosis and infected human dendritic cells and macrophages.* PLoS One, 2008. **3**(1): p. e1403.

119. Talaat, A.M., et al., *Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling.* J Bacteriol, 2007. **189**(11): p. 4265-74.

120. Voskuil, M.I., et al., *Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program.* J Exp Med, 2003. **198**(5): p. 705-13.

121. Talaat, A.M., et al., *The temporal expression profile of Mycobacterium tuberculosis infection in mice.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4602-7.

122. Weinstein, E.A., et al., *Inhibitors of type II NADH:menaquinone oxidoreductase represent a class of antitubercular drugs.* Proc Natl Acad Sci U S A, 2005. **102**(12): p. 4548-53.

123. Eisenreich, W., et al., *Carbon metabolism of intracellular bacterial pathogens and possible links to virulence.* Nat Rev Microbiol, 2010. **8**(6): p. 401-12.

124. Eoh, H. and K.Y. Rhee, *Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2013. **110**(16): p. 6554-9.

125. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry.* 7th ed. 2012, New York: W.H. Freeman. xxxii, 1054, 43, 41, 48 p.

126. McKinney, J.D., et al., *Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase.* Nature, 2000. **406**(6797): p. 735-8.

127. Sharma, V., et al., *Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis.* Nat Struct Biol, 2000. **7**(8): p. 663-8.

128. Av-Gay, Y. and M. Everett, *The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis.* Trends Microbiol, 2000. **8**(5): p. 238-44.

129. Narayan, A., et al., *Serine threonine protein kinases of mycobacterial genus: phylogeny to function.* Physiol Genomics, 2007. **29**(1): p. 66-75.

130. Alam, M.S., S.K. Garg, and P. Agrawal, *Molecular function of WhiB4/Rv3681c of Mycobacterium tuberculosis H37Rv: a [4Fe-4S] cluster co-ordinating protein disulphide reductase.* Mol Microbiol, 2007. **63**(5): p. 1414-31.

131. Larsson, C., et al., *Gene expression of Mycobacterium tuberculosis putative transcription factors whiB1-7 in redox environments.* PLoS One, 2012. **7**(7): p. e37516.

132. Sala, A., P. Bordes, and P. Genevaux, *Multiple toxin-antitoxin systems in Mycobacterium tuberculosis.* Toxins (Basel), 2014. **6**(3): p. 1002-20.

133. Ramage, H.R., L.E. Connolly, and J.S. Cox, *Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution.* PLoS Genet, 2009. **5**(12): p. e1000767.

134. Calamita, H., et al., *The Mycobacterium tuberculosis SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence.* Cell Microbiol, 2005. **7**(2): p. 233-44.

135. Mehra, S., et al., *The Mycobacterium tuberculosis stress response factor SigH is required for bacterial burden as well as immunopathology in primate lungs.* J Infect Dis, 2012. **205**(8): p. 1203-13.

136. Sachdeva, P., et al., *The sigma factors of Mycobacterium tuberculosis: regulation of the regulators.* FEBS J, 2010. **277**(3): p. 605-26.

137. Ewann, F., et al., *Transient requirement of the PrrA-PrrB two-component system for early intracellular multiplication of Mycobacterium tuberculosis.* Infect Immun, 2002. **70**(5): p. 2256-63.

138. Parish, T., et al., *The senX3-regX3 two-component regulatory system of Mycobacterium tuberculosis is required for virulence.* Microbiology, 2003. **149**(Pt 6): p. 1423-35.

139. Zahrt, T.C., et al., *Functional analysis of the Mycobacterium tuberculosis MprAB two-component signal transduction system.* Infect Immun, 2003. **71**(12): p. 6962-70.

140. Perez, E., et al., *An essential role for phoP in Mycobacterium tuberculosis virulence.* Mol Microbiol, 2001. **41**(1): p. 179-87.

141. Hett, E.C. and E.J. Rubin, *Bacterial growth and cell division: a mycobacterial perspective.* Microbiol Mol Biol Rev, 2008. **72**(1): p. 126-56, table of contents.

142. Rustad, T.R., et al., *Hypoxia: a window into Mycobacterium tuberculosis latency.* Cell Microbiol, 2009. **11**(8): p. 1151-9.

143. Leistikow, R.L., et al., *The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy.* J Bacteriol, 2010. **192**(6): p. 1662-70.

144. Park, H.D., et al., *Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis.* Mol Microbiol, 2003. **48**(3): p. 833-43.

145. Rustad, T.R., et al., *The enduring hypoxic response of Mycobacterium tuberculosis.* PLoS One, 2008. **3**(1): p. e1502.

146. Boon, C. and T. Dick, *How Mycobacterium tuberculosis goes to sleep: the dormancy survival regulator DosR a decade later.* Future Microbiol, 2012. **7**(4): p. 513-8.

147. Ortega, C., et al., *Mycobacterium tuberculosis Ser/Thr protein kinase B mediates an oxygen-dependent replication switch.* PLoS Biol, 2014. **12**(1): p. e1001746.

148. Selvaraj, S., et al., *In silico analysis of DosR regulon proteins of Mycobacterium tuberculosis.* Gene, 2012. **506**(1): p. 233-41.

149. Sherman, D.R., et al., *Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding alpha -crystallin.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7534-9.

150. Dasgupta, N., et al., *Characterization of a two-component system, devR-devS, of Mycobacterium tuberculosis.* Tuber Lung Dis, 2000. **80**(3): p. 141-59.

151. Boon, C., et al., *Proteins of Mycobacterium bovis BCG induced in the Wayne dormancy model.* J Bacteriol, 2001. **183**(8): p. 2672-6.

152. Boon, C. and T. Dick, *Mycobacterium bovis BCG response regulator essential for hypoxic dormancy.* J Bacteriol, 2002. **184**(24): p. 6760-7.

153. Honaker, R.W., et al., *Unique roles of DosT and DosS in DosR regulon induction and Mycobacterium tuberculosis dormancy.* Infect Immun, 2009. **77**(8): p. 3258-63.

154. Kumar, A., et al., *Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor.* Proc Natl Acad Sci U S A, 2007. **104**(28): p. 11568-73.

155. Chao, J.D., et al., *Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in Mycobacterium tuberculosis.* J Biol Chem, 2010. **285**(38): p. 29239-46.

156. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.

157.   Kumar, A., et al., *Heme oxygenase-1-derived carbon monoxide induces the Mycobacterium tuberculosis dormancy regulon.* J Biol Chem, 2008. **283**(26): p. 18032-9.

158.   Shiloh, M.U., P. Manzanillo, and J.S. Cox, *Mycobacterium tuberculosis senses host-derived carbon monoxide during macrophage infection.* Cell Host Microbe, 2008. **3**(5): p. 323-30.

159.   Sardiwal, S., et al., *A GAF domain in the hypoxia/NO-inducible Mycobacterium tuberculosis DosS protein binds haem.* J Mol Biol, 2005. **353**(5): p. 929-36.

160.   Hu, Y., et al., *Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo.* Infect Immun, 2006. **74**(2): p. 861-8.

161.   Black, G.F., et al., *Immunogenicity of novel DosR regulon-encoded candidate antigens of Mycobacterium tuberculosis in three high-burden populations in Africa.* Clin Vaccine Immunol, 2009. **16**(8): p. 1203-12.

162.   Commandeur, S., et al., *Double- and monofunctional CD4(+) and CD8(+) T-cell responses to Mycobacterium tuberculosis DosR antigens and peptides in long-term latently infected individuals.* Eur J Immunol, 2011. **41**(10): p. 2925-36.

163.   Leyten, E.M., et al., *Human T-cell responses to 25 novel antigens encoded by genes of the dormancy regulon of Mycobacterium tuberculosis.* Microbes Infect, 2006. **8**(8): p. 2052-60.

164.   Malhotra, V., et al., *Disruption of response regulator gene, devR, leads to attenuation in virulence of Mycobacterium tuberculosis.* FEMS Microbiol Lett, 2004. **231**(2): p. 237-45.

165.   Minch, K.J., et al., *The DNA-binding network of Mycobacterium tuberculosis.* Nat Commun, 2015. **6**: p. 5829.

166.   Rustad, T.R., et al., *Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network.* Genome Biol, 2014. **15**(11): p. 502.

167.   Keren, I., et al., *Characterization and transcriptome analysis of Mycobacterium tuberculosis persisters.* MBio, 2011. **2**(3): p. e00100-11.

168.   Rachman, H., et al., *Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis.* Infect Immun, 2006. **74**(2): p. 1233-42.

169.   Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* Science, 2010. **329**(5991): p. 533-8.

170.   Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.

171.   Foss, E.J., et al., *Genetic variation shapes protein networks mainly through non-transcriptional mechanisms.* PLoS Biol, 2011. **9**(9): p. e1001144.

172.   Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence.* Infect Immun, 1996. **64**(6): p. 2062-9.

173.   Wayne, L.G., *Dormancy of Mycobacterium tuberculosis and latency of disease.* Eur J Clin Microbiol Infect Dis, 1994. **13**(11): p. 908-14.

174.   Wayne, L.G. and K.Y. Lin, *Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions.* Infect Immun, 1982. **37**(3): p. 1042-9.

175. Edwards, D.I., *Mechanisms of selective toxicity of metronidazole and other nitroimidazole drugs.* Br J Vener Dis, 1980. **56**(5): p. 285-90.

176. Wayne, L.G. and H.A. Sramek, *Metronidazole is bactericidal to dormant cells of Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 1994. **38**(9): p. 2054-8.

177. Edwards, D.I., *Nitroimidazole drugs--action and resistance mechanisms. I. Mechanisms of action.* J Antimicrob Chemother, 1993. **31**(1): p. 9-20.

178. Wayne, L.G. and C.D. Sohaskey, *Nonreplicating persistence of mycobacterium tuberculosis.* Annu Rev Microbiol, 2001. **55**: p. 139-63.

179. NIAID, *Metronidazole for Pulmonary Tuberculosis (South Korea). In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US).2000- [cited 2015 Jan 15] Available from: https://clinicaltrials.gov/ct2/show/study/NCT00425113 NLM Identifier: NCT00425113*

180. Lin, P.L., et al., *Metronidazole prevents reactivation of latent Mycobacterium tuberculosis infection in macaques.* Proc Natl Acad Sci U S A, 2012. **109**(35): p. 14188-93.

181. Upton, A.M., et al., *In Vitro and In Vivo Activities of the Nitroimidazole TBA-354 against Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2015. **59**(1): p. 136-44.

182. Kussell, E., et al., *Bacterial persistence: a model of survival in changing environments.* Genetics, 2005. **169**(4): p. 1807-14.

183. Imlay, J.A. and S. Linn, *Bimodal pattern of killing of DNA-repair-defective or anoxically grown Escherichia coli by hydrogen peroxide.* J Bacteriol, 1986. **166**(2): p. 519-27.

184. Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch.* Science, 2004. **305**(5690): p. 1622-5.

185. Wakamoto, Y., et al., *Dynamic persistence of antibiotic-stressed mycobacteria.* Science, 2013. **339**(6115): p. 91-5.

186. Kint, C.I., et al., *New-found fundamentals of bacterial persistence.* Trends Microbiol, 2012. **20**(12): p. 577-85.

187. Peyron, P., et al., *Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence.* PLoS Pathog, 2008. **4**(11): p. e1000204.

188. Russell, D.G., et al., *Foamy macrophages and the progression of the human tuberculosis granuloma.* Nat Immunol, 2009. **10**(9): p. 943-8.

189. Bacon, J., et al., *Non-replicating Mycobacterium tuberculosis elicits a reduced infectivity profile with corresponding modifications to the cell wall and extracellular matrix.* PLoS One, 2014. **9**(2): p. e87329.

190. Daniel, J., et al., *Mycobacterium tuberculosis uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages.* PLoS Pathog, 2011. **7**(6): p. e1002093.

191. Low, K.L., et al., *Triacylglycerol utilization is required for regrowth of in vitro hypoxic nonreplicating Mycobacterium bovis bacillus Calmette-Guerin.* J Bacteriol, 2009. **191**(16): p. 5037-43.

192. Deb, C., et al., *A novel in vitro multiple-stress dormancy model for Mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen.* PLoS One, 2009. **4**(6): p. e6077.

193. Daniel, J., et al., *Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in Mycobacterium tuberculosis as it goes into a dormancy-like state in culture.* J Bacteriol, 2004. **186**(15): p. 5017-30.

194. Nyka, W., *Method for staining both acid-fast and chromophobic tubercle bacilli with carbolfuschsin.* J Bacteriol, 1967. **93**(4): p. 1458-60.

195. Nyka, W., *Studies on the effect of starvation on mycobacteria.* Infect Immun, 1974. **9**(5): p. 843-50.

196. Xie, Z., N. Siddiqi, and E.J. Rubin, *Differential antibiotic susceptibilities of starved Mycobacterium tuberculosis isolates.* Antimicrob Agents Chemother, 2005. **49**(11): p. 4778-80.

197. Rodriguez, G.M., et al., *ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response.* Infect Immun, 2002. **70**(7): p. 3371-81.

198. Loebel, R.O., E. Shorr, and H.B. Richardson, *The Influence of Adverse Conditions upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli.* J Bacteriol, 1933. **26**(2): p. 167-200.

199. Loebel, R.O., E. Shorr, and H.B. Richardson, *The Influence of Foodstuffs upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli.* J Bacteriol, 1933. **26**(2): p. 139-66.

200. Deretic, V., et al., *Mycobacterium tuberculosis is a natural mutant with an inactivated oxidative-stress regulatory gene: implications for sensitivity to isoniazid.* Mol Microbiol, 1995. **17**(5): p. 889-900.

201. Master, S.S., et al., *Oxidative stress response genes in Mycobacterium tuberculosis: role of ahpC in resistance to peroxynitrite and stage-specific survival in macrophages.* Microbiology, 2002. **148**(Pt 10): p. 3139-44.

202. Vandal, O.H., C.F. Nathan, and S. Ehrt, *Acid resistance in Mycobacterium tuberculosis.* J Bacteriol, 2009. **191**(15): p. 4714-21.

203. Ramakrishnan, S.R., et al., *Integrating shotgun proteomics and mRNA expression data to improve protein identification.* Bioinformatics, 2009. **25**(11): p. 1397-403.

204. Dahan, O., H. Gingold, and Y. Pilpel, *Regulatory mechanisms and networks couple the different phases of gene expression.* Trends Genet, 2011. **27**(8): p. 316-22.

205. Maier, T., et al., *Quantification of mRNA and protein and integration with protein turnover in a bacterium.* Mol Syst Biol, 2011. **7**: p. 511.

206. Jayapal, K.P., et al., *Uncovering genes with divergent mRNA-protein dynamics in Streptomyces coelicolor.* PLoS One, 2008. **3**(5): p. e2097.

207. Kaebernick, M., et al., *Multiple alternate transcripts direct the biosynthesis of microcystin, a cyanobacterial nonribosomal peptide.* Appl Environ Microbiol, 2002. **68**(2): p. 449-55.

208. Wagner, S.D. and J.A. Berglund, *Alternative pre-mRNA splicing.* Methods Mol Biol, 2014. **1126**: p. 45-54.

209. Peng, Z., et al., *Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome.* Nat Biotechnol, 2012. **30**(3): p. 253-60.

210. Wilson, R.C. and J.A. Doudna, *Molecular mechanisms of RNA interference.* Annu Rev Biophys, 2013. **42**: p. 217-39.

211. du Toit, A., *Translation: When ribosomes don't stop.* Nat Rev Mol Cell Biol, 2014. **15**(1): p. 4-4.

212. Ikeuchi, Y., et al., *molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition.* Mol Cell, 2005. **19**(2): p. 235-46.

213. Dumelin, C.E., et al., *Discovery and biological characterization of geranylated RNA in bacteria.* Nat Chem Biol, 2012. **8**(11): p. 913-9.

214.   Kowtoniuk, W.E., et al., *A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA.* Proc Natl Acad Sci U S A, 2009. **106**(19): p. 7768-73.

215.   Armengod, M.E., et al., *Enzymology of tRNA modification in the bacterial MnmEG pathway.* Biochimie, 2012. **94**(7): p. 1510-20.

216.   Arragain, S., et al., *Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenosine in tRNA.* J Biol Chem, 2010. **285**(37): p. 28425-33.

217.   Miyauchi, K., S. Kimura, and T. Suzuki, *A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification.* Nat Chem Biol, 2013. **9**(2): p. 105-11.

218.   Machnicka, M.A., et al., *MODOMICS: a database of RNA modification pathways--2013 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D262-7.

219.   Czerwoniec, A., et al., *MODOMICS: a database of RNA modification pathways. 2008 update.* Nucleic Acids Res, 2009. **37**(Database issue): p. D118-21.

220.   Cantara, W.A., et al., *The RNA Modification Database, RNAMDB: 2011 update.* Nucleic Acids Res, 2011. **39**(Database issue): p. D195-201.

221.   Steinberg, S., A. Misch, and M. Sprinzl, *Compilation of tRNA sequences and sequences of tRNA genes.* Nucleic Acids Res, 1993. **21**(13): p. 3011-5.

222.   Grosjean, H., *DNA and RNA modification enzymes: structure, mechanism, function and evolution.* 2009: Landes Bioscience Austin, TX.

223.   Bjork, G.R., et al., *A primordial tRNA modification required for the evolution of life?* EMBO J, 2001. **20**(1-2): p. 231-9.

224.   Carell, T., et al., *Structure and function of noncanonical nucleobases.* Angew Chem Int Ed Engl, 2012. **51**(29): p. 7110-31.

225.   Karr, J.R., et al., *A whole-cell computational model predicts phenotype from genotype.* Cell, 2012. **150**(2): p. 389-401.

226.   Karr, J.R., et al., *WholeCellKB: model organism databases for comprehensive whole-cell models.* Nucleic Acids Res, 2013. **41**(Database issue): p. D787-92.

227.   Koonin, E.V., *How many genes can make a cell: the minimal-gene-set concept.* Annu Rev Genomics Hum Genet, 2000. **1**: p. 99-116.

228.   Forster, A.C. and G.M. Church, *Towards synthesis of a minimal cell.* Mol Syst Biol, 2006. **2**: p. 45.

229.   Dong, H., et al., *2'-O methylation of internal adenosine by flavivirus NS5 methyltransferase.* PLoS Pathog, 2012. **8**(4): p. e1002642.

230.   Batista, P.J., et al., *m(6)A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells.* Cell Stem Cell, 2014. **15**(6): p. 707-19.

231.   Chionh, Y.H., et al., *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Res, 2013. **41**(17): p. e168.

232.   Björk, G., *The role of modified nucleosides in tRNA interactions.* Transfer RNA in protein synthesis. CRC Press, Boca Raton, Fla, 1992: p. 23-85.

233.   Suzuki, T., et al., *Mass spectrometric identification and characterization of RNA-modifying enzymes.* Methods Enzymol, 2007. **425**: p. 211-29.

234.   Davis, D.R., *Stabilization of RNA stacking by pseudouridine.* Nucleic Acids Res, 1995. **23**(24): p. 5020-6.

235. Motorin, Y. and M. Helm, *tRNA stabilization by modified nucleotides.* Biochemistry, 2010. **49**(24): p. 4934-44.
236. Sylvers, L.A., et al., *A 2-thiouridine derivative in tRNAGlu is a positive determinant for aminoacylation by Escherichia coli glutamyl-tRNA synthetase.* Biochemistry, 1993. **32**(15): p. 3836-41.
237. Alexandrov, A., et al., *Rapid tRNA decay can result from lack of nonessential modifications.* Mol Cell, 2006. **21**(1): p. 87-96.
238. Ohira, T. and T. Suzuki, *Retrograde nuclear import of tRNA precursors is required for modified base biogenesis in yeast.* Proceedings of the National Academy of Sciences, 2011. **108**(26): p. 10502-10507.
239. Kadaba, S., et al., *Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae.* Genes & development, 2004. **18**(11): p. 1227-1240.
240. Urbonavicius, J., et al., *Improvement of reading frame maintenance is a common function for several tRNA modifications.* EMBO J, 2001. **20**(17): p. 4863-73.
241. Bjork, G.R., et al., *Transfer RNA modification: influence on translational frameshifting and metabolism.* FEBS Lett, 1999. **452**(1-2): p. 47-51.
242. Novoa, E.M., et al., *A role for tRNA modifications in genome structure and codon usage.* Cell, 2012. **149**(1): p. 202-13.
243. Kohrer, C., et al., *Life without tRNAIle-lysidine synthetase: translation of the isoleucine codon AUA in Bacillus subtilis lacking the canonical tRNA2Ile.* Nucleic Acids Res, 2014. **42**(3): p. 1904-15.
244. Agris, P.F., F.A. Vendeix, and W.D. Graham, *tRNA's wobble decoding of the genome: 40 years of modification.* J Mol Biol, 2007. **366**(1): p. 1-13.
245. Nasvall, S.J., P. Chen, and G.R. Bjork, *The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo.* RNA, 2004. **10**(10): p. 1662-73.
246. Weixlbaumer, A., et al., *Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines.* Nat Struct Mol Biol, 2007. **14**(6): p. 498-502.
247. Katz, A. and O. Orellana, *Protein synthesis and the stress response.* Cell-free protein synthesis. InTech, Croatia, 2012: p. 111-134.
248. Helm, M. and J.D. Alfonzo, *Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland.* Chem Biol, 2014. **21**(2): p. 174-85.
249. Yi, C. and T. Pan, *Cellular dynamics of RNA modification.* Acc Chem Res, 2011. **44**(12): p. 1380-8.
250. Takano, Y., et al., *A gene involved in modifying transfer RNA is required for fungal pathogenicity and stress tolerance of Colletotrichum lagenarium.* Mol Microbiol, 2006. **60**(1): p. 81-92.
251. Tomikawa, C., et al., *N7-Methylguanine at position 46 (m7G46) in tRNA from Thermus thermophilus is required for cell viability at high temperatures through a tRNA modification network.* Nucleic Acids Res, 2010. **38**(3): p. 942-57.
252. Reader, J.S., et al., *Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine.* J Biol Chem, 2004. **279**(8): p. 6280-5.
253. Garcia, G.A. and J.D. Kittendorf, *Transglycosylation: a mechanism for RNA modification (and editing?).* Bioorg Chem, 2005. **33**(3): p. 229-51.

254. Kersten, H. and W. Kersten, *Biosynthesis and Function of Queuine and Queuosine tRNAs. In "Chromatography and Modification of Nucleosides Part B"(KCT Kuo, ed.), pp. B69–B108*. 1990, Elsevier, Amsterdam.

255. Okada, N., et al., *Detection of unique tRNA species in tumor tissues by Escherichia coli guanine insertion enzyme.* Proc Natl Acad Sci U S A, 1978. **75**(9): p. 4247-51.

256. Langgut, W., et al., *Modulation of mammalian cell proliferation by a modified tRNA base of bacterial origin.* FEBS Lett, 1993. **336**(1): p. 137-42.

257. Szabo, L. and W. Farkas, *Possible involvement of queuine in oxidative metabolism*, in *Prostaglandin and Lipid Metabolism in Radiation Injury*. 1987, Springer. p. 109-114.

258. Marks, T. and W.R. Farkas, *Effects of a diet deficient in tyrosine and queuine on germfree mice.* Biochem Biophys Res Commun, 1997. **230**(2): p. 233-7.

259. Hurt, J.K., S. Olgen, and G.A. Garcia, *Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro.* Nucleic Acids Res, 2007. **35**(14): p. 4905-13.

260. Hortner, S.R., et al., *Potent inhibitors of tRNA-guanine transglycosylase, an enzyme linked to the pathogenicity of the Shigella bacterium: charge-assisted hydrogen bonding.* Angew Chem Int Ed Engl, 2007. **46**(43): p. 8266-9.

261. Reisser, T., W. Langgut, and H. Kersten, *The nutrient factor queuine protects HeLa cells from hypoxic stress and improves metabolic adaptation to oxygen availability.* Eur J Biochem, 1994. **221**(3): p. 979-86.

262. Meier, F., et al., *Queuosine modification of the wobble base in tRNAHis influences 'in vivo' decoding properties.* EMBO J, 1985. **4**(3): p. 823-7.

263. Hatfield, D.L., et al., *Translational suppression in retroviral gene expression.* Adv Virus Res, 1992. **41**: p. 193-239.

264. Begley, U., et al., *Trm9-catalyzed tRNA modifications link translation to the DNA damage response.* Mol Cell, 2007. **28**(5): p. 860-70.

265. Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.

266. Elf, J., et al., *Selective charging of tRNA isoacceptors explains patterns of codon usage.* Science, 2003. **300**(5626): p. 1718-22.

267. dos Reis, M., R. Savva, and L. Wernisch, *Solving the riddle of codon usage preferences: a test for translational selection.* Nucleic Acids Res, 2004. **32**(17): p. 5036-44.

268. Pechmann, S. and J. Frydman, *Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding.* Nat Struct Mol Biol, 2013. **20**(2): p. 237-43.

269. Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.

270. Patil, A., et al., *Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications.* RNA Biol, 2012. **9**(7): p. 990-1001.

271. Patil, A., et al., *Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response.* Cell Cycle, 2012. **11**(19): p. 3656-65.

2. Mycobacterial RNA isolation optimized for non-coding RNA: High fidelity isolation of 5S rRNA from *Mycobacterium bovis* BCG reveals novel post-transcriptional processing and a complete spectrum of modified ribonucleosides[1]

2.1. Abstract

A major challenge in the study of mycobacterial RNA biology is the lack of a comprehensive RNA isolation method that overcomes the unusual cell wall to faithfully yield the full spectrum of non-coding RNA (ncRNA) species. Here, we describe a simple and robust procedure optimized for the isolation of total ncRNA, including 5S, 16S and 23S ribosomal RNA (rRNA) and tRNA, from mycobacteria, using *Mycobacterium bovis* BCG to illustrate the method. Based on a combination of mechanical disruption and liquid and solid-phase technologies, the method produces all major species of ncRNA in high yield and with high integrity, enabling direct chemical and sequence analysis of the ncRNA species. The reproducibility of the method with BCG was evident in bioanalyzer electrophoretic analysis of isolated RNA, which revealed quantitatively significant differences in the ncRNA profiles of exponentially growing and non-replicating hypoxic bacilli. The method also overcame an historical inconsistency in 5S rRNA isolation, with direct sequencing revealing a novel post-transcriptional processing of 5S rRNA to its functional form and with chemical analysis revealing seven post-transcriptional ribonucleoside modifications in the 5S rRNA. This optimized RNA isolation procedure thus

---

provides a means to more rigorously explore the biology of ncRNA species in mycobacteria.

2.2. Introduction

Infection with *Mycobacterium tuberculosis* (Mtb) represents one of the most widespread microbial diseases, with nearly one-third of the world's population showing signs of exposure, more than 20 million people actively infected, and almost 80% of the population of some countries testing positive in tuberculin tests [1, 2]. This rate of infection is due to both a paucity of diagnostic tools [3-6] and ineffective chemotherapy in the face of emerging drug-resistance [7, 8], both of which reflect poor understanding of mycobacterial biology and host-pathogen interactions [9, 10]. One feature of mycobacterial biology that has hampered investigations is a thick, waxy cell wall consisting of a network of peptidoglycans, arabinogalactans, mycolic acids, and polysaccharides [11, 12], which makes mycobacteria resistant to lysis by most commercial chaotropic or cell lysis reagents and poses challenges to the rigorous purification of cellular biomolecules. We are concerned here with the isolation of non-coding RNA (ncRNA). The importance of rigorous ncRNA purification is illustrated by recent advances in RNA sequencing and bioinformatics that have led to the discovery of disease-relevant ncRNA species in mycobacteria [13, 14], while critical features of modified ribonucleosides in transfer RNA (tRNA) and ribosomal RNA (rRNA) are known to play a role in adaptive responses to stress [15-18]. In all cases, the systems-level analysis of ncRNA requires unbiased isolation of RNA with sequence integrity and relative quantity intact. Numerous methods for mycobacterial RNA isolation have been developed that include liquid or solid-phase extraction following cell lysis

by either sonication, enzymatic hydrolysis, chemical treatment, French pressure cell rupture, or bead-beating [19-22]. However, there has been no rigorous optimization of mycobacterial RNA isolation techniques to ensure purification of the full spectrum of ncRNA species with quantitative and qualitative fidelity. Furthermore, in addition to recognized size- and sequence-dependent biases in the isolation of specific non-coding RNA species [23], these methods require large quantities of cells or have time-consuming steps that can lead to degradation or enzymatic modification of the RNA [24]. To address these problems, we developed an optimized method for efficient isolation of all types of ncRNA from mycobacteria with high biological fidelity. Using *Mycobacterium bovis* Bacille Calmette-Guérin (BCG) as the model mycobacterial species, the method represents a combination of bead-beating with phenol-chloroform and solid-phase extraction steps optimized for both yield and quality of tRNA, 5S, 16S and 23S rRNA, as well as mRNA for seven genes. Application of the method to BCG revealed hypoxia-induced alterations of the relative quantities of 16S and 23S rRNA, a novel post-transcriptional processing of 5S rRNA, and the first complete analysis of the full set of modified ribonucleosides in mycobacterial 5S rRNA.

2.3. Material and methods

2.3.1. Bacterial cultures

For exponentially growing mycobacteria, *Mycobacterium bovis* Bacille Calmette-Guérin (str. Pasteur 1173P2; BCG) was grown at 37 $^{\circ}$C in a shaking incubator in Middlebrook 7H9 broth (Difco, BD Diagnostics, Sparks, MD) to an OD_{600} of 0.6 – 0.8. BCG cells in a hypoxia-induced non-replicative state were obtained using the Wayne model adapted from Low *et al*. [25]. Briefly, a

culture of exponentially growing BCG in Dubos broth (Difco) was diluted to an OD_{600} of 0.05 and placed in a tightly-sealed (latex-lined cap) 1 L glass flask (Duran, Mainz, Germany) with stirring at 170 rpm and air headspace volume of 450 mL. With a plateau in cell growth at 7 d (as determined by OD_{600}), induction of a non-replicative state was confirmed by the absence of change in OD_{600} for the next 14 d, at which point the cells were harvested (**Supplementary Figure 2.1**).

2.3.2. Development of the RNA isolation method

The optimized RNA isolation protocol was developed around a simple framework consisting of lysis by bead-beating in a denaturing extraction solvent, with subsequent RNA retrieval from the solvent. Using this framework, we systematically tested a variety of conditions and parameters, including extraction reagent, bead beating conditions, and RNA retrieval method, to define an optimal protocol to extract all types of ncRNA from BCG. Two nucleic acid extraction reagents were considered: TRIzol (Invitrogen, Carlsbad, CA) and phenol:chloroform:isoamyl alcohol (Qiagen, Hilden, Germany). While harshly denaturing, these reagents are insufficient to lyse mycobacterial cells. To assist in lysing cells, mechanical disruption by bead beating on a reciprocal shaker was chosen, with optimization of bead size, extraction solvent volumes, and duration. For retrieval of RNA from the extraction solvent following mechanical disruption, we compared simple isopropanol precipitation of RNA to solid-phase isolation of RNA using the Purelink miRNA Isolation Kit (Invitrogen, Carlsbad, CA). A summary of all conditions tested and their respective nucleic acid yields before DNase I treatment as assessed by UV spectroscopy is detailed in **Supplementary Table 2.1**; the analyses were performed in three biological replicates.

Supplementary Figure 2.2 illustrates the final optimized ncRNA isolation method, which was applied to BCG cultures growing exponentially or in a hypoxia-induced non-replicative state. Cells are harvested from culture medium by centrifugation at 4,000 x g for 5 min at 4 $^\circ$C. Pellets containing 10^9 cells were washed in phosphate-buffered saline (PBS) containing 0.5% Tween 80 to remove media debri that could interfere with RNA isolation and resuspended in 1 mL of ice-cold phenol:chloroform:isoamyl alcohol (volume ratio 25:24:1; saturated with 10 mM Tris, 1 mM EDTA, pH 8) and 400 µL of 2 M sodium acetate (pH 5.2). The suspension was combined with 100 µL of 0.1 mm zirconia-silica beads (Biospec, Bartesville, OK) in a 2 mL screw-capped tube and vigorously shaken on a Qiagen Tissuelyser II (Qiagen, Hilden, Germany) with chambers pre-chilled to -20 °C to account for heating to ambient temperature by the end of the 30 min run at a frequency of 30 Hz. After centrifugation (12,000 x g, 15 min, 4 $^\circ$C), the aqueous phase (~400 µL) was transferred into a new tube to which 215 µL of 100% ethanol was added with mixing and the solution loaded onto a Purelink miRNA Isolation Kit (Invitrogen, Carlsbad, CA) column (column #1) to retain large RNA species. Following centrifugation of the column (12,000 x g, 1 min), the flow-through containing small RNAs was collected, placed on ice, combined with 700 µL of 100% ethanol, and loaded onto a second Purelink miRNA Isolation column (column #2) to retain the small RNA species. The column was again centrifuged (12,000 x g, 1 min) and the flow-through discarded. Both columns (#1 and #2) were washed twice with washing buffer provided with the kit. RNase-free water (50 µL) was added to the columns, which were incubated at ambient temperature for 2 min followed by centrifugation (12,000 x g, 1 min). The eluates from columns #1 and #2 contained mainly large and small RNA, respectively, and these were combined to recreate a total RNA preparation. Aliquots of each RNA sample were adjusted to a concentration of

1 ng/µL and thee integrity and relative quantities of the RNA species in the sample were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Column eluates were stored at -80 $^{\circ}$C.

2.3.3. HPLC purification of individual ncRNA species

Individual RNA species were purified by size-exclusion (SEC) high-performance liquid chromatography (HPLC), as described in a recent publication [26]. Following addition of 30 units of DNAse I (Qiagen, Hilden, Germany) to each sample, the total RNA preparation in 20 µL was injected onto an Agilent Bio SEC3 column (300 Å, 7.8 x 100 mm; Agilent, Foster City, CA) attached to an Agilent 1200 HPLC system to resolve tRNA and 5S rRNA species. For 16S and 23S rRNA, an Agilent Bio SEC5 column (1000 Å, 7.8 x 100 mm) was used. Both columns were eluted with an isocratic gradient with 100 mM ammonium acetate (pH 7.5) at 0.5 mL/min and 60 $^{\circ}$C with RNA elution monitored by UV absorbance. Peak resolution (R) was calculated using the following equation where tRA and tRB are the retention times for peaks A and B, respectively, and WA and WB are the peak base widths for peaks A and B, respectively:

$$R = 2\left(\frac{tRB - tRA}{WA + WB}\right)$$

Fractions containing individual RNA species were analyzed on an Agilent 2100 Bioanalyzer. The 5S rRNA fraction was concentrated and desalted using a 2,000 Da molecular weight cut-off spin filter (Sartorius-Stedium, Goettingen, Germany). Purified 5S rRNA was quantified by integrating the UV absorbance for the 5S rRNA peak in HPLC profiles and interpolating concentration from linear standard curves prepared with purified 28S rRNA (CCRF-SB cells), as shown in **Supplementary Figure 2.3**.

2.3.4. Sequencing of BCG 5S rRNA

To confirm the identity of the purified 5S rRNA, the 3' end of 25 pmol of the RNA was ligated to 50 pmol of a pre-adenylated DNA adaptor (IDT, Coralville, IA) using 1.5 µg of RNA ligase 2 (truncated K227Q, NEB, Beverly, MA) with incubation at 22 °C for 16 h. The ligated product was purified by SEC3 HPLC and then reverse transcribed with 200 U of Superscript III RT (Invitrogen, Carlsbad, CA). The resulting cDNA was ligated to 50 pmol of another pre-adenylated 3' adaptor (IDT, Coralville, IA) using 10 U of T4 RNA ligase 1 (NEB, Beverly, MA) with incubation at 4 °C for 16 h. Following SEC3 HPLC purification, the ligated cDNA was PCR amplified, TA cloned (Genewiz, South Plainfield, NJ), and sequenced to confirm its identity (Genewiz, South Plainfield, NJ). Sequence analysis was performed with Nucleotide BLAST (NCBI) against the annotated full genome sequence of the same BCG strain (NC_008769.1). The cDNA sequence from mature 5S rRNA is available on GeneBank (Accession number KC203333).

2.3.5. Analysis of mRNA by quantitative real-time polymerase chain reaction (qPCR)

To compare the mRNA yield from the optimized RNA isolation approach to that of the conventional TRIzol method, RNA was isolated from exponentially growing BCG cultures using TRIzol or the optimized RNA isolation method and DNA was removed using the TURBO DNA-free Kit (Ambion, Life Technologies, Carlsbad, CA). RNA (50 ng/µL) was reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA) according to the manufacturer's instructions. The reverse transcription program was run as follows: 25 $^{\circ}$C for 5 min, 42 $^{\circ}$C for 30 min and 85 $^{\circ}$C for 5 min, followed by a cooling step at 4 $^{\circ}$C. Two-step real-time qPCR was then performed using the

SsoAdvanced universal SYBR Green supermix (Bio-Rad, Hercules, CA).

Primer sequences and melting temperatures can be found in **Supplementary Table 2.3**. The qPCR program was run as follows: 95 $^{\circ}$C for 30 s followed by 40 cycles of denaturation at 95 $^{\circ}$C for 15 s and annealing/extension at 60 $^{\circ}$C for 30 s. A melting curve analysis consisting of 0.5 $^{\circ}$C increments from 65 to 95 $^{\circ}$C was performed for all reactions to ascertain the specificity of the primers.

2.3.6. Identification and characterization of modified ribonucleosides in BCG 5S rRNA

The full spectrum of modified ribonucleosides in the purified 5S rRNA was characterized by analysis of ribonucleoside hydrolysates of the RNA by chromatography-coupled quadrupole time-of-flight and triple quadrupole mass spectrometry, using methods described previously [15, 26, 27]. Briefly, purified BCG 5S rRNA was enzymatically hydrolyzed in the presence of antioxidants and deaminase inhibitors and the ribonucleosides resolved by HPLC using a Hypersil GOLD aQ Analytical HPLC Column (100 × 2.1 mm, 1.9 μm particle size) (Thermo Scientific, Wilmington, DE) coupled to either an Agilent 6460 triple quadrupole mass spectrometer or an Agilent 6520 quadrupole time-of-flight mass spectrometer (LC-MS/MS) with ESI ionization operated in positive ion mode. The identity of individual ribonucleosides was confirmed by comparison with synthetic standards for HPLC retention time, exact molecular weight, and collision-induced dissociation (CID) fragmentation patterns. In cases where a standard was not available, a putative structure was inferred from exact molecular weight and collision-induced dissociation (CID) fragmentation patterns**.**

2.3.7. Statistical analysis

Statistical analysis for comparison of qPCR C_T values for the RNA isolation methods and quantities of RNA species in exponentially growing and hypoxic, non-replicative (Wayne day 21) cultures was performed using Student's t-test with significance set at $p \leq 0.05$.

2.4. Results

Given the challenges of working with mycobacteria, we sought to develop a method for purification of ncRNA that optimized the cell lysis, RNA extraction and RNA purification steps to yield the full spectrum of intact ncRNA species in quantities that reflect their abundances *in vivo*. The workflow shown in **Supplementary Figure 2.2** represents the culmination of a series of optimization studies in which different procedures were assessed for the yield of RNA, the relative quantities of the various ncRNA species and the quality of the RNA in terms of degradation. Subsequent analysis of the isolated ncRNA species from BCG revealed several novel features of RNA biology in mycobacteria.

2.4.1. Optimization of RNA isolation parameters

With all of the optimization data shown in **Supplementary Table 2.1**, optimization of the ncRNA isolation method was initiated with a comparison of two standard extraction reagents, TRIzol and the 25:25:1 mixture of phenol, chloroform, and isoamyl alcohol. An extraction reagent was added to the cells prior to bead beating to quench adventitious enzymatic reactions during cell lysis, including nucleases and RNA modification enzymes. Given the evidence for biased isolation of miRNA with TRIzol [23], we compared the

yields of the major ncRNA species in BCG with TRIzol and phenol:chloroform:isoamyl alcohol. A 77% increase in nucleic acid recovery was obtained with phenol:chloroform:isoamyl alcohol compared to TRIzol (**Supplementary Table 2.1**). Since phenol:chloroform:isoamyl alcohol saturated with 10 mM Tris, pH 8.0, also extracts DNA, it is possible that part of the observed increase in nucleic acid mass results from DNA contamination. However, the yield of longer RNA species, in particular 16S and 23S rRNA, was significantly reduced using traditional acid phenol or TRIzol-based RNA extraction methods (**Supplementary Figure 2.4**). Hence, it was preferable to use the phenol:chloroform:isoamyl alcohol approach to obtain all cellular RNA species in high yield and in proportions that truly reflect their abundances *in vivo*, with contaminating DNA subsequently removed by DNAse I digestion or HPLC purification, without sacrificing longer RNA species **(Supplementary Figure 2.5)**.

Next, we optimized for cell lysis, comparing the RNA yield obtained using 0.5 mm and 1.0 mm glass beads and 0.1 mm zirconia-silica beads. While 0.5 mm glass beads offered a 61% increase in nucleic acid yield against 1.0 mm glass beads, we observed an additional 37% increase in RNA yield with 0.1 mm zirconia-silica beads (**Supplementary Table 2.1**). Using 0.1 mm zirconia-silica beads, we optimized for the duration of bead beating. A duration of 10 min was sufficient for lysis of mycobacterial cells using the TissueLyzer II, while a duration of 30 min was required for the TissueLyzer LT. Use of a French press was ruled out due to the time required for the lysis process (>15 min) in the absence of denaturants, which allows enzymatic reactions to affect the spectrum of tRNA and rRNA modifications or cause RNA degradation.

Finally, to recover RNA from the extraction solvents, we chose the Purelink miRNA Isolation Kit system over isopropanol-based RNA precipitation. While the Purelink columns were optimized for use with TRIzol, we modified the manufacturer's protocol to use phenol:chloroform:isoamyl alcohol instead, which allowed for the maximal recovery of both large and small RNA. Using 35% ethanol, larger RNA species were recovered on column #1 while smaller RNA species were recovered on column #2 using 70% ethanol.

2.4.2. Qualitative assessment of the optimized ncRNA isolation method

To assess the performance of the optimized RNA isolation method and the quality of the isolated ncRNA, the method was applied to BCG cells growing exponentially and to cells arrested in a hypoxia-induced non-replicative state that mimics the granuloma of pulmonary mycobacterial infections. The quality and relative quantities of the various ncRNA species in total RNA extracted from BCG were assessed using a Bioanalyzer with either the Agilent Pico Chip for large RNA (16S, 23S) and the Agilent Small RNA Chip for small RNA species (<200 nt; miRNA-sized, tRNA, 5S rRNA). As shown in **Figure 2.1**, the peaks corresponding to individual ncRNA species were sharp and distinguishable. From the exponentially growing cultures, the eluate from column #1 contained 23S rRNA, 16S rRNA, 5S rRNA and tRNA (**Figures 2.1A, B**), whereas column #2 eluate contained predominantly tRNA (**Figures 2.1C, D**). As a numeric index of the quality of the isolated RNA, the RNA integrity number (RIN) determined from the Bioanalyzer profile ranges from 10 for theoretically intact RNA to 1 for completely degraded RNA [28]. For exponentially growing BCG, the RIN for RNA eluted from column #1 in four independent extractions was 9.0 ± 0.5, while the RIN for the combined eluates from columns #1 and #2 from five independent extractions was 8.5 ±
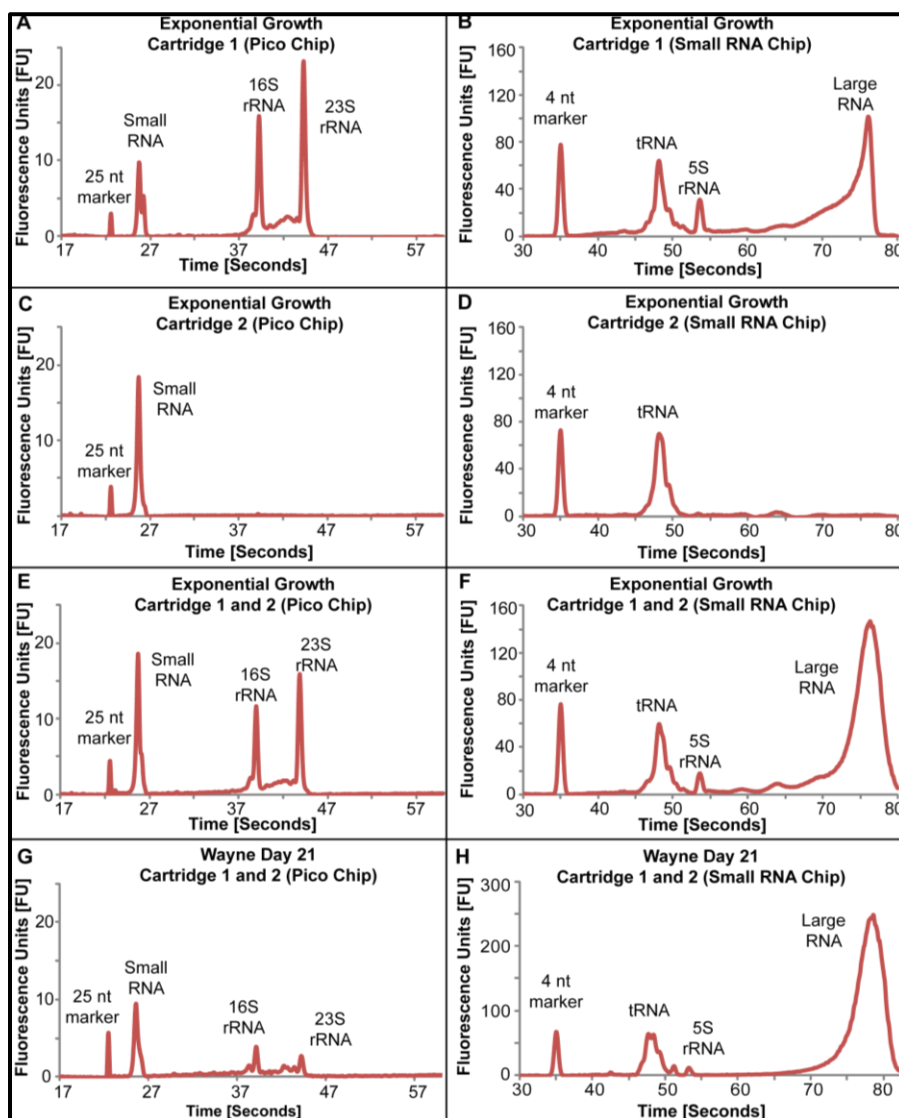
Figure 2.1. Representative Bioanalyzer electropherograms for BCG RNA recovered from Purelink miRNA Isolation columns #1 and #2. Electropherograms from column #1 RNA resolved on (**A**) a Pico Chip and (**B**) a Small RNA Chip, and from column #2 RNA resolved on (**C**) a Pico Chip and (**D**) a Small RNA Chip. Electropherograms for the combined RNA from columns #1 and #2 resolved on (**E**) a Pico Chip and (**F**) Small RNA Chip. RNA isolated from BCG subjected to hypoxic stress (combined columns #1 and #2) was resolved on (**G**) a Pico Chip and (**H**) a Small RNA Chip. The electropherogram profiles presented here are representative of three independent experiments. Note that the small RNA chip on the Agilent Bioanalyzer 2100 system is not optimized to read large RNA species (>150 nucleotides). Thus the amount of RNA in the 'large RNA' peak cannot be accurately quantified.

0.3. Interestingly, Bioanalyzer RNA profiles of the hypoxic cultures displayed lower relative intensities of the 16S and 23S rRNA peaks compared to those from exponentially growing cultures (**Figure 2.1G**), with comparable quantities of tRNA under both conditions (**Figure 2.1H**). The RIN for the combined eluates from columns #1 and #2 from four independent RNA extractions of hypoxic bacilli was 5.9 ± 0.2, which suggested a degree of degradation of both 23S and 16S rRNA peaks (**Figure 2.1G**).

To assess the performance of the Purelink miRNA Isolation columns, the proportions of ncRNA species in column eluates were calculated by manual integration of electropherogram peaks with Bioanalyzer 2100 Expert software. A comparison of the TRIzol and phenol:chloroform:isoamyl alcohol methods is summarized in **Supplementary Table 2.2**. Using this approach, large ncRNAs (16S and 23S, identified by size) were almost entirely recovered from column #1 along with a significant portion of the small ncRNAs (<200 nt), while small ncRNAs represented >95% of the RNA eluted from column #2. The small ncRNAs, whose identities were assigned based on size, consisted of three major species: miRNA-like species, tRNA and 5S rRNA. This retention of small ncRNAs on column #1 points to potential bias in the small ncRNAs recovered from column #2 as "small RNA". However, by combining the eluates from both columns, the yield of all ncRNA species was maximized (**Figures 2.1E, F**), with small RNA, and mainly tRNA, as the dominant species of ncRNA in BCG (**Supplementary Table 2.2A**). For TRIzol however, small RNA was the predominant species of ncRNA, with tRNA consisting of more than 80% of the small RNA species **(Supplementary Table 2.2B)**.

Considering both the RIN and the recovery of all ncRNAs from the columns using the phenol:chloroform:isoamyl alcohol approach, the results with

exponentially growing BCG RNA suggest that the degradation apparent in hypoxic, non-replicative BCG is not the result of the RNA isolation procedure and instead reflects physiological processing of the RNA. These differences were next assessed more quantitatively.

2.4.3. Application of the mycobacterial RNA isolation method in isolating mRNA

While optimized for the isolation of ncRNA, we investigated the possibility of utilizing our approach for mRNA isolation. Here, we performed qPCR against seven gene transcripts on samples derived from RNA isolation with our approach and samples derived from the conventional TRIzol RNA isolation approach. Analysis revealed that there were no significant differences in C_T values against five gene transcripts as illustrated in **Table 2.1**. The remaining two transcripts showed that our RNA isolation method yielded a lower C_T value (i.e. higher concentration of mRNA) compared to the TRIzol approach.

2.4.4. Application of the mycobacterial RNA isolation method: Quantitative comparison of ncRNA species in non-replicative and exponentially growing BCG

Given the differences in the relative quantities of 23S and 16S rRNAs between exponential and non-replicative cultures noted in the Bioanalyzer electropherograms (**Figures 2.1E, G**), we undertook a more rigorous quantitative analysis by HPLC resolution of total RNA extracts from BCG during large-scale purification of ncRNA species. As shown in **Figures 2.2A** and **2B**, we were able to achieve baseline resolution (R = 3.9 ± 0.02) of 16S and 23S rRNAs using an Agilent size-exclusion HPLC column (SEC5 1000 Å), while tRNA and 5S rRNA were well resolved (R = 4.1 ± 0.04) using a solid phase with a smaller pore size (SEC3 300Å). The absolute quantities of

Table 2.1. C_T values obtained from qPCR analysis of samples derived from TRIzol and the optimized RNA isolation approach

| Gene | Average C_T Value | | Student's T Test |
	P:C:IAA (Optimized Approach)	TRIzol	P-value
hspX	21.38 ± 0.20	23.60 ± 0.91	0.014
ethA	27.34 ± 0.11	27.74 ± 1.13	0.608
fdxA	20.62 ± 0.31	20.45 ± 0.56	0.663
relA	23.82 ± 0.33	25.08 ± 0.89	0.081
sigA	24.92 ± 0.15	24.70 ± 0.52	0.093
dosR	23.73 ± 0.36	25.13 ± 0.82	0.053
dosS	24.92 ± 0.32	26.53 ± 0.76	0.028

An unpaired Student's t-test was performed for the samples derived from the two approaches. Data represent mean ± SD for three biological replicates.
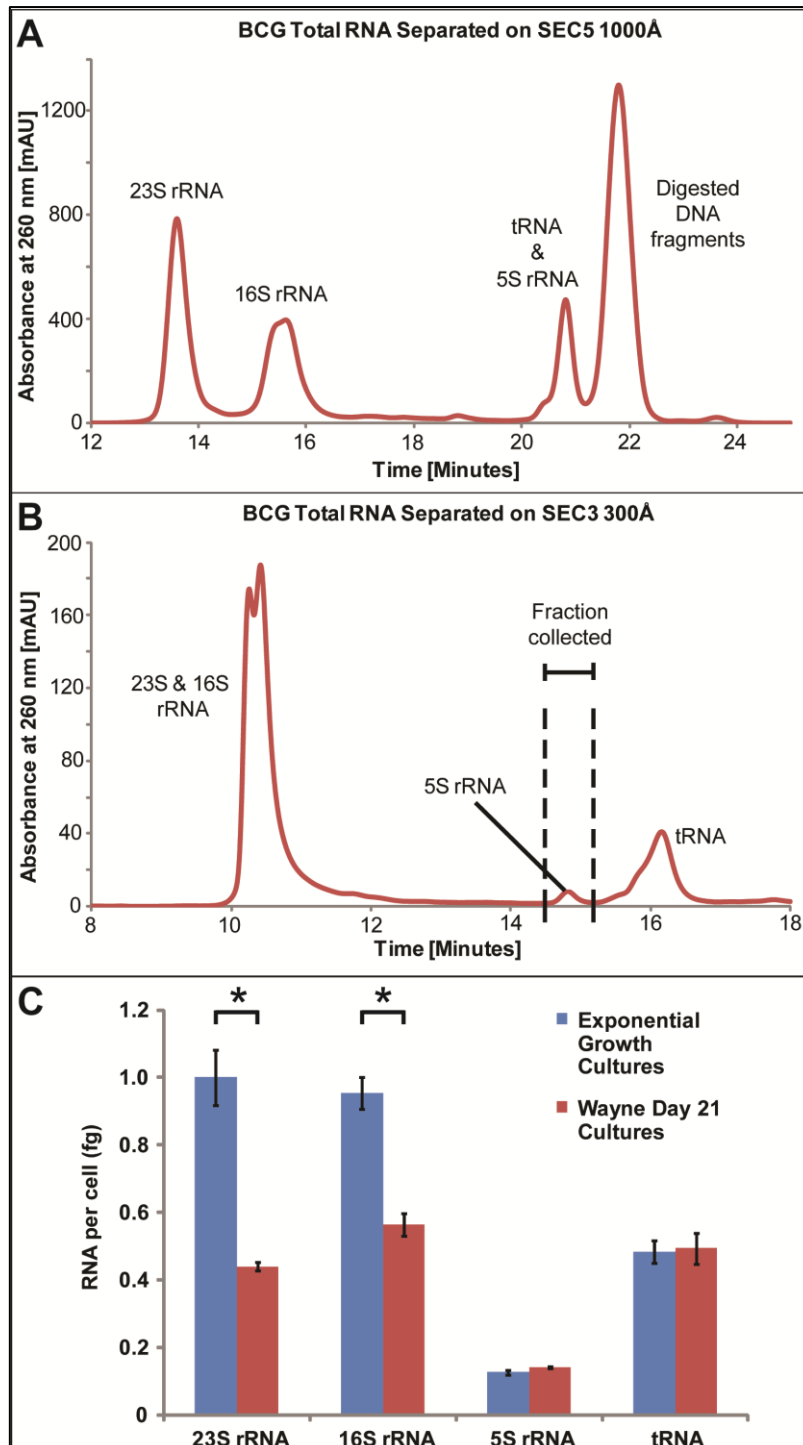
Figure 2.2. Size-exclusion HPLC chromatograms for individual RNA species. BCG total RNA was resolved on an Agilent Bio Size Exclusion Column SEC5, 1000Å (**A**), or on an Agilent Bio Size Exclusion Column SEC3, 300Å (**B**). Panels **A** and **B** are representative of six independent HPLC runs. (**C**) Cellular quantities of 5S, 16S and 23S rRNA and tRNA species based on HPLC UV absorbance interpolated from a standard curve based on purified human 28S rRNa. An asterisk indicates a significant difference in the quantity of RNA from exponentially growing cells compared to hypoxic cells on day 21 (unpaired T-test, $p \leq 0.05$). Data represent mean ± SD for three biological replicates.

individual ncRNA species were calculated by interpolating the UV absorbance peak areas for HPLC-resolved ncRNA species with a standard curve prepared from purified 28S rRNA. These quantities were then combined with cell counts from colony forming assays for cultures used in the RNA isolation to calculate the quantity of each rRNA species on a per cell basis. As shown in **Figure 2.2C**, the quantities of 16S and 23S rRNA from exponentially growing BCG were significantly higher than in hypoxia-induced non-replicative BCG ($p \leq 0.05$), yet there was no difference in the quantities of 5S rRNA and tRNA for the two growth conditions. These data for cellular quantities of ncRNA are consistent with theoretical estimates of mycobacterial molecular composition [29].

2.4.5. Sequence of BCG 5S rRNA

Using the optimized RNA isolation method with BCG, we were able to reproducibly detect an RNA species with a size consistent with 5S rRNA. Though relatively abundant when compared to other bacterial small RNAs, the observation of 5S rRNA is notable in light of the fact that this species has never been isolated for direct sequencing. In other studies of mycobacterial ncRNA, small RNA were identified by genomic analysis, by microarray detection of non-coding transcripts, by shotgun cloning of small RNAs, or by co-purification with RNA-binding proteins [30-34]. Here, pure 5S rRNA-containing fractions were collected from the HPLC resolution and the 5S rRNA subjected to sequencing to confirm its identity and to define the sequence of its mature, functional form. Analysis revealed a sequence of 109 nt that, in a Blast search, yielded an expect value (E) of $4 \times e^{-49}$ and a maximal identity of 100% for several *Mycobacterium* species, including BCG, *M. tuberculosis*, *M. canettii*, and *M. africanum.* Comparison of the annotated BCG 5S rRNA sequence (*M. bovis BCG str. Pasteur 1173P2*, Gene ID:

4698678) and the cDNA sequence of our purified 5S rRNA revealed several differences in the 3'- and 5'-ends, as shown in **Figure 2.3B.** The purified 5S rRNA lacked 6 nt: positions 1 to 5 and position 115 in the annotated 5S rRNA sequence. These results suggest discrete processing of 5S rRNA from its genomic transcript as a 9S precursor to the functional sequence by a mycobacterial enzyme such as RNase E/G (MycRne), as illustrated in **Figure 2.3C**.

2.4.6. The spectrum of modified ribonucleosides in BCG 5S rRNA

To take advantage of the ability to isolate 5S rRNA using the optimized method, we further explored 5S rRNA maturation by analysing post-transcriptional modifications. Here, we used a highly sensitive chromatography-coupled mass spectrometric platform to identify modified ribonucleosides in RNA [15, 26, 27]. As shown in **Table 2.2**, application of this platform to BCG 5S rRNA revealed seven modified ribonucleosides: N^1-methyladenosine (m^1A), N^6,N^6-dimethyladenosine ($m^{6,6}A$), N^7-methylguanosine (m^7G), N^1-methylguaonsine (m^1G), 2'-*O*-methylguanosine (Gm) and inosine (I), each with distinct mass-to-charge transition ratios and retention times as shown in **Supplementary Figure 2.6**. The structures of six of the modified ribonucleosides were confirmed with synthetic standards, while the structure of the seventh is tentatively assigned as N^4-methyl-2'-*O*-methylcytidine (m^4Cm) based on collision-induced dissociation fragmentation patterns.

2.5. Discussion

While there are established approaches for isolating specific types of RNA from mycobacteria [19-22], there are no methods that have been optimized to

yield the complete set of commonly recognized ncRNA species (tRNA and 5S, 16S, and 23S rRNA). The value of preserving the fidelity of a cellular RNA population by isolating the full complement of ncRNA is illustrated in recent discoveries of many new ncRNA species in *M. tuberculosis* based on genomic analyses [13, 14] and with the application of quantitative RNA-seq methods to both coding and ncRNA in mycobacteria [35]. Furthermore, the systems-level analysis modified ribonucleosides in mycobacterial ncRNA also requires isolation of the full spectrum of ncRNA species [15-17]. To address these challenges, we used *M. bovis* BCG to develop an optimized method for mycobacterial RNA isolation that combines accessible reagents and approaches with commercial kits to achieve rapid isolation of the full spectrum of ncRNA in high yield with minimal artefacts. The method yielded high quality ncRNA and led to the discovery of novel post-transcriptional processing of 5S rRNA in BCG.

A step-by-step analysis of the RNA isolation method provides insights into the importance of preserving both the quantitative and qualitative fidelity of the isolated RNA. Perhaps the most challenging step in mycobacterial RNA isolation involves cell lysis, owing to the unique biochemistry of mycobacterial cell walls, which renders them refractory to typical chaotropic or mild detergent-based cell lysis solutions used with other bacteria and eukaryotic cells. Several methods to overcome this problem have been developed, including enzymatic spheroplasting followed by sonication [19], detergent-based lysis [20], mechanical disruption by bead-beating [21], and mechanical shearing with a French pressure cell [22]. However, in the absence of denaturants, all of these methods provide an opportunity for adventitious enzymatic reactions to occur during the lysis step, including nuclease degradation and alteration of modified ribonucleosides in the RNA. Our
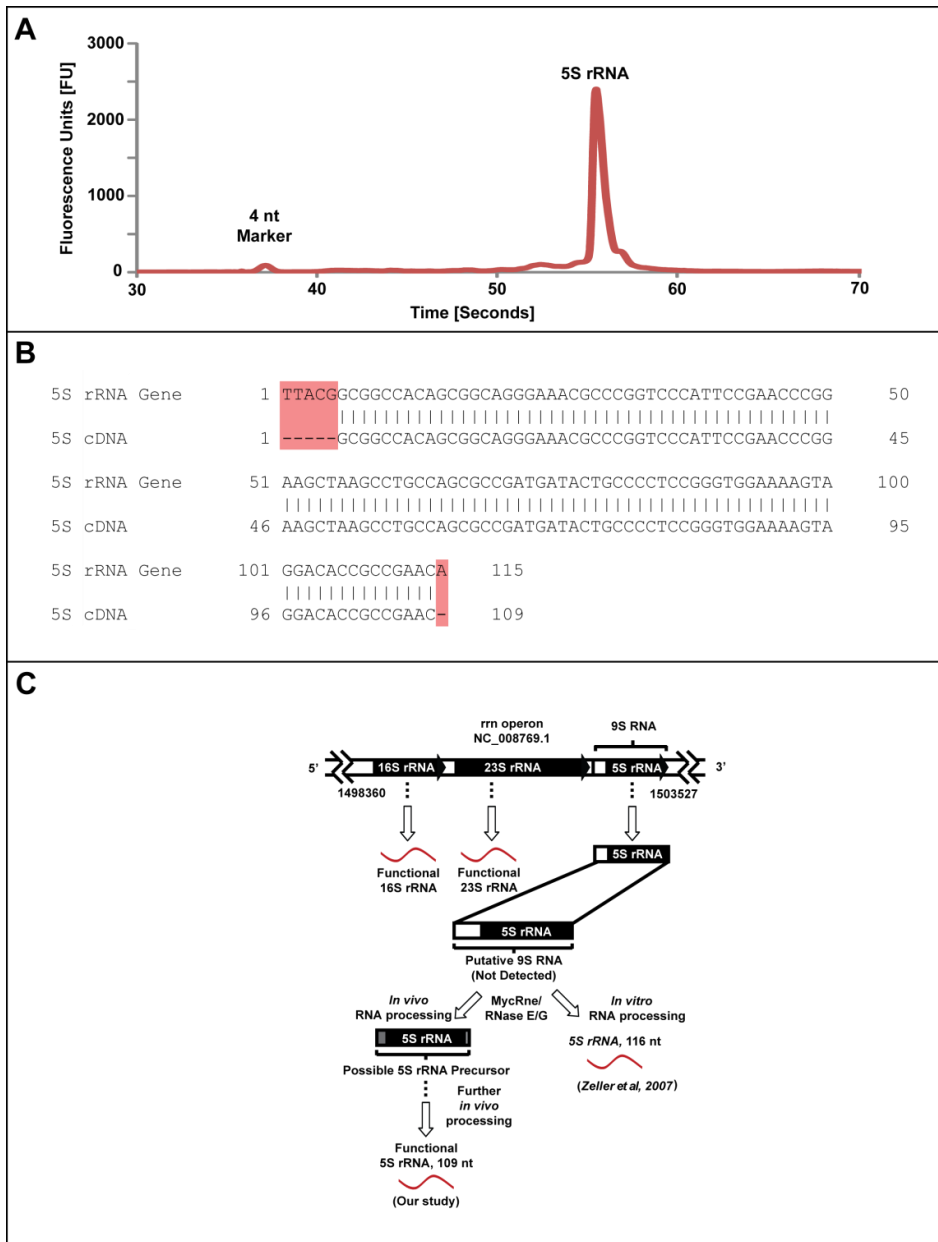
Figure 2.3. Purification and sequencing of BCG 5S rRNA and a proposed model for 5S rRNA processing. (**A**) A Bioanalyzer electropherogram for HPLC-purified 5S rRNA shows high purity. (**B**) Comparison of sequenced 5S rRNA against its annotated gene sequence. A total of six nucleotides are absent from the sequenced 5S rRNA (highlighted in red). (**C**) Possible fates of the putative 9S RNA, a precursor of functional 5S rRNA. Zeller *et al.* demonstrated *in vitro* processing of 9S rRNA to a 116 nt 5S rRNA by MycRNE/RNase E/G [36], while our analysis reveals different or further *in vivo* processing to yield a 109 nt functional 5S rRNA molecule.

Table 2.2. List of modified bases detected in BCG 5S rRNA.

Name	Abbreviation	Formula	Precursor (m/z)	Product (m/z)	Transition (m/z)	Retention Time (Minutes)
Adenosine	A	$C_{10}H_{13}N_5O_4$	268.1	136.1	132	5.09
Uracil	U	$C_4H_4N_2O_2$	245.1	113.1	132	2.45
Guanosine	G	$C_{10}H_{13}N_5O_5$	284.0	152.0	132	6.44
Cytidine	C	$C_4H_5N_3O$	244.1	112.1	132	1.65
Inosine	I	$C_{10}O_5N_4H_{12}$	269.1	137.1	132	5.38
1-methyladenosine	m1A	$C_{11}O_4N_5H_{15}$	282.1	150.1	132	2.53
N6,N6-dimethyladenosine	m6,6A	$C_{12}O_4N_5H_{17}$	296.1	164.1	132	20.49
7-methylguanosine	m7G	$C_{11}O_5N_5H_{17}$	298.1	166.1	132	4.69
1-methylguanosine	m1G	$C_{11}O_5N_5H_{15}$	298.1	166.1	132	13.72
2'-O-methylguanosine	Gm	$C_{11}O_5N_5H_{15}$	298.1	152.1	146	14.54
*N4,2'-O-dimethylcytidine	m4Cm	$C_{11}O_5N_3H_{17}$	272.1	112.1	160	22.64

Identity of modifications determined by comparisions of accurate mass measurements on QTOF and retention times of chemical standards on the LC-QTOF and LC-QqQ. *Based on accurate mass of precursor ion on QTOF and neutral loss of 2'O methyl ribose on QqQ; no standards available.

approach of bead-beating in the presence of a strong denaturant such as phenol:chloroform:isoamyl alcohol minimizes artefacts by reducing the time required for lysis and by quenching all physiological and biochemical processes.

It is also important to consider the choice of denaturing extraction solvents as a determinant the fidelity of the RNA population. The introduction of the 'single-step' method of RNA isolation using acid guanidinium thiocynate-phenol-chloroform (*i.e.*, TRIzol) provided versatility and ease in retrieving RNA from a wide variety of sources [37]. However, Kim *et al*. recently described poor recovery of small RNA with low GC content when utilizing the isopropanol precipitation step in the TRIzol approach, challenging the assumption that total RNA preparations recover all RNA species with equal efficiency [23]. We also observed that TRIzol yields a smaller proportion of 16S and 23S rRNA from mycobacteria compared to the phenol, chloroform, isoamyl alcohol mixture (**Supplementary Figures 2.2B and 2.4**). The alternative use of cetyltrimethylammonium bromide with phenol and bead-beating for RNA extraction from mycobacteria by Cheung *et al*. [38] proved to be problematic [20]. For retrieval of RNA from the extraction solvents, an isopropanol precipitation step is frequently used, though this results in biases against small RNAs. Given these issues of extraction and recovery, we found that the combination of phenol:chloroform:isoamyl alcohol extraction with bead-beating followed by solid-phase extraction columns for RNA retrieval provided optimal RNA yields in terms of quantity and quality. As the use of phenol:chloroform:isoamyl alcohol at an alkaline pH increases the co-purification of genomic DNA, a DNA removal step, such as addition of DNase I or HPLC purification, should be employed for downstream applications involving PCR amplification or chemical analysis of RNA. The solid phase

extraction technology is based on nucleic acid binding to silica-based polymers with RNA retention dependent on the concentration of salt and ethanol in the aqueous phase [39]. Lower percentages of ethanol (*e.g.* 35%) allow retention of larger RNA while higher percentages of ethanol (*e.g.*, 70%) retain small RNA. While there is significant loss of small RNAs, most notably tRNA, on the column supposedly retaining large RNAs (column #1 in the current method), quantitative recovery of all RNA species is achieved by combining the eluates from both columns.

This optimized approach of using denaturing extraction solvents during bead-beating to lyse mycobacteria, followed by solid-phase recovery of RNA allows for rapid, efficient isolation of the complete set of ncRNA species with minimal degradation artefacts. The efficiency of the method is evident in the yield of DNA-free RNA, which amounted to a reproducible 2.6 μg of RNA from 10^9 bacilli (*i.e.*, 2.6 fg per cell). This amount of cellular RNA is consistent with recent studies on the molecular composition of mycobacteria that gave a theoretical total content of RNA in a single BCG bacillus at 2.2-10 fg with dependence on the growth rate [29]. Given our isolation results and the agreement with theoretical RNA amounts, previous claims of 20 fg RNA per cell for mycobacteria are likely over-estimates possibly resulting from DNA or other contamination [20, 38]. For the highest degree of purity needed for chemical analysis of RNA, for example, individual RNA species can purified by one-, two- or even three-dimensional chromatography combining size-exclusion with reversed-phase matrices [26]. This step provides an opportunity to remove contaminating DNA and DNase I. In our experiments, we show essentially complete digestion of DNA by DNase I in order to minimize DNA contamination in purification steps (**Supplementary Figure 2.5**).

The immediate application of this method to the *M. tuberculosis* surrogate, *M. bovis* BCG, which is also the tuberculosis vaccine strain, led to three important observations: pathophysiological shifts in the relative quantities of ncRNA species, and novel post-transcriptional processing of the elusive 5S rRNA in terms of sequence trimming and an expanded set of modified ribonucleosides. Among a variety of environmental changes that occur in the hallmark pathology of *M. tuberculosis* infection, oxygen depletion *in vitro* and granulomas have been shown to induce a non-replicative persistent state that mimics the dormant state of tuberculosis bacilli [40]; this phenotype is shared by BCG and the faster growing *M. smegmatis* [41]. Consistent with the observation of a hypoxia-induced decrease in nucleic acid synthesis [42] and loss of both 50S and 30S ribosomal subunits [41], we observed significant decreases in the relative quantities of 16S and 23S at the fully developed non-replicative state at 21 days of hypoxia (**Figure 2.2C**). Surprisingly, the levels of small RNAs did not change significantly in the hypoxic state, either in terms of Bioanalyzer RNA profiles or HPLC quantification. Furthermore, the decrease in the relative quantity of 16S rRNA raises questions about the feasibility of using bacterial 16S rRNA a housekeeping gene in bacterial gene expression studies, particularly those examining stress conditions. While possible, it unlikely that the thicker cell wall of the hypoxic, non-replicative state could lead to less cell wall disruption during the cell lysis step and cause differential release of small and large ncRNA species.

The optimized RNA isolation method also yielded insights into the maturation and processing of 5S rRNA in mycobacteria. In 1978, using temperature-sensitive *rne E. coli* mutants, several small RNA molecules were discovered, including 9S RNA that contained the 5S rRNA sequence [43]. Further studies

demonstrated that processing of 9S RNA by RNase E and/or RNase G yields what was thought to be a mature 5S rRNA [43, 44]. An RNase E/G homologue was identified from *M. tuberculosis* and was characterized as a 5'-endoribonuclease able to cleave putative 9S RNA *in vitro* as well as *in vivo* [36]. This RNase E/G homologue, termed MycRne, was shown to cleave 9S RNA in proximity to, but upstream of, the annotated genomic 5S rRNA sequence, with the difference attributed to the association of ribosomal proteins. Using the RNA isolation method with HPLC purification, we were able to obtain a homogeneous population of 5S rRNA molecules for sequence analysis and chemical analysis of modified ribonucleosides. Sequencing of purified 5S rRNA uncovered two sites of truncation with respect to the BCG 5S rRNA gene, as shown in **Figure 2.3B**. Our findings strongly support the idea of a cleavage downstream from, but in proximity to, the start of the annotated genomic sequence, as well as a cleavage one nucleotide upstream from the end sequence to yield a functional 5S rRNA. To determine the cleavage site of 9S RNA, the previously mentioned group incubated PCR-amplified putative 9S sequences with a truncated form of MycRne protein that was overexpressed in *E. coli.* Additionally, their group utilized primer extension of total RNA to determine the position of the 5' end of 5S rRNA. As the sequences differed by one or two nucleotides, they attributed the differences to the presence of ribosomal proteins [36]. The uniformity of the observed 5S rRNA sequence in our studies rigorously establishes the functional form of this RNA.

In addition to sequence alterations, ncRNAs undergo maturation in form of post-transcriptional insertion of dozens of different modified ribonucleosides, especially tRNA. To date, however, there have been very few observations of modified ribonucleosides in 5S rRNA in either eukaryotes or prokaryotes.

Modified ribonucleosides in 5S rRNA have been described in archea and to a smaller extent, eukaryotes [45, 46]. Although Yan *et al.* claim to have detected Am, Cm, m^{3}C, m^{6}A in mouse liver 5S rRNA, and while Cm, ac^{4}C, ac^{4}Cm have been detected in *Sulfolobus solfataricus, Pyrodictium occultum*, and *Haloferax volcanii* 5S rRNA [46, 47], the Albany Modification Database lists only four modifications in 5S rRNA: Cm, ac^{4}C, ac^{4}Cm, and Y [48]. Our analysis of BCG 5S rRNA thus adds seven more modified ribonucleosides to the 5S rRNA roster: m^{1}A, m$_6^6$A, m^{7}G, m^{1}G, Gm, I and a putative m^{4}Cm (**Table 2.2**). Given the importance of rRNA modifications in 16S and 23S rRNA in ribosome assembly and translational fidelity, as well as in mechanisms of resistance to aminoglycoside, macrolide and linezolid antibiotics [49-52], it is reasonable to assume that 5S rRNA modifications have potential effects with respect to bacterial growth and survival.

Given that the use of phenol:chloroform:isoamyl alcohol at an alkaline pH partitions nucleic acids in the organic phase (including mRNA), we sought to determine if our method was suitable for mRNA isolation or gene expression studies. We thus probed for the presence of seven mRNA transcripts on samples derived from our RNA isolation approach and samples derived from the conventional TRIzol RNA isolation approach. Analysis showed that the optimized phenol:chloroform:isoamyl alcohol approach resulted in a similar yield of mRNA when compared to TRIzol-based RNA isolation, while the RIN data attests to the quality of the total RNA preparation. We therefore believe that total RNA derived from our approach can be used for gene expression studies.

The results presented here demonstrate the utility of a mycobacterial RNA isolation method that maximally preserves the quantitative and qualitative

fidelity of ncRNA species. While optimized for ncRNA, the method also

provides high fidelity isolation of mRNA from mycobacteria.

2.6. Supplementary material

2.6.1 Supplementary figures
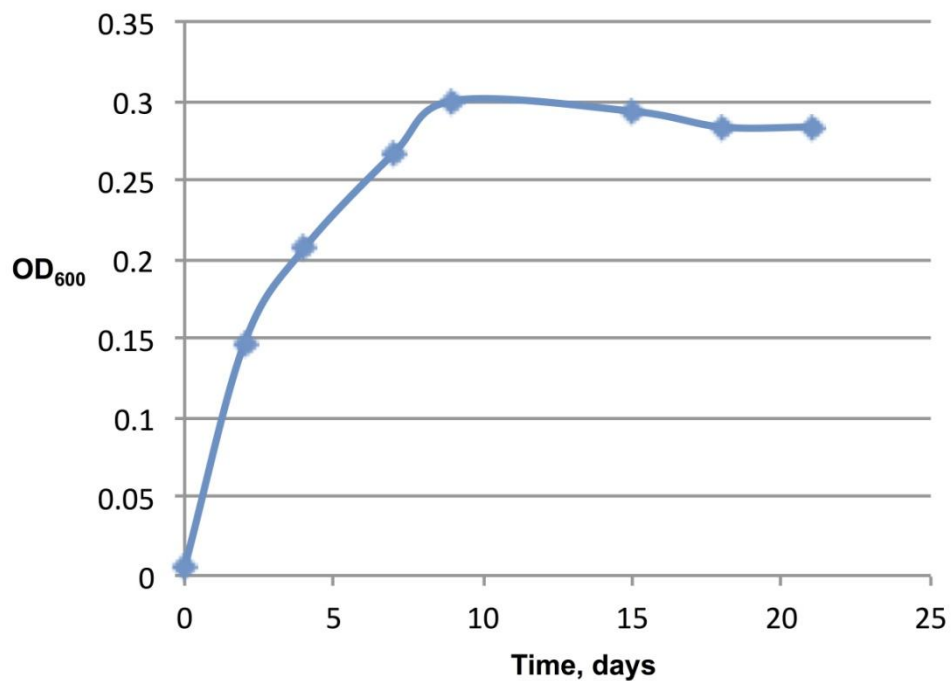
Supplementary Figure 2.1. Growth of *M. bovis* BCG under hypoxia.
Cultures of BCG were sealed on day = 0 and the OD_{600} determined regularly
for 21 days. The plateau in growth represents the hypoxia-induced non-
replicative state.
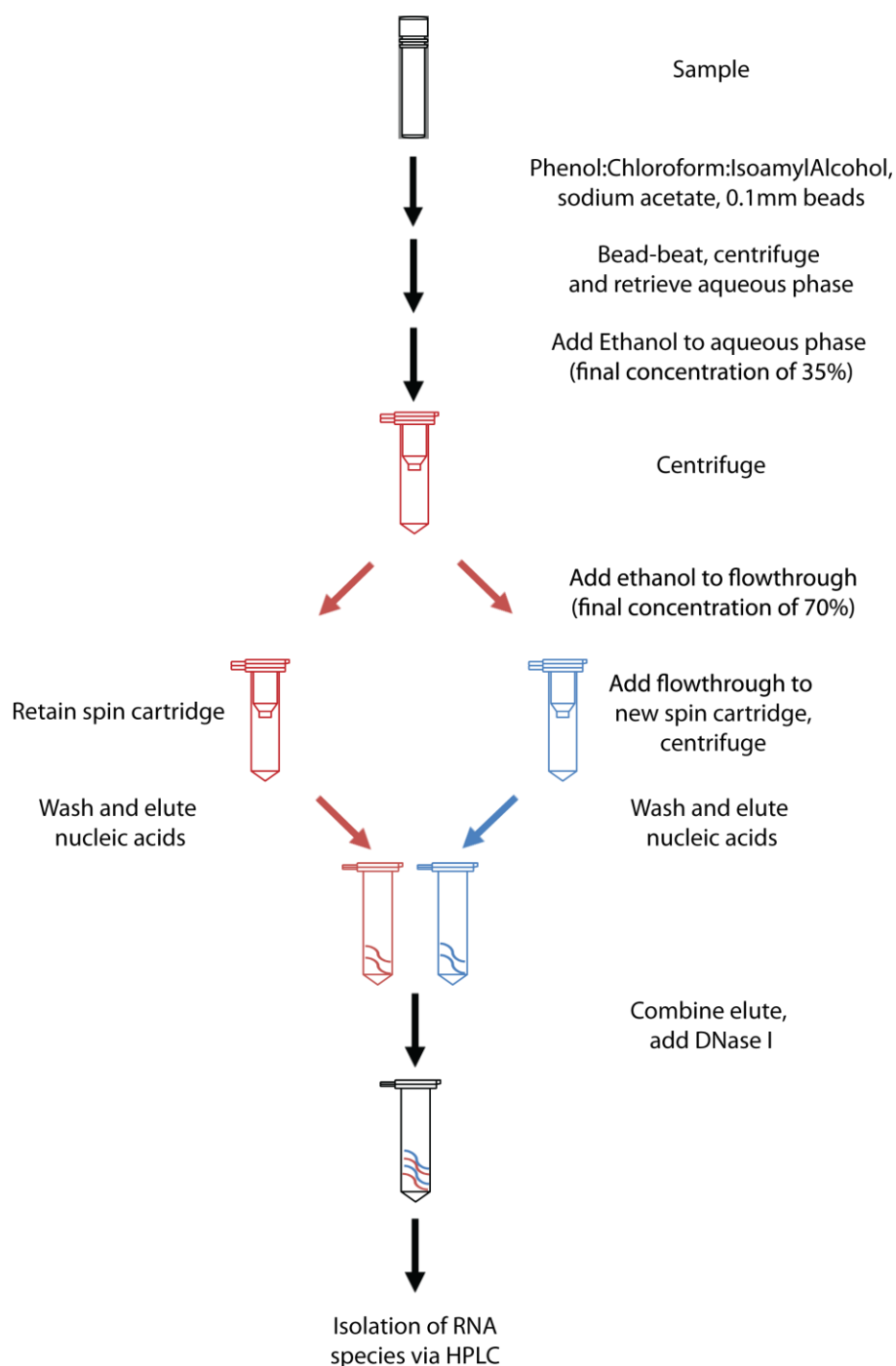
Supplementary Figure 2.2. Workflow for the isolation of total ncRNA from BCG. Bacilli are lysed by reciprocal shaking with 0.1 mm silica beads in phenol/chloroform/isoamyl alcohol and the RNA retrieved using a solid-phase extraction system (Purelink miRNA Isolation Kit). RNA from both columns is eluted and combined to yield total RNA. The ncRNA species are subsequently isolated by size-exclusion HPLC.
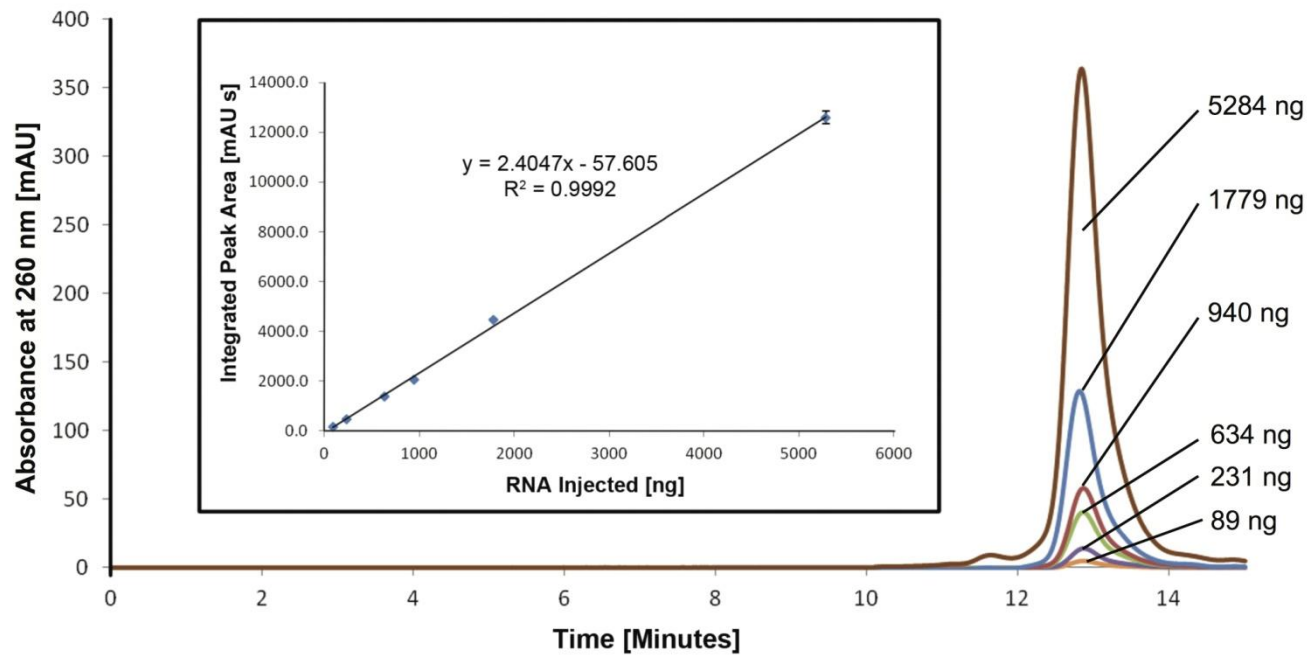
Supplementary Figure 2.3. Standard curve (R^2 = 0.999) for RNA quantification as determined by HPLC analysis of 28S rRNA standards of varying concentrations.
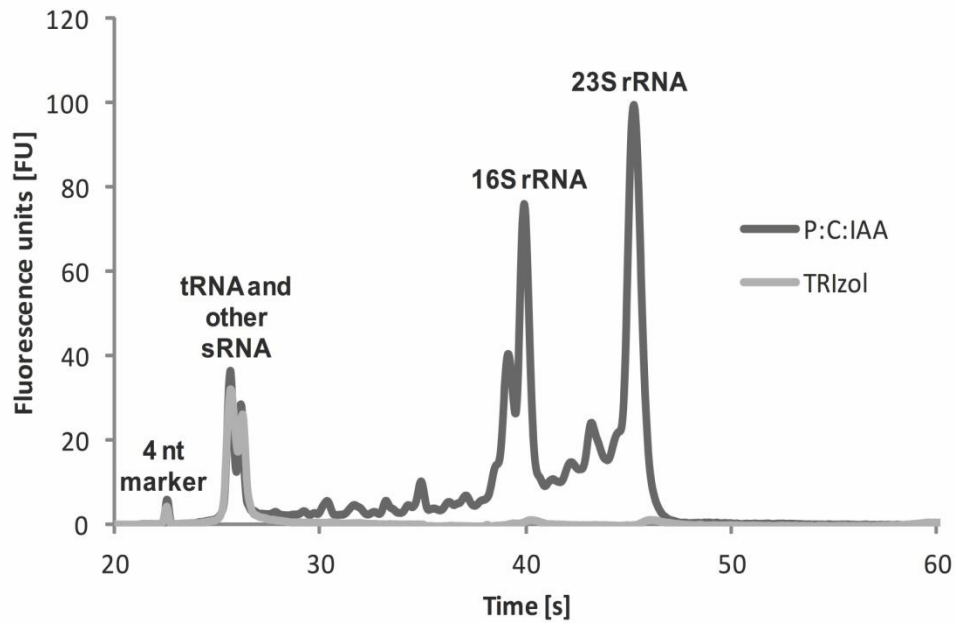
Supplementary Figure 2.4. **Comparison of RNA extractions with phenol:chloroform:isoamyl alcohol 25:24:1 (P:C:IAA), saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Black) and TRIzol (grey).** When used on BCG, significant less 16S and 23S rRNA are extracted with TRIzol when compared to phenol:chloroform:isoamyl alcohol.
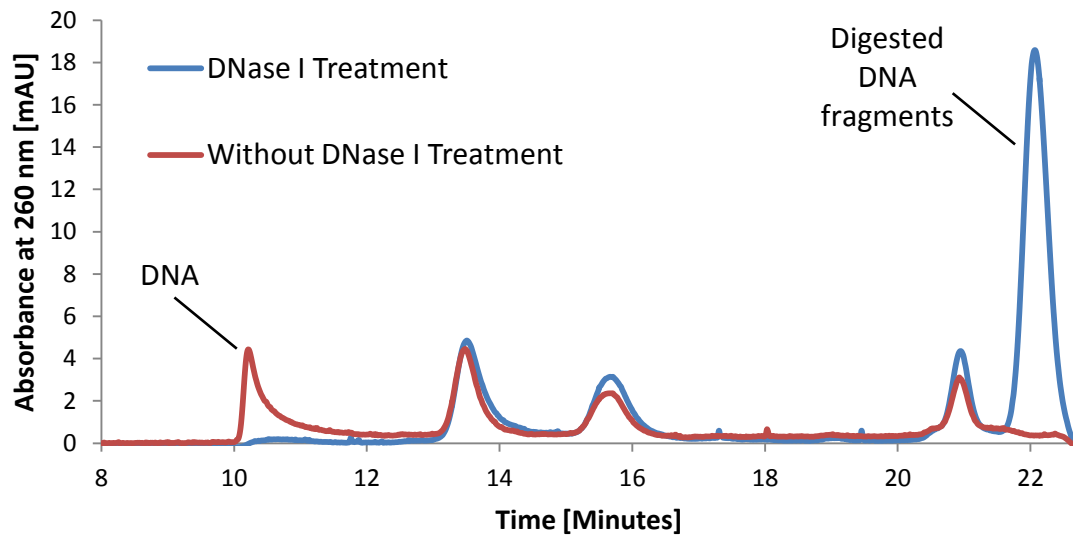
Supplementary Figure 2.5. Size-exclusion HPLC analysis showing the removal of DNA from BCG total RNA after treatment with DNase I (SEC5 1000Å column).
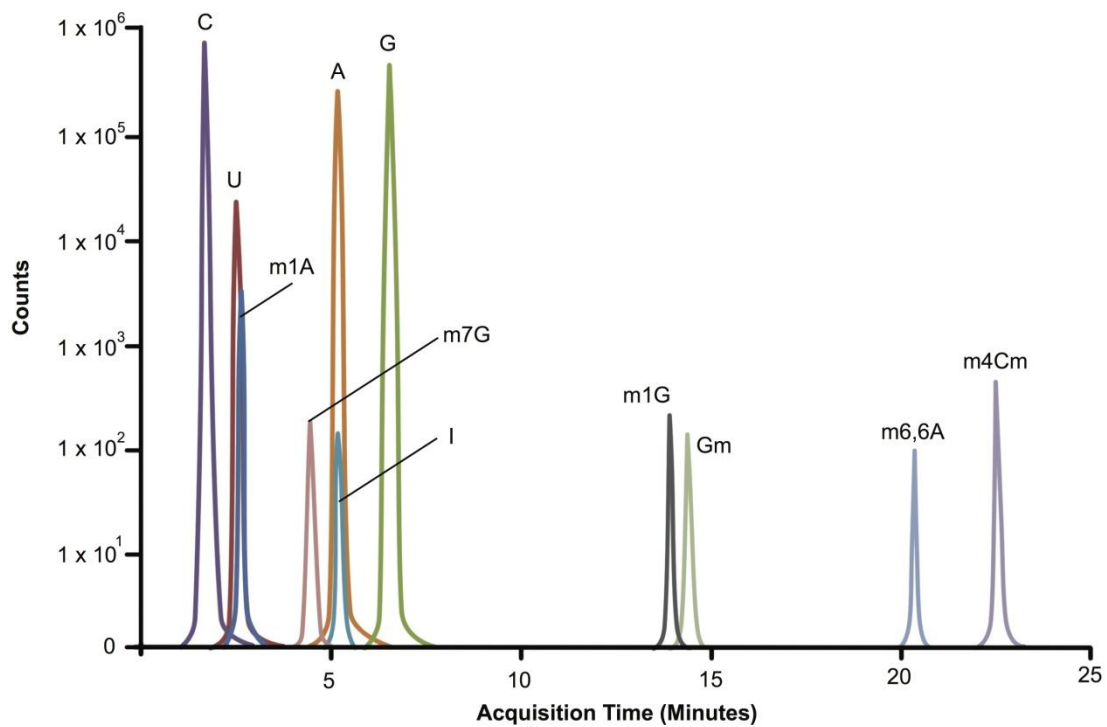
Supplementary Figure 2.6. Composite extracted ion chromatograms of modified ribonucleosides in hydrolyzed BCG 5S rRNA.

2.6.2. Supplementary tables

(next page)

Supplementary Table 2.1. Conditions used for the optimization of mycobacterial RNA extractions.

No. of cells	Comparision	Conditions — Extraction reagent	Bead Material	Vol. of beads (µl)	Bead diameter (mm)	Bead-beating Duration	RNA retrieval	Concentration (ng/ul)	Percentage Difference (%)	Individual Results	Compiled Results	Remarks
4.5 X 10^8	P:C:IAA VS TRIzol	P:C:IAA	Glass	50	1.0	1 min beat, 1 min ice (10X)	IPA	120.1	n.a.	Use P:C:IAA	Use P:C:IAA	Loss of large ncRNA using TRIzol and TRIzol Max but not when P:C:IAA was used
		TRIzol						28.1				
		TRIzol Max						49.1				
4.5 X 10^8	Addition of detergents during lysis	P:C:IAA+10%SDS	Glass	50	1	1 min beat, 1 min ice (10X)	IPA	98.7	n.a.	Addition of SDS gives the best yields	No detergent required for lysis	Addition of SDS yields more large / Ratio of organic: aqueous affected by addition of detergents
		P:C:IAA+10%Tween 20						20				
		P:C:IAA+10%TritonX100						22.3				
		P:C:IAA+5%SDS + 5% Tween 20						74.5				
		P:C:IAA+10%Tween20 + 10% Triron X100						21.5				
		P:C:IAA+5%SDS + 5% Triron X100						107.35				
		P:C:IAA+3.3%SDS + 3.3% Tween 20 +3.3% Triron X100						90.95				
4.5 X 10^8	0.5mm VS 1.0mm Glass Beads in TRIzol	TRIzol	Glass	50	0.5	1 min beat, 1 min ice (10X)	IPA	23.6 ± 4.1	No difference	No difference	No difference	-
					1.0			25.1 ± 4.2				
4.5 X 10^8	0.5mm VS 1.0mm Glass Beads in P:C:IAA	P:C:IAA	Glass	50	0.5	1 min beat, 1 min ice (10X)	IPA	258.4 ± 53.7	60.9	Use 0.5mm glass beads	Use 0.5mm glass beads	Smaller bead diameter preferred
					1.0			101.0 ± 27.0				
1.0 X 10^9	0.1mm Glass VS 0.1mm Silica Beads on Receprocal Bead-beating	P:C:IAA	Glass	50	0.1	1 min beat, 1 min ice (10X)	Purelink Cartridge 1	30.7 ± 6.0	37.2	Use 0.1mm silica beads	Use 0.1mm Silica Beads	Silica beads are better at 0.1mm
			Silica					48.9 ± 0.5				
			Glass			1 min beat, 1 min ice (10X)	Purelink Cartridge 2	12.0 ± 7.5	No difference	No difference		
			Silica					29.5 ± 7.6				
1.0 X 10^9	Pulse bead beating 1 min beat, 1 min ice (10X) VS continuous beating 30 min	P:C:IAA	Silica	50	0.1	1 min beat, 1 min ice (10X)	Purelink Cartridge 1	48.9 ± 0.5	225.0	Use continuous beating	Use continuous beating	Bead beating chambers prechilled.
						15 mins continuous		158.8 ± 7.2				
						1 min beat, 1 min ice (10X)	Purelink Cartridge 2	29.5 ± 7.6	No difference	No difference		No loss of RNA intergity observed
						15 mins continuous		27.2 ± 3.9				
1.0 X 10^9	Beating duration 30 mins VS 40 mins	P:C:IAA	Silica	50	0.1	30 mins continuous	Purelink Cartridge 1	206.78 ± 10.6	No difference	No difference	30 mins continous beating sufficient	Bead beating chambers prechilled
						40 mins continuous		220.2 ± 51.3				
						30 mins continuous	Purelink Cartridge 2	36.6 ± 6.2	No difference	No difference		No loss of RNA intergity observed
						40 mins continuous		31.2 ± 6.8				
1.0 X 10^9	Acid P:C:IAA VS P:C:IAA	Acid P:C:IAA	Silica	50	0.1	30 mins continuous	Purelink Cartridge 1	64.0 ± 17.0	96.8	Use P:C:IAA	Use P:C:IAA	Acid P:C:IAA similar to TRIzol; loss of longer rRNA species
		P:C:IAA						1995.0 ± 360.8				
		Acid P:C:IAA					Purelink Cartridge 2	241.0 ± 157.4	No difference	No difference		
		P:C:IAA						241.8 ± 10.4				
1.0 X 10^9	50µl Beads VS 100µl Beads	P:C:IAA	Silica	50	0.1	30 mins continuous	Purelink Cartridge 1	92.4 ± 2.9	55.6	Use 100µl silica beads	Use 100µl Beads	-
				100				207.9 ± 7.1				
				50			Purelink Cartridge 2	12.1	50.8	Use 100µl silica beads		
				100				24.6				

P:C:IAA = Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (v/v/v); saturated with 10 mM Tris, pH 8.0, 1 mM EDTA
Acid P:C:IAA = Phenol: chloroform: isoamyl125:24:1 (v/v/v); saturated with 2 M NaOAc pH 4
Where possible standard errors for each series of extractions are provided

Supplementary Table 2.2. RNA integrity numbers (RIN) and proportions of non-coding RNA as determined using the Agilent RNA 6000 Pico Kit and Agilent Small RNA kit as determined by the Agilent Bioanalyzer 2100 system. **(A)** the phenol;chloroform:isoamyl alcohol approach and **(B)** the conventional TRIzol approach.

A	Bioanalyzer Chip		RNA Integrity Number	Percentage of RNA Eluted From Column		
				23S rRNA	16S rRNA	Small RNA
	Agilent RNA 6000 Pico Kit	Column #1	9.0 ± 0.5	20.7 ± 20.7	17.8 ± 3.3	27.0 ± 3.6
		Column #2	-	-	-	95.8 ± 5.3
		Combined columns #1, #2	8.5 ± 0.3	16.6 ± 2.6	14.9 ± 1.7	38.9 ± 4.1

Bioanalyzer Chip		RNA Integrity Number	Percentage of RNA Eluted From Column		
			5S rRNA	tRNA	miRNA
Agilent Small RNA kit	Column #1	-	17 ± 1	50 ± 3	33 ± 3
	Column #2	-	2 ± 1	86 ± 2	11 ± 1
	Combined columns #1, #2	-	12 ± 2	71 ± 5	17 ± 4

B	Bioanalyzer Chip	RNA Integrity Number	Percentage of RNA Eluted From Column		
			23S rRNA	16S rRNA	Small RNA
	Agilent RNA 6000 Pico Chip	-	-	13.9 ± 1.4	10.5 ± 1.7	75.6 ± 3.0

Supplementary Table 2.3. qPCR Primers for *M. bovis* BCG str. Pasteur 1173P2.

Gene	Direction	Sequence (5' - 3')	Melting Temperature (°C)
hspX	F	GACGAGATGAAAGAGGGGCG	90.0
	R	GTCGTCCTCGTCAGCACCTA	
ethA	F	CGAGGCCGACGTTCTACTTAT	90.0
	R	GCGACTTCGACACTGGTTGC	
fdxA	F	AGTGAGTGCGTGGATGTGATG	88.5
	R	TGGTGTTGATCGTCGGGTAGA	
relA	F	ATTGCCACCAGAAACACCGA	91.5
	R	GGTTCCGGGCGATGTGATTA	
sigA*	F	CGATGAGCCGGTAAAACGC	91.0
	R	GAGCCACTAGCGGACTTCGC	
dosR	F	AGTGAGTGCGTGGATGTGATG	88.5
	R	TGGTGTTGATCGTCGGGTAGA	
dosS	F	ATTGCCACCAGAAACACCGA	91.5
	R	GGTTCCGGGCGATGTGATTA	

* Primer sequence obtained from Cappelli et al. [53]

2.7. Acknowledgements

2.8. References

1. Jasmer, R.M., P. Nahid, and P.C. Hopewell, *Clinical practice. Latent tuberculosis infection.* N Engl J Med, 2002. **347**(23): p. 1860-6.
2. World Health Organization, G.T.P., *Global tuberculosis control : WHO report.* 2009, Global Tuberculosis Programme: Geneva.
3. Bailey, S.L., et al., *Missed opportunities for tuberculosis diagnosis.* Int J Tuberc Lung Dis, 2011. **15**(2): p. 205-10, i.
4. Pan, W., et al., *Comparison of mantoux and tine tuberculin skin tests in BCG-vaccinated children investigated for tuberculosis.* PLoS One, 2009. **4**(11): p. e8085.
5. Niemz, A., T.M. Ferguson, and D.S. Boyle, *Point-of-care nucleic acid testing for infectious diseases.* Trends Biotechnol, 2011. **29**(5): p. 240-50.
6. Raviglione, M., et al., *Scaling up interventions to achieve global tuberculosis control: progress and new developments.* Lancet, 2012. **379**(9829): p. 1902-13.
7. Lienhardt, C., et al., *Global tuberculosis control: lessons learnt and future prospects.* Nat Rev Microbiol, 2012. **10**(6): p. 407-16.
8. Almeida Da Silva, P.E. and J.C. Palomino, *Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs.* J Antimicrob Chemother, 2011. **66**(7): p. 1417-1430.
9. Partnership, S.T., *The Global Plan to Stop TB 2011–2015: Transforming the fight towards elimination of tuberculosis*, in *World Health Organization, Global Tuberculosis Programme*, G.T.P. World Health Organization, Editor. 2010, World Health Organization: Geneva.
10. Laang, H. *EC investment in TB research in FP7.* in *Tuberculosis 2012: Biology, Pathogenesis, Intervention strategies*. 2012. Institut Pasteur, Paris, France.

11.   Brennan, P.J., *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2003. **83**(1-3): p. 91-7.

12.   Hoffmann, C., et al., *Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure.* Proc Natl Acad Sci U S A, 2008. **105**(10): p. 3963-7.

13.   Arnvig, K. and D. Young, *Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis.* RNA Biol, 2012. **9**(4).

14.   Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.

15.   Chan, C.T., et al., *Identification of N6,N6-dimethyladenosine in transfer RNA from Mycobacterium bovis Bacille Calmette-Guerin.* Molecules, 2011. **16**(6): p. 5168-81.

16.   Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.

17.   Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.

18.   Phizicky, E.M. and A.K. Hopper, *tRNA biology charges to the front.* Genes Dev, 2010. **24**(17): p. 1832-60.

19.   Akhtar, S., et al., *A method to extract intact and pure RNA from mycobacteria.* Anal Biochem, 2011. **417**(2): p. 286-8.

20.   Mangan, J.A., et al., *An effective method of RNA extraction from bacteria refractory to disruption, including mycobacteria.* Nucleic Acids Res, 1997. **25**(3): p. 675-6.

21.   Vandeventer, P.E., et al., *Mechanical disruption of lysis-resistant bacterial cells by use of a miniature, low-power, disposable device.* J Clin Microbiol, 2011. **49**(7): p. 2533-9.

22.   Kinger, A.K., A. Verma, and J.S. Tyagi, *A method for the isolation of pure intact RNA from mycobacteria.* Biotechniques, 1993. **14**(5): p. 724-5.

23.   Kim, Y.K., et al., *Short Structured RNAs with Low GC Content Are Selectively Lost during Extraction from a Small Number of Cells.* Mol Cell, 2012. **46**(6): p. 893-5.

24.   Havelund, J.F., et al., *Identification of 5-hydroxycytidine at position 2501 concludes characterization of modified nucleotides in E. coli 23S rRNA.* J Mol Biol, 2011. **411**(3): p. 529-36.

25.   Low, K.L., et al., *Triacylglycerol utilization is required for regrowth of in vitro hypoxic nonreplicating Mycobacterium bovis bacillus Calmette-Guerin.* J Bacteriol, 2009. **191**(16): p. 5037-43.

26.   Chionh, Y.H., et al., *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Res, 2013. **41**(17): p. e168.

27.   Su, D., et al., *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry.* Nat Protoc, 2014. **9**(4): p. 828-41.

28.   Schroeder, A., et al., *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Mol Biol, 2006. **7**: p. 3.

29.   Beste, D.J., et al., *Compiling a molecular inventory for Mycobacterium bovis BCG at two growth rates: evidence for growth rate-mediated regulation of ribosome biosynthesis and lipid metabolism.* J Bacteriol, 2005. **187**(5): p. 1677-84.

30. Liu, J.M., et al., *Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing.* Nucleic Acids Res, 2009. **37**(6): p. e46.

31. Altuvia, S., *Identification of bacterial small non-coding RNAs: experimental approaches.* Curr Opin Microbiol, 2007. **10**(3): p. 257-61.

32. Sharma, C.M. and J. Vogel, *Experimental approaches for the discovery and characterization of regulatory small RNA.* Curr Opin Microbiol, 2009. **12**(5): p. 536-46.

33. Zhang, Y., et al., *Identifying Hfq-binding small RNA targets in Escherichia coli.* Biochem Biophys Res Commun, 2006. **343**(3): p. 950-5.

34. Straub, J., et al., *Small RNAs in haloarchaea: identification, differential expression and biological function.* RNA Biol, 2009. **6**(3): p. 281-92.

35. Dedrick, R.M., et al., *Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles.* Mol Microbiol, 2013. **88**(3): p. 577-89.

36. Zeller, M.E., et al., *Quaternary structure and biochemical properties of mycobacterial RNase E/G.* Biochem J, 2007. **403**(1): p. 207-15.

37. Chomczynski, P. and N. Sacchi, *The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on.* Nat Protoc, 2006. **1**(2): p. 581-5.

38. Cheung, A.L., K.J. Eberhardt, and V.A. Fischetti, *A method to isolate RNA from gram-positive bacteria and mycobacteria.* Anal Biochem, 1994. **222**(2): p. 511-4.

39. Rio, D.C., et al., *Guidelines for the use of RNA purification kits.* Cold Spring Harb Protoc, 2010. **2010**(7): p. pdb ip79.

40. Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence.* Infect Immun, 1996. **64**(6): p. 2062-9.

41. Trauner, A., et al., *The Dormancy Regulator DosR Controls Ribosome Stability in Hypoxic Mycobacteria.* J Biol Chem, 2012. **287**(28): p. 24053-63.

42. Wayne, L.G. and C.D. Sohaskey, *Nonreplicating persistence of mycobacterium tuberculosis.* Annu Rev Microbiol, 2001. **55**: p. 139-63.

43. Ghora, B.K. and D. Apirion, *Structural analysis and in vitro processing to p5 rRNA of a 9S RNA molecule isolated from an rne mutant of E. coli.* Cell, 1978. **15**(3): p. 1055-66.

44. Ow, M.C., T. Perwez, and S.R. Kushner, *RNase G of Escherichia coli exhibits only limited functional overlap with its essential homologue, RNase E.* Mol Microbiol, 2003. **49**(3): p. 607-22.

45. Szymanski, M., et al., *5S ribosomal RNA database Y2K.* Nucleic Acids Res, 2000. **28**(1): p. 166-7.

46. Bruenger, E., et al., *5S rRNA modification in the hyperthermophilic archaea Sulfolobus solfataricus and Pyrodictium occultum.* FASEB J, 1993. **7**(1): p. 196-200.

47. Grosjean, H., et al., *RNomics and Modomics in the halophilic archaea Haloferax volcanii: identification of RNA modification genes.* BMC Genomics, 2008. **9**: p. 470.

48. Rozenski, J., P.F. Crain, and J.A. McCloskey, *The RNA Modification Database: 1999 update.* Nucleic Acids Res, 1999. **27**(1): p. 196-197.

49. Long, K.S. and B. Vester, *Resistance to linezolid caused by modifications at its binding site on the ribosome.* Antimicrob Agents Chemother, 2012. **56**(2): p. 603-12.

50.   Doi, Y. and Y. Arakawa, *16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides.* Clin Infect Dis, 2007. **45**(1): p. 88-94.

51.   Demirci, H., et al., *Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function.* RNA, 2010. **16**(12): p. 2319-24.

52.   Vester, B. and S. Douthwaite, *Macrolide resistance conferred by base substitutions in 23S rRNA.* Antimicrob Agents Chemother, 2001. **45**(1): p. 1-12.

53.   Cappelli, G., et al., *Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function.* Res Microbiol, 2006. **157**(5): p. 445-55.

3. A multi-dimensional platform for the purification of non-coding RNA species[1]

3.1. Abstract

A renewed interest in non-coding RNA (ncRNA) has led to the discovery of novel RNA species and post-transcriptional ribonucleoside modifications, and an emerging appreciation for the role of ncRNA in RNA epigenetics. While much can be learned by amplification-based analysis of ncRNA sequence and quantity, there is a significant need for direct analysis of RNA, which has led to numerous methods for purification of specific ncRNA molecules. However, no single method allows purification of the full range of cellular ncRNA species. To this end, we developed a multi-dimensional chromatographic platform to resolve, isolate, and quantify all canonical ncRNAs in a single sample of cells or tissue, as well as novel ncRNA species. The applicability of the platform is demonstrated in analyses of ncRNA from bacteria, human cells and plasmodium-infected reticulocytes, as well as a viral RNA genome. Among the many potential applications of this platform are a systems level analysis of the dozens of modified ribonucleosides in ncRNA, characterization of novel long ncRNA species, enhanced detection of rare transcript variants, and analysis of viral genomes.

3.2. Introduction

The renewed interest in RNA modifications and the discovery of many new non-coding RNA species (ncRNA)[1-5] has increased the demand for methods to purify RNA species. However, while eukaryotic mRNA is readily purified by exploiting a polyA tail, studies of ncRNA structure and

---

[1] Featured in Nucleic Acids Res. 2013 Sep;41(17):e168. Reproduced with permission from Oxford University Press.

biochemistry, post-transcriptional processing (capping, end-processing), ribonucleoside modification, and biological function have been limited by the inability to obtain RNA in pure form [6-11]. This is especially important given the emerging interest in studying RNA biology on a systems level, with coordinated analysis of modified ribonucleosides present in virtually all forms of RNA [4, 5, 12, 13]. Existing RNA purification methods are limited by specificity, size range, or yield. While affinity purification approaches are specific to a unique sequence, more general size-based gel electrophoretic approaches are hampered by a narrow size range, gel contaminants from RNA elution, and high losses during purification [14-16]. Liquid chromatography (LC) approaches employing all types of stationary phase have solved some of these problems, with a wider size range but lower resolving power than gel electrophoresis [17-22]. Improvements in capacity and specificity have been achieved by specialized combinations of chromatography with affinity purification [23, 24] and electrophoresis [25]. Though useful for isolating specific RNA classes or RNA species, these methods do not permit size-fractionation of total RNA or isolation of all classes of ncRNA from single sample for a systems-level analysis of modified ribonucleosides, for example.

To address this unmet need in RNA biology, we report a comprehensive, multi-dimensional, high performance liquid chromatography (HPLC) platform that can be used to purify all major classes of ncRNA from a single sample of total RNA. The method takes advantage of the strengths of two types of HPLC, thus increasing the peak capacity and the resolution of ncRNA across a wide size range. Such an approach is well developed in proteomics and has proven valuable in fractionating complex biological mixtures [26, 27]. To resolve ncRNA, we combined two ranges of size-exclusion chromatography

(SEC) with ion-pair reverse-phase chromatography (IP RPC) to achieve a complete separation of RNA species ranging from 20 to >10,000 nucleotides (nt), including viral RNA genomes, large and small subunit rRNAs, tRNA and miRNA. The approach is demonstrated for both individual HPLC steps (1-dimensional, 1-D) and for 2-dimensional (2-D) HPLC resolution of total RNA from human, plasmodium parasite and bacterial cells, as well as a dengue viral RNA genome.

3.3 Material and methods

3.3.1. Chemicals and reagents

RPMI 1640, fetal bovine serum and Penicillin Streptomycin (Pen-Strep) for cell cultures were purchased from Gibco, Invitrogen (Carlsbad, CA). RiboGreen and PicoGreen kits for RNA and DNA quantitation respectively were purchased from Molecular Probes, Invitrogen (Eugene, OR). Chemicals unless otherwise specified were purchased from Sigma Chemical Co. (St. Louis, MO).

3.3.2. Bacterial and mammalian cell culture

E. coli strain DH5α, was grown for 3 h in Luria-Bertani (LB) medium (Becton, Dickinson and Company, Franklin Lakes, NJ) at 37°C with shaking (250 rpm) until an optical density (600 nm; OD600) of 0.6 was reached at mid-exponential growth phase. *M. bovis* BCG strain Pasteur 1173P2 bacilli (ATCC) were grown in 7H9 culture media (4.9 g of 7H9 powder, 10 mL of 50% glycerol, 2.5 mL of 20% TWEEN 80, 900 mL of water and 100 mL of ADS solution) at 37 ^{o}C in an 850 cm^2 polystyrene roller bottle (Corning, NY, USA) at 10 rpm to an OD600 = 0.6, at which point the concentration of cells was ~5x10^7/mL. Mononuclear, B lymphoblastoid CCRF-SB and TK6 cells (ATCC) were grown in suspension cultures in RPMI complete medium (90%

RPMI 1640 medium, 10% fetal bovine serum with added Pen-Strep) to a density of 10^7 cells/mL.

3.3.3. *In vitro* transcription of dengue viral RNA from plasmid DNA template

Genome-length 10.7 kb viral RNA (vRNA) of DENV-1 was *in vitro* transcribed from full-length cDNA plasmid linearized by *SacII* [28]. The vector is then electroporated into BHK-21 cells and transfected cells resuspended in 20 mL of DMEM medium[29]. The transfected cultures were subsequently subjected to viral production and specific infectivity assays as communicated previously [28]. RNA was isolated with Trizol reagent (Invitrogen) as directed by the manufacturer and re-purified on an Agilent Bio SEC-5 2000Å (i.d. 7.8 mm; length 300 mm) column (Agilent Technologies, Foster City, CA) with an isocratic elusion as described below.

3.3.4. Rodent infection with *Plasmodium berghei* and isolation of schizont-infected murine reticulocytes

Male BALB/c mice and male WISTAR rats of 6–8 weeks old were obtained from Sembawang Laboratory Animal Center, National University of Singapore, and subsequently bred under specific pathogen free (SPF) conditions at Nanyang Technological University Animal Holding Unit. Infection and handling of laboratory animals with the blood-stage *Plasmodium berghei* ANKA strain were performed using the methods described elsewhere [30]. Blood was collected from infected rats at a parasitemia of 1-3% by cardiac puncture with heparin (Sigma). Blood was then filtered by leukocyte filter (Plasmodipur filter, Euro-Diagnostica) and cultured until the schizont stage in complete RPMI 1640 containing 20% FBS with gentle shaking at 37 °C. Schizont-infected cells were separated from uninfected cells by 50%–60%

Nycodenz (Sigma-Aldrich) gradient centrifugation. Schizonts were purified by lysing rat erythrocytes with 0.15% saponin and washed three times with cold PBS before flash freezing in liquid nitrogen.

3.3.5. Ethics statement

The study was carried out in strict accordance with the recommendations of the NACLAR (National Advisory Committee for Laboratory Animal Research) guidelines under the Animal and Birds (Care and Use of Animals for Scientific Purposes) Rules of Singapore. The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of the Nanyang Technological University of Singapore (Approval number: ARFSBS/ NIE A002). All efforts were made to minimize the suffering.

3.3.6. Total RNA extraction

Prior to lysis cells were washed with ice-cold PBS again. BCG pellets were lysed by French high-pressure homogenization (Thermo Scientific). Ice-cold phenol:chloroform:isoamyl alcohol (25:24:1) was added directly to the lysate and total RNA obtained with PureLink miRNA Isolation Kit (Invitrogen), as described by the manufacturer. For *E. coli*, the lysis was performed with 100 µL of lysozyme (1 mg/mL) for 20 min at ambient temperature. *E. coli* total RNA was extracted with Trizol reagent (Invitrogen), precipitated with ice-cold isopropyl alcohol and washed with 75% ethanol. CCRF-SB, TK6 and *P. berghei* infected reticulocytes were directly lysed in Trizol reagent and eukaryotic total RNA extracted in an identical fashion. RNA pellets were air dried and re-dissolved in RNAse-free water. DNase I (Qiagen, Valencia, CA) treatment was performed for 15 min at 37 °C. The quality and quantity of RNA was determined using Agilent Bioanalyzer RNA 6000 Nano chips (Agilent

Technologies, Santa Clara, CA). Only samples with a RNA Integrity Number
(RIN) of 8.0 or greater were used for further experiments.

3.3.7. Size-exclusion chromatography of total RNA

One-dimensional SEC was performed with either a Bio SEC-3 300Å (i.d.
7.8mm; length 300mm) or a Bio SEC-5 1000Å (i.d. 7.8mm; length 300mm)
column (Agilent Technologies, Foster City, CA), with 100 mM ammonium
acetate (pH 7.0) as the mobile phase. Isocratic separations were performed
at 1 mL/min for 20 min under partially denaturing conditions at 60 °C on the
Agilent 1200 HPLC system. 2-D SEC was performed with Bio SEC-3 300Å
and either Bio SEC-5 1000Å or Bio SEC-5 2000Å (id 7.8 mm, length 300 mm)
columns. A second Agilent 1260 Infinity system is connected with the valve
configuration described in **Supplementary Figure 3.2** and isocratic
separations were performed with 100 mM ammonium acetate at 0.5 ml/min
rates for 40 min at 60 °C. Chromatograms were recorded at every 10 nm from
200 to 300 nm. Peaks were collected with an Agilent 1260 Infinity auto-
sampler by time segments. For separations of BCG total RNA, a Bio SEC-5
1000Å column was connected to the column 1 position while a Bio SEC-3
300Å column was connected to the column 2 position. From 0-0.1 min column
1 is connected to the detector, from 0.1- 18.5 min, column 2 is connected and
this is switched back to column 1 at 18.5 min. For separations of TK6 total
RNA, a Bio SEC-5 2000Å column was connected to the column 1 position
while a Bio SEC-3 300Å column was connected to the column 2 position.
From 0-0.1 min, column1 is connected to the detector, and from 0.1-20 min,
column 2 is connected and switched back to column 1 at 20 min. For
separations of total RNA from *P. berghei*-infected rat reticulocytes, a Bio
SEC-3 300Å column was connected to the column 1 position while a Bio
SEC-5 1000Å column was connected to the column 2 position; the column

switching scheme is identical as that used for TK6 total RNA. For the analysis of TK6 total RNA, 100 mM triethylammonium acetate (TEAA) (pH 7.0) instead of 100 mM ammonium acetate was used as the mobile phase. In addition, a Bio SEC-5 2000 Å column was used. All other conditions were kept constant.

3.3.8. Ion-pair reversed-phase chromatography of total RNA

CCRF-SB and TK6 total RNA was analyzed by IP RPC on the Agilent 1200 system using a Source 5RPC 4.6/150 column (GE Healthcare Life Sciences, Piscataway, NJ). HPLC was performed using a two eluent buffer system in which buffer A consists of an aqueous solution of 100 mM TEAA, pH 7.0 with 2% acetonitrile and buffer B consists of 100% acetonitrile under partially denaturing conditions at 60 °C with chromatograms recorded and as described above. IP RPC were performed with the following gradient conditions: flow rate 1 mL/min, 5-5.5% B over 6 min, to 13% B over 45 min, to 16.5% B over 9 min, to 100% B over 10 min, held at 100% B for 10 min, brought back to 5% B over 10 min.

3.3.9. Analysis of isolated RNA species

2D-SEC-purified DENV-110.7 kb vRNA, 28S, and 18S rRNA from CCRF-SB and TK6, 23S and 16S rRNA from *E. coli* and BCG, and *P. berghei* 28S, 18S and 800nt rRNA were concentrated and desalted with an Amicon Ultracel 10k MWCO spin column (Millipore, Bedford, MA), while 5.8S, 5S, tRNA and miRNA from the aforementioned total RNA preparations were concentrated and desalted with a Vivacon-500 2k MWCO spin column (Sartorius, Goettingen, Germany). The efficiency of the isolation process, quality and quantity of each RNA species after HPLC purification and desalting was assessed using the Agilent Bioanalyzer (Agilent). RNA 6000 Pico chips were used to evaluate 28S, 18S, 23S, 16S rRNA and *P. berghei* 800 nt rRNA while

5S, 5.8S rRNA, tRNA and miRNA were evaluated on Small RNA chips. The presence of contaminating DNA was determined using a PicoGreen dye-based quantitation assay using calf-thymus DNA (Sigma) as a calibration standard.

The theoretical maximum composition of each RNA species from their respective total RNA preparations was estimated by the product of the percentage area under the curve for each chromatographic peak with the quantity of total RNA injected into the chromatographic system. We then established the yields of each RNA species that was recovered from the RNA isolation process by calculating the ratios between the quantities of purified RNA against its corresponding theoretical maximum yield.

3.3.10. RiboGreen assay for species-specific fluorometric responses

Purified RNA samples were diluted to stock solutions of 20 µg/mL RNA in TE buffer, (10 mM Tris, 1 mM EDTA, pH 7.8) and their concentrations were confirmed with a Nanodrop Spectrophotometer (Thermo Scientific, Wilmington, DE). A 96-well based RiboGreen RNA quantitation assay was performed as indicated by the manufacturer instructions with the stocks of each purified RNA species used in place of the RNA standard provided. Scans for RiboGreen fluorescence were also performed from 450 to 650 nm at 10 nm intervals. Fluorometric measurements were conducted on a BioTek Synergy 2 Multi-Mode Microplate Reader (Winooski, VT) with 485/20 nm and 528/20 nm bandpass filters for excitation and emission, respectively. All readings were taken at 25 °C.

3.3.11. Detection and relative quantification of ribonucleosides from BCG tRNA by chromatography-coupled mass spectrometry

Purified BCG tRNA (5 μg) from 3 independent exponentially growing cultures of BCG was hydrolyzed enzymatically as described previously [31]. Hydrolyzed RNA was resolved on a Thermo Hypersil aQ column (100 x 2.1 mm, 1.9 μm particle size) in a two buffer eluent system with buffer A consisting of water with 0.1% (v/v) formic acid and buffer B consisting of acetonitrile with 0.1% (v/v) formic acid. HPLC was performed at a flow rate of 0.3 mL/min. The gradient of acetonitrile in 0.1% formic acid was as follows: 0-12 min, held at 0%; 12-15.3 min, 0%-1%; 15.3-18.7 min, 1%-6%; 18.7-20 min, held at 6%; 20-24 min, 6%-100%; 24-30 min, held at 100%; 30-35 min, 100%-0%; 35-40 min, 0%. The HPLC column was maintained at 25 ºC and directly connected to an Agilent 6460 triple quadrupole mass spectrometer (LC-MS/MS) with ESI Jetstream ionization operated in positive ion mode. Multiple reaction monitoring (MRM) was performed to detect and quantify the ribonucleosides with the parameters of retention time, *m/z* of the transmitted parent ion, and *m/z* of the monitored product ion as noted in Supplementary Table 2. Fragmentor voltages of 89 V and collision energies of 18 V were used. The dwell time for each ribonucleoside was 100 ms and 18 ions were monitored from 0.5 to 28 min. The voltages and source gas parameters were as follows: gas temperature, 350 ºC; gas flow, 5 L/min; nebulizer, 40 psi; sheath gas temperature, 325 ºC; sheath gas flow, 7 L/min and capillary voltage, 3500 V.

3.3.12. MS^2 Structural characterization of *N^6,N^6*-dimethyladenosine in BCG tRNA

The ribonucleoside-like species eluting at 20.14 min and possessing an [M+H]$^+$ ion with *m/z* of 296.1356 was subjected to structural characterization by collision-induced dissociation (CID) performed on the LC-QTOF system using the same column and chromatographic system as described above.

The mass spectrometer was operated in positive ion mode with the following voltages and source gas parameters: gas temperature, 325 ⁰C; drying gas, 7 L/min; nebulizer, 15 psig; capillary voltage, 4000 V. The *m/z* detection range for parent ions was 100 to 800 and that for product ions was 50 to 500. The fragmentor voltage was set at 85 V and the collision energy was 25 V.

3.3.13. Data graphing and statistical analysis

Analysis of Covariance (ANCOVA) of RNA species-specific RiboGreen calibration curves was performed with Graphpad Prism 5.0. Multiple regressions were not performed. Instead, relative pair-wise comparisons were made between each purified RNA species and the total RNA from which it was isolated from. The 3-D and contour plots were made with Matlab R2012a (Mathworks, Natick, MA). Peaks in the raw chromatograms corresponding to miRNA, tRNA, 5S rRNA, sn/snoRNA, 5.8S rRNA, 18S rRNA and 28S rRNA were manually assigned based on their respective sizes as determined by Bioanalyzer 2100 LabChips. In the 3-D plot, the region on the X-Y basal plane that represents a particular RNA species was determined by the characteristic retention time of that species in both LC methods; the height of peaks in the Z-direction, each representing a particular RNA species, denotes the geometric average of the absorbance values of that specific RNA in both LC methods, which was calculated as the square root of the product of the two absorbance values. This assignment of Z-axis values takes into account the absorbance measurements by both LC methods and therefore accurately reflects the relative intensity or abundance of the different RNA species present within one cellular extract. For all areas on the X-Y basal plane other than those designated with a particular RNA species, the Z-axis values were set to be zero so as to remove the artifact peaks generated by juxtaposing data from two orthogonal LC methods.

3.4. Results

The overall strategy for developing the RNA purification method involved defining the parameters for the individual chromatographic elements for purification of ncRNA and then combining the two size-exclusion steps into a 2-D system.

3.4.1. 1-D size exclusion chromatography of eukaryotic and prokaryotic total RNA

Given the observation that virtually all classes of RNA, except mRNA, can be defined by their sequence lengths and molecular sizes, we coordinated two different size-range SEC columns to resolve total RNA. 1-D SEC of total cellular RNA has been used to separate small ncRNA (tRNA; 5S, 5.8S rRNA) from large ncRNA (16S, 18S, 23S, 26S rRNA), but large rRNA species were not resolved [21, 22, 32]. We empirically tested several commercial SEC columns and found that two second-generation columns (Agilent Bio SEC-5 1000 Å pore size and Bio SEC-3 300 Å pore size) provided the best resolution after optimizing for temperature (60 °C; evaluated range: ambient-70 °C), eluent ionic strength (10 mM ammonium acetate; evaluated range: 0-100 mM) and flow rate (0.5 mL/min; evaluated range: 0.5-1 mL/min). Separations of total RNA from *E. coli* and human B lymphoblastic CCRF-SB cells on both columns are shown in **Figure 3.1A-D**. Baseline separation (resolution R_s > 1.5) of large rRNAs was achieved with the SEC-5 1000 Å column for both 23S and 16S rRNA from *E. coli* (**Figure 3.1A**) and for 28S and 18S rRNA from CCRF-SB (**Figure 3.1B**). While low molecular weight rRNAs (5.8S, 5S), tRNA and smaller ncRNA eluted together on this column (**Figure 3.1A, B**), the SEC-3 300 Å column achieved resolution of tRNA from 5.8S/5S rRNA (**Figure 3.1C, D**).
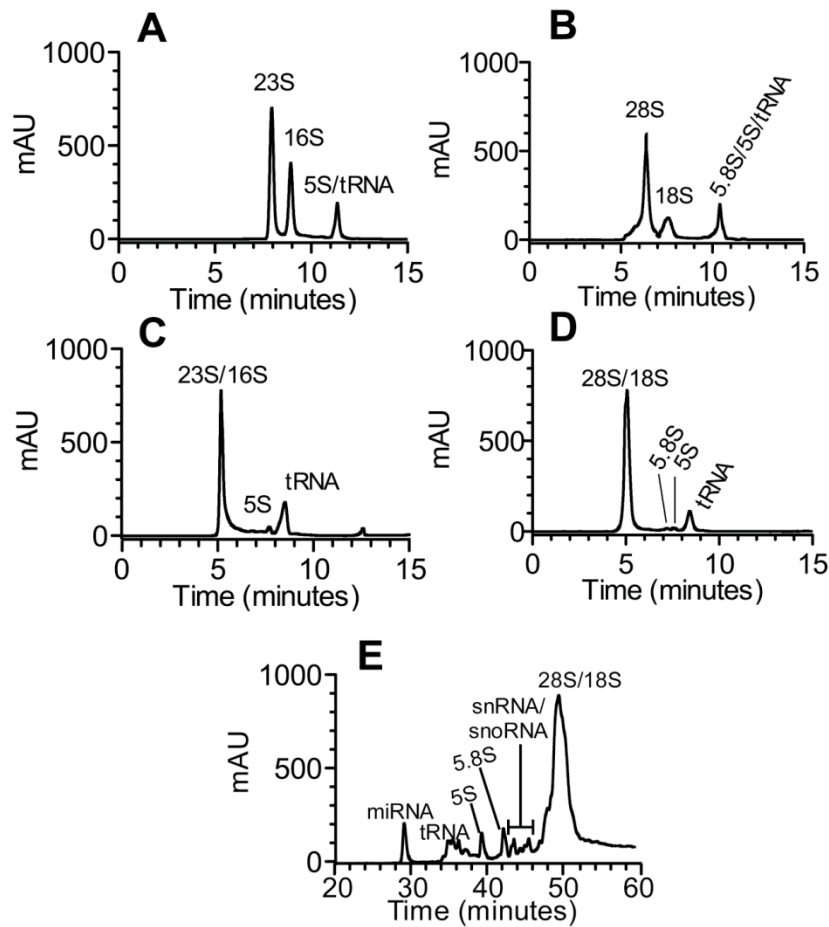
Figure 3.1. Separation of CCRF-SB and *E. coli* total RNA by SEC HPLC. (**A**) Typical profile of *E. coli* total RNA consisting of 23S, 16S rRNAs and co-eluting 5S rRNA and tRNA obtained on a Bio SEC-5 1000Å column. (**B**) Typical profile of CCRF-SB total RNA consisting of 28S, 18S rRNAs and co-eluting 5.8S, 5S rRNA and tRNA obtained on a Bio SEC-5 1000Å column. (**C**) Typical profile of *E. coli* total RNA consisting of 5S rRNA, tRNAs and co-eluting 16S and 23S rRNAs obtained on a Bio SEC-3 300Å column. (**D**) Typical profile of CCRF-SB total RNA consisting of 5.8S, 5S rRNA, tRNAs and co-eluting 18S and 28S rRNAs obtained on a Bio SEC-3 300Å column. The chromatograms show the analysis of 10 μg of total RNA extracted using Trizol reagent (Material and Methods). (**E**) Separation of miRNA, tRNAs, 5.8S, 5S rRNAs, putative snRNAs and snoRNAs, and co-eluting 18S and 28S rRNAs from human lymphoblastic cell line CCRF-SB total RNA by IP RP HPLC obtained on a SOURCE 5RPC ST 4.5/150 column. The identity and purity of the RNAs collected in each fraction was validated with Bioanalyzer RNA 6000 Pico and Small RNA LabChips (**Supplementary Figures. 3.1 and 3.3**).

For preparative SE-HPLC isolations of RNA at microgram quantities, a fraction collector (Agilent 1260 FC-AS) is placed in-line after the detector. Fractions are collected by time segments, and the purity and integrity of each fraction were analyzed on an Agilent 2100 Bioanalyzer with appropriate RNA microfluidic chips (**Supplementary Figure 3.1**). For each fraction, mRNA and other non-canonical ncRNA species were not detected above the 50 pg/μl limit of detection of the RNA 6000 Pico (total RNA) and Small RNA chips. It should be noted that for these small, micro-fluidic separations, it is not always possible to fully discern degradation products or co-purified RNA species with similar electrophoretic properties. The small fronting or tailing peaks observed in several of the electropherograms (**Supplementary Figures 3.1D, 3.1G, 3.1H and 3.1L**) may be attributed to these RNAs. Therefore, in these cases, we recommend a second purification using the same 1-D SEC system with strict time segment cut-offs to improve the purity of the targeted RNA species.

Apart from miRNA and sRNAs, our isolation method allows recovery ≥80% for all other RNA species (**Supplementary Figure 3.2A, 3.2B**) on both SEC-3 300 Å and SEC-5 1000 Å columns. In addition, we observed no change in the linearity of the absorption isotherm for each targeted RNA species for up to 100 µg of injected total RNA (peak symmetries 1.0-1.4), with detector saturation at ~490 µg of total RNA injected. These parameters allow the user to optimize the injected load of RNA for purposes of recovery, purity or throughput. Equally important, we observed no volume overloading effects at sample volume injections up to 100 µl.

3.4.2. 1-D ion-pair, reversed-phase chromatography for complete resolution of small RNA species

Ion-pair, reversed-phase chromatography (IP RPC) was employed as the third dimension of separation to resolve all small ncRNA species, including 5S and 5.8S rRNA [18, 33, 34]. Using a SOURCE 5RPC ST 4.5/150 column (5 μm particle size; 4.6 x 150 mm) and triethylammonium acetate (TEAA) for charge neutralization, we achieved complete resolution of small ncRNA from CCRF-SB human lymphoblast cells (**Figure 3.1E**), including separation of miRNA-sized fragments from tRNA and 5S from 5.8S rRNA. The size-based identity of these RNA species was verified by Bioanalyzer (**Supplementary Figure 3.3**). We observed multiple RNA peaks between 31 and 48 min (**Figure 3.1E**), which proved to be ~60-80 nt in apparent length by Bioanalyzer analysis (**Supplementary Figure 3.3E**). Given reports of sequence dependency in IP RPC single-stranded RNA separations [35], we believe that these peaks represent groups of tRNA isoforms. In addition, from 42 to 45 min (**Figure 3.1E**), another series of peaks with apparent lengths between 70-130 nt was observed (**Supplementary Figure 3.3H**), which match the size of several snoRNA and snRNA species [36]. A final noteworthy feature of these studies is that baseline separation of 18S and 28S rRNA was achieved with SEC but not IP RPC (**Figure 3.1A**; **Supplementary Figures 3.1G, 3.1H and 3.3B**). Hence, the two complementary approaches provide complete resolution of all major ncRNA species.

We validated this chromatographic system for preparative isolations with RNA loads up to 40 μg of total RNA. At this scale, we recovered ≥70-80% of the targeted RNA species (**Supplementary Figure 3.2C**) with no significant loss of peak quality and resolution.

3.4.3. 2-D SEC design and validation with total RNA from *Mycobacterium bovis* BCG

The results with 1-D HPLC systems indicates that a 2-D separation utilizing both SEC-5 and SEC-3 columns would allow the complete resolution of most ncRNA under non-denaturing conditions, with the exception of 5S and 5.8S rRNA. To this end, we designed a valve-switching scheme that facilitates an online tandem SEC approach to the separation of total RNA (**Supplementary Figure 3.4**). In this non-orthologous 2D HPLC scheme, high molecular weight (MW) RNA (> 500 kDa; *e.g.*, 28S, 23S, 18S, 16S rRNA) is resolved on the SEC-5 column, while low MW RNA (<40 kDa; *e.g.*, 5.8S/5S rRNA, tRNA, miRNA) is resolved on the SEC-3 column. When total cellular RNA is injected onto the SEC-5 column, high MW species are resolved from each other and from the retained low MW RNAs, which elute as a group directly onto the SEC-3 column where they are fractionated. Conversely, high MW RNA species separated on the SEC-5 column bypass the SEC-3 column and elute directly. This arrangement is completely modular and the column positions can be swapped while achieving the same level of separation – with the low MW RNA eluting before the high MW species. We performed 2-D SEC on total RNA extracted from *Mycobacterium bovis* Bacille Calmette-Guérin (BCG) and separated its entire range of major ncRNA species within a single 40 min run (**Figure 3.2A**). Reproducible baseline separations were achieved for tRNA and 5S rRNA ($R_s = 4.4 \pm 0.1$), 5S rRNA and 16S rRNA ($R_s = 22 \pm 0.2$), and 16S rRNA and 18S rRNA ($R_s = 3.9 \pm 0.1$) for total RNA extracted from three independent BCG cultures. Additionally, the chromatography did not affect the spectrum or quantity of modified ribonucleosides in tRNA (**Figure 3.3B**; **Supplementary Table 3.1**), relative to our published values [31].
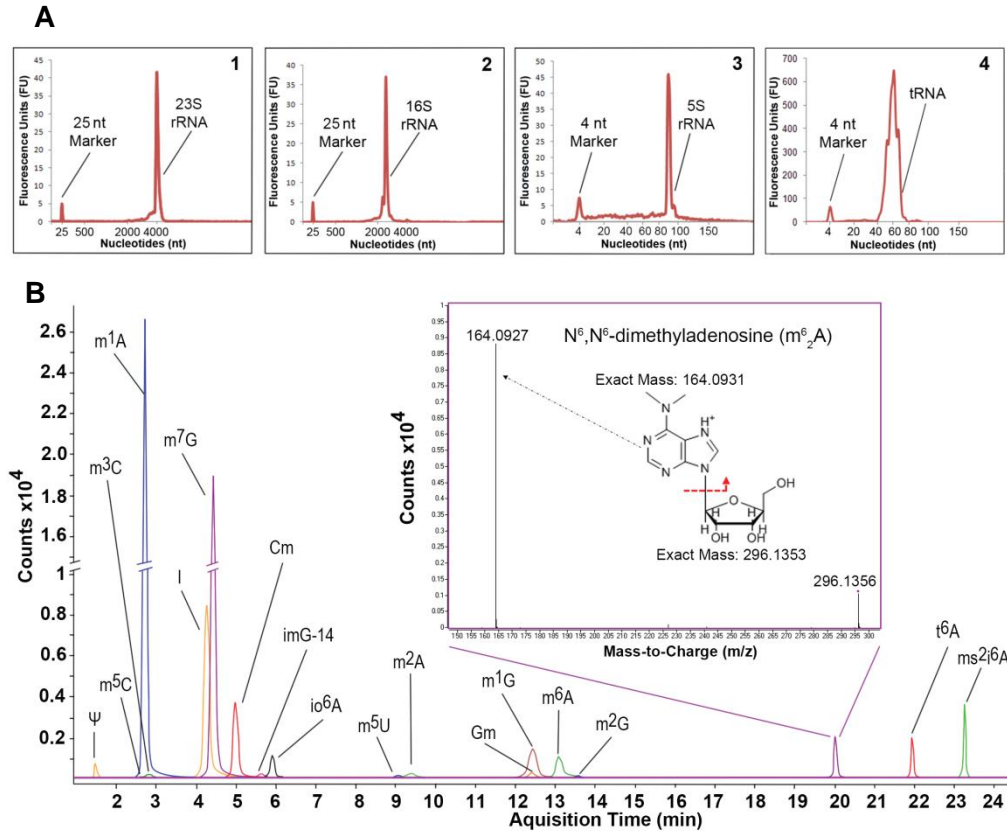
Figure 3.2. 2D-SEC of BCG total RNA preserves the native post-transcriptional ribonucleoside modifications in purified tRNA. (**A**) Baseline separation of BCG total RNA to its component 23S (insert 1), 16S (insert 2), 5S rRNAs (insert 3) and tRNA (insert 4) obtained on a Bio SEC-5 1000 Å (column 1) and Bio SEC-3 300 Å (column 2) two-column online HPLC system (ID 7.8mm, 300mm for each column) demonstrating the purity and quality of each RNA species based on sequence lengths on the Bioanalyzer RNA 6000 Pico and Small RNA LabChips. (**B**) Extracted ion chromatograms of naturally occurring modified ribonucleosides in hydrolyzed BCG tRNA identified by HPLC-coupled triple quadrupole mass spectrometry in multiple reaction monitoring mode. The peaks of m^{1}A (2.87 min) and m^{7}G (4.53 min) are marked with '//' to indicate that they are in different scales from other peaks. Identities of the 18 post-transcriptional modifications detected are shown in Supplementary Table 2. Inset: MS/MS validation of N^6,N^6-dimethyladenosine (m$^6_2$A). Ribonucleoside with *m/z* of 296.1356 was targeted for CID fragmentation on a high mass accuracy quadrupole time-of-flight mass spectrometer yielding a daughter ion with *m/z* 164.0927.

3.4.4. Limitations of 2-D SEC

To ascertain the biologically relevant exclusion limits of the 2-D SEC platform,

we performed studies covering the size range of 22 to 10,700 nt. Calibration

curves with synthetic RNA MW markers [37] would not provide information

about the exclusion and penetration limits for the columns for biological RNA species given the presence of RNA secondary structures that affect the molecular dimensions and thus chromatographic behavior of biological RNA molecules. Therefore, we used viral RNA and 22 nt miRNA-like molecules to define the exclusion limits of the 2-D system. We first spiked *in vitro* transcribed, full-length dengue viral RNA genome (DENV vRNA) measuring 10.7 kb into total RNA extracted from CCRF-SB cells. Although a separation with a resolution of 3.7 was achieved for 28S rRNA and vRNA using an Agilent SEC5 column (**Supplementary Figure 3.5A**), comparison of this chromatogram with that of DENV vRNA alone reveals a large tailing from the vRNA peak that spreads into the 28S rRNA peak. This implies that pure fractions of vRNA but not 28S rRNA can be purified from this mixture using the current 2-D SEC setup. Similarly, to determine the biologically relevant penetration limit of the 2-D SEC platform, we analyzed the resolution of a mixture of random 22 nt synthetic oligos, to simulate miRNA, and purified tRNA from CCRF-SB cells on an Agilent SEC3 column (**Supplementary Figure 3.5B**). Although a partial separation could be achieved between these two small ncRNA species, given the low cellular abundance of miRNA relative to tRNA, isolation of pure fractions of miRNA, solely based on SEC, would not be possible under these 2-D SEC conditions. An additional IP RPC step could solve this problem (see **Figure 3.1E**).

To demonstrate the resolving power of this multi-dimensional (2-D SEC, IP RPC) chromatographic approach to RNA purification, we reconstructed the RNA landscape of TK6 total RNA using the retention times of 2-D SEC and IP RPC chromatograms to define spatial localizations on the X-Y plane and peak area as the Z-axis (**Figure 3.3A**). As the relative intensities in the small RNA

region are two orders-of-magnitude smaller than those of large ribosomal

RNA, we rescaled this region in **Figure 3.3B**.
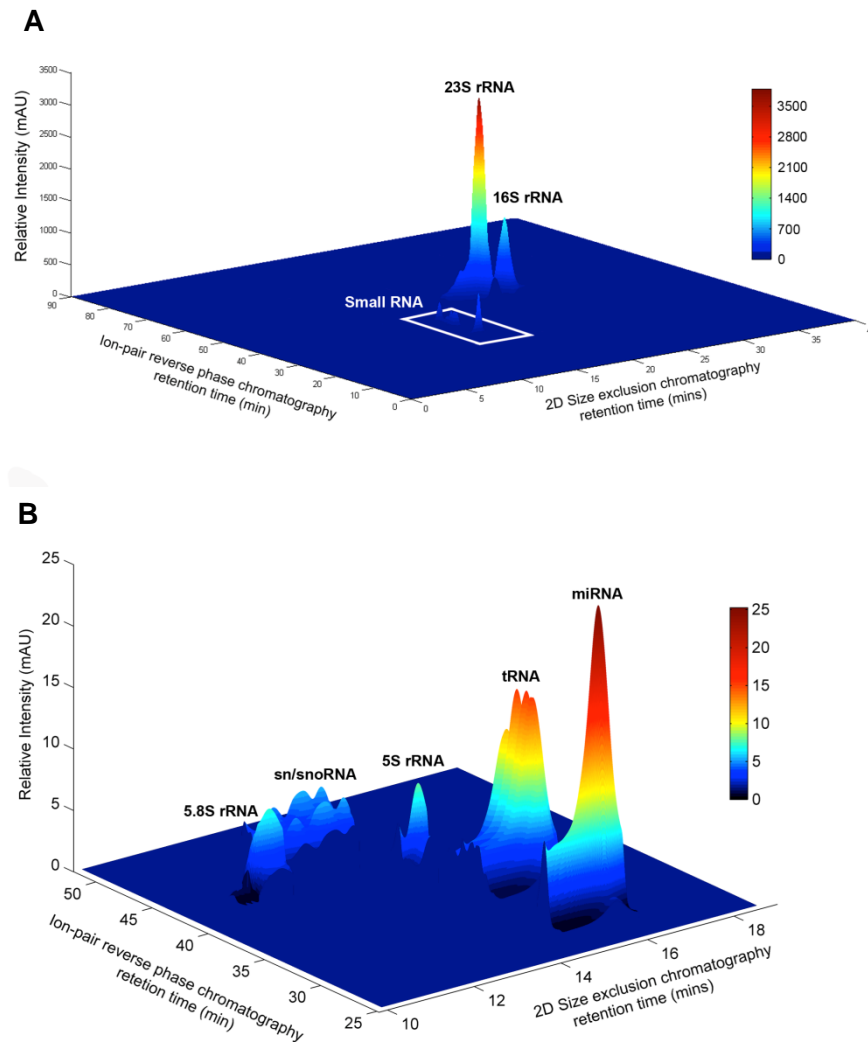
Figure 3.3. Reconstruction of the RNA landscape of EBV transformed TK6 cells by 2D SEC and IP RPC. (**A**) The 3-D surface plot shows the analysis of 500 ng of total RNA extracted from TK6 cells using both 2-D SEC (Column 1: Bio SEC-3 300 Å; Column 2: Bio SEC-5 2000 Å) and IP RPC (SOURCE 5RPC ST 4.5/150). (**B**) Enhanced view of the smaller RNA species consisting of miRNA, tRNA, 5s and 5.8s rRNAs and putative snRNA and/or snoRNA. The identity and purity of the component RNA from orthographic projections of 2-D SEC and IP RPC chromatograms are shown in **Supplementary Figure 3.7**.

3.4.5. Application of 2-D SEC for isolation of *Plasmodium berghei* ncRNA from infected reticulocytes

To assess the resolving power of the HPLC platform, we applied it to the particularly challenging problem of separating host RNA species from those of an infecting microbial pathogen, using the example of ncRNA from malaria parasites infecting mammalian red blood cells (RBC). *Plasmodium berghei* is a protozoan parasite that causes malaria in rodents and is used as an *in vivo* disease model for human malaria [38]. *P. berghei* preferably infects reticulocytes in which residual host RNA still exists. As a result, trace amounts of host RNA could contaminate preparations of *P. berghei* schizonts from lysed infected reticulocytes, as shown in **Figure 3.4A**. We injected total cellular RNA from purified *P. berghei* schizonts onto the 2-D SEC system and resolved the variety of small and large ncRNA species (**Figure 3.4B**), with Bioanalyzer analysis of the fractions collected for each peak confirming the predicted length of the isolated RNA species (**Supplementary Figure 3.6**). In particular, pure fractions of tRNA, 5.8S rRNA, and 5S rRNA were isolated, as well as putative snoRNA and snRNA species. We were able to completely resolve the unique ~800 nt partner of another ~3000 nt fragment that combine functionally to form the 28S rRNA of *P. berghei* (**Figure 3.4A**) [39, 40]. Baseline separation of the ~3000 nt 28S rRNA fragment and putative 18S rRNA was not achieved due to their close proximity in time (**Figure 3.4B**). In addition, no miRNA-length small RNA species were detected.

3.4.6. Fluorometric quantification of purified RNA

Of the many commonly used spectroscopic methods for quantifying RNA, fluorescent RNA-binding dyes have proven to be highly sensitive and specific when a simple correction is applied for DNA contamination [41] and they
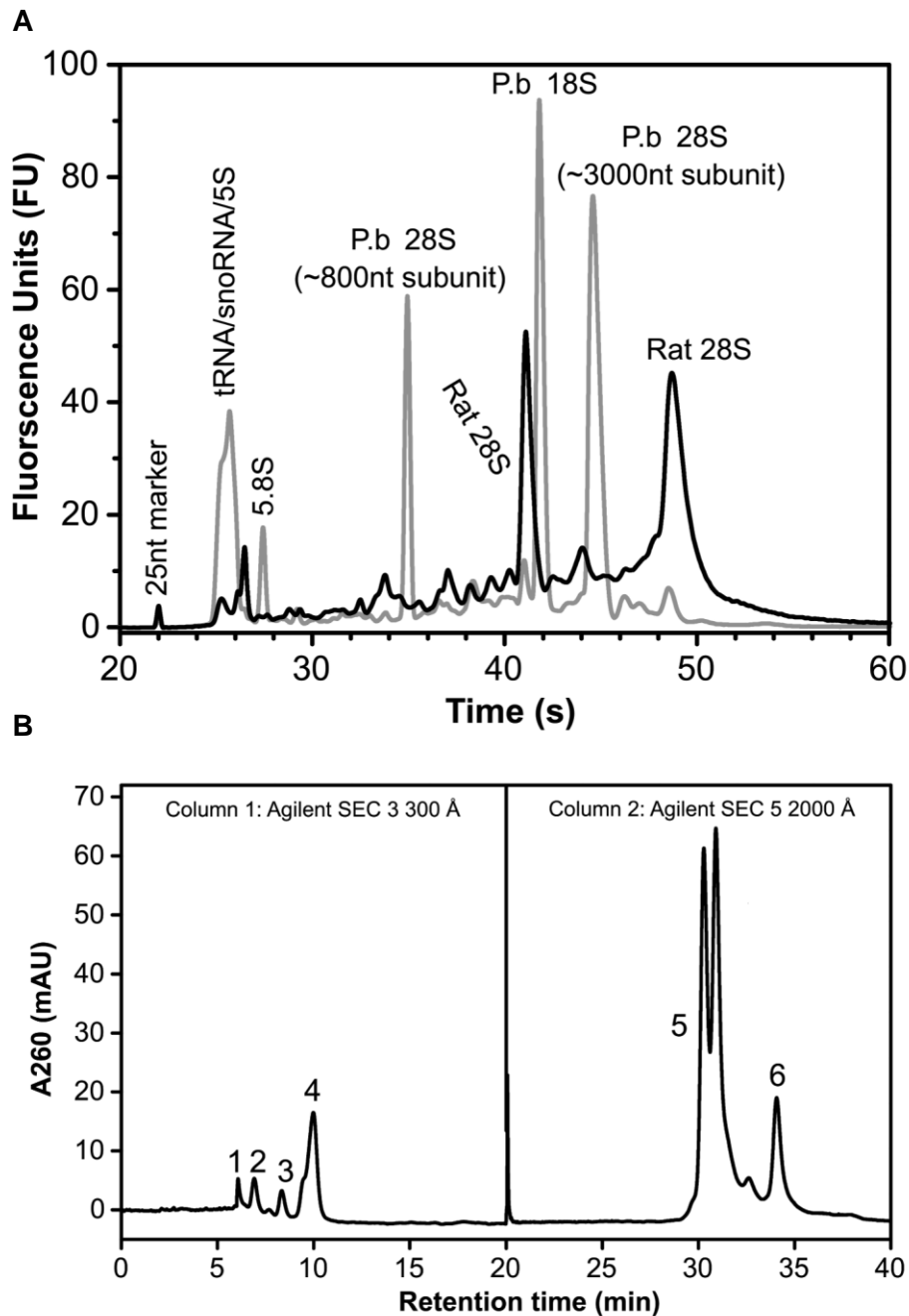
Figure 3.4. Isolation of ncRNA from *P. berghei*-infected rodent reticulocytes. (**A**) Bioanalyzer analysis of total RNA extracted from whole blood from uninfected rats (black line) and total RNA extracted from *P. berghei* schizonts purified from infected rat erythrocytes following lysis of the RBC and washing of the schizonts, as described in Online Methods (gray lane). (**B**) 2D-SEC purification of *P. berghei* ncRNA. Total RNA from schizonts purified from lysed rodent reticulocytes was resolved on the 2-D HPLC system, with Bio SEC-3 300 Å resolution of 5.8S rRNA (1), 5s rRNA (2), putative snRNA/snoRNA (3) and tRNA (4), and Bio SEC-5 1000 Å resolution of the 28s 800nt rRNA fragment (6) from the co-eluting 18s rRNA and 28s 3000nt rRNA fragment (5). Individual RNA species were collected from the 2-D HPLC elution and the purity and identity of each fraction was evaluated by Bioanalyzer analysis as shown in **Supplementary Figure 3.6**.

avoid the problems of protein and DNA contamination associated with absorbance at 260 nm. However, the variety of secondary structures in different RNA species could affect the efficiency of fluorescence emission upon dye binding and thus affect quantitative accuracy. To resolve this problem, we used the 2-D HPLC platform to purify specific ncRNA species and then determined fluorescence correction factors for each type of RNA. Purified fractions of 28S rRNA, 18S rRNA and tRNA from CCRF-SB, and 23S rRNA, 16S rRNA and tRNA from *E. coli* DH5α were concentrated, desalted, qualitatively assessed (Bioanalyzer 6000 Nano LabChip) and quantified by absorbance at 260 nm. Following addition of Ribogreen dye, fluorescence emissions were quantified in dilutions of each RNA standard over concentrations ranging from 20-1000 ng/mL. As shown in **Figure 3.5**, there are significant differences in the fluorescence emission profiles of each of the ncRNA species when compared to a standard using total cellular RNA. Particularly, the fluorescence signal for RiboGreen increases by 10% for 28S rRNA compared to CCRF-SB total RNA, while it decreases by up to 11% for bacterial tRNA compared to *E. coli* total RNA (Supplementary Tables 3.2A, 3.2B). Application of these weighting factors thus allows the use of readily purified total RNA from target cells as a standard for accurate quantification of a specific ncRNA using fluorescence dye assays.

3.5. Discussion

To facilitate the study of the chemistry and biology of ncRNA, we developed a multi-dimensional HPLC platform that reproducibly and efficiently purifies the major species of ncRNA across the complete size range up to 10,000 nt. This universal platform allows resolution, purification and quantification of ncRNA species in a cell under non-denaturing conditions and with post-transcriptional ribonucleoside modifications intact. The method was validated by purifying
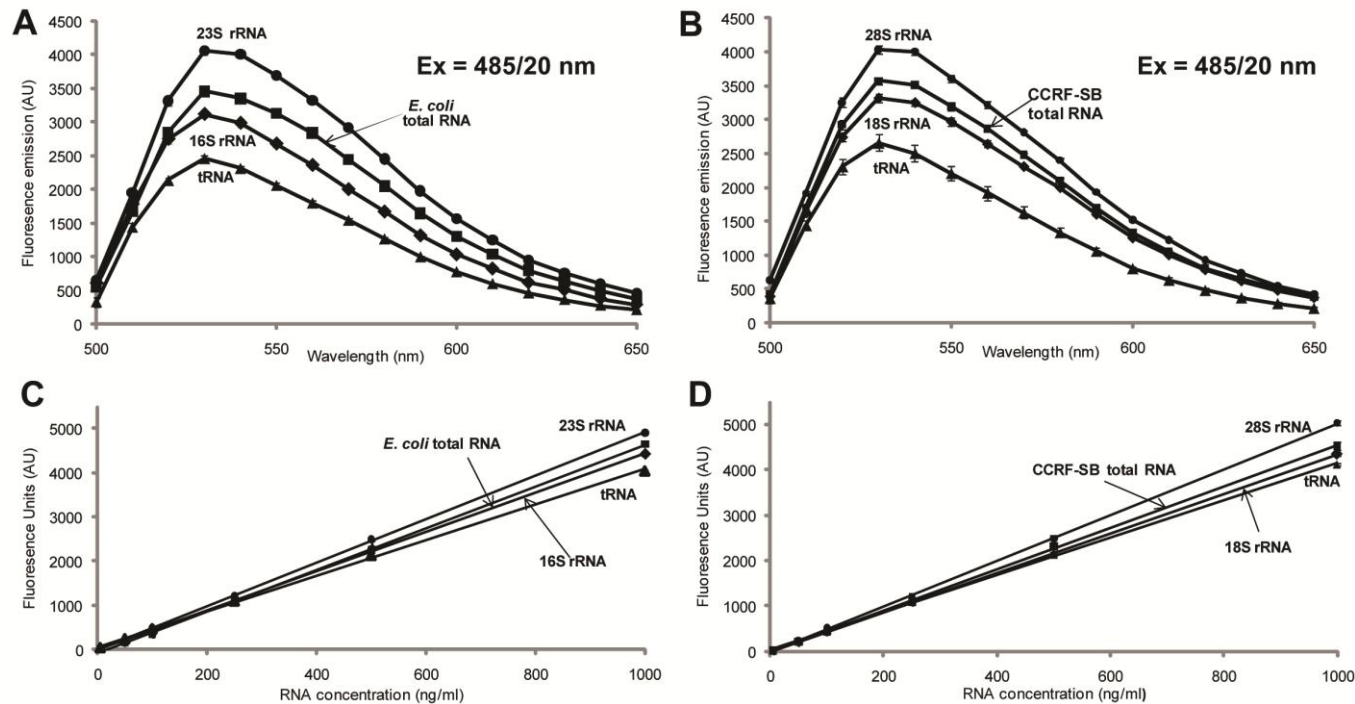
Figure 3.5. Fluorometric quantitation of purified RNA using RiboGreen with adjustments for species-specific responses. (**A**) Fluorescence enhancement of RiboGreen upon binding to *E. coli* total RNA (■) and HPLC purified *E. coli* 23S rRNA (●), 16S rRNA (◆) and tRNA (▲). (**B**) Fluorescence enhancement of RiboGreen upon binding to CCRF-SB total RNA (■) and HPLC purified CCRF-SB 28S rRNA (●), 18S rRNA (◆) and tRNA (▲). (**C**) RNA concentration curves of *E. coli* total RNA (■) and HPLC purified *E. coli* 23S rRNA (●), 16S rRNA (◆) and tRNA (▲). (**D**) RNA concentration curves of CCRF-SB total RNA (■) and HPLC purified CCRF-SB 28S rRNA (●), 18S rRNA (◆) and tRNA (▲). The fluorescence of each RNA species with concentrations from 0 to 1000 ng/ml stained with RiboGreen was measured for fluorescence at 528/20nm.

putative tRNA, small and large subunit rRNAs, and other small RNAs from *E. coli*, mycobacteria, human cells and plasmodium-infected rat reticulocytes, as well as *in vitro* transcribed, full-length viral RNA. As shown in **Figure 3.3**, the successful resolution of the major ncRNA species by the combined 2-D SEC and IP RPC provides a complete picture of the RNA landscape within a cell or tissue and allows both sequencing of purified ncRNA and chemical interrogation of the pool of ncRNA species, such as an analysis of the stress-induced reprogramming of the spectrum of modified ribonucleosides in all ncRNA species in a cell [12, 13, 31].

This multi-dimensional chromatographic approach offers significant advantages over other methods for purifying ncRNA. Conventional preparative polyacrylamide gel electrophoresis is time-consuming (≥ 2 d), denaturing, and limited to RNA sizes less than 600 nt [42], with eluted RNA contaminated with acrylamide oligomers [15] and any dyes used to visualize the RNA. For purification of tRNA, denaturing (urea) polyacrylamide gels [43] have the advantage that larger amounts of RNA can be loaded into a single gel, with 1-3 mg of total RNA resolved on a 10% polyacrylamide gel (8 M urea) yielding 240 ± 60 ug of tRNA [13]. With tRNA representing ~14% of total RNA in yeast (based on areas under the curve for UV absorbance signals from HPLC chromatograms), this yield of tRNA represents ~70% of input tRNA. However, this does not account for losses in subsequent removal of polyacrylamide fragments by filtration and/or purification by SEC HPLC. Similarly, agarose gels can be used for resolution and purification of larger RNAs (> 600 nt) [44], with recovery by a variety of methods [45-48] that suffer from the same contamination problems. In contrast, RNA chromatography is faster, can be performed under non-denaturing conditions, utilizes simple

solvent mobile phases that do not contaminate the sample, and is more sensitive and easily automated for fraction collection [49].

The multi-dimensional chromatography platform developed here also has advantages over existing chromatographic methods that target individual ncRNA species [18, 22, 32, 50]. The newer Agilent SEC columns solve difficulties with the separation of large molecular weight rRNA observed in earlier work with RNA SEC [22, 32]. This limitation was only partially relieved in more recent efforts that utilize IP RPC for RNA separation [18, 50]. Conversely, purification of RNA species in the low molecular weight size ranges (<100 nt) often requires an additional enrichment step either through the use of commercially available small RNA kits or with solid phase extraction processes due to their low natural abundance and inefficiencies in extracting them with the commonly used acidic phenol chloroform method [51, 52] without further enrichment on silica gel or glass fiber matrices [53]. Although commercial systems such as LabChip XT (Caliper) and Pippin Prep (Sage Science) have recently been developed for size-fractionation of DNA and analysis of quality of cDNA, an equivalent system for the analysis and separation of RNA that covers the entire range of biologically relevant lengths is not available.

We solved these technical problems using multi-dimensional HPLC. By coupling two different SEC columns together with a comprehensive offline IP RPC column, we were able to purify all major ncRNA species from miRNA to large 28s rRNA from a single batch of total RNA. The system was also capable of purifying viral genomic RNA intact from the host cell mixture of ncRNA. Interestingly, we were able to purify miRNA from CCRF-SB and TK6 B lymphoblast cell lines but not from *P. berghei*-infected rat reticulocytes, which is consistent with recent work suggesting that plasmodium parasites do

not express miRNA [54]. These results demonstrate the utility of this HPLC

platform for analysis and purification of the various ncRNA species within a

cell. The sequences of the isolated RNA species can be collaborated their

genomic sequences by either RNA-seq or northern blotting. It is also

important to point out that, while 2-D chromatography provides the most

efficient means to purify individual RNA species, the individual 1-D HPLC

systems can be used separately to achieve the same results as the 2-D

system in the event that the 2-D HPLC technology is not available.

As with other chromatographic methods, the major limitation of this HPLC

approach to RNA purification lies in the contamination of specific ncRNA

fractions with similarly sized degradation products from larger RNA species.

The problem presented by such contamination is illustrated with the analysis

of modified ribonucleosides in RNA. For example, the nucleoside m^6_2A has

been identified in the 16S and 23S rRNA of archaea and bacteria and the 18S

rRNA of eukarya, but never in tRNA. However, we identified m^6_2A as an

abundant modified ribonucleoside in tRNA from BCG [31] – detailed in

Chapter 4, which raised the question of rRNA contamination of the tRNA

fraction. m^6_2A was not detected in HPLC-purified tRNA from yeast, rat liver

and human cells, which partially ruled out contamination with rRNA fragments

[31], and Bioanalyzer analysis further confirmed a lack of degradation of

rRNA in all of the samples [31]. Along with such comparative analysis,

contamination of small RNA fractions with fragments from larger RNA species

can be assessed by RNA sequencing methods [55, 56] and can potentially be

removed by repurification of RNA fractions by an orthogonal chromatographic

system that provides another basis for separating the contaminating RNA

species, such as SEC-purified tRNA analyzed on IP RPC.

Another critical parameter for many applications of purified ncRNA species involves accurate quantification of the isolated RNA. For example, analysis of stress-induced reprogramming of modified ribonucleosides in tRNA [12, 13, 31] requires highly accurate quantification of input tRNA to control variance and allow meaningful comparisons of small changes. Our observation of variable emission spectra when RiboGreen is bound to different purified ncRNA species illustrates the impact of RNA composition on fluorescence-based quantification methods. Interestingly, it has been reported that PicoGreen, a related dye used for DNA quantitation, yields divergent fluorescence emissions when bound to genomic DNA from different organisms (57). To obviate the host of structure- and sequence-based causes of this discrepancy (58-60), we took the pragmatic approach of empirically characterizing fluorescent emission for many types of ncRNA and presenting the data as correction factors (Supplementary Table 1) for the major ncRNAs in E. coli and human cells (CCRF-SB) when using total cellular RNA as the calibration standard.

Among the most obvious applications for this RNA purification method are targeted RNA sequencing, characterization of modified ribonucleosides, purification of novel RNA species, and analysis of viral RNA genomes. Targeted purification of RNA is essential for structural and functional characterization of novel RNA species, such as the large number of regulatory long ncRNA . The functional forms of these ncRNA after post-transcriptional processing events can then be correlated with their genomic sequences. The resulting enrichment for a specific transcript increases the sequencing power and efficiency for rare variants. Purification of individual ncRNA species is also critical for identifying and quantifying the dozens of chemically modified ribonucleosides incorporated post-transcriptionally into

both coding (5) and ncRNA (10) species as part of the mechanism of translational control of cell response (12,13). The analysis of viral RNA genomes would also be aided by the ability to purify the virus RNA free from host RNA species. For example, it was recently discovered that the 11 kb RNA genome of the dengue virus is subject to adenosine methylation, which warrants purification of the viral RNA for analysis of other modified ribonucleosides. Hence, this platform provides a general strategy for preparing RNA for large-scale sequencing and composition analysis, and for downstream functional elucidation projects.

3.6. Supplementary material

3.6.1 Supplementary figures

(next page)
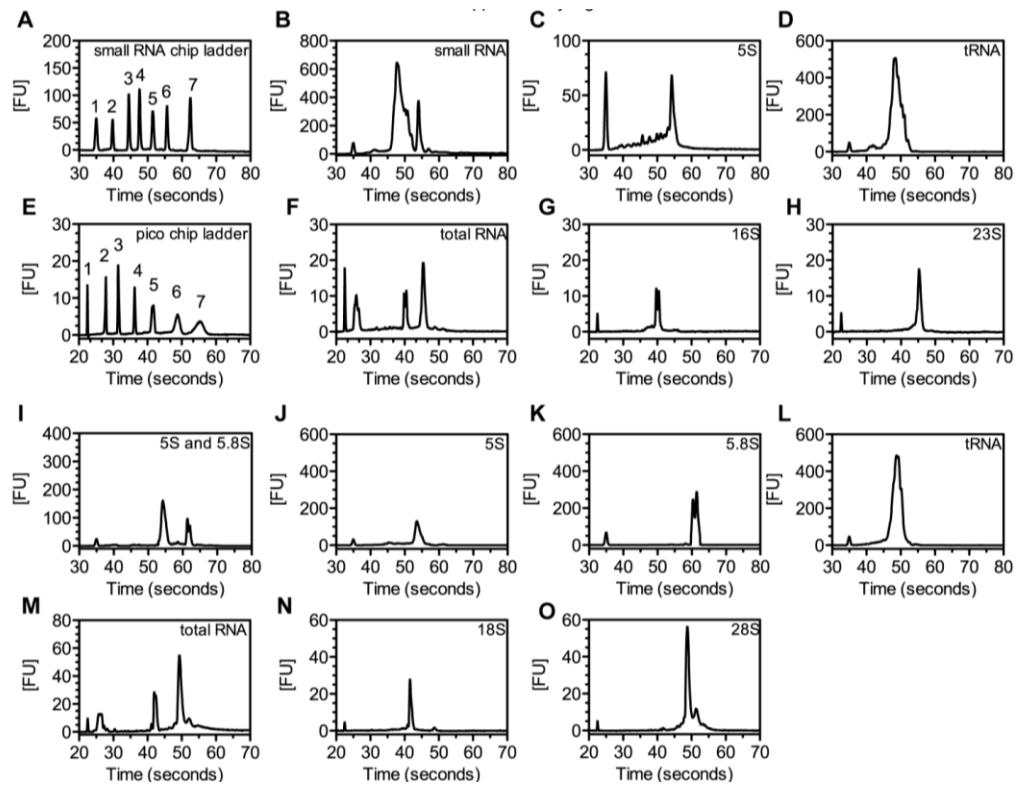
Supplementary Figure 3.1. Validation of RNA identity and purity of *E. coli* and CCRF-SB RNA species isolated using SEC by Bioanalyzer LabChip analysis. *Top row*: (**A**) Small RNA chip ladder standards (1, 4 nt; 2, 20 nt; 3, 40 nt; 4, 60 nt; 5, 80 nt; 6, 100 nt; 7, 150 nt). (**B**) *E. coli* small RNA consisting of 5s rRNA and tRNA isolated using a Bio SEC-3 300 Å column. (**C**) *E. coli* 5s rRNA isolated using a Bio SEC-3 300 Å column. (**D**) *E. coli* tRNA isolated using a Bio SEC-5 1000 Å column. *Second row*: (**E**) RNA 6000 Pico chip ladder standards (1, 25 nt; 2, 200 nt; 3, 500 nt; 4, 1000 nt; 5, 2000 nt; 6, 4000 nt; 7, 6000 nt). (**F**) *E. coli* total RNA. (**G**) *E. coli* 16s rRNA isolated using a Bio SEC-5 1000 Å column. (**H**) *E. coli* 23s rRNA isolated using a Bio SEC-5 1000 Å column. *Third row*: (**I**) CCRF-SB small RNA consisting of 5.8s and 5s rRNAs isolated using a Bio SEC-5 1000 Å column. (**J**) CCRF-SB 5s rRNA isolated using a Bio SEC-3 300 Å column. (**K**) CCRF-SB 5.8s rRNA isolated using a Bio SEC-3 300 Å column. (**L**) CCRF-SB tRNA isolated using a Bio SEC-3 300 Å column. *Bottom row*: (**M**) CCRF-SB total RNA. (**N**) CCRF-SB 18s rRNA isolated using a Bio SEC-5 1000 Å column. (**O**) CCRF-SB 28s rRNA isolated using a Bio SEC-5 1000 Å column.
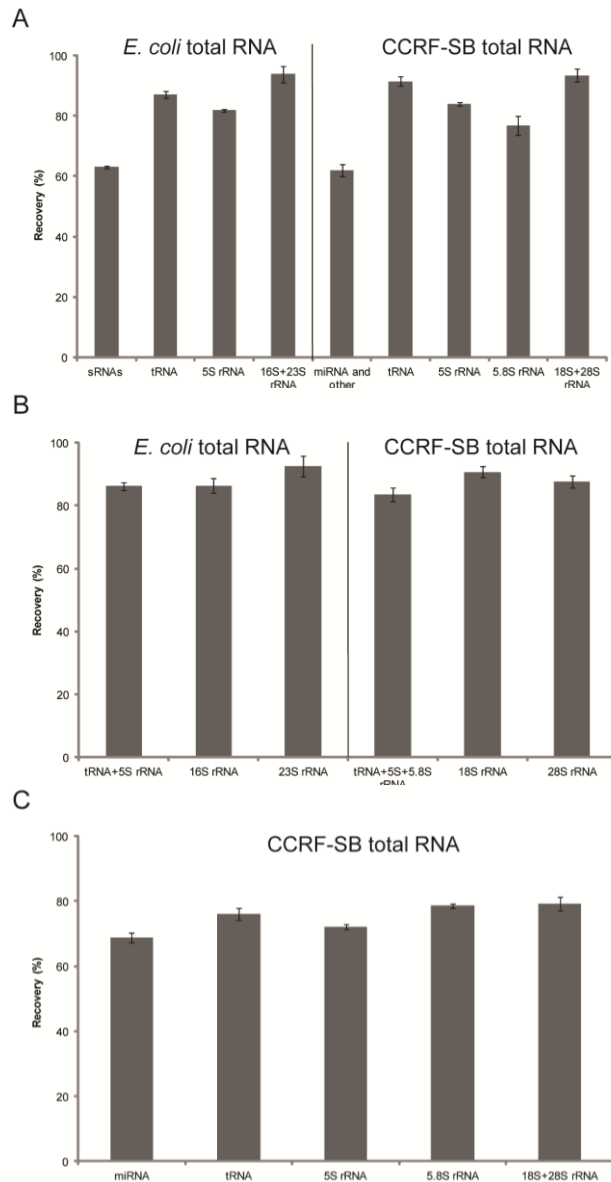
Supplementary Figure 3.2. Yields of ncRNA purified by SEC or IP-RPC.
Fraction collection was automated by use of a fraction collector and RNA
collected by custom time segments corresponding to the retention times and
peak widths of each RNA species. Samples clean-up was then performed as
stated in Material and Methods. (**A**) Recoveries (%) of sRNA (of lengths
smaller than tRNA), tRNA, 5S rRNA and co-eluding 16S and 23S rRNA from
E. coli total RNA (left) and miRNA and other sRNA, tRNA, 5S rRNA, 5.8S
rRNA and co-eluding 18S and 28S rRNA from CCRF-SB total RNA (right) off
the Bio-SEC3 300Å HPLC system. (**B**) Recoveries (%) of co-eluding 5S rRNA
and tRNA, 16S rRNA and 23S rRNA from *E. coli* total RNA (left) and co-
eluding 5S, 5.8S rRNA and tRNA, 18S rRNA and 28S rRNA from CCRF-SB
total RNA (right) off the Bio-SEC5 1000 Å HPLC system. 10-100 µg of total
RNA are injected unto the SEC systems. (**C**) Recoveries (%) miRNA, tRNA,
5S rRNA, 5.8S rRNA and co-eluding 18S and 28S rRNA from the Source
RPC 4.6/150 HPLC system. 2-40 µg of total RNA are injected into the IP RPC
system.
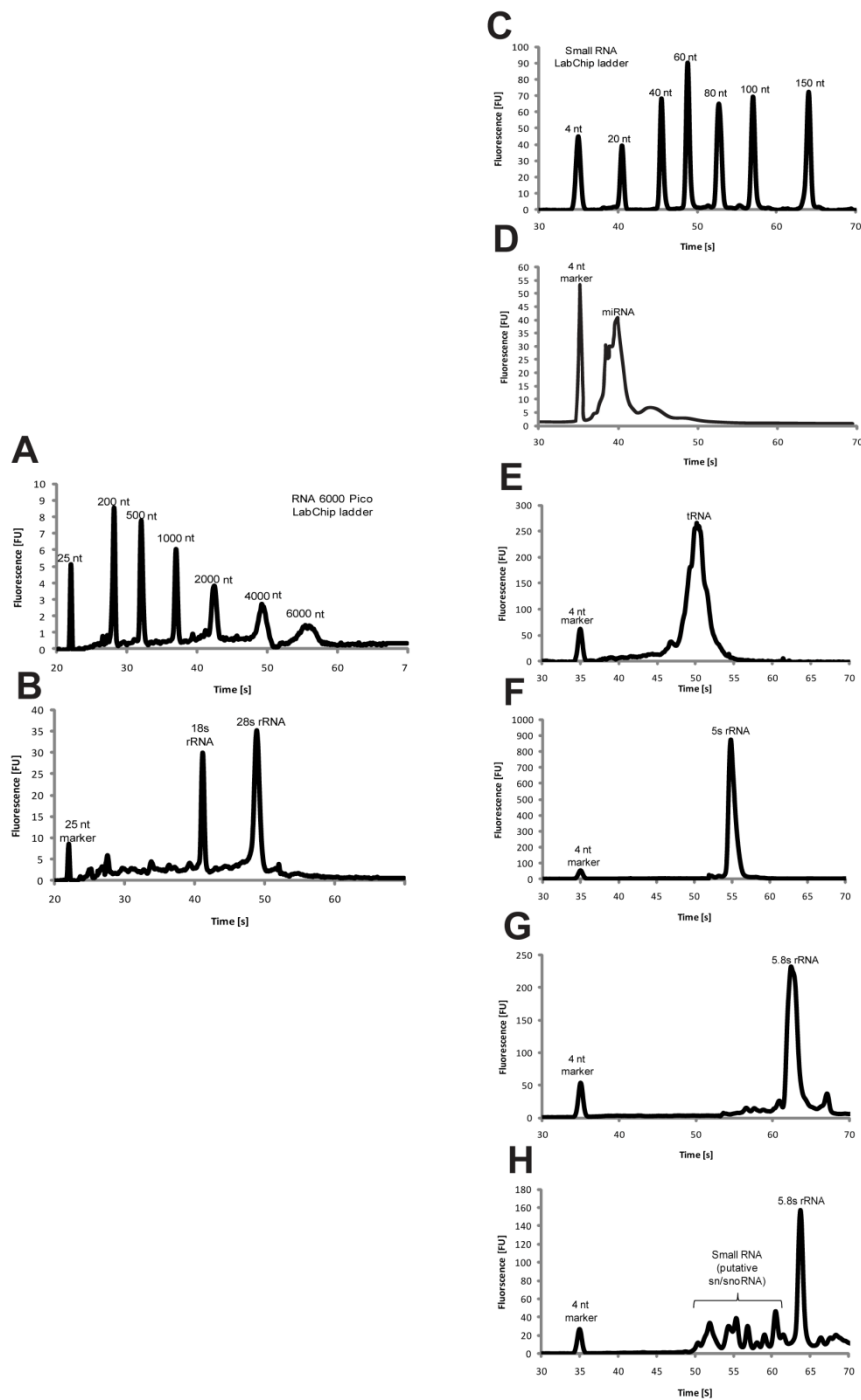
Supplementary Figure 3.3. Validation of RNA identity and purity of CCRF-SB RNA species isolated using IP RPC on a Source 5RPC 4.6/150 column by Bioanalyzer LabChip analysis. (**A**) RNA 6000 Pico chip ladder standards (From left to right: 25 nt; 200 nt; 500 nt; 1000 nt; 2000 nt; (**B**) Co-eluding 18S and 28S rRNA analyzed on a RNA 6000 Pico chip. 4000 nt; 6000 nt). (**C**) Small RNA chip ladder standards (From left to right: 4 nt; 20 nt; 40 nt; 60 nt; 80 nt; 100 nt; 150 nt). (**D**) miRNA, (**E**) tRNA, (**F**) 5S rRNA, (**G**) 5.8S rRNA, (**H**) putative sn/sno RNA analyzed on a small RNA chip.
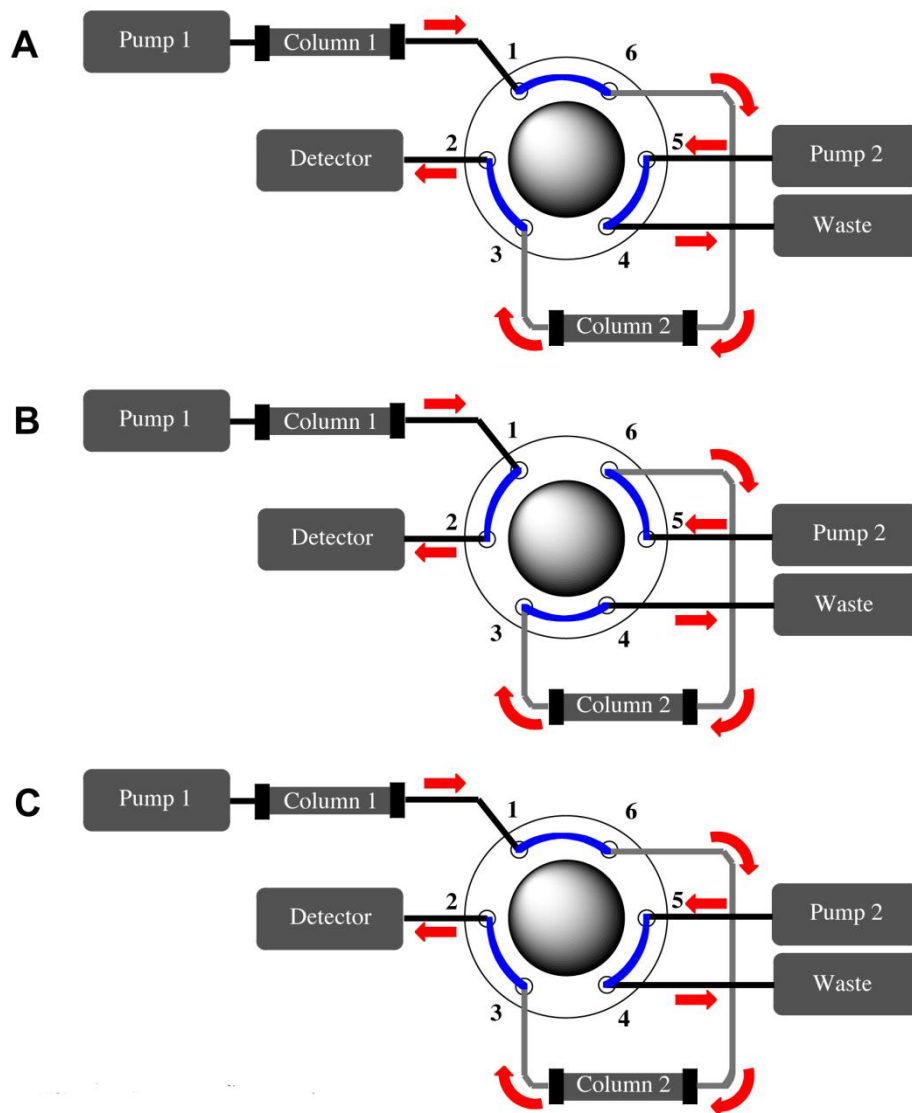
Supplementary Figure 3.4. Valve configuration for 2-D SEC. Either Bio SEC-3 or Bio SEC-5 columns can be connected into the column 1 or column 2 positions. (**A**) Step 1: Columns are in tandem and detector is in-line. If Bio SEC-5 is in the column 1 position, small RNAs are fractioned from large molecular weight rRNAs on the Bio SEC-5 and are eluted onto the Bio SEC-3 column (column 2) for separation. Alternatively, If Bio SEC-3 is in the column 1 position, large molecular weight rRNAs are eluted onto the Bio SEC-5 column (column 2) for separation. (**B**) Step 2: Column 2 is taken off-line (switched to pump 2) while the detector is in-line with column 1. Hence, if Bio SEC-5 is in the column 1 position, large molecular weight rRNAs are quantified by absorbance and collected. Alternatively, if Bio SEC-3 is in the column 1 position, small RNAs are quantified and collected. With column 2 in-line with pump 2, RNA separation can continue at the same or lower flow rates depending on the run time of column 1. (**C**) Step 3: Column 1 and 2 are connected in tandem and the detector is connected in-line with column 2. RNAs from column 2 (small RNA if Bio SEC-5 is in the column 1 position, and large molecular weight rRNAs if Bio SEC-3 is in the column 1 position) are quantified and collected
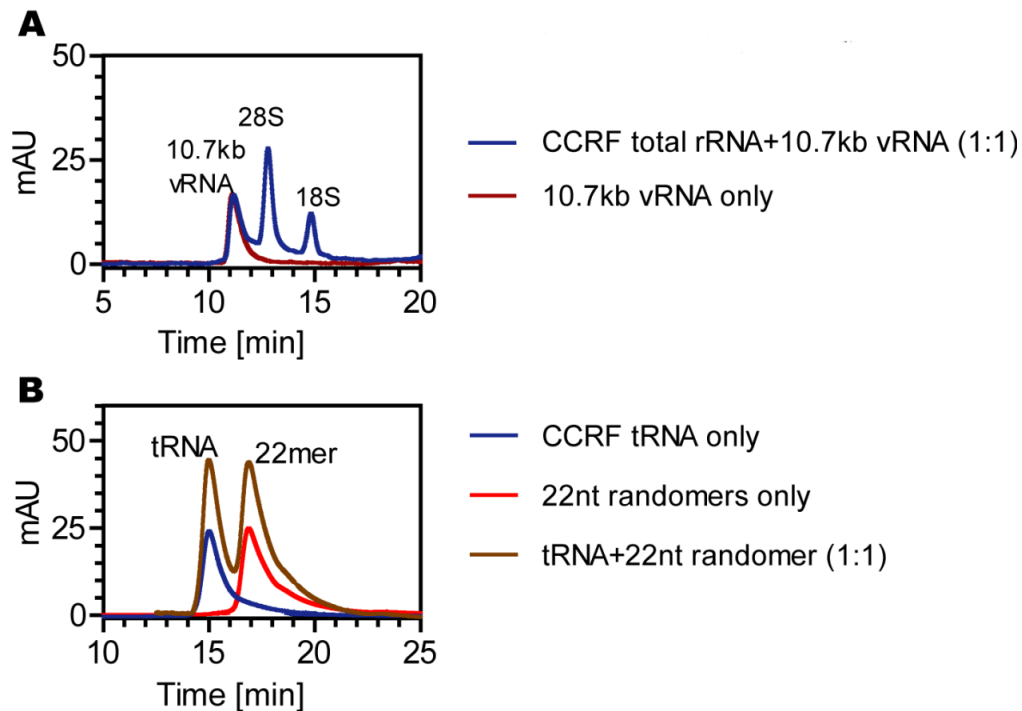
Supplementary Figure 3.5. Biologically relevant exclusion and permeation limit of SEC HPLC. (**A**) Exclusion limit of Bio SEC-5 1000 Å (ID 7.8mm, 300mm) assessed with 10.7 kb *in vitro* transcribed full length DENV vRNA. Chromatograms of 50 ng of DENV 10.7 kb vRNA alone (red) and 50 ng of 10.7 kb vRNA spiked into 50 ng of CCRF-SB total RNA (blue) performed with 100 mM ammonium acetate buffer (pH 7.0) at a flow rates of 0.5 ml/min. (**B**) Permeation limit of Bio SEC-3 300 Å (ID 7.8mm, 300mm) assessed with synthetic 22 nt random oligonucleotides. Chromatograms of 100 ng of 22 nt random oligonucleotides alone (red), 100 ng of CCRF-SB tRNA alone (blue) and 100 ng of 22 nt random oligonucleotides spiked into 100 ng of CCRF-SB tRNA (brown) performed with 100 mM ammonium acetate buffer (pH 7.0) at a flow rate of 0.5 ml/min.
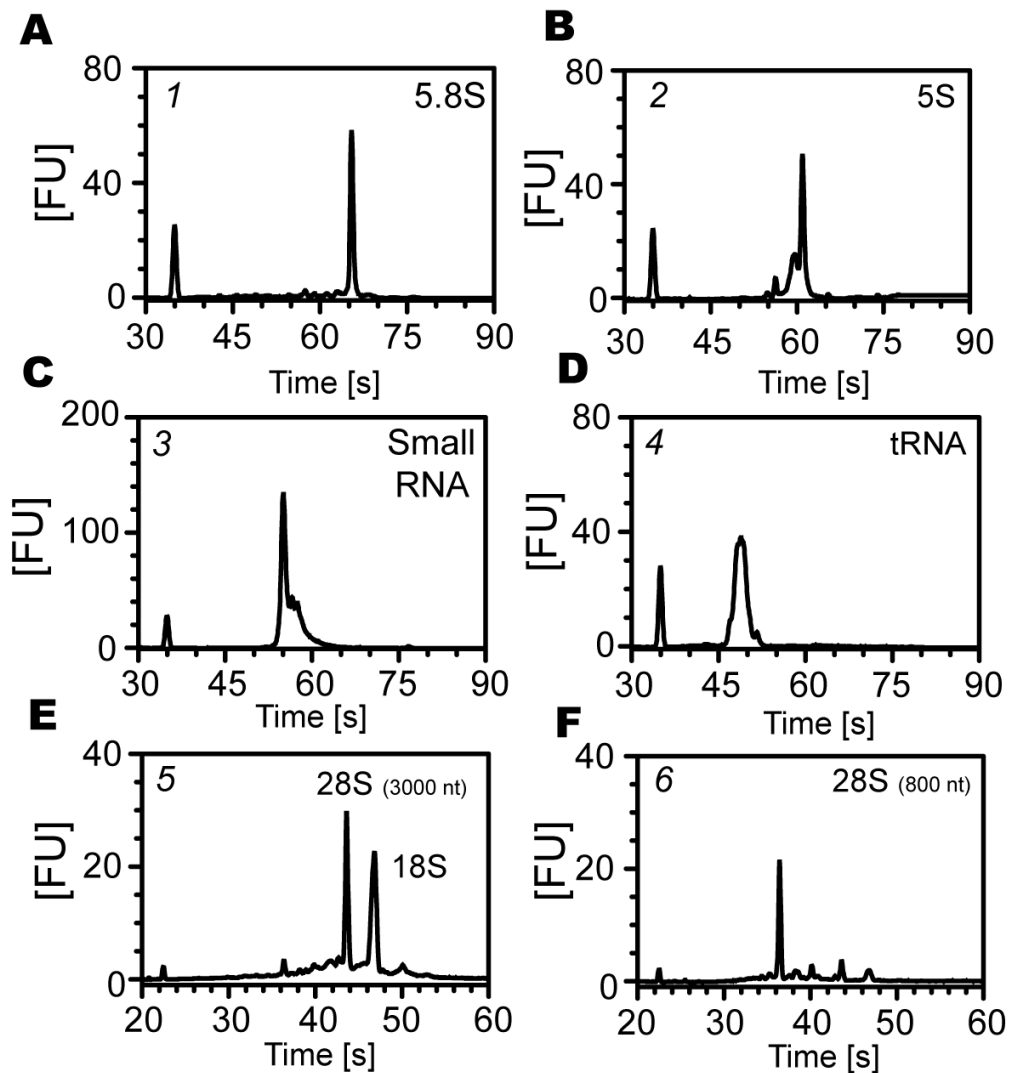
Supplementary Figure 3.6. Bioanalyzer LabChip validation of RNA identity and purity of _P. berghei_ RNA species isolated using 2-D SEC. (**A**) 5.8S rRNA (fraction 1). (**B**) 5S rRNA (fraction 2). (**C**) Small RNA (putative snoRNA and snRNA; fraction 3). (**D**) tRNA (fraction 4). (**E**) 28S rRNA 3000 nt fragment and 18S rRNA isolated from fraction 5 as labeled in **Figure 3.4b**. (**F**) 28S rRNA 800 nt fragment isolated from fraction 6. Fractions 1-4 are analyzed on a Bioanalyzer Small RNA chip while Fraction 5 and 6 are analyzed on a Bioanalyzer RNA 6000 Pico chip. Sizing ladders are shown in **Supplementary Figures 3.1A and 3.1E**.

A

2-D SE HPLC

B

IP-RP HPLC
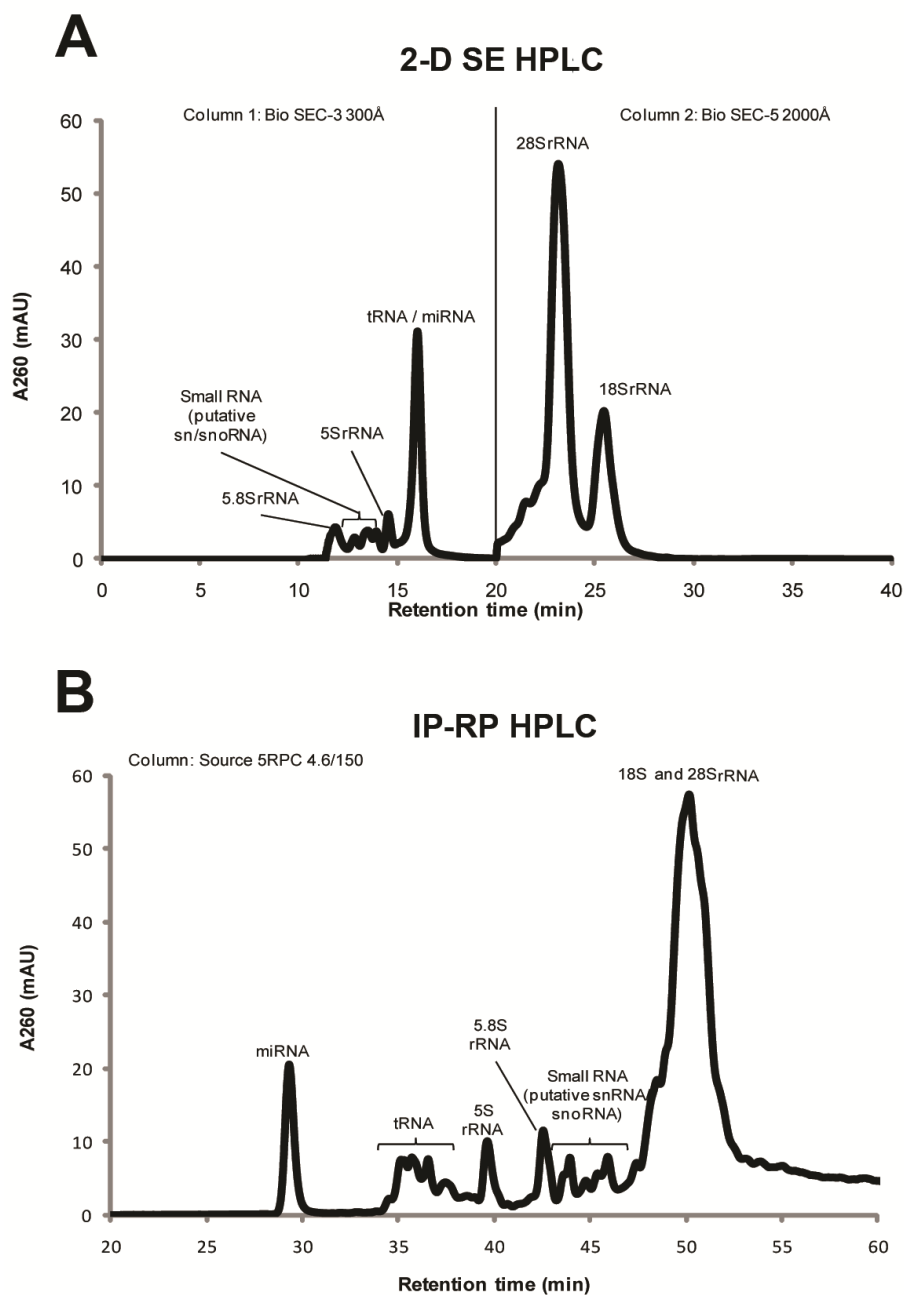
Supplementary Figure 3.7. Orthogonal separations of TK6 total RNA by 2-D SEC and IP RPC. (**A**) 2-D SEC chromatogram of TK6 total RNA consisting of co-eluting miRNA and tRNA, putative snRNA and snoRNA, 5S, 5.8S, 18S, 28S rRNAs separated obtained on a Bio SEC-3 300 Å (column 1) and Bio SEC-5 2000 Å (column 2) arranged as an in-line two-column system. Both separations were performed in 100 mM TEAA (pH 7.0) buffer at a flow rate of 1 ml/min at 60 °C. (**B**) IP RP HPLC of TK6 total RNA consisting of miRNA, tRNA, putative snRNA and snoRNA, 5S, 5.8S rRNAs, and co-eluting 18S and 28S rRNAs separated on a Source 5RPC 4.6/150 column, performed at a flow rate of 1 ml/min at 60 °C. Buffer A: 100 mM TEAA, pH 7.0 with 2% acetonitrile. Buffer B: 100% acetonitrile. Elution gradient as stated in Materials and Methods.

3.6.2. Supplementary Tables

Supplementary Table 3.1. Modified ribonucleosides detected by mass spectrometric analysis of BCG tRNA hydrolysates[1]					
Identity[2]	Short name[2]	Retention Time, min	Precursor ion, m/z	Product ion, m/z	Signal intensity[3]
Pseudouridine	Ψ	1.41	245.1	125.1	1179
5-Methylcytidine	m^5C	2.52	258.1	126.1	76
3-Methylcytidine	m^3C	2.87	258.1	126.1	786
1-Methyladenosine	m^1A	2.87	282.1	150.1	1452735
Inosine	I	4.39	269.1	137.1	328908
7-Methylguanosine	m^7G	4.53	298.1	166.1	977212
2′-O-Methylcytidine	Cm	4.98	258.1	112.1	70613
4-Demethylwyosine	*imG-14*	*5.68*	*322*	*190*	*866*
N^6-(cis-Hydroxyisopentenyl)adenosine	io^6A	5.89	352	220	10166
5-Methyluridine	m^5U	9.13	259.1	127.1	134
2-Methyladenosine	m^2A	9.49	282.1	150.1	1577
2′-O-Methylguanosine	Gm	12.5	298.1	152.1	2326
1-Methylguanosine	m^1G	12.52	298.1	166.1	93811
6-Methyladenosine	m^6A	13.16	282.1	150.1	52194
2-Methylguanosine	m^2G	13.62	298.1	166.1	772
N^6,N^6-Dimethyladenosine	m^6_2A	20.14	296.1	164.1	28173
N^6-Threonylcarbamoyladenosine	t^6A	21.94	413.1	281.1	26426
2-Methylthio-N^6-isopentenyladenosine	*ms^2i^6A*	*23.25*	*382.1*	*250*	*40357*

[1] Multiple reaction monitoring (MRM) experiments were performed with a triple quadrupole mass spectrometer, as described in the Methods section.

[2] RNA modifications noted in *italics* are tentative identifications based on collision-induced dissociation (CID) fragmentation on a quadrupole time-of-flight mass spectrometer. All other modified ribonucleosides were corroborated with synthetic standards.

[3] Mean intensity from analysis of 3 biological replicates.

Supplementary Table 3.2A: ANCOVA of linear regressions of CCRF-SB RNA species-specific RiboGreen responses.

RNA Species	Slope[1], $AU.ml.ng^{-1}$	Goodness of fit, r^2	Y-intercept[1], AU	F[2], DFn, DFd	p	Correction[3], %
E. coli total RNA	4.645 (4.51, 4.779)	0.9992	-47.67 (-102.4, 7.063)	N/A	N/A	0
23S rRNA	4.915 (4.876, 4.955)	0.9999	-0.9157 (-16.8, 14.97)	17.2139 (1, 60)	0.00011	+6
16S rRNA	4.454 (4.392, 4.515)	0.9998	5.472 (-19.44, 30.38)	8.55627 (1, 60)	0.0049	-4
tRNA	4.048 (3.934, 4.161)	0.9992	35.18 (-10.99, 81.35)	77.5287 (1, 60)	<0.0001	-13

Supplementary Table 3.2B: ANCOVA of linear regressions of the *E. coli* RNA species-specific RiboGreen responses.

RNA Species	Slope[1], $AU.ml.ng^{-1}$	Goodness of fit, r^2	Y-intercept[1], AU	F, DFn, DFd[2]	p	Correction[3], %
CCRF-SB total RNA	4.554 (4.423, 4.684)	0.9992	-13.33 (-66.41, 39.74)	N/A	N/A	0
28S rRNA	5.021 (4.948, 5.093)	0.9998	-10.03 (-39.63, 19.58)	41.312 (1, 52)	<0.0001	+10
18S rRNA	4.345 (4.293, 4.397)	0.9999	-2.469 (-23.73, 18.80)	8.159 (1, 52)	0.006	-5
tRNA	4.137 (4.012, 4.261)	0.9991	26.01 (-24.61, 76.63)	35.511 (1, 52)	<0.0001	-9

[1] 95% confidence intervals are displayed in parenthesis.

[2] Degrees of freedom for calculation of *F* are displayed in parenthesis.

[3] Correction factors were calculated based on comparisons of slopes between each purified RNA species and *E. coli* total RNA and CCRF-SB total RNA (**Figure 3.2C, 3.2D** respectively). These represent the percentage change in fluorescence signals observed with the indicated purified RNA species relative to the total RNA the RNA species was isolated from.

3.7. Acknowledgements

3.8. References

1. Lujambio, A. and S.W. Lowe, *The microcosmos of cancer.* Nature, 2012. **482**(7385): p. 347-55.
2. Bernstein, E. and C.D. Allis, *RNA meets chromatin.* Genes Dev, 2005. **19**(14): p. 1635-55.
3. Galasso, M., M. Elena Sana, and S. Volinia, *Non-coding RNAs: a key to future personalized molecular therapy?* Genome Med, 2010. **2**(2): p. 12.
4. Dominissini, D., et al., *Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq.* Nature, 2012. **485**(7397): p. 201-6.
5. Jia, G., et al., *N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO.* Nat Chem Biol, 2011. **7**(12): p. 885-7.
6. Jahn, C.E., A.O. Charkowski, and D.K. Willis, *Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation.* J Microbiol Methods, 2008. **75**(2): p. 318-24.
7. Anderson, A.C., et al., *HPLC purification of RNA for crystallography and NMR.* Rna, 1996. **2**(2): p. 110-7.
8. Perez-Novo, C.A., et al., *Impact of RNA quality on reference gene expression stability.* Biotechniques, 2005. **39**(1): p. 52, 54, 56.
9. Sharp, P.A., *The centrality of RNA.* Cell, 2009. **136**(4): p. 577-80.
10. Phizicky, E.M. and A.K. Hopper, *tRNA biology charges to the front.* Genes Dev, 2010. **24**(17): p. 1832-60.
11. Li, M., et al., *Widespread RNA and DNA sequence differences in the human transcriptome.* Science, 2011. **333**(6038): p. 53-8.
12. Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.
13. Chan, C.T.Y., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.

14. Wyatt, J.R., M. Chastain, and J.D. Puglisi, *Synthesis and purification of large amounts of RNA oligonucleotides.* Biotechniques, 1991. **11**(6): p. 764-9.

15. Lukavsky, P.J. and J.D. Puglisi, *Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides.* Rna, 2004. **10**(5): p. 889-93.

16. Seong, B.L. and U.L. RajBhandary, *Escherichia coli formylmethionine tRNA: mutations in GGGCCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop.* Proc Natl Acad Sci U S A, 1987. **84**(2): p. 334-8.

17. Easton, L.E., Y. Shibata, and P.J. Lukavsky, *Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography.* Rna, 2010. **16**(3): p. 647-53.

18. Waghmare, S.P., et al., *Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions.* J Chromatogr A, 2009. **1216**(9): p. 1377-82.

19. Likic, S., G. Rusak, and M. Krajacic, *Separation of plant viral satellite double-stranded RNA using high-performance liquid chromatography.* J Chromatogr A, 2008. **1189**(1-2): p. 451-5.

20. Kim, I., et al., *Rapid purification of RNAs using fast performance liquid chromatography (FPLC).* Rna, 2007. **13**(2): p. 289-94.

21. Parvez, H., Y. Kato, and S. Parvez, *Gel permeation and ion-exchange chromatography of proteins and peptides.* 1985, Utrecht, The Netherlands: VNU Science Press. viii, 223 p.

22. Kato, Y., et al., *Operational variables in high-performance gel filtration of DNA fragments and RNAs.* J Chromatogr, 1983. **266**: p. 341-9.

23. Batey, R.T. and J.S. Kieft, *Improved native affinity purification of RNA.* Rna, 2007. **13**(8): p. 1384-9.

24. Miyauchi, K., T. Ohara, and T. Suzuki, *Automated parallel isolation of multiple species of non-coding RNAs by the reciprocal circulating chromatography method.* Nucleic Acids Res, 2007. **35**(4): p. e24.

25. Cunningham, L., K. Kittikamron, and Y. Lu, *Preparative-scale purification of RNA using an efficient method which combines gel electrophoresis and column chromatography.* Nucleic Acids Res, 1996. **24**(18): p. 3647-8.

26. Wu, Q., et al., *Recent advances on multidimensional liquid chromatography-mass spectrometry for proteomics: from qualitative to quantitative analysis--a review.* Anal Chim Acta, 2012. **731**: p. 1-10.

27. Nagele, E., et al., *2D-LC/MS techniques for the identification of proteins in highly complex mixtures.* Expert Rev Proteomics, 2004. **1**(1): p. 37-46.

28. Zou, G., et al., *Functional analysis of two cavities in flavivirus NS5 polymerase.* J Biol Chem. **286**(16): p. 14362-72.

29. Shi, P.Y., et al., *Infectious cDNA clone of the epidemic west nile virus from New York City.* J Virol, 2002. **76**(12): p. 5847-56.

30. Janse, C.J., et al., *Plasmodium berghei: in vivo generation and selection of karyotype mutants and non-gametocyte producer mutants.* Exp Parasitol, 1992. **74**(1): p. 1-10.

31. Chan, C.T., et al., *Identification of N6,N6-dimethyladenosine in transfer RNA from Mycobacterium bovis Bacille Calmette-Guerin.* Molecules, 2011. **16**(6): p. 5168-81.

32.   Uchiyama, S., et al., *Separation of low molecular weight RNA species by high-speed gel filtration.* J Biochem, 1981. **90**(3): p. 643-8.

33.   Azarani, A. and K.H. Hecker, *RNA analysis by ion-pair reversed-phase high performance liquid chromatography.* Nucleic Acids Res, 2001. **29**(2): p. E7.

34.   Noll, B., et al., *Characterization of small interfering RNA by non-denaturing ion-pair reversed-phase liquid chromatography.* J Chromatogr A, 2011. **1218**(33): p. 5609-17.

35.   Gilar, M., et al., *Ion-pair reversed-phase high-performance liquid chromatography analysis of oligonucleotides: retention prediction.* J Chromatogr A, 2002. **958**(1-2): p. 167-82.

36.   Matera, A.G., R.M. Terns, and M.P. Terns, *Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 209-20.

37.   Wu, C.-s., *Handbook of size exclusion chromatography and related techniques.* 2nd ed. 2004, New York: Marcel Dekker. xiv, 694 p.

38.   Janse, C.J., et al., *High efficiency transfection of Plasmodium berghei facilitates novel selection procedures.* Mol Biochem Parasitol, 2006. **145**(1): p. 60-70.

39.   Dame, J.B. and T.F. McCutchan, *The four ribosomal DNA units of the malaria parasite Plasmodium berghei. Identification, restriction map, and copy number analysis.* J Biol Chem, 1983. **258**(11): p. 6984-90.

40.   Dame, J.B., M. Sullivan, and T.F. McCutchan, *Two major sequence classes of ribosomal RNA genes in Plasmodium berghei.* Nucleic Acids Res, 1984. **12**(14): p. 5943-52.

41.   Jones, L.J., et al., *RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization.* Anal Biochem, 1998. **265**(2): p. 368-74.

42.   Nilsen, T.W., *Gel purification of RNA.* Cold Spring Harb Protoc, 2013. **2013**(2).

43.   Barciszewska, M.Z., et al., *Rapid separation of tyrosine-specific tRNA from white lupin.* Acta Biochim Pol, 1992. **39**(2): p. 223-6.

44.   Brown, A.J., *Preparation of total RNA.* Methods Mol Biol, 1996. **53**: p. 269-76.

45.   Duro, G., V. Izzo, and R. Barbieri, *Methods for recovering nucleic acid fragments from agarose gels.* J Chromatogr, 1993. **618**(1-2): p. 95-104.

46.   Hammann, C. and M. Tabler, *Quantitative recovery of nucleic acids from excised gel pieces by isotachophoresis.* Biotechniques, 1999. **26**(3): p. 422-4.

47.   Krowczynska, A.M., K. Donoghue, and L. Hughes, *Recovery of DNA, RNA and protein from gels with microconcentrators.* Biotechniques, 1995. **18**(4): p. 698-703.

48.   Fremont, P., F.T. Dionne, and P.A. Rogers, *Recovery of biologically functional messenger RNA from agarose gels by passive elution.* Anal Biochem, 1986. **156**(2): p. 508-14.

49.   Gjerde, D.T., L. Hoang, and D. Hornby, *RNA purification and analysis : sample preparation, extraction, chromatography.* 2009, Weinheim: Wiley-VCH. xi, 195 p.

50.   Dickman, M.J. and D.P. Hornby, *Enrichment and analysis of RNA centered on ion pair reverse phase methodology.* Rna, 2006. **12**(4): p. 691-6.

51. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.* Anal Biochem, 1987. **162**(1): p. 156-9.
52. Chomczynski, P. and N. Sacchi, *The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on.* Nat Protoc, 2006. **1**(2): p. 581-5.
53. Rio, D.C., et al., *Guidelines for the use of RNA purification kits.* Cold Spring Harb Protoc, 2010. **2010**(7): p. pdb ip79.
54. Xue, X., et al., *Identification and characterization of novel microRNAs from Schistosoma japonicum.* PLoS One, 2008. **3**(12): p. e4034.
55. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
56. Ozsolak, F., et al., *Direct RNA sequencing.* Nature, 2009. **461**(7265): p. 814-8.

4. Quantitative analysis of modified ribonucleoside by HPLC-coupled mass spectrometry reveals N^6, N^6-dimethyladenosine as a novel tRNA modification in *Mycobacterium bovis* Bacille Calmette-Guérin [1]

4.1. Abstract

Post-transcriptional modification of RNA is an important determinant of RNA quality control, translational efficiency, RNA-protein interactions, and stress response. This is illustrated by the observation of toxicant-specific changes in the spectrum of tRNA modifications in a stress response mechanism involving selective translation of codon-biased mRNA for critical proteins. To facilitate systems-level studies of RNA modifications, we developed a liquid chromatography-coupled mass spectrometry (LC-MS) technique for the quantitative analysis of modified ribonucleosides in tRNA or other RNA species. The protocol includes tRNA purification by HPLC, enzymatic hydrolysis, reversed-phase HPLC resolution of the ribonucleosides, and identification and quantification of individual ribonucleosides by LC-MS using dynamic multiple reaction monitoring. This approach enabled the quantification of modified ribonucleosides in several micrograms of tRNA, or other RNA, in a 15-minute LC-MS run. We coupled this LC-MS platform with the RNA isolation and purification methods developed in Chapters 2 and 3. Our initial scan using this approach revealed 17 modified ribonucleoside BCG tRNA, of which 12 were definitively identified based on comparisons to synthetic standards and 5 were tentatively identified by exact mass comparisons to RNA modification databases. Among these was one not

---

[1] Featured in Nat Protoc. 2014 Apr;9(4):828-41 and Molecules. 2011 Jun 21;16(6):5168-81. Reproduced with permission from Nature Publishing Group and Multidisciplinary Digital Publishing Institute.

previously described in tRNA, which we characterized as N^6, N^6-dimethyladenosine.

4.2. Introduction

While the four canonical nucleobases in DNA are adequate to encode the genome, the functional diversity of RNA is greatly enhanced by the presence of over 100 different targeted structural modifications of the nucleobase and ribose sugar in RNA across all organisms [1, 2]. Transfer RNA (tRNA) and ribosomal RNA (rRNA) are the most frequently modified forms of RNA, though other species such as messenger RNA (mRNA) and microRNAs are also known to possess specific ribonucleoside modifications, such as the N^7-methylguanosine cap on mRNA.

Emerging evidence points to important roles for RNA modifications in translational control of cellular responses to stress and other stimuli [3-6]. For example, we recently observed that exposure of the yeast *Saccharomyces cerevisiae* to cytotoxic chemicals causes a reprogramming of the spectrum of two dozen RNA modifications in tRNA, with a unique ribonucleoside signature for each agent [7]. These observations point to the importance of fully understanding the structure and function of ribonucleoside analogs in both eukaryotic and prokaryotic organisms.

The discovery of this emergent property of tRNA modifications has been enabled in part by application of liquid chromatography-coupled mass spectrometry (LC-MS) to quantify the complete set of tRNA modifications and how its composition varies in response to different stresses. Traditional approaches to quantifying individual modified ribonucleosides, such as two-dimensional thin-layer chromatography [8, 9], are semi-quantitative at best and not suitable for studying all tRNA modifications at once, in addition to

being labor and time intensive and employing radioactive labeling. Mass spectrometry-based methods have recently emerged as powerful tools for identifying and quantifying RNA modifications.[10-19] For example, we recently developed a rigorously quantitative platform for LC-MS analysis of the system of modified ribonucleosides in the population of tRNA molecules in a cell or in individual, purified tRNA species (**Figure 4.1**), with the method applied to studies in bacteria, yeast and human cells [7, 20-23]. Application of this approach to the identification and quantification of 25 tRNA ribonucleoside modifications in *Saccharomyces cerevisiae* led to the discovery of a direct link between cell stress and translation of stress-response proteins [7, 21]. As illustrated in **Figure 4.1**, there are 5 stages to this process. (1) exposure of cells to stress (see Chapters 5 and 6), (2) tRNA isolation, purification, quantification and its enzymatic hydrolysis to ribonucleosides; (3) reversed-phase HPLC resolution of hydrolysed ribonucleosides, (4) identification and quantification of ribonucleosides by tandem quadrupole mass spectrometry (QQQ); and (5) multivariate statistical analysis of patterns of change in the set of modified ribonucleosides caused by the stress or stimulus.

Aiming to characterize the RNA modifications in the clinically important pathogens from the *Mtb* complex, we used *Mycobacterium bovis* Bacille Calmette-Guérin (BCG) as a surrogate for *Mycobacterium tuberculosis* (*Mtb*). Infection with *Mtb* represents one of the most widespread diseases in the world, with nearly one-third of the world's population showing signs of exposure, more than 20 million people actively infected, and almost 80% of the population of some developing countries testing positive in tuberculin tests [24, 25]. BCG is a mycobacterial species closely related to Mtb and is widely used as a vaccine [26]. Here, I present results from our initial

characterization of the modified ribonucleosides in BCG (the fully repertoire is presented in Chapter 6 and Appendix 1), with the discovery of N^6, N^6-dimethyladenosine (m^6_2A) as a novel tRNA modification.
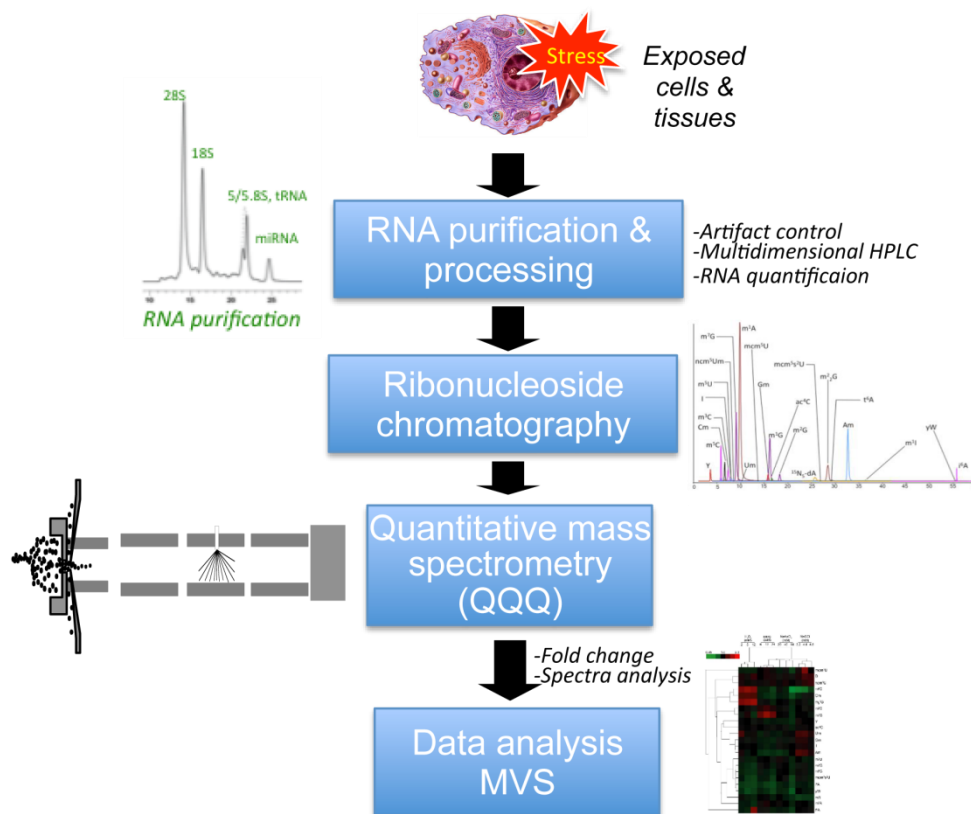
Figure 4.1. Workflow for the quantitative analysis of modified ribonucleosides in tRNA. Stages include (1) exposure of cells to stress, (2) tRNA isolation, purification, quantification and its hydrolysis to ribonucleosides; (3) reversed-phase HPLC resolution of hydrolysed ribonucleosides, (4) identification and quantification of ribonucleosides QQQ mass spectrometry; (5) multivariate statistical analysis (MVS).

4.3. Material and methods

4.3.1. Reagents and instrumentation

OADC solution, 7H9 culture media powder, and 7H11 agar powder were purchased from Biomed Diagnostics (White City, OR). Trizol reagent and PureLink miRNA Isolation Kit was purchased from Invitrogen (Carlsbad, CA). 2′-O-Methyluridine (Um), pseudouridine (Y), N^1-methyladenosine (m^1A), N^2,N^2-dimethylguanosine (m^2_2G), N^6,N^6-dimethyladensoine (m^6_2A), and 2′-O-

methylguanosine (Gm) were purchased from Berry and Associates (Dexter, MI). N^6-threonylcarbamoyladenosine (t^6A) was purchased from Biolog (Bremen, Germany). N^6-isopentenyladenosine (i^6A) was purchased from International Laboratory LLC (San Bruno, CA). 2′-O-Methyladenosine (Am), N^4-acetylcytidine (ac^4C), 5-methyluridine (m^5U), inosine (I), 2-methylguanosine (m^2G), N^7-methylguanosine (m^7G), 2′-O-methylcytidine (Cm), 3-methylcytidine (m^3C), 5-methylcytidine (m^5C), alkaline phosphatase, RNase A, ammonium acetate, geneticine, bovine serum albumin, deferoxamine mesylate, butylated hydroxytoluene, glucose, sodium chloride, nuclease P1, formic acid, and 20% Tween80 solution were purchased from Sigma Chemical Co. (St. Louis, MO). Glycerol was purchased from SinoChem Corp. (Beijing, China). Phosphodiesterase I was purchased from USB (Cleveland, OH). RNase A, RNase V1 and RNase T1 were purchased from Ambion Inc. (Austin, TX). HPLC-grade water, acetonitrile, and chloroform were purchased from Mallinckrodt Baker (Phillipsburg, NJ). *M. bovis BCG* was gifted by Sylvie Alonso (Department of Microbiology, National University of Singapore) and *S. cerevisiae* were purchased from American Type Culture Collections (Manassas, VA). Rat liver tissue (discarded) was obtained from Laura J. Trudel (Department of Biological Engineering, Massachusetts Institute of Technology). Filters with a 10 KDa MW cut-off were purchased from Pall Life Sciences (Port Washington, NY). Experiments were performed with Thermo FP120 Bead beater (Two Rivers, WI), Qiagen TissueRuptor (Valencia, CA), Agilent Bioanalyzer series 2100 (Santa Clara, CA), Agilent LC/QQQ 6460 (Santa Clara, CA), Agilent LC/TOF G6210A (Santa Clara, CA) and Agilent LC/QTOF 6520 (Santa Clara, CA).

4.3.2. Preparation of BCG culture media

The 7H9 media were prepared by mixing 4.9 g of 7H9 powder, 10 mL of 50% glycerol, 2.5 mL of 20% TWEEN 80, 900 mL of double-deionized water, and 100 mL of ADS solution. The ADS solution was prepared by mixing 50 g of BSA, 20 g of glucose, and 8.1 g of NaCl in 950 ml of double-deionized water. The 7H11 agar plates were prepared by mixing 4.2 g of 7H11 agar powder, 2 mL of 50% glycerol, 180 ml of double-deionized water, and 20 mL of OADC solution. The solution was then heated in a microwave oven until a clear solution was achieved and the solution was transferred to petri dishes. The agar plates were cooled and solidified at ambient temperature.

4.3.3. BCG cultures

BCG cells were grown in 7H9 culture media (see Supporting Information) at 37 ^{o}C in an incubator with 5 % CO2. After the culture reached an optical density of OD600 ~0.6, at which point the concentration of cells was ~3x107/mL, the cells were harvested by centrifugation at 12000xg for 10 min at 4 ^{o}C. Cell pellets were snap-frozen in liquid nitrogen and stored at -80 ^{o}C. For glycerol stocks, the post-centrifugation cell pellet was resuspended in 1 mL of 7H9 media with 25% glycerol. The solution was then further diluted to a final concentration with OD600~1 and the stocks stored at -80 ^{o}C. To determine the quantity of living cells in a glycerol stock or a culture, colony counting was performed with each cell sample with 100 µL of serial dilutions plated on 7H11 agar and incubated at 37 ^{o}C with 5% CO2.

4.3.4. tRNA isolation

tRNA was isolated from several organisms, including BCG (~10^8 cells), S. cerevisiae (5x10^7 cells), human B lymphoblastoid TK6 cells (3x10^7 cells), and rat liver (~150 mg). Cells or tissues were suspended in 1.5 mL of Trizol reagent with 5 mg/mL coformycin, 50 µg/mL tetrahydrouridine, 0.1 mM

deferoxamine mesylate, and 0.5 mM butylated hydroxytoluene to prevent nucleoside modification artifacts [11, 12]. BCG was lysed as indicated in Chapter 2. *S. cerevisiae* cells were lysed by 3 cycles of bead beating in a Thermo FP120 Bead Beater set at 6.5 m/s for each 20 s cycle, with 1 min of cooling on ice between cycles. The TK6 cells and rat liver tissue were lysed with a Qiagen TissueRuptor. Following cell or tissue disruption, all lysates were warmed to ambient temperature for 5 min and extracted with 0.3 mL volume of chloroform, with subsequent incubation at ambient temperature for 3 min. The solutions were centrifuged at 12000xg for 15 min at 4 °C and the aqueous phase was collected. Absolute ethanol was added to the aqueous phase (35% v/v) and small RNA species were then isolated using the PureLink miRNA Isolation Kit according to manufacturer's instructions. The quality and concentration of the resulting small RNA mixture was assessed by Bioanalyzer analysis (Agilent Small RNA Kit), with tRNA comprising >95% of the small RNA species present in the mixture, with no detectable 5S rRNA (Figure 1). tRNA from BCG was further purified from small RNA isolates by size-exclusion HPLC using an Agilent SEC3 300 Å, 7.8 x 300 mm column eluted with 10 mM ammonium acetate as described in Chapter 3. The tRNA fraction was collected and desalted using an Ambion Millipore 5K MWCO column.

4.3.5. Enzymatic hydrolysis of BCG tRNA

Samples of purified tRNA (6 µg) were lyophilized and redissolved in 100 µL of 10 ng/mL RNase A, 0.01 U/mL RNase T1, 0.001 U/mL RNase V1, 0.15 U/mL nuclease P1, 2.5 mM deferoxamine mesylate, 10 ng/mL coformycin, 50 µg/mL tetrahydrouridine, 0.5 mM butylated hydroxytoluene, and 1xRNA Structure Buffer from Ambion. The solution was incubated at 37 °C for 3 h, after which alkaline phosphatase was added to a final concentration of 0.1

U/mL. The sample was incubated at 37 °C overnight, followed by removal of proteins by YM10 filtration. The resulting filtrate was used directly for mass spectrometric analysis.

4.3.6. Identification of ribonucleosides in small RNA from BCG

Hydrolyzed RNA was resolved on a Thermo Hypersil aQ column (100 x 2.1 mm, 1.9 μm particle size) with an acetonitrile gradient (HPLC system A) in 0.1% (v/v) formic acid in water as mobile phase at a flow rate of 0.3 mL/min. The gradient of acetonitrile in 0.1% formic acid was as follows: 0-15.3 min, 0%; 15.3-18.7 min, 1%; 18.7-20 min, 6%; 20-24 min, 6%; 24-27.3 min, 100%; 27.3-41 min, 0%. The HPLC column was directly connected to a triple quadrupole mass spectrometer (Agilent LC/QQQ 6460) in positive ion, neutral loss mode for loss of m/z 132 and 146 in the range of m/z 200-700. The voltages and source gas parameters were as follows: gas temperature, 300 ^{o}C; gas flow, 6 L/min; nebulizer, 15 psi; and capillary voltage, 4000 V. The ions detected in neutral loss mode were selected for identification with the LC-MS/MS system in multiple reaction monitoring (MRM) mode using the same HPLC and mass spectrometer parameters. Since the C-C glycosidic bond in pseudouridine (Ψ) does not readily fragment to produce neutral loss of a ribose residue, the presence of Ψ was verified by CID fragmentation of the ribose to yield the nucleobase with the ribose C1 methylene group attached (m/z 125) [27].

4.3.7. High mass-accuracy mass spectrometric analysis of candidate ribonucleosides

The exact molecular weights of candidate ribonucleosides were determined by HPLC-coupled high mass-accuracy quadrupole time-of-flight mass spectrometry (Agilent LC/QTOF 6520). Hydrolyzed RNA samples were

resolved on a Thermo Hypersil aQ column (150 x 2.1 mm, 3 μm particle size) eluted with a mobile phase as noted earlier with the following solvent schedule for acetonitrile in 0.1% formic acid: 0-23 min, 0%; 23-28 min, 2%; 28 -36 min, 7%; 36-47 min, 100%; 47-67 min, 0% (HPLC system, B). The HPLC system was coupled to an Agilent QTOF 6520 Mass Spectrometer operated in positive ion mode and scanning for ions from m/z 100-1700 with the following parameters: gas temperature, 325 ^{o}C; drying gas, 5 L/min; nebulizer, 30 psi; and capillary voltage, 3500 V.

4.3.8. Structural characterization of N^6,N^6-dimethyladenosine in BCG small RNA

The ribonucleoside-like species eluting at 20.1 min and possessing an [M+H]$^+$ ion with m/z of 296.1337 was subjected to structural characterization by collision-induced dissociation (CID) using both MS^2 and pseudo-MS^3 (*i.e.*, in-source fragmentation) performed on the LC-QTOF system using a Thermo Hypersil aQ column (100 x 1 mm, 3 μm particle size) at a flow rate of 90 µL/min using the same mobile phase described earlier, with a solvent schedule for the 0.1% formic acid in acetonitrile was as follows: 0-18 min, 0%; 18-30 min, 7%; 30-44 min, 100%; and 44-60 min, 0% (HPLC system C). The mass spectrometer was operated in positive ion mode with the following voltages and source gas parameters: gas temperature, 325 ^{o}C; drying gas, 8 L/min; nebulizer, 30 psi; capillary voltage, 3500 V. The m/z detection range for parent ions was 100 to 800 and that for product ions was 50 to 800. For MS^2 analysis, the fragmentor voltage was 85 V, the collision energy was 10 V and the target ion for the unknown was m/z 296.1, while the fragmentor voltage was increased to 250 V for MS^3 analysis, which caused an in-source fragmentation of m/z 296.1 to give m/z 164.1 for further CID analysis. The

m/z 164.1 ion was fragmented with collision energies of 0 V, 20 V, 30 V, and 60V.

4.3.9. Quantification of tRNA modifications

Perhaps the most critical stage in this LC-MS platform for quantitative comparison of changes in modified ribonucleosides across samples is the quantification of RNA in each sample. Standard spectroscopic methods (*i.e.,* A_{260}) are entirely inadequate for this purpose, as they are heavily affected by contamination with even small amounts of protein, DNA, buffers and other chemicals. However, even highly precise and accurate RNA quantification methods, such as fluorescent dye binding,[28] will not provide accurate information about the RNA content of the sample if the method is applied too early in the sample processing and needs proper calibration standards (See Chapter 3). Notably, loss of RNA in subsequent HPLC purification and enzymatic digestion steps will cause artifacts in RNA quantification and wide variations in data from replicate. The optimal solution to these issues is to quantify the canonical ribonucleosides (*i.e.,* cytidine, uridine, adenosine, and guanosine) in the fully processed sample at the time of LC-MS analysis, was achieved using an in-line UV absorbance detector, with an external calibration curve. To avoid problems caused by day-to-day variation in instrument performance, the precision of relative quantification of modifications across different analysis sessions may be enhanced by spiking into the fully processed analyte an internal standard, such as $[^{15}N]_5$-2-deoxyadenosine ($[^{15}N]$-dA) to a final concentration of 40 nM immediately prior to MS analysis. 4 pmol of $[^{15}N]$-dA was added to 4 µg of tRNA and the samples were subjected to enzymatic hydrolysis as described earlier. Following volume adjustment to achieve final concentrations of ~40 nM $[^{15}N]$-dA and ~40 ng/µl ribonucleosides, 10 µl of sample was analyzed by LC-MS/MS.

Ribonucleosides were resolved on a Thermo Hypersil aQ column (100 x 2.1 mm, 1.9 μm particle size) with 0.1% (v/v) formic acid in water and in acetonitrile as mobile phase and a flow rate of 0.3 mL/min. The solvent schedule for acetonitrile in 0.1% formic acid was as follows: 0-10 min, 5%; 10-12 min, 30%; 12 min, 95%. Electrospray ionization MS/MS analysis of the HPLC eluant was performed in positive ion mode with the following parameters for voltages and source gas: gas temperature, 350°C; gas flow, 10 L/min; nebulizer, 20 psi; and capillary voltage, 3500 V. The mass spectrometer was operated in MRM mode to quantify two ribonucleosides with the following parameters (retention time, m/z of the transmitted parent ion, m/z of the monitored product ion, fragmentor voltage, collision energy): [^{15}N]-dA -- 3.0 min, m/z 257→141, 90 V, 10 V; and m$^6_2$A -- 11.1 min, m/z 296→164, 90 V, 15 V. The dwell time for each ribonucleoside was 200 ms and these two ions were monitored throughout the whole HPLC run.

Absolute quantification m$^6_2$A was achieved by preparing external calibration curves using synthetic standard. Linear calibration curves were obtained using a fixed concentration (40 nM) of [^{15}N]-dA and varying concentrations of m$^6_2$A (5, 10, 50, 100, 500 nM) (**Supplementary Figure 4.2**).

4.4. Results

4.4.1. Identification of ribonucleosides in BCG tRNA

As a critical feature of our platform for studying the complete set of ribonucleosides in an organism [7], liquid chromatography-coupled mass spectrometry has previously been demonstrated to be a powerful tool for characterizing the structure of ribonucleosides and for quantifying them in biological systems [10, 11]. Ribonucleosides from hydrolyzed tRNA were initially characterized using LC-MS/MS with neutral loss analysis. During

CID, there is characteristic cleavage of the glycosidic bond between the nucleobase and either the ribose or 2'-O-methylribose moiety, which causes a loss of either 132 or 146 Da, respectively. We used this property to search for modified nucleosides, with parallel analysis of a sample of prepared without added RNA to account for artifacts. It should be noted that, since the C-C glycosidic bond in pseudouridine (Ψ) does not readily fragment to produce neutral loss of a ribose residue, the presence of Y was verified by CID fragmentation of the ribose to yield the nucleobase with the ribose C1 methylene group attached (m/z 125) [27].

This analysis revealed several ribonucleoside candidates (**Figure 4.2**). Of these, 12 were definitely identified by comparison of retention time, exact mass and CID fragmentation to synthetic standards, while another 5 were tentatively identified on the basis of exact mass comparisons to RNA modification databases [1, 2], with no specific structures that can be assigned to the isobaric methylated ribonucleosides (**Table 4.1**). In Chapter 6, this basic analysis was further enhanced with the addition of dynamic MRM scheduling, neutral loss scans and molecular feature extraction from high mass resolution scans on the quadrupole time-of-flight mass spectrometer to reveal the final repertoire of 40 distinct modified ribonucleosides (**6.3 Material and methods**, **Figure 6.1A** and **Appendix 1**).

4.4.2. Structural characterization of the ribonucleoside with m/z 296.1350

One of the ribonucleosides identified in neutral loss analysis had an *m/z* value of 296.1350 ± 0.0011 (mean ± SD; **Supplementary Figures 4.1A and B**), which yields a chemical formula of $C_{12}H_{18}N_5O_4^+$ (*m/z* 296.1359). This formula
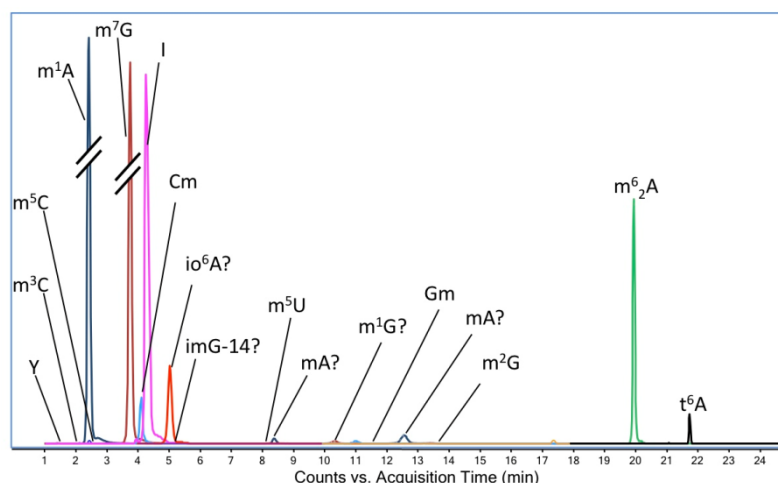
Figure 4.2. Extracted ion chromatogram of ribonucleoside candidates in hydrolyzed BCG tRNA identified by LC-MS/MS in MRM mode. The peaks of m^{1}A (2.5 min) and m^{7}G (3.9 min) are marked with '//' to indicate that they are in different scales from other peaks. The identities of ribonucleosides marked with '?' are tentative and have not been confirmed against standards. 'mA' denotes a monomethylated adenosine.

corresponded to an adenosine with 2 methyl groups or 1 ethyl group. Subsequent MS^2 analysis confirmed the neutral loss of ribose to yield a fragment with *m/z* 164.0924 (**Figure 4.3**), which corresponds to an adenine nucleobase with an additional C_2H_4 (*m/z* 164.0936). Pseudo-MS^3 analysis consisted of in-source fragmentation-induced loss of ribose to yield an ion with *m/z* 164.09, which was selected for CID in the second quadrupole. The resulting mass spectrum shown in **Figure 4.4A** revealed a variety of fragment ions, many of which had *m/z* values associated with fragmentation of methylated adenosine [29]. Based upon this dissociation model and the fragmentation pattern, we concluded that the structure of the ion with *m/z* 296.1350 was *N^6,N^6*-dimethyladenosine (m$^6_2$A). This was confirmed by comparison to synthetic m$^6_2$A, which produced identical values for retention time, exact mass, and MS^2 and pseudo-MS^3 fragmentation (**Figure 4.4B**).

Table 4.1. Ribonucleosides identified by mass spectrometric analysis of BCG tRNA hydrolysates[1]

Retention time, min[1]	Precursor ion, *m/z*	Product ion, *m/z*	Signal Intensity	Identity[2]
1.43	245.1	125.1	110	**Ψ**
2.15	258.1	126.1	70	**m^5C**
2.45	258.1	126.1	279	**m^3C**
2.47	282.1	150.1	500000	**m^1A**
3.85	298.1	166.1	500000	**m^7G**
4.22	258.1	112.1	2500	**Cm**
4.4	269.1	137.1	80000	**I**
5.16	352	220	16000	io^6A?
5.27	322	190	24000	imG-14?
8.22	259.1	127.1	30	**m^5U**
8.6	282.1	150.1	7000	mA, Am?
10.69	298.1	166.1	3500	m^1G?
11.4	298.1	152.1	450	**Gm**
12.9	282.1	150.1	17000	mA, Am?
13.91	298.1	166.1	600	**m^2G**
20.1	296.1	164.1	25000	**m^6_2A**
21.88	413.1	281.1	2500	**t^6A**

[1] Neutral loss analysis (except for Y) was performed with a triple quadrupole mass spectrometer, as described in the Experimental section.
[2] RNA modifications noted in bold font were corroborated with synthetic standards. "?" denotes tentative identification; "mA, Am" denotes a monomethylated adenosine.
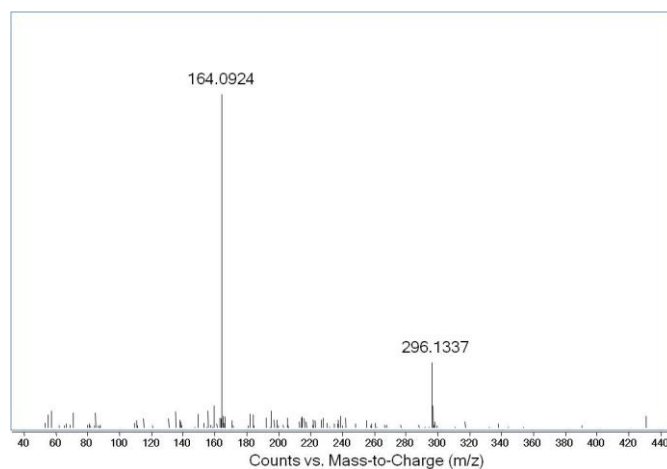
Figure 4.3. MS[2] fragmentation of the ribonucleoside with *m/z* 296.1350.
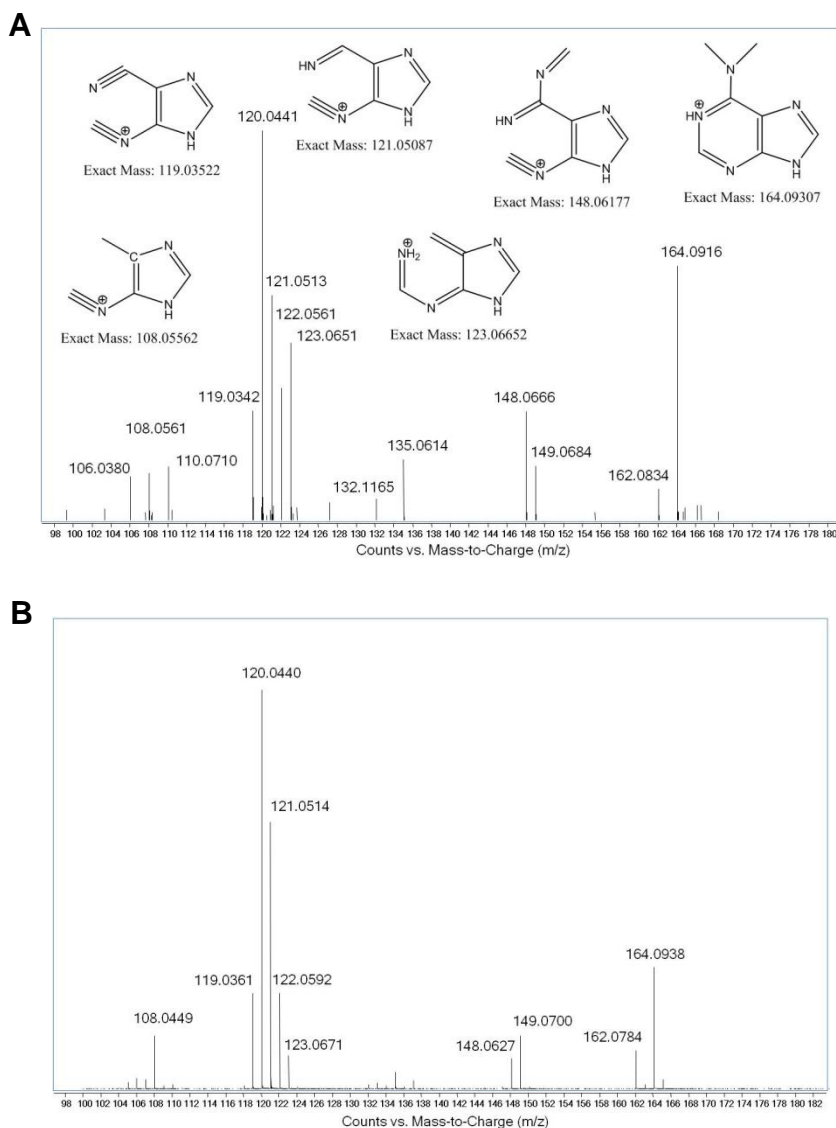
Figure 4.4. Pseudo-MS³ fragmentation analysis of candidate ribonucleosides. (**A**) Fragmentation of *m/z* 164.0924 derived from the *m/z* 296.1350 ribonucleoside. Tentative structures of fragment ions are based on a model proposed by Nelson and McCloskey [29]. (**B**) Pseudo-MS³ fragmentation of *m/z* 164.0924 derived from the *m/z* 296.1350 ion of synthetic m$^6_2$A.

4.4.3. Quantification of m$^6_2$A in tRNA from different organisms

Since m$^6_2$A had been described in rRNA previously but not in tRNA, we quantified this modification by LC-MS/MS in samples of tRNA from BCG, rat liver tissue, TK6 cell line, and yeast by using external calibration against [^{15}N]-dA (**Supplementary Figures 4.2 and 4.3**). In BCG, the level of m$^6_2$A was

0.88 pmol per µg of tRNA, while the level of m^6_2A in rat liver, and TK6 and yeast cells was below the detection limit of the assay. On average, the size of mycobacterial tRNA is ~65 nt, which suggests that m^6_2A occurs once in every ~51 tRNA molecules. The data in **Table 4.2** also suggest that the presence of m^6_2A in the small RNA isolates is not due to contamination with rRNA fragments, since we would have expected to detect m^6_2A in the small RNA isolates from the other species if rRNA fragment contamination had occurred. To further confirm this, tRNA was purified from samples of BCG small RNA by size-exclusion HPLC (**Supplementary Figure 4.4**) and m^6_2A was again detected in the fraction containing only tRNA.

Table 4.2. Level of m^6_2A in tRNA from BCG, human, rat and yeast.

	BCG	Human TK6	Rat liver	*S. cerevisiae*
MS signal[1]	5.8 ± 0.9	<0.006[2]	<0.006[2]	<0.006[2]
pmol/µg tRNA	0.88 ± 0.14	< 0.0009[2]	< 0.0009[2]	< 0.0009[2]

[1] MS signal normalized to total tRNA (0.4 µg)

[2] Less that the limit of quantification

4.5. Discussion

Using an LC-MS platform developed for analysis of ribonucleoside modifications in yeast [7], we have begun to systematically characterize the spectrum of modified ribonucleosides in a variety of microorganisms, starting with *Mycobacterium bovis* BCG. The platform that we have developed addresses all levels of highly quantitative analysis, including isolation of high quality RNA (Chapter 3: **Figure 3.2**), rigorous quantification of the RNA, and RNA processing under conditions that obviate modification artifacts caused by adventitious oxidation (addition of deferoxamine and butylated hydroxytoluene) and deamination (addition of coformycin and

tetrahydrouridine). Subsequent resolution of the complete set of ribonucleosides by reversed-phase HPLC provides good separation (Figure 2) for mass spectrometric characterization for both high mass accuracy and fragmentation. Using this approach, we were able definitively identify 12 modified ribonucleosides in BCG tRNA: Ψ, m^5C, m^3C, m^1A, m^7G, Cm, I, m^5U, Gm, m^2G, t^6A, and m^6_2A. All of these species have been described previously in either tRNA or rRNA from other organisms [30, 31], which suggests conservation of function in mycobacteria. Of these ribonucleosides, only 1-methyladenosine (m^1A) has been previously identified in a mycobacterial species (*Mtb)*, where is occurs at position 58 of tRNA [31, 32]. While m^1A and m^7G are highly abundant in both BCG and yeast tRNA, the relative proportions of other ribonucleosides common to both organisms differ significantly, as judged from mass spectrometric signal strength for identical analytical conditions (**Figure 4.2**).

In addition to the 12 defined ribonucleosides, we tentatively identified 5 other ribonucleosides on the basis of CID molecular transitions and exact mass comparisons with RNA modification databases: N^6-(cis-hydroxyisopentenyl)adenosine (io^6A; *m/z* 352.1616; *m/z* 352→220), 4-demethylwyosine (imG-14; *m/z* 322.1146; *m/z* 322→190), two adenosine species in which the nucleobase is methylated (*m/z* 282.1197; *m/z* 282→150) and a species in which the guanosine nucleobase is methylated (*m/z* 298.1146; *m/z* 298→166). The latter is likely to be m^1G since no other mono-methylated guanosines other than Gm, m^2G and m^7G have been described [30, 31]. Possible candidates for mono-methylated adenosine include Am, m^2A m^8A, m^6A; m^7A has not been described previously [30, 31]. These modifications have been observed in tRNA in other organisms [30, 31].

Our initial analysis of ribonucleosides in BCG tRNA revealed an abundant ribonucleoside not previously identified in tRNA (**Figure 4.2**): an adenosine species with either an ethyl group or two methyl groups attached to the nucleobase. Among the possible candidates for this species were 1,2-O-dimethyladenosine (m^1Am), N^6,2-O-dimethyladenosine (m^6Am) and N^6,N^6-dimethyladenosine (m^6_2A) [30, 31]. As shown in **Figures 4.3 and 4.4**, MS^2 and pseudo-MS^3 fragmentation suggested that the structure was m^6_2A, a conclusion corroborated by comparison to a synthetic m^6_2A standard (**Figure 4.4B**). The nucleoside m^6_2A was discovered by Littlefield and Dunn in *Bacterium coli*, *Aerobacter aerogenes*, yeast, and rat liver tissues [33]. It was identified in the 16S and 23S rRNA of archaea and bacteria and the 18S rRNA of eukarya [34-38]. However, m^6_2A has not been identified in tRNA from any organism or cell types other than BCG, including yeast, rat liver and human cells (**Table 4.2**). That the m^6_2A did not arise from contamination of small RNA isolates with 5S rRNA or fragments of larger rRNA species is supported by the results in Table 2, the Bioanalyzer results in Figures 1 and S4, and the analysis of HPLC-purified tRNA from BCG.

While m^6_2A has been observed previously only in rRNA, several features of its structure and biosynthesis may provide insights into its presence in tRNA. An examination of ribonucleoside databases reveals that adenosine species with N^6 modifications (*e.g.,* t^6A, i^6A) tend to be located at position 37 of tRNA, which suggests a possible location for the similarly hydrophobic m^6_2A in tRNA in BCG and other mycobacterial species. Orthologs of the methyltransferase KsgA catalyze the formation of m^6_2A in the 3'-ends of the small subunit rRNAs in most organisms [39], while members of the Erm family of methyltransferases catalyze m^6_2A formation in 23S rRNA in many bacteria [40, 41]. Homologs of both Erm and KsgA are present in BCG and *Mtb* [42,

43]. Given the precedent for m^5U formation in both tRNA and 16S rRNA by the tRNA (m^5U54) methyltransferase [44, 45], it is possible that the KsgA or Erm homologs also catalyze m^6_2A formation in BCG tRNA.

In conclusion, we have employed a chromatography-coupled mass spectrometric approach to systematically define the spectrum of modified ribonucleosides in *Mycobacterium bovis* BCG. This approach revealed a variety of ribonucleoside candidates in tRNA from BCG, of which 12 were definitively identified based on comparisons to synthetic standards and 5 were tentatively identified by exact mass comparisons to RNA modification databases. Among the ribonucleosides observed in BCG tRNA was one not previously described in tRNA, which we have now characterized as N^6,N^6-dimethyladenosine.

4.6. Supplementary material

4.6.1. Supplementary figures

(next page)
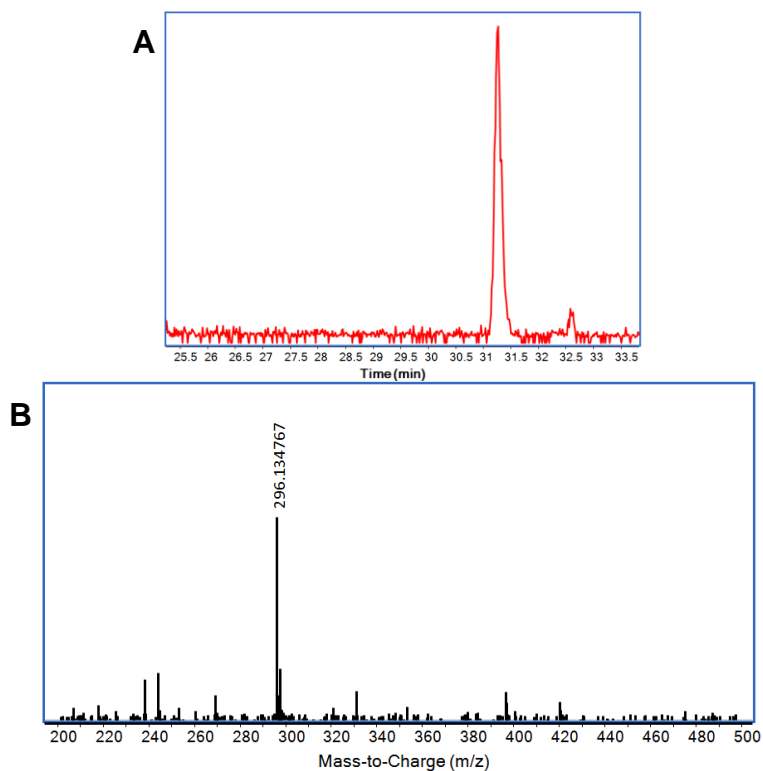
Supplementary Figure 4.1. Preliminary structural characterization of novel ribonucleoside candidate _m/z_ 296.13. (**A**) Extracted ion chromatogram of ions with _m/z_ 296.13 from the LC/TOF scan of the hydrolyzed tRNA. (**B**) A background-subtracted mass spectrum at time = 31.3 min.
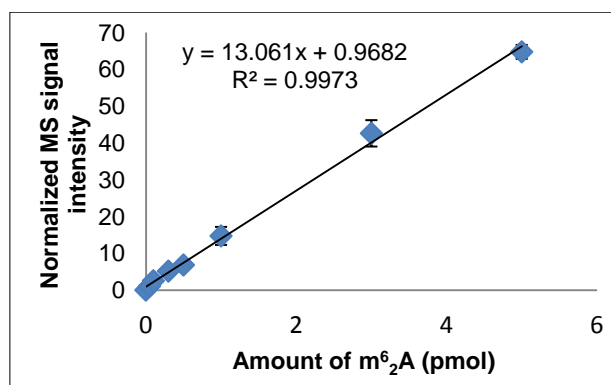
Supplementary Figure 4.2. External calibration curve for quantifying m$^6_2$A; performed as described in 4.3.Materials and methods.
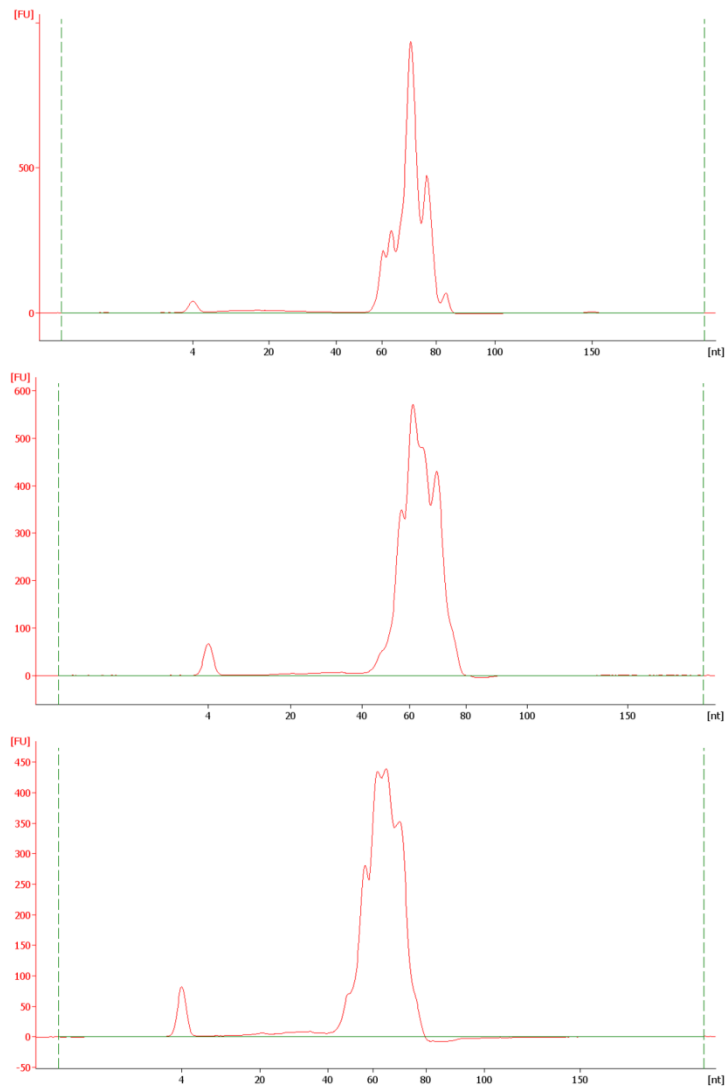
Supplementary Figure 4.3. Analysis of small RNA isolated from the yeast *S. cerevisiae*, rat liver, and human B lymphoblastoid TK6 cells (respectively, from top to bottom). Samples were analyzed with an Agilent Bioanalyzer small RNA chips as described in Chapter 3
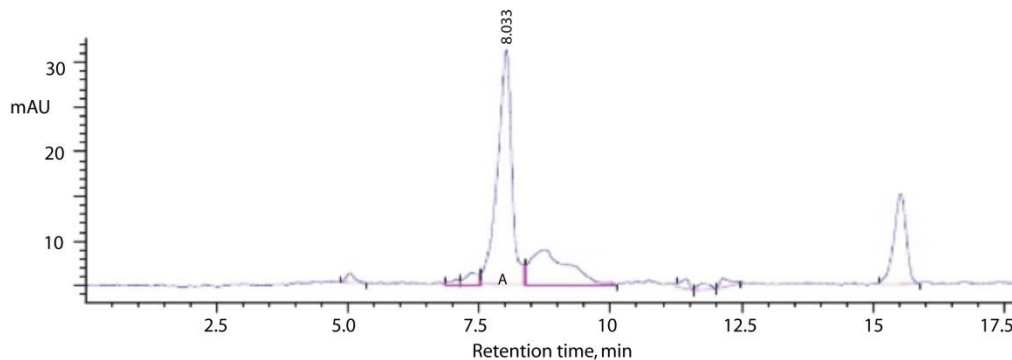
Supplementary Figure 4.4: **Purification of BCG tRNA from small RNA isolates by size-exclusion HPLC**. Fraction A (at 8.033 min) was collected and the tRNA analyzed for m^6_2A content as described in the text.

4.7. Acknowledgements

4.8. References

1. Czerwoniec, A., et al., *MODOMICS: a database of RNA modification pathways. 2008 update.* Nucleic Acids Res, 2009. **37**(Database issue): p. D118-21.
2. McCloskey, J.A. and J. Rozenski, *The Small Subunit rRNA Modification Database.* Nucleic Acids Res, 2005. **33**(Database issue): p. D135-8.
3. Bennett, C.B., et al., *Genes required for ionizing radiation resistance in yeast.* Nat Genet, 2001. **29**(4): p. 426-34.
4. Kalhor, H.R. and S. Clarke, *Novel methyltransferase for modified uridine residues at the wobble position of tRNA.* Mol Cell Biol, 2003. **23**(24): p. 9283-92.
5. Alexandrov, A., E.J. Grayhack, and E.M. Phizicky, *tRNA m7G methyltransferase Trm8p/Trm82p: evidence linking activity to a growth phenotype and implicating Trm82p in maintaining levels of active Trm8p.* Rna, 2005. **11**(5): p. 821-30.
6. Begley, U., et al., *Trm9-catalyzed tRNA modifications link translation to the DNA damage response.* Mol Cell, 2007. **28**(5): p. 860-70.
7. Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.
8. Grosjean, H., et al., *Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates.* Methods in Enzymology, 2007. **425**: p. 55-101.
9. Köhrer, C. and U.L. Rajbhandary, *The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases.* Methods, 2008. **44**(2): p. 129-38.
10. Suzuki, T., et al., *Mass spectrometric identification and characterization of RNA-modifying enzymes.* Methods Enzymol, 2007. **425**: p. 211-29.
11. Meng, Z. and P.A. Limbach, *Mass spectrometry of RNA: linking the genome to the proteome.* Brief Funct Genomic Proteomic, 2006. **5**(1): p. 87-95.

12.   Pomerantz, S.C. and J.A. McCloskey, *Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry.* Methods Enzymol, 1990. **193**: p. 796-824.

13.   Limbach, P.A., P.F. Crain, and J.A. McCloskey, *Summary: the modified nucleosides of RNA.* Nucleic Acids Res, 1994. **22**(12): p. 2183-96.

14.   Rozenski, J., P.F. Crain, and J.A. McCloskey, *The RNA Modification Database: 1999 update.* Nucleic Acids Res, 1999. **27**(1): p. 196-197.

15.   Mandal, D., et al., *Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2872-7.

16.   Ikeuchi, Y., et al., *Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea.* Nature chemical biology, 2010. **6**(4): p. 277-82.

17.   Suzuki, T. and K. Miyauchi, *Discovery and characterization of tRNAIle lysidine synthetase (TilS).* FEBS Lett, 2010. **584**(2): p. 272-7.

18.   Miyauchi, K., S. Kimura, and T. Suzuki, *A cyclic form of N(6)-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification.* Nature chemical biology, 2012.

19.   Crain, P.F., *Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry.* Methods Enzymol, 1990. **193**: p. 782-90.

20.   Begley, U., et al., *A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1-alpha.* EMBO Mol Med, 2013. **5**(3): p. 366-83.

21.   Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.

22.   Patil, A., et al., *Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications.* RNA Biol, 2012. **9**(7): p. 990-1001.

23.   Patil, A., et al., *Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response.* Cell Cycle, 2012. **11**(19): p. 3656-3665.

24.   Jasmer, R.M., P. Nahid, and P.C. Hopewell, *Clinical practice. Latent tuberculosis infection.* N Engl J Med, 2002. **347**(23): p. 1860-6.

25.   World Health Organization, G.T.P., *Global tuberculosis control : WHO report.* 2009, Global Tuberculosis Programme: Geneva.

26.   Fine, P.E. and L.C. Rodrigues, *Modern vaccines. Mycobacterial diseases.* Lancet, 1990. **335**(8696): p. 1016-20.

27.   Dudley, E., et al., *Study of the mass spectrometric fragmentation of pseudouridine: comparison of fragmentation data obtained by matrix-assisted laser desorption/ionisation post-source decay, electrospray ion trap multistage mass spectrometry, and by a method utilising electrospray quadrupole time-of-flight tandem mass spectrometry and in-source fragmentation.* Rapid Commun Mass Spectrom, 2005. **19**(21): p. 3075-85.

28.   Chionh, Y.H., et al., *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Res, 2013. **in press**.

29.   Nelson, C.C.M., J. A., *Collision-Induced Dissociation of Adenine.* J Am Chem Soc, 1992. **114**(1): p. 3661 to 3668.

30.   Brahmachari, V. and T. Ramakrishnan, *Studies on 1-methyl adenine transfer RNA methyltransferase of Mycobacterium smegmatis.* Arch Microbiol, 1984. **140**(1): p. 91-5.

31. Varshney, U., et al., *Mycobacterium tuberculosis Rv2118c codes for a single-component homotetrameric m1A58 tRNA methyltransferase.* Nucleic Acids Res, 2004. **32**(3): p. 1018-27.

32. Bujnicki, J.M., *In silico analysis of the tRNA:m1A58 methyltransferase family: homology-based fold prediction and identification of new members from Eubacteria and Archaea.* FEBS Lett, 2001. **507**(2): p. 123-7.

33. Littlefield, J.W. and D.B. Dunn, *Natural occurrence of thymine and three methylated adenine bases in several ribonucleic acids.* Nature, 1958. **181**(4604): p. 254-5.

34. Brown, G.M. and G. Attardi, *Methylation of nucleic acids in HeLa cells.* Biochem Biophys Res Commun, 1965. **20**(3): p. 298-302.

35. Dubin, D.T. and A. Günalp, *Minor nucleotide composition of ribosomal precursor, and ribosomal, ribonucleic acid in Escherichia coli.* Biochim Biophys Acta, 1967. **134**(1): p. 106-123.

36. Fox, G.E., et al., *Classification of methanogenic bacteria by 16S ribosomal RNA characterization.* Proc Natl Acad Sci U S A, 1977. **74**(10): p. 4537-41.

37. Lai, C.J., J.E. Dahlberg, and B. Weisblum, *Structure of an inducibly methylatable nucleotide sequence in 23S ribosomal ribonucleic acid from erythromycin-resistant Staphylococcus aureus.* Biochemistry, 1973. **12**(3): p. 457-60.

38. Starr, J.L. and R. Fefferman, *The Occurrence of Methylated Bases in Ribosomal Ribonucleic Acid of Escherichia Coli K12 W-6.* J Biol Chem, 1964. **239**: p. 3457-61.

39. Inoue, K., S. Basu, and M. Inouye, *Dissection of 16S rRNA methyltransferase (KsgA) function in Escherichia coli.* J Bacteriol, 2007. **189**(23): p. 8510-8.

40. Weisblum, B., et al., *Erythromycin-inducible resistance in Staphylococcus aureus: requirements for induction.* J Bacteriol, 1971. **106**(3): p. 835-47.

41. Thakker-Varia, S., A.C. Ranzini, and D.T. Dubin, *Ribosomal RNA methylation in Staphylococcus aureus and Escherichia coli: effect of the "MLS" (erythromycin resistance) methylase.* Plasmid, 1985. **14**(2): p. 152-61.

42. Buriankova, K., et al., *Molecular basis of intrinsic macrolide resistance in the Mycobacterium tuberculosis complex.* Antimicrob Agents Chemother, 2004. **48**(1): p. 143-50.

43. Brosch, R., et al., *Genome plasticity of BCG and impact on vaccine efficacy.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5596-601.

44. Gu, X., J. Ofengand, and D.V. Santi, *In vitro methylation of Escherichia coli 16S rRNA by tRNA (m5U54)-methyltransferase.* Biochemistry, 1994. **33**(8): p. 2255-61.

45. Auxilien, S., et al., *Specificity shifts in the rRNA and tRNA nucleotide targets of archaeal and bacterial m5U methyltransferases.* Rna, 2011. **17**(1): p. 45-53.

5. Ketogenesis: An Achilles' heel of persistent mycobacteria

5.1. Abstract

Many bacteria, including mycobacteria, respond to environmental stresses by entering a non-replicating antibiotic-tolerant state of persistence, yet the mechanistic underpinnings remain poorly understood. We used a data-driven approach to characterize the response of three mycobacteria, *M. tuberculosis, M. bovis* BCG and *M. smegmatis*, to starvation and subsequent resuscitation in terms of growth behavior, antibiotic susceptibility, lipid metabolism, nutrient utilization, reactive oxygen species (ROS) generation and global gene expression. Applying multivariate analysis to integrate the datasets, we found strong correlations among multi-drug tolerance, ROS production and a metabolic shift to triacylglycerol utilization and consequent ketogenesis, with strong Mg^{2+}-dependence for persistence for all three bacteria. This novel ketone body metabolic pathway is mediated by cytochrome P450s that paradoxically generate significant ROS yet confer hypersensitivity to hydrogen peroxide by P450 inactivation. These results are especially pertinent for developing therapeutics against latent infections with high risk of post-treatment relapse.

5.2. Introduction

Having established an analytical platform for the isolation, purification, characterization and quantification of cellular RNA (Chapters 2-4), we now turn our attention to defining the stress response of mycobacteria; starting with nutrient statvation.

One of the most effective adaptations of bacterial pathogens to the stress of infection involves the ability to enter a persistent state characterized by a lack of cell growth and resistance to antibiotics [1]. The antibiotic tolerance of persisters is a non-heritable, reversible resistance to antibiotic-mediated killing that correlates with physiological adaptation to environmental stresses [2, 3] and often necessitates complicated and prolonged therapeutic regimens [4, 5]. Emerging evidence suggests that bacterial persistence results from active responses to physiological stress, rather than passive quiescence [6, 7]. However, the challenge to effectively treating persistent bacteria lies in our lack of understanding of the molecular mechanisms underlying persistence and the multiplicity of overlapping and redundant pathways by which bacteria adapt to different environmental stresses. Here, we focus on the molecular mechanisms underlying mycobacterial persistence caused by nutrient deprivation.

As a pathogen thought to asymptomatically infect one-third of the world's population [8], *Mycobacterium tuberculosis* (*Mtb*) characteristically enters a state of persistence when challenged by a variety of stresses imposed by the human immune response, including nitric oxide exposure, hypoxia, and nutrient deprivation [9]. *Mtb* likely shares a number of features of stress response with other bacteria, including programmed shifts in metabolism and altered sensitivity to reactive oxygen (ROS) and nitrogen species (RNS) [9, 10]. For example, the stringent response is an extensively studied bacterial reaction to carbon, amino acid or fatty acid deprivation. Central to this response is alarmone signaling mediated by guanosine pentaphosphate or tetraphosphate ((p)ppGpp), This signaling pathway, regulated by Rsh (RelA/SpoT homologs), coordinates starvation stress responses and tolerance to antibiotic killing by affecting the expression and activities of RNA

polymerases, sigma factors, toxin-antitoxin systems, chaperones and cAMP signaling [11-13]. Starvation of *Mtb* has been shown to cause global transcriptional reprogramming [14], modified metabolism and virulence through the induction of toxin-antitoxin modules [15], altered permeability to drugs and metabolites [16, 17] and phenotypic resistance to a wide range of first- and second-line antibiotics with varied mechanisms of action [18, 19]. Of particular relevance to our current studies is the observation that oxidative stress plays a definitive role in the killing of mycobacteria by antibiotics [20]. Yet, despite new viewpoints on the precise mechanism of antibiotic killing and its association to ROS [21, 22], the series of physiological, metabolic and gene regulation events required for mycobacterial persistence and antibiotic tolerance remains unclear.

In the studies detailed in this chapter, we combined comprehensive measurements of ROS levels with biochemical assays, multiplexed phenotypic metabolic screens, and global analysis of gene expression before, during, and in recovery from prolonged nutrient deprivation to deduce that starved mycobacteria switch to a cytochrome P450 (CYP)-dependent ketogenic metabolic state. *Cyp135A1*, in particular, proved essential for surviving nutrient deprivation. Further, we show that these starvation-induced persister populations are sensitive to killing by hydrogen peroxide due to disruption of CYP-dependent metabolism and, paradoxically, an associated reduction in ROS production. Metadata analysis of these results suggests that the antibiotic tolerance of mycobacterial persistence is a multifactorial trait that integrates efflux pump expression, drug target down-regulation, and metabolism of xenobiotics. Mycobacterial ketogenesis thus represents both an adaptive strength and a potentially exploitable weakness for overcoming persistence.

5.3. Material and methods

5.3.1. Bacteria strains and culture conditions

M. tuberculosis CDC 1551*, M. bovis* BCG str. Pasteur 1173P2 and *M. smegmatis* mc^{2}155 were grown in roller bottles with 7H9 broth or PBS (with 0.05% (v/v) tyloxapol) at 5 rpm and 37 °C. Serial dilution plating was used for colony-forming units (CFU) determination on 7H11 agar for *Mtb* and 7H10 agar for BCG and SMG. Exponentially growing cultures (OD$_{600}$ of 0.4-0.8) were starved by washing pellets twice with and resuspended in PBS-tyloxapol (0.05%). At indicated time points after starvation cultures were resuscitated for up to 10 days by re-suspension in 7H9. Specific compositions of 7H9 broth, PBS, 7H11 and 7H10 agars are as follows: Middlebrooke7H9 (BD) supplemented with 0.5% (w/v) albumin, 0.2% (w/v) glucose, 0.085% (w/v) NaCl, 0.2% (v/v) glycerol and 0.05% (v/v) Tween 80 as nutrient replete media; unless specified nutrient starvation media was phosphate buffered saline (PBS; 137 mM NaCl , 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$) with 0.05% v/v tyloxapol, a non-hydrolysable detergent, adjusted to pH 7.4; where specified, Dulbecco's PBS (DPBS, Sigma-Aldrich) containing 0.5 mM MgCl$_2$, 0.9 mM CaCl$_2$, 137 mM NaCl , 2.7 mM KCl, 8 mM Na$_2$HPO$_4$, 1.5 mM KH2PO$_4$ was supplemented with 0.05% v/v tyloxapol and used as nutrient starvation media. For CFU determinations, serial dilutions were plated on 7H10 or 7H11 agar supplemented with 10% (v/v) oleic acid-albumin-dextrose-catalase (OADC). Prolonged starvation cultures were routinely assessed for contamination by microscopy (Gram, Ziehl-Neelsen, or auramine-rhodamine staining) and streaking on blood agar.

5.3.2. Antibiotic, azole, formaldehyde and hydrogen peroxide susceptibility testing

Log, S4, S10, S20 and R6 cultures were treated at OD_{600} 0.1 with streptomycin, rifampicin, ethambutol, isoniazid, clotrimazole, econazole, miconazole, methanol-free formaldehyde or hydrogen peroxide at indicated concentrations in 96-well plates for up to 5 days. Technical replicates were sacrificially assayed for OD_{600} and CFU. CFU were obtained from plating serial dilutions at the various indicated time points post-treatment on 7H10 agar after 3-4 weeks of incubation at 37 °C for *Mtb* and BCG and 2-3 days for SMG.

5.3.3. Flow Cytometry of Cellular Physiology

At indicated time points bacilli were stained for esterase activity, ROS generation or intracellular pH in 96-well plates. Bacilli were stained with CFDA (10 ng/mL, Life Technologies) or CellROX® green (500 nM, Life Technologies) for 30 minutes at 37°C followed by fixation in PBS+1% paraformaldehyde with DAPI (1 mg/mL, Life Technologies). Intracellular pH was measured using SNARF®-5F 5-(and-6)-carboxylic acid, acetoxymethyl ester, acetate (10 µM, Life Technologies) according to manufacturer's instructions. Stained cells were analyzed on a LSRII HTS flow cytometer.

For staining, 20 µl of sample was added to a well of a v-bottom 96-well plate containing 180 µl of 500nM CellROX™ Green (Life Technologies) in PBS then incubated at 37 °C for 30 minutes. Following incubation, bacteria were stained for DNA and fixed by addition of 50 µl of 4% PFA containing DAPI (1 µg/mL) for 10 minutes. Samples were analyzed on an LSRII HTS flow cytometer (BD Biosciences) within an hour of staining. Data was analyzed using FlowJo software (TreeStar Inc). ROS[hi] bacteria were identified based on a sequential gating scheme for 1) size (FSC and SSC, **Supplementary Figures 5.2C and 5.2G**), 2) DNA content (DAPI, **Supplementary Figures**

5.2D and 5.2H), 3) ROS detection dye uptake (CellROX® signal above unstained, **Supplementary Figures 5.2E and 5.2I**) and 4) ROS^{hi} above basal (CellROX® signal above untreated, **Supplementary Figures 5.2F and 5.2J**). CFDA^{hi} cells were identified via a similar staining protocol (1 ng/mL CFDA, Life Technologies) and gating scheme (**Supplementary Figures 5.2K-N**).

For pH determination, cells were pelleted, washed once with 100 mM phosphate buffer, pH 7.0, and resuspended at ~ 2 x 10^6 cells/ml in carboxy SNARF 5F-AM acetate (SNARF, 10 µM, Life Technologies) in 100 mM phosphate buffer, pH 7.0 or buffer only for 30 min. Additional aliquots of cells were used for a calibration curve in which SNARF plus nigericin (10 µM, Life Technologies) in 100 mM phosphate buffers at pH 5.5 to 8.0 was added to cells. Every condition and flow analysis had a separate calibration curve. DAPI (1 µg/mL, Life Technologies) was added to all cells for at least 10 min prior to flow cytometry. SNARF was excited at 561 nm with emission at 585/15 and 610/20. Bacteria were identified by size and DNA content, as described above. pH was calculated according to manufacturer's instructions using the ratios of median fluorescent intensity (MFI) of SNARF at the two wavelengths. Photon multiplier tube (PMT) voltage settings for the LSRII were SSC (250), FSC (400), DAPI ex 355/em 450/50 (550), CellROX® and CFDA ex 488nm em 525/50 (400), SNARF 585/15 (550), and SNARF 610/20 (550). A minimum of 50,000 events were collected for analysis from each sample.

5.3.4. RNA Extraction and Composition Analysis

Despite our findings on RNA yields with different extraction mixtures in Chapter 2, Trizol instead of Tris-EDTA saturated phenol-chloroform was used for bacterial lysis RNA extracted as it was compatible with downstream rRNA depletion and linker ligation processes. To ensure complete lysis of

mycobacteria, ~10^9 pelleted cells were combined with 1mL of TRIzol reagent (Life Technologies) and 100 µL of 0.1 mm silica/zirconium beads in a 2 mL screw-capped tube and shaken vigorously on a Qiagen Tissuelyser II (Qiagen) in pre-chilled chambers for 15 min at a frequency of 30 Hz. Lysates were chilled at -20 °C and a second beat-beating cycle was performed in the presence of 16% chloroform (v/v). Phase separation and RNA extraction was performed using the PureLink® RNA Mini Kit (Life Technologies, Carlsbad, CA). A 2 column separation strategy was used. 35% ethanol (v/v) on the first column to enrich for long (> ~150 nt) RNA species. 70% ethanol (v/v) was used on the second to trap small (~150 nt - ~20 nt) RNA species. Concentration was then quantified by UV spectroscopy at 260/280nm. RNA integrity and composition was determined using the appropriate Bioanalyzer RNA chips (Agilent Technologies) (See Chapter 3). Thereafter, we performed DNA elimination using a TURBO DNA-free™ Kit (Ambion, Life Technologies**).** For RNA-seq, rRNA is depleted using the Ribo-Zero™ Magnetic Kit for Bacteria (Epicentre, Illumina) as per manufacturer's instructions.

5.3.5. Biochemical plate assays

For catalase determination, pellets from 10 mL of culture at indicated time points were resuspended in 0.5 mL of 50 mM phosphate buffer plus 0.1 mM phenylmethylsulfonyl flouride PMSF and added to ~100 uL of 0.1 mm silica/zirconium beads in 2 mL screw-cap tubes with o-rings. Cells were bead-beaten with a Tissuelyser II (Qiagen) with pre-chilled chambers at ambient temperature for 10 min at 50 Hz. Lysate (supernatant) was collected after centrifuging at 12,000 g for 15 min at 4 °C to pellet the beads and cellular debris. Lysate was aliquoted and stored at -80 °C until use. Catalase activity in the lysate was measured using a catalase assay kit (Sigma-Aldrich) adapted for a 96-well plate. Protein content in the lysates was measured the

same day as catalase using a BCA assay according to manufacturer's protocol (Pierce, Thermo Scientific). Catalase levels were normalized to the protein content of each lysate.

ATP levels were quantified in cell lysates as mentioned above. ATP levels in lysates were compared with a standard curve of ATP (Promega) with the BacTiter Glo Assay (Promega) following manufacturer's instructions. ATP levels were normalized to CFU in the culture.

Levels of 3HB were quantified using a b-Hydroxybutyrate (Ketone Body) Colormetric Assay Kit (Cayman Chemical). Lysates were prepared as described except the pellets were resuspended in assay buffer (100 mM Tris-HCl, pH 8.5) plus 10% ethanol for bead-beating. 3HB levels were measured directly after bead-beating and normalized to CFU in the culture.

5.3.6. Triacylglycerol analysis by thin layer chromotography

Thin layer chromatography was used for semi-quantitative analysis of lipid content as described in [23]. To extract lipids for triacylglyceride analysis, lyophilized and weighed cell pellets were lysed in methanol/petroleum ether (2:1) with 0.02% (w/v) sodium chloride with the volume proportional to dry-weight. Soluble extracts were dried, dissolved in petroleum ether then spotted onto silica TLC plates with hexane/ethyl ether/acetic acid (45:5:1) as running buffer and iodine vapor as developing agent. Densitometry was performed using ImageJ (v1.49) (NIH), calculating pixel density based on the R_f of TAG standards.

5.3.7. Metabolic phenotype assay development

To adapt the standard GENIII microplate protocol (Biolog) to provide relative quantifications of metabolite usage, we first optimized for the survivability and

tetrazolium reduction of BCG in either IF-C or IF-B media (Biolog) in the presence or absence of 7H9 (0-20% (v/v)) at various cell densities (from starting OD_{600} of 0.01 – 0.2) and for periods from 1-7 days at 37 ^{o}C based on reported protocols [24, 25]. Full UV-VIS spectroscopic scans from 300 nm – 700 nm were performed on a Synergy 4 Multi-Mode Microplate Reader (BioTek) at 25 ^{o}C every 8 hours to track the conversion of tetrazolium to formazan and changes in culture turbidity. Data was loaded unto Origin (v8.5) (OriginLab) for background substraction, normalization, deconvolution (between tetrazolium reduction and optical density) and peak integration. Design of Experiments (DOE) was computed on Unscrambler® X (Camco) and a starting OD_{600} of 0.05 with 5 days of incubation was optimal while A_{565} and OD_{660} were determined to minimize spillover from between tetrazolium reduction to optical density. Test plates were then performed to ascertain the sensitivity (limit of detection, LOD) for A_{565} and OD_{660} based on their initial values prior to incubation, by spiking in known volumes of reduced dye or varying the cell density. Metabolites that induced had A_{565} readings below LOD were arbitrarily assigned a null value, while those that had OD_{660} at or below the initial values, but A_{565} readings above LOD were considered to induce abiotic reduction of the dye.

5.3.8. Metabolic phenotype screens

Utilization of metabolites was measured with Gen III Microplates (Biolog) for BCG at indicated time points. A custom phenotypic fingerprinting assay was developed to compare metabolite usage of different carbon sources (see above). Cultures were resuspended in assay media at OD_{600} 0.05. 100 µL of the bacilli suspension was inoculated onto Gen III Microplates. Each plate sealed and placed in an air-tight container with wet napkins to prevent evaporation. After 5 days incubation at 37 ºC the color density at A_{565} and

OD$_{660}$ was determined. Technical duplicated were performed for each samples (in biological triplicate for each time point), content of wells were plated onto 7H10 agar after incubation and A$_{565}$ and OD$_{660}$ reads to determine culture survivability. We noted that several wells which assayed for chemical sensitivity gave A$_{565}$ readings below that of the negative control and a design decision was made to exclude these data from subsequent statistical modeling.

5.3.9. RNA sequencing and transcriptome analysis

Coding sequences were enriched through the depletion of rRNAs and tRNAs and converted into template libraries using TruSeq® RNA Sample Preparation v2 kit (Illumina) as described by the manufacturer. Pre-sequencing quality controls for fragment sizes, dsDNA concentrations and loading molarities were performed. Samples were then pooled, loaded onto MiSeq® Reagent Kit v2 cartridges (Illumina) and ran on a MiSeq® System set at paired ends 151 base pair reads. Data from technical replicates were pooled, and post-sequencing quality controls of transcriptome coverage, RNA purity and quality were determined (**Supplementary Table 5.2**). Sequencing reads were mapped using the BWA-MEM algorithm (http://bio-bwa.sourceforge.net) referenced against Mycobacterium bovis BCG Pasteur 1173P2, complete genome (GenBank: AM408590.1). Gene expression level was determined based on sequencing coverage of the gene boundaries defined.

5.3.10. Quantitative real-time PCR (qPCR)

DNA-free RNA was diluted to a concentration of 50 ng/µl and reverse-transcribed using the iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. The reverse transcription program was run as

follows: 25°C for 5 minutes, 42°C for 30 minutes and 85°C for 5 minutes, followed by a cooling step at 4°C. Two-step real-time qPCR was performed by using the Sso Advanced Universal SYBR Green Supermix (Bio-Rad). The list of primers and their respective sequences and melting temperatures can be found in Table S3. The qPCR program was run as follows. 95°C for 30 seconds followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 30 seconds. A melt curve analysis consisting of 0.5°C increments from 65°C to 95°C was performed for all reactions to ascertain the specificity of the primers. Sample C_T values were normalized against endogenous sigJ expression (whose levels were shown to be invariate across the starvation time course) and analyzed using the comparative C_T method (see Chapter 2).

5.3.11. Data handling, processing and statistical methods

Data from survival and susceptibility assays were normalized as percentages of their respective untreated or initial CFU as determined by experimental design. Data from biochemical measurements, metabolite utilization and gene expression assays were deflated as fold changes against Log values. Logarithmic transformations were used for datasets with skewed or wide distributions and indicated in their respective figures legends. Single cell features quantified by flow cytometry were summarized as median fluorescence intensities for their respective gated subpopulations. Gated features represented as percentages of the total population. Comparisons between 2 samples were made using the appropriate two-tailed t tests after equality of variances were tested using *F*-tests. Comparisons between multiple samples were determined using One-way or Two-way ANOVA when comparing 1 or 2 factors respectively. Bonferroni's or Dunnett's Multiple Comparison or Tukey's HSD post-hoc tests were used where it best reflects

how sample means were compared and stated in their respective figure legends. Trends in antibiotic tolerances were determined using Mann-Kendall trend test. Unless otherwise stated, all data are represented as arithmetic means± SE. To aid interpretation on statistical significances, $p < 0.05$, $p < 0.01$ and $p < 0.001$ are denoted as *, ** and *** respectively.

Descriptive, univariate, one- and two-way ANOVA were performed on Prism v.5.1 (GraphPad). Hierarchical clustering was performed using background-subtracted, mean-centered data of tetrazolium dye reduction based on colormetric measures at A_{565}. Heat-maps plotted using MultiExperiment Viewer (Dana-Farber Cancer Institute). One and two-sample Kolmogorov-Smirnov test, Mann-Kendall Trend analysis, Grubb's Test for Outliers, Box-Cox transformations and RDA were performed using XLSTAT (Addinsoft). PCA and PLS-DA were performed using Unscrambler® X.

For metabolic phenotypes, data was computed as fold-changes against the mean values of each respective metabolite across all Log cultures. Samples were assigned groupings based on their respective timepoints. To better quantify the contributions of each PLS-DA predictor, orthogonal projections to latent structures discriminate analysis (OPLS-DA) was performed using SIMCA-P+ v12 (Umetrics) and the VIP values presented in Table S1 [26, 27].

Differential expression analyses were carried out using DeSeq (http://genomebiology.com/2010/11/10/r106). The 3 replicates in each timepoint were compared against 3 control replicates as independent DeSeq statistical tests. Local fit option was used to estimate dispersions in order to reach convergence. Genes of interest are defined with at least 1.5 fold up- or down-regulation, and with multi-testing adjusted p values ≤0.05.

GSEA (Broad Institute) was performed for targets identified by RNA-seq by pair-wise comparisons of Log samples against S4, S10, S20 and R6 samples. 1000 permutations were performed. The genes, which are associated with the ketone body-CYP reactome are identified by RNA-seq based on a combination of significant gene expression change upon starvation (absolute fold change > 1.5 at any one time point; $p < 0.05$) and comprehensive functional annotation searches against the DAVID (http://david.abcc.ncifcrf.gov/), Reactome (http://www.reactome.org/), TbDB (http://www.tbdb.org/), Tuberculist (http://tuberculist.epfl.ch/), BioCys (http://biocyc.org/) and KEGG (http://www.genome.jp/kegg/) databases. Genes whose functions and expressions fits previously reported phenotypes were defined as the high-confidence ketone body-CYP reactome genes and used as a gene set.

To integrate the various datasets for exploratory analysis, gene expression and metabolic phenotype and antibiotic tolerance data were expressed as fold changes against the means of Log samples. Fold change of gene expressions was further $\log_2$ transformed, while fold change of antibiotic tolerances were $\log_{10}$ transformed. Due to negative values after compensation, cell populations detected with flow cytometry were Box-Cox transformed. Prior to PCA, Kolmogorov-Smirnov test and Grubb's Test for Outliers were used to identify and remove samples with values deviating from the norm or group that likely came from technical errors.

Cluster analysis was performed using two-way hierarchical clustering with Euclidean distances and complete linkages. Exploratory data analysis was performed with PCA using the NIPALS method. Metabolic utilization predictors for cell state were determined using PLS-DA and tested on independent S30 and Log samples. Multiple regression was performed using

RDA after parsimonious canonical correspondence analysis was used to determine the minimal number of genes that explain observed variances in population phenotypes. Gradient lengths of variables were estimated by detrended correspondence analysis (DCA).

Screening PCAs were performed with gene expression of all 3138 genes significantly up-or down- regulated as the explanatory variable and antibiotic tolerance and population characteristics (measured by flow cytometry) as the response variable. The "find outlier" function in Unscrambler® X was used to eliminate genes that caused over-fitting while a second application on the remaining genes identified putative influencers. There were manually inspected using residue-variance plots and T^2 statistics to remove erroneous data. 235 genes remain. Similarly, metabolic phenotypes were used to explain antibiotic tolerance and population characteristics. 7 metabolic phenotypes were selected. Thereafter, to characterize antibiotic tolerance and population characteristics and metabolite utilization in terms of gene expression a gradient analysis was performed with DCA. A PCA triplot was made using these datasets using centered and standardized data on the same measurement scales (**Supplementary Figure 5.5M**).

The 235 genes were then subjected to RDA. The amount of variation in observed phenotypes uniquely explained by each variable in the model and the amount of variation explained by all explanatory variables in common were assessed with the method of 'variation partitioning' according to the procedure of Borcard [28]. Significant variables were selected using the Canoco 5 (Microcomputing Power) with $p < 0.05$ as the significance threshold over 1000 Monte Carlo Permutations. 68 genes which together explained 96% of the observed variance were selected. To aid visualization, a RDA triplot was plotted with the observed antibiotic tolerance, metabolic utilization

and population phenotypes as independent variables and gene expression as the dependent variable.

5.4. Results

5.4.1. A data-driven approach to characterize starvation-induced persistence in mycobacteria

How does starvation lead to persistence in mycobacteria? We addressed this problem with a three-pronged approach that started with defining growth kinetics and cytotoxicity of rapid nutrient deprivation in three species of mycobacteria – the epidemic *Mtb* strain CDC1551 [29], the attenuated vaccine *M. bovis* BCG (Pasteur 1173P2; BCG) and the fast-growing *M. smegmatis* (mc^{2}155; SMG) – in simple phosphate-buffered saline (PBS). Using these models, we then pursued a data-driven analysis of the mycobacterial stringent response – hallmarks of a coordinated reprogramming of metabolic pathways [6, 30]. This data-driven approach was divided into exploratory and confirmatory phases (**Figure 5.1A**). The exploratory phase involved the acquisition, mining and analysis of flow cytometry to assay enzyme activity and ROS levels in persister populations (McBee et al., unpublished), phenotype arrays to determine metabolic potential [31, 32], and RNAseq to characterize gene expression. An integrative analysis revealed previously uncharacterized pathways and generated several testable hypotheses. In the confirmatory phase, these hypotheses were tested by targeted measurements of physiological and biochemical properties of starved mycobacteria.

5.4.2. Evaluating mycobacterial models for starvation-induced persistence

From its inhalation as aerosols to phagocytosis by alveolar macrophages in the lung and subsequent containment in granulomas during latent infection, mycobacteria are exposed to drastic changes in their microenvironments including nutrient availability. Compared to well-defined hypoxic stress models [33, 34] few studies have addressed biochemical characterization and phenotypic profiles of starved mycobacteria outside of the *Mtb* H37Rv strain [19, 35]. To this end, we defined growth kinetics and cytoxicity of rapid nutrient deprivation *Mtb*, BCG and SMG in PBS. As shown in **Figures 5.1B-H**, starvation in PBS caused a classic biphasic decline similar to that observed in other bacteria [36, 37]; [38, 39]. This is defined by a rapid decrease in both optical density (OD_{600}) and culturable colony forming units (CFU) in the initial phase (~4 d for *Mtb*, BCG; ~2 d for SMG) and subsequent stabilization at ~16%, 10% and 31% of initial CFUs for *Mtb* (days 10-22), BCG (days 10-30) and SMG (days 2-6), respectively (**Figures. 5.1B-5.1G**). The steep decrease in OD_{600} and CFU in the first phase of starvation led us to address cannibalism as a source of biomass for bacterial population maintenance in nutrient deprivation models [36]. Unlike *Bacillus subtilis*, in which starvation induced cannabalism [40], we found no evidence of this in terms of additional cell death caused by washing starved bacteria and resuspending them in fresh PBS (see **5.5.1. Extended results** and **Supplementary Figure 5.1A**). While previous reports suggested no loss of *Mtb* viability during starvation in vaguely defined medium [14, 19], we found that this phenotype was strictly dependent on the presence of Mg^{2+} and Ca^{2+} ions, which are components of Dulbecco's PBS (DPBS) but not true PBS used here (see **5.5.1. Extended results** and **Supplementary Figures 5.1B, C**). Divalent cations thus represent a major determinant of survival of mycobacteria under nutrient starvation in vitro and possibly *in vivo*.

Throughout the starvation period, *Mtb*, BCG and SMG all retained their ability to regrow when transferred in nutrient replete 7H9 media (**Figure 5.1H**). To rule out a stress-induced transition into a viable but non-culturable (VBNC) state [1, 41], we compared growth profiles of exponentially growing BCG (denoted as Log) to BCG starved for 4, 10, 20 and 30 d (denoted as S4, S10, S20 and S30) and then, for each time point, placed in 7H9 medium for 2, 4, 6, and 10 d of resuscitation growth (denoted as R2, R4, R6 and R10). As shown in **Figure 5.1H**, we observed no significant changes in doubling times, which ranged from 23 to 29 h (p > 0.05). This suggested a minimal contribution of VBNC bacteria to the overall starvation-induced persistent phenotype.

To address the second feature of persistence, we assessed antibiotic tolerance in BCG cultures with 4 clinically relevant antibiotics: isoniazid (INH), rifampicin (RIF), streptomycin (STM) and ethambutol (EMB). Compared to log-phase cells, starved BCG at S4, S10 and S20 were refractory to antibiotic killing by INH, RIF, STM to EMB across a wide range of antibiotic concentrations (**Figures 5.2A-D,** ranges of drug concentrations tested provided in figure legends). Their non-replicating nature and high levels of tolerance – often beyond the solubility of these antibiotics in PBS (within non-cytotoxic concentrations of DMSO) – made direct comparisons by MIC_{50} and MBC_{90} impracticable. Nonetheless, there was a striking, time-dependent increase in antibiotic tolerance as cells adapted to nutrient deprivation. When treated with a minimal bactericidal dose of INH (MBC_{99}), 96% of killing occurs within 24h. By 48h, another 75% of the surviving bacteria are lost. In contrast, at S10 and S20, only 39% of the viable cells were lost on the first day and only an additional 17% were lost on the second day. Cells at S4 displayed an intermediate phenotype. Similar trends can be observed with RIF, STM and
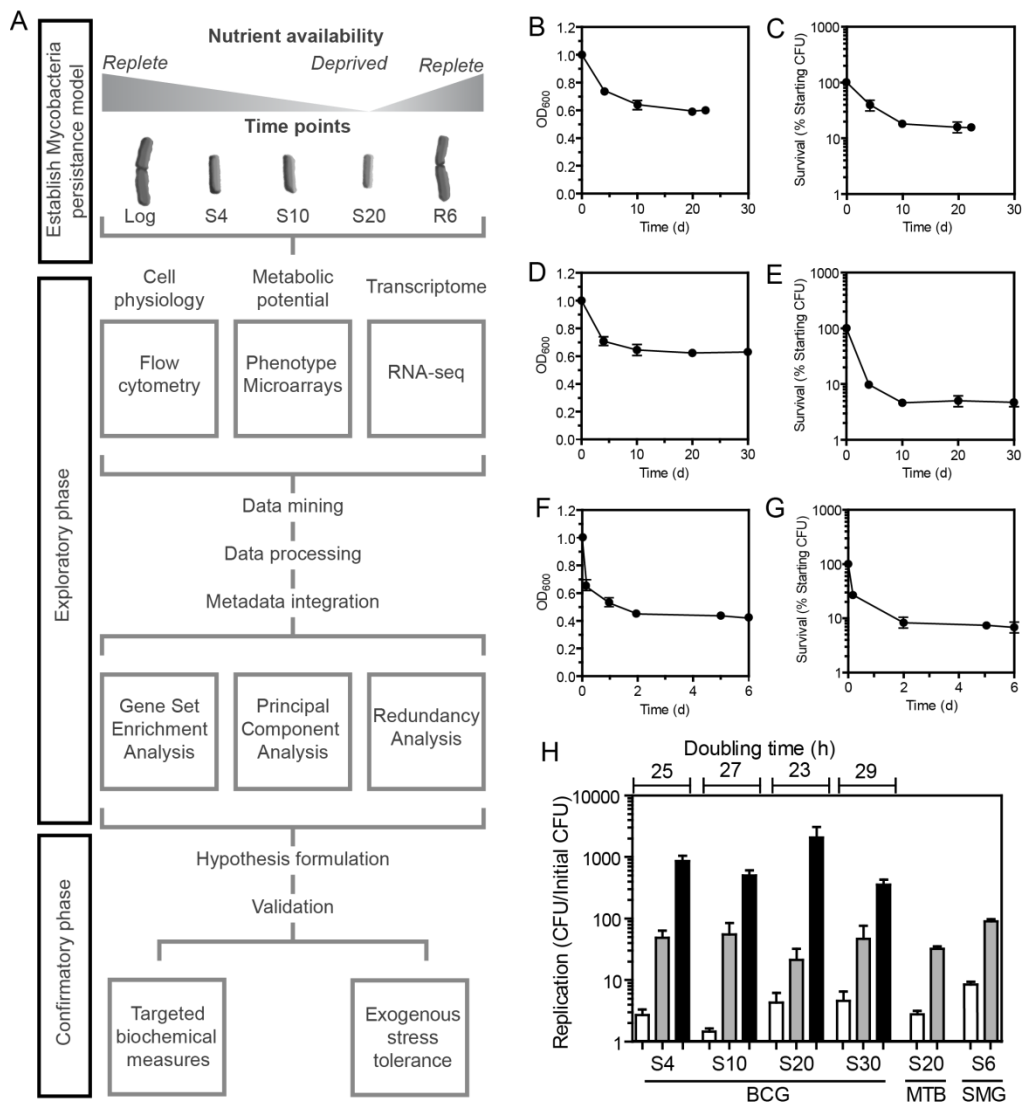
Figure 5.1. Mycobacterial persistence study design based on the survival and recovery of *Mtb,* BCG and SMG during and after starvation. (**A**) Experimental workflow starting with the development of a starvation-induced mycobacteria persistence model, followed by characterizations of cellular physiology, metabolite utilization and gene expression. Data was pooled and exploratory data analyses performed. Hypotheses were generated, validated on targeted biochemical measures and explicitly tested for predicted stress susceptibilities. (**B-G**) Survival profiles (OD_{600} and CFU) of *Mtb* (**B and C**), BCG (**D and E**), SMG (**F and G**) under nutrient deprivation in PBS. (**H**) Recovery profiles of BCG after 4, 10, 20 and 30 days of starvation in PBS (denoted as S4, S10, S20 and S30 respectively) followed by 2, 6 and 10 days of resuscitation in 7H9 broth (R2 ▭, R6 ▭ and R10 ▪); *Mtb* after 20 days of starvation (S20) and followed by 2 and 6 days of resuscitation (R2 ▭ andR6 ▭); SMG after 6 days of starvation (S6) and followed by 9 and 24 hours of resuscitation (R9h ▭ and R36h ▭). No significant differences in doubling times observed; $p > 0.05$ (one-way ANOVA with post-hoc Tukey's HSD). Data shown as arithmetic means of ≥ 3 independent experiments ± SE.

EMB exposures (**Supplementary Figures 5.1D-G**). These results indicated that starvation shifts this dynamic toward a drug tolerant phenotype prior to antibiotic exposure, rather than as a response to the exposure. Of relevance to tuberculosis relapse, we tested whether the drug tolerant state was reversible. Indeed, the susceptibility to all 4 drugs reverts in resuscitated (R6) cells (**Figures 5.2A-D**). Thus, starvation induces a state of reversible persistence with phenotypic drug tolerance in mycobacteria.

5.4.3. Biphasic modulation of the molecular hallmarks of starvation survival

Having established the mycobacterial models of starvation-induced persistence, we next interrogated key modulators of the stringent response in relation to the biphasic mycobacterial growth behavior. While nutrient deprivation induces characteristic alterations in physiology through the up-regulation of Rsh homologs (Rel_{BCG} in BCG) that regulate the intracellular levels of pp(p)Gpp [13, 42, 43], there are several ubiquitous hallmarks of starvation stress including loss of rRNA, toxin-antitoxin induction, and cessation of cell division. Consistent with these findings, we observed that the stringent response dominated the initial adaptive phase to nutrient deprivation (Log to S4) in BCG. Using RNAseq and qPCR, we observed that rel_{BCG} transcription peaked at S4 (**Figure 5.2E**) and was associated with a concomitant 40% reduction in total RNA content compared with log-growing bacilli. After S4, no further decreases in total RNA occurred ($p < 0.05$ for Log vs S4, S10 and S20; $p > 0.05$ for S4 vs S10 and S20). However, RNA returned to pre-starvation steady-state levels within 6 d of resuscitation (R6) (**Figure 5.2F**). These changes in total RNA could be parsed into individual species of non-coding RNA, with rRNAs (23S, 16S, 5S) all exhibiting the same pattern (**Supplementary Figures 5.1H-J**). However, tRNA levels only

decreased at S10 and S20 (**Supplementary Figure 5.1K**). Similar to rel_{BCG}, synchronous induction of toxin-antitoxin modules (*relBE2, mazEF6, vapBC3, vapBC31*) and repression of *ftsZ*, which encodes a major cytoskeletal protein essential for cell division [44], occurred at S4 (**Figure 5.2E**). In line with these observations, we detected a decrease in intracellular ATP at S4 but not at S10, S20 or R6 (**Supplementary Figure 5.1L**). Thus, we refer to the "adaptive phase" of starvation as the initial decrease in cell viability, the activation of the stringent response, and the development of drug tolerance. The "persistent phase" is characterized by population stability, greater drug tolerance, and a rebound in intracellular ATP (**Supplementary Figure 5.1L**). Interestingly, subpopulations of cells had increases in basal intracellular ROS (CellROX^{hi}) generation of ~8% at S4 and S10 and ~40% above Log and R6 basal levels at S20 (**Figure 5.2G, Supplementary Figures 5.2C-J**). Additionally, no accompanying changes in catalase activities or *katG* expression were detected (**Supplementary Figure 5.1M, Figure 5.6C**). Collectively, these results suggest a transition from the adaptive phase leading to an altered metabolic state in the persistent phase.

5.4.4. Lipid catabolism and ketone body usage define the metabolic transition from the adaptive to the persistent phase

We next sought to unravel the progression of metabolic events leading from the adaptive to the persistent phase of starvation-induced persistence. Here we used flow cytometry for analysis of intracellular metabolism at single-cell resolution and phenotypic arrays for population-based assessment of cellular metabolic potential. Flow cytometric analysis supported the notion of a dynamic shift in persisters formation during starvation, with coalescence in the population forward- and side-scatter parameters (**Supplementary Figures 5.2A and B**) implying a greater uniformity in cell size and granularity
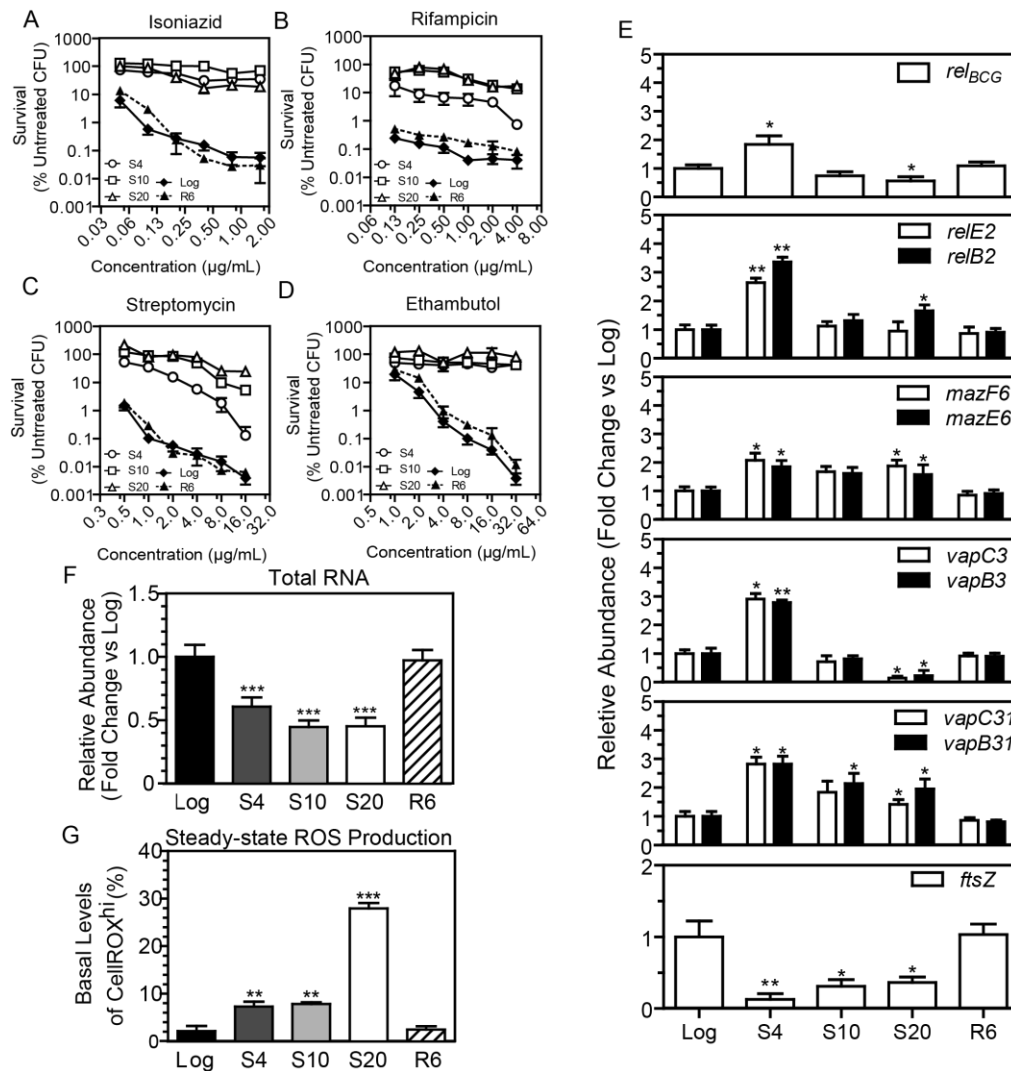
Figure 5.2. Development of antibiotic tolerance coincides with stringent response induction and increased basal ROS levels. (**A-D**) Starvation of BCG induces viable but reversible persisters. Antibiotic susceptibility profiles of Log, S4, S10, S20 and R6 BCG after 48h of exposures to 0.05 - 1.6 µg/mL of isoniazid (**A**), 0.13 - 4.0 µg/mL rifampicin (**B**), 0.5 - 16 µg/mL streptomycin (**C**) and 1.0 - 32 µg/mL ethambutol (**D**). Starved (S4, S10 and S20) bacilli tend to be drug tolerant, seldom killing > 90% within these ranges of drug concentrations (n ≥ 6). (**E**) Induction of the stringent response, mediated by rel_{BCG}, its associated effectors including toxin-antitoxin systems *relBE2*, *mazEF6*, *vapBC3* and *vapBC31* and repression of its targets such as cell division marker *ftsZ* measured by RNA-seq. (**F**) Reduction in total cellular RNA levels at S4, S10 and S20 and its restoration at R6. Changes in major non-coding RNA components are found in **Supplementary Figures 5.1H-K**. (**G**) Percentage of cells with elevated steady-state ROS production (CellROX[hi]) during and after starvation. For (**E-G**), n ≥ 3; *, ** and *** denotes p-values (see Experimental Procedures) determined by One-way ANOVA with Dunnett's test vs Log.

in the starved state. Moreover, starvation enriched for cells possessing high

esterase activity (CFDA[hi]; Log – 16%, S4 – 75%, S20 – 82%), which reverted

upon resuscitation (**Figure 5.3A, Supplementary Figures 5.1K-Q**).

Contributing to the increased esterase activity are lipases active on

triacylglycerols (TAGs), which are a known carbon source during resuscitation

from a hypoxia-induced non-replicative state [23, 45]. As shown in **Figure 5.4**,

we detected starvation-induced up-regulation of several lipases, including

PE/PPE family lipases *lipY* and *lipX*, as previously reported [14], with LipY

active on TAGs [46]. These observations were further confirmed using thin

layer chromatography (TLC) to reveal that TAG stores were depleted within 1

day of starvation and replenished within 2 days of resuscitation (**Figure 5.3B,

Supplementary Figure 5.1N**). TAGs are thus a potential carbon source in

the adaptive phase of mycobacterial starvation.

Since TAG metabolism generates toxic intermediates [47] and TAG depletion

early in starvation raises questions about metabolism in the persistent phase,

we assessed the capability of persistent BCG to metabolize various

substrates using phenotype arrays. We surveyed 71 different metabolites for

growth and metabolic activity and found that 41 were metabolized and 32 of

these also supported growth (**Figure 5.3C**). Hierarchical clustering (**Figure

5.3C**) and principle component analysis (**Figures 5.3D and E**) of the

metabolic phenotypes again revealed the biphasic behavior in starvation

survival observed in earlier studies, with the adaptive phase (S4) and

persistent phase (S10 and S20) clearly distinguished from exponentially

growing cells (Log, R6) based on their metabolite utilization capabilities

(**Figures 5.3C-E**). One notable feature of the persistent phase bacilli was

their ability to utilize the ketone bodies acetoacetate (AcAc) and 3-

hydroxybutyrate (3HB) for growth. Partial least squares discriminate analysis

(PLS-DA) of the 41 metabolitically active substrates selected AcAc and 3HB

utilization as predictors of the persistent phase (S10 and S20) (**Figure 5.3E**;

see **Supplementary Table 5.1** for contributions from each predictor).

Furthermore, the model successfully discriminated between Log and S30

samples obtained from independent experiments (Log-4 and S30-1 in **Figure**

5.3E). The validity of this model was confirmed in direct assays of intracellular

3HB, which showed accumulation in S20 persisters (**Supplementary Figure**

5.1O).
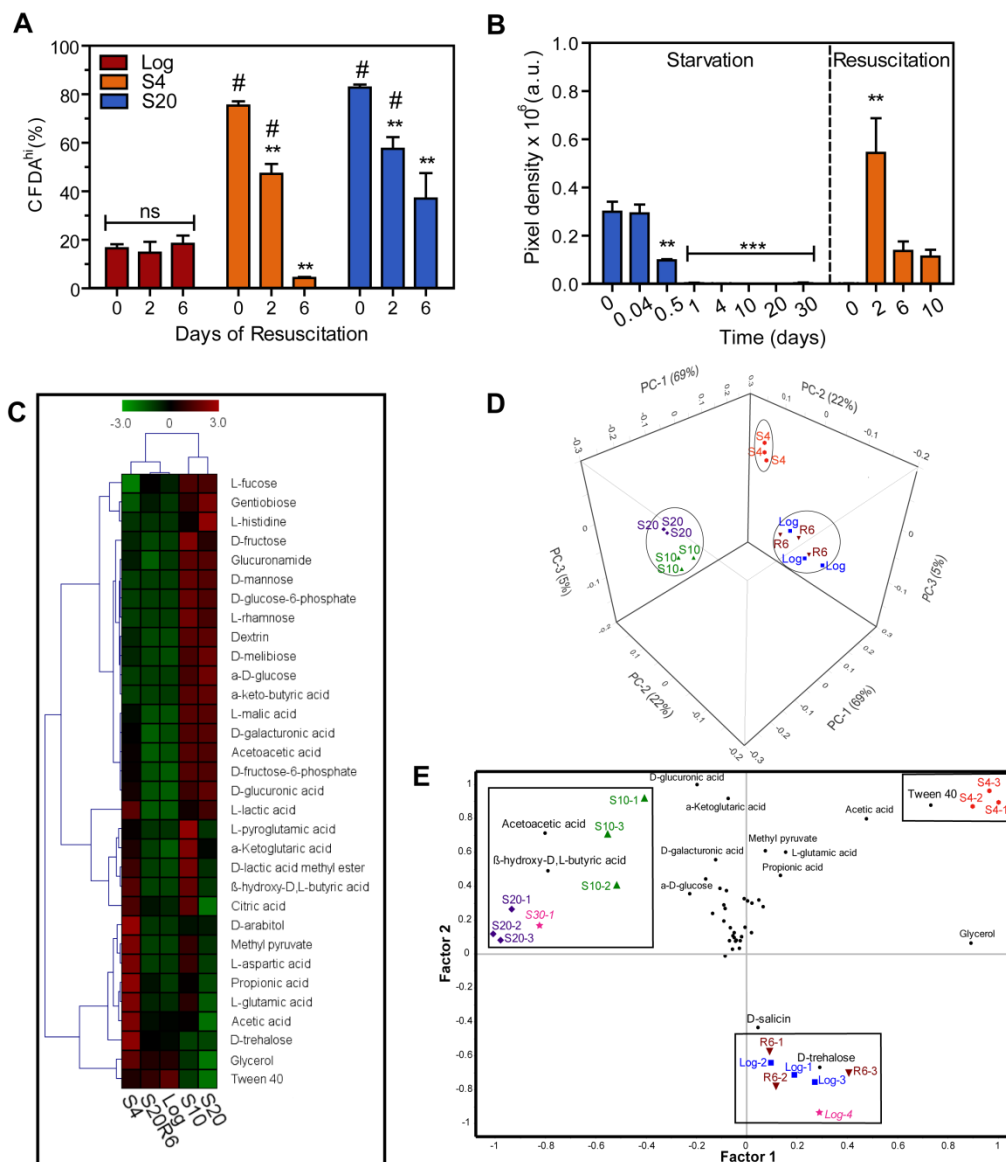
Figure 5.3. Starvation induces shifts in lipid and ketone body metabolism. (**A**) Percentage of cells with elevated esterase activity (CFDA[hi]) during starvation and resuscitation. A greater number of cells in S4 and S20 BCG cultures (day 0) show significant shifts in esterase activity (assayed by flow cytometry). n = 6; #: $p < 0.001$; starvation accounts for 46.9% of total variance (two-way ANOVA with Bonferroni post-tests comparing starvation

and time interactions). Esterase activity changes upon resuscitation for S4 and S20 but not Log cultures (n = 6; ns: not significant; **, one-way ANOVA with post-hoc Tukey's HSD). (**B**) TAG content of BCG during starvation (■) and resuscitation (■) analyzed by TLC. Densitometric means of 3 independent experiments ± SE. Representative TLC images in Figure S1N. For (A) and (B) **$p < 0.01$, one-way ANOVA with Dunnett's test (sample means vs day 0). (**C**) Hierarchical cluster analysis of the metabolic phenotypes of Log, S4, S10, S20 and R6 cultures on carbon sources that induced growth. Carbon source utilization measured by the reduction of tetrazolium dye. Red (higher), black (equal) or green (lower) blocks represent dye reduction relative to the positive control. (**D**) PCA scores plot and (**E**) Bi-plot of PLS-DA scores and loadings of metabolic phenotype based on carbon sources with significant dye reduction (one-way ANOVA with Bonferroni post-test). The positions of eigenvectors (■ Log, ● S4, ▲ S10, ◆ S20 and ▼ R6) in relation to each other have similar metabolic phenotypes (Log and R6; S10 and S20). S4 cultures are dissimilar to both groups. 96% of observed variance can be explained by 3 principle components (n = 3, PC-1: 69%, PC-2: 22%, PC-3: 5%). The PLS-DA model shown in Panel E is based on correlation coefficients between PCA scores and loadings contributions by carbon sources (●), wherein proximities of phenotypic eigenvectors (scores) and metabolite utilization activities (loadings) represent the contributions (relative weights) of the carbon source utilizations that characterize the Log/R6, S4, S10/S20 states. These predictors were used to successfully differentiate independent ★Log and ★S30 samples.

5.4.5. RNAseq identifies novel ketone body metabolic pathways in nutrient-deprived BCG

To gain further insights into the molecular mechanisms underlying lipid metabolism in starvation-induced persistence, we performed RNAseq on Log, S4, S10, S20 and R6 BCG to characterize changes in gene expression caused by starvation. Appending the reads from independent technical duplicates of each of 3 biological replicates, 1-2 million read pairs with a 3'-to-5' ratio of 1.02 ± 0.03 for each sample and >90% coverage of the transcriptome for all time points were obtained (**Supplementary Table 5.2**). Amidst the global reduction of transcriptional activity **(Supplementary Figure 5.3A)**, we noted characteristic changes several well-studied genes altered by carbon starvation, including *rel*$_{BCG}$, sigma factors *sigA* and *sigF* (*rpoS* homolog) and molecular chaperones *hspX* and *dnaK* (**Figure 5.4**) [14, 48, 49]. We also observed contributions from WhiB-family transcription factors,

cAMP signaling, Lsr2 repressor, PhoPR, TrcPR and KdpDE two-component signaling, which could facilitate global transcriptional reprogramming (**Figure 5. 4**). Further, we performed DAVID functional annotation and gene functional classification on 511 up-regulated genes (fold-change ≥1.5 and p ≤0.05) and 680 down-regulated genes (fold-change ≤0.66 and p ≤0.05) and cross-validated these pathway enrichments using KEGG Pathway Mapper [50, 51]. Specifically, analysis of core sets of genes either up-regulated during starvation and down-regulated during resuscitation, or down-regulated during starvation and up-regulated during resuscitation, revealed altered carbon and nitrogen metabolism, substrate translocation across membranes, environmental and intracellular signal transduction, increased DNA repair and RNA, protein, lipid and xenobiotic degradation (**Supplementary Figures 5.3B and C**). Of particular interest here was the selective enrichment of pathways involved in ketone body synthesis and degradation during starvation (**Supplementary Figure S3C**). Ketone bodies are products of fatty acid β-oxidation and their accumulation causes a toxic ketoacidosis (3HB, pK_a =4.70; AcAc, pK_a = 3.58; cytoplasmic pH = 6.1-7.2) [52]. Given the utilization of intracellular lipid stores during starvation, ketone body generation was not surprising.

Ketone body detoxification has not been characterized in mycobacteria or annotated in databases. An analysis of the RNA-seq and other data, along with literature precedent, suggested a novel pathway for detoxification of ketone bodies in mycobacteria. The analysis starts by considering two annotated pathways operant in other prokaryotes but not mycobacteria. First, AcAc can be avoided by conversion of acetoacetyl-CoA (AcAc-CoA) to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) through the mevalonate pathway or by breakdown of AcAc-CoA to acetyl-coa (Ac-CoA) that then enters the

TCA cycle via the action of acetyl-CoA C-acetyltransferase (thiolase II; FadA - A6). However, mycobacteria do not utilize the mevalonate pathway and acetyl-CoA acetyltransferase also catalyzes the reverse Claisen condensation (Ac-CoA to AcAc-CoA) in a substrate-dependent manner [53, 54]. Second, 3HB can be polymerized into poly-3-hydroxybutyrate (PHB) granules [55]. However, the RNA-seq data do not show any changes in expression of very long chain acyl-CoA dehydrogenases and synthases that would catalyze this reaction, or changes in expression of homologs of PHB depolymerase that would catalyze the subsequent breakdown of PHB in BCG. We propose a novel third detoxification pathway in mycobacteria, in which AcAc is decarboxylated to acetone that is further oxidized by cytochrome P450s (CYPs) to acetol. Baeyer-Villager monooxygenases (BVMO) then convert acetol to hydroxymethylene acetate, which decomposes spontaneously into formaldehyde and acetic acid. Formaldehyde is further detoxified by formaldehyde dehydrogenases into formate, which is utilized in tetrahydrofolate (THF) based one-carbon metabolism (**Figure 5.4**).

This ketone body metabolic pathway is supported by our experimental data. First, there is strong evidence for up-regulation of the key metabolic enzymes, including significant up-regulation in the expression of 3 class II thiolases, 11 CYPs, 3 P450-linked oxidoreductases and 5 BVMOs (**Figure 5.4**; **Supplementary Figure 5.4**), previous reports on mycobacterial acetol metabolism [56, 57], and current observations on 3HB utilization and accumulation (**Figure 5.3C**; **Supplementary Figure 5.1O**). Moreover, metabolism of alcohols and acetone by mammalian CYPs is well established and known to produce superoxide anion radicals [58-62], which fits our observation of increased basal ROS production in persisters (**Figure 5.2G**).
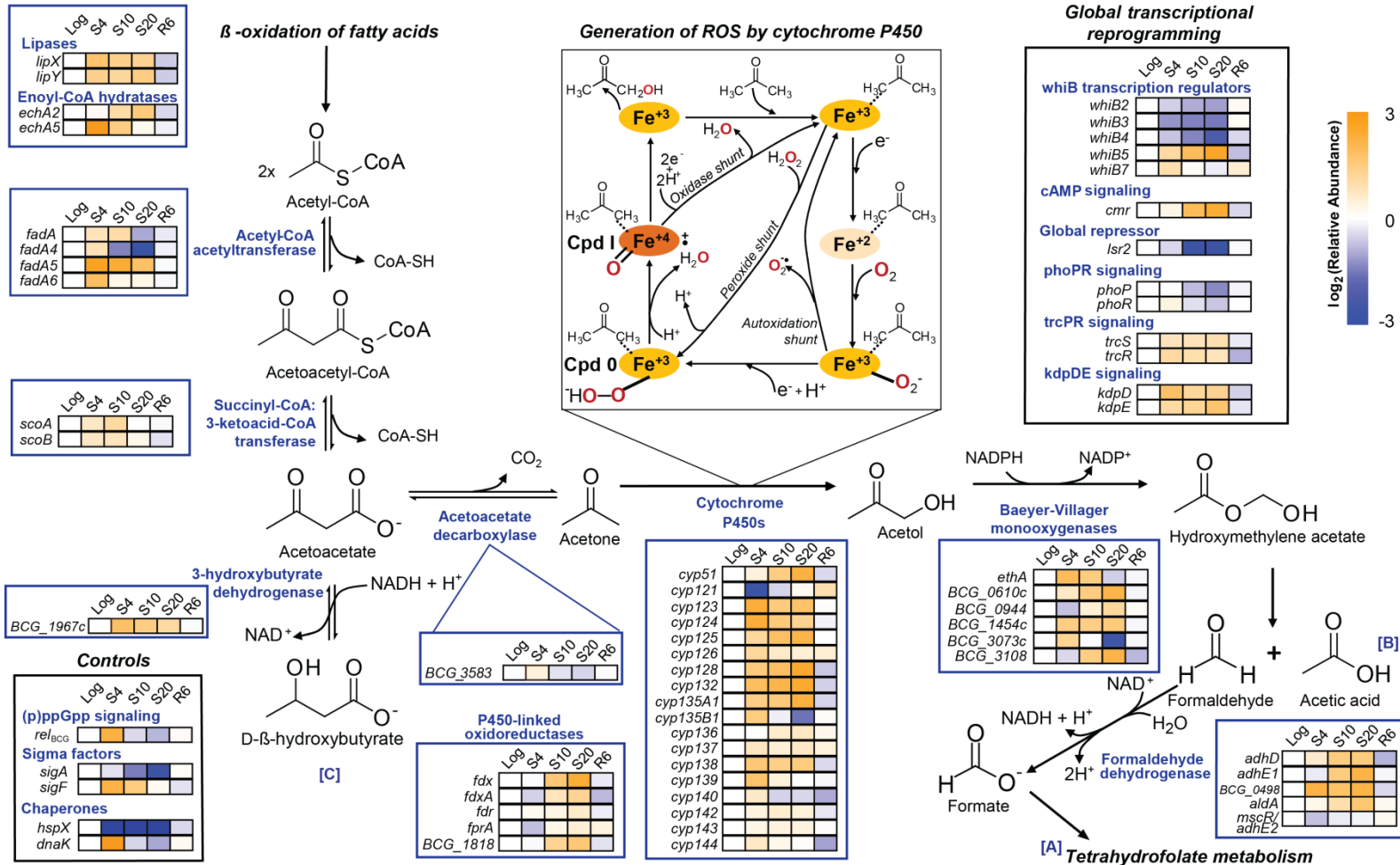
Figure 5.4. Transcriptome analysis reveals ketone body metabolic pathways linking β-oxidation of fatty acids to C1 cycling by tetrahydrofolate. Starvation induced global transcriptional reprogramming. Canonical effectors, such as rel_{BCG}, sigma factors (sigA and sigF) and chaperones (hspX and dnaK) together with transcriptional regulators (*whiB*-family proteins, *cmr* and *lsr2*) and two-component signalers (*phoPR*, *trcPR* and *kdpDE*) are implicated. Transcript levels of a number of enzymes involved in fatty acid β-oxidation increased transiently. *echA2*, *echA5*, *fadA*, *fadA4*, *fadA5* and *fadA6*, involved in the hydration and thiolysis steps are shown. Correspondingly, transitory up-regulations of succinyl-CoA:3-ketoacid-CoA transferase, putative 3-hydroxybutyrate dehydrogenase, putative acetoacetate decarboxylase, CYPs, BVMOs and putative formaldehyde dehydrogenases were observed. Metabolism of formate and acetate (Pathway **[A]** and **[B]**) are further supported by up-regulations of enzymes functioning in interacting pathways. 3HB alternately, could be oxidized directly by CYPs (Figure S4A) and participate in lactate metabolism (**Supplementary Figure 5.4B**, Pathway **[C]**). Heat maps show expression levels of various enzymes at Log, S4, S10, S20 and R6. Orange, blue and white blocks represent up-regulation, down-regulation and no change in transcript levels relative to Log respectively. Putative functions of *BCG_1967c*, *BCG_3583*, *BCG_1818*, *BCG_0610c*, *BCG_0944*, *BCG_1454c*, *BCG_3073c*, *BCG_3019* and *BCG_0498* predicted through homology.

Unlike the metabolism of AcAc to methylglyoxyl in mammalian liver microsomes, mycobacterial BVMOs prevent the formation of this bactericidal electrophile [63]. Furthermore, we detected a coordinated increase in acetyl-CoA synthase (*acs*), and alcohol and aldehyde dehydrogenases (*adhD*, *adhE1*, *aldA* and *BCG_0498*) that metabolize acetate and formaldehyde, the breakdown products of acetol, to acetyl-CoA and formate, respectively (**Figure 4**; **Supplementary Figure 5.4**). The RNA-seq data also revealed up-regulation of *folD,* a multi-functional enzyme that sequesters formate for one-carbon transfer reactions (**Supplementary Figure 5.4**). Thus, the data point to a coordinated gene regulatory response for a CYP-based ketone body detoxification pathway in the persistent phase of mycobacterial starvation.

5.4.6. Meta-data analysis builds consensus for a model of starvation-induced ketosis

The exploratory phase of the studies (**Figure 5.1A**) continued with meta-analyses of the various datasets. To determine if this ketone body detoxification pathway and its associated interactions were significantly enriched during starvation, we mapped the pathway onto the intermediary metabolism reactome, and performed Gene Set Enrichment Analysis (GSEA) [64]. This revealed a gradual enrichment of 45 of the 63 genes in the reactome as high confidence targets, cumulating at S20 (NES = 1.6, FDR = 0.055) (**Figures 5.5A-C**). Three CYP genes, *cyp51*, cyp125, cyp128, *cyp132* and *cyp135A1*, were featured in the core enrichment with rank metric scores (signal-to-noise ratios) greater than 3. In addition to confirming the previously reported shutdown of the TCA cycle [14], we also noted transient enrichment of the SigF and Lsr2 regulons peaking at S10 and S20, respectively (see **5.6.1. Extended results** and **Supplementary Figures 5.5A-K**). Furthermore, all gene set enrichments reverted at R6 (**Figure 5.5D** and **Supplementary**

Figures 5.5D, H and L), demonstrating the plasticity of transcription control mechanisms of the starvation stress response.

We next pursued a multivariate analysis of the data. Unconstrained principle component analysis (PCA) of 3138 genes deferentially up- or down-regulated during starvation (>1.5 fold change; $p < 0.05$) identified *cyp124*, *cyp128*, *cyp132* and *cyp135A1* as strong influencers of the starvation phenotypes (**Supplementary Figure 5.5M**). Not only was cell state discriminated based on metabolic utilization, but antibiotic tolerance was correlated with the ketone body metabolism, increased basal ROS production, and esterase activities (**Figure 5.5E** and **Supplementary Figure 5.5M**). Importantly, we determined the genes that contributed to the drug tolerant phenotypes. When 235 highly influential genes (ascertained from PCA leverage and residue plots) were regressed against the observed phenotypes, 68 genes were correlated (squared cosines of > 0.5) with ethambutol, isoniazid, rifampicin and streptomycin tolerance (**Figure 5.5E**). These include known markers of drug resistance (*inhA, embB, rpoB, rpsL, gidB, aphD,* and *katG*), efflux transporters (*mmpS5*, *mmpL5*, *BCG_1316c*, *BCG_2699c* and *BCG_2700c*), enzymes involved in xenobiotic biodegradation (*BCG_0610c, BCG_0944* and *BCG_1454c*), and CYPs (*cyp51*, *cyp125*, *cyp128*, *cyp132* and *cyp135A1*). Multivariate regression by redundancy analysis (RDA) showed that *cyp128* and *cyp135A1* were strongly correlated with both increased basal ROS production and 3HB metabolism (**Figure 5.5E**). Thus, by consensus, CYPs, particularly *cyp128* and *cyp135A1,* play a role in ketone body degradation and antibiotic tolerance in persistence.
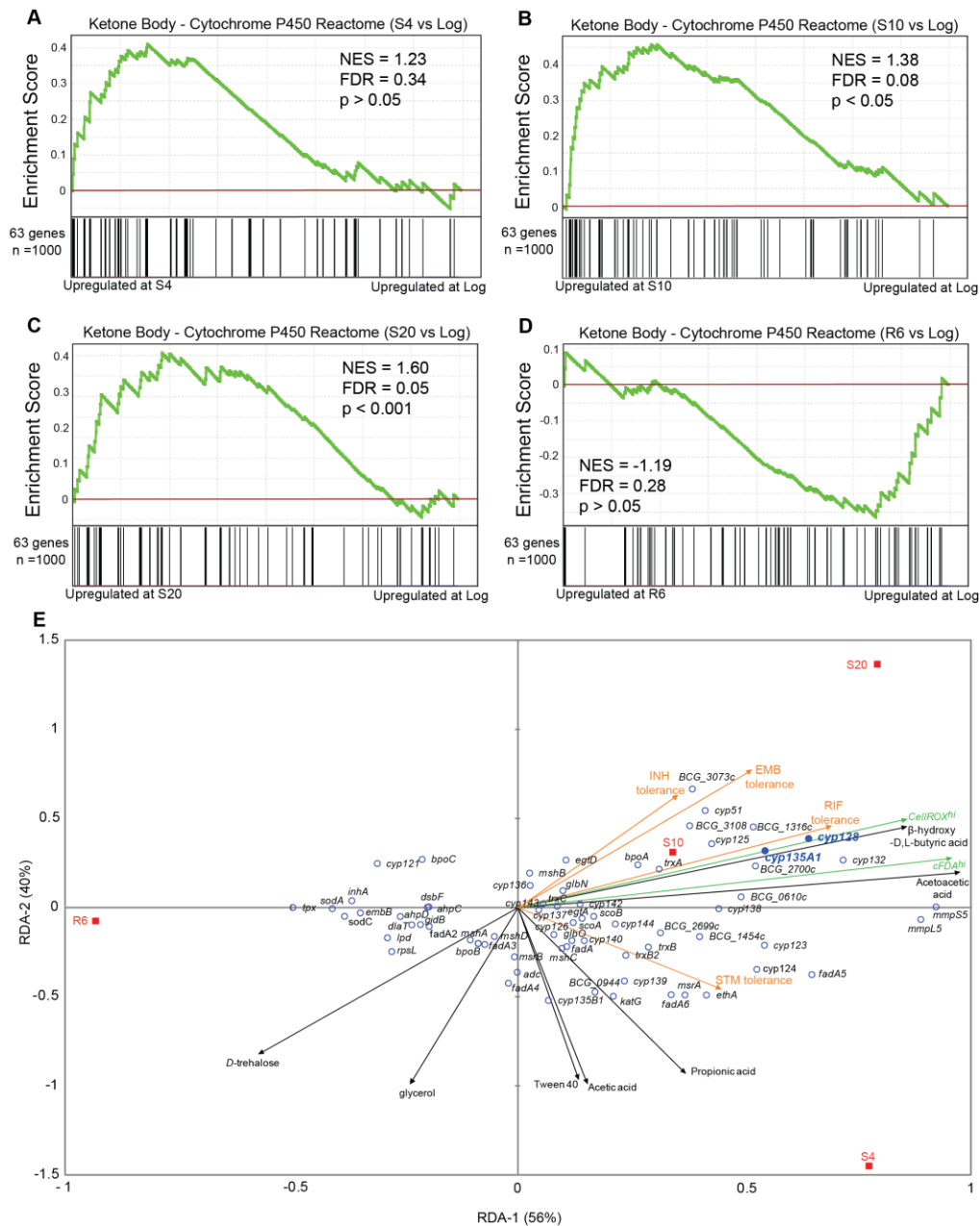
Figure 5.5. Ketone bodies utilization is associated with ROS production and CYP up-regulation. (**A-D**) Gene Set Enrichment Analysis (GSEA) determining the correlation between ketone body metabolism and gene expression change upon starvation. The GSEA plot indicates the degree to which members of the predicted ketone body-CYP reactome are over-represented at the extreme left, up-regulated at S4 (**A**), S10 (**B**), S20 (**C**) or R6 (**D**), or on the extreme right (up-regulated at Log) of the entire gene set. Solid bars represent bound genes. NES: normalized enrichment score. FDR: false discovery rate. (**E**) Redundancy analysis (RDA) correlation triplot showing the relationships among population characteristics (CellROX[hi], CFDA[hi] : size of subpopulation with greater than basal steady-state ROS production or esterase activity; green arrows), tolerance to INH, RIF, EMB and STM (orange arrows), metabolite utilization for *D*-trehalose, glycerol, Tween 40, acetic acid, propionic acid, AcAc and 3HB (black arrows) and gene expression (blue circles). *cyp135A1* and *cyp128* (●) are highlighted through consensus to play a role in ROS generation (discussed in the text). Axes represent the distribution of samples (S4, S10, S20 and R6). 96% of total variance is explained by canonical axes RDA-1 (56%) and RDA-2 (40%).

5.4.7. Multivariate regression correlates antibiotic exposure, ROS production, and starvation

To explore the relationship between antibiotic tolerance, ROS production and starvation, we extended the RDA mentioned earlier. In RDA correlation triplots, the cosine of the angle between two vectors is approximately equal to the correlation between the corresponding variables [65]. The cosine between the vectors representing INH, EMB, RIF and STM tolerance in starvation and starvation-induced CellROX^{hi} were 0.89, 0.92, >0.99, and 0.71, respectively, suggesting that RIF is more likely than STM to induce ROS, especially in starved bacilli (**Figure 5.5E**). The predictions were assessed by measuring steady-state levels of ROS in antibiotic treated Log cells, which revealed that treatment with RIF, INH, and EMB, but not STM, induced cells to generate higher than basal levels of ROS in a dose-dependent manner (**Figure 5.6A**). Interestingly, only RIF induced ROS generation above basal levels in S20 BCG (**Figure 5.6B**). S4 and S10 cells behaved in a similar manner (**Supplementary Figures 5.6A-D**). These findings are of biological significance as they parallel previous observations that antibiotic susceptibility is linked to oxidative stress [20, 66]. Indeed, 24 of the 68 genes selected for RDA by parsimony are involved in antioxidant defense, including thioredoxins, superoxide dismutases and enzymes involved in mycothiol and ergothioneine biosynthesis (**Figure 5.6C**). However, these genes explain only ~40% of the observed variance in the CellROX^{hi} phenotype. Therefore, differential responses in ROS production upon antibiotic treatments could not be explained solely by the regulation of these antioxidant genes, indicating the involvement of other metabolic factors in the generation of ROS under antibiotic stress in starved mycobacteria.
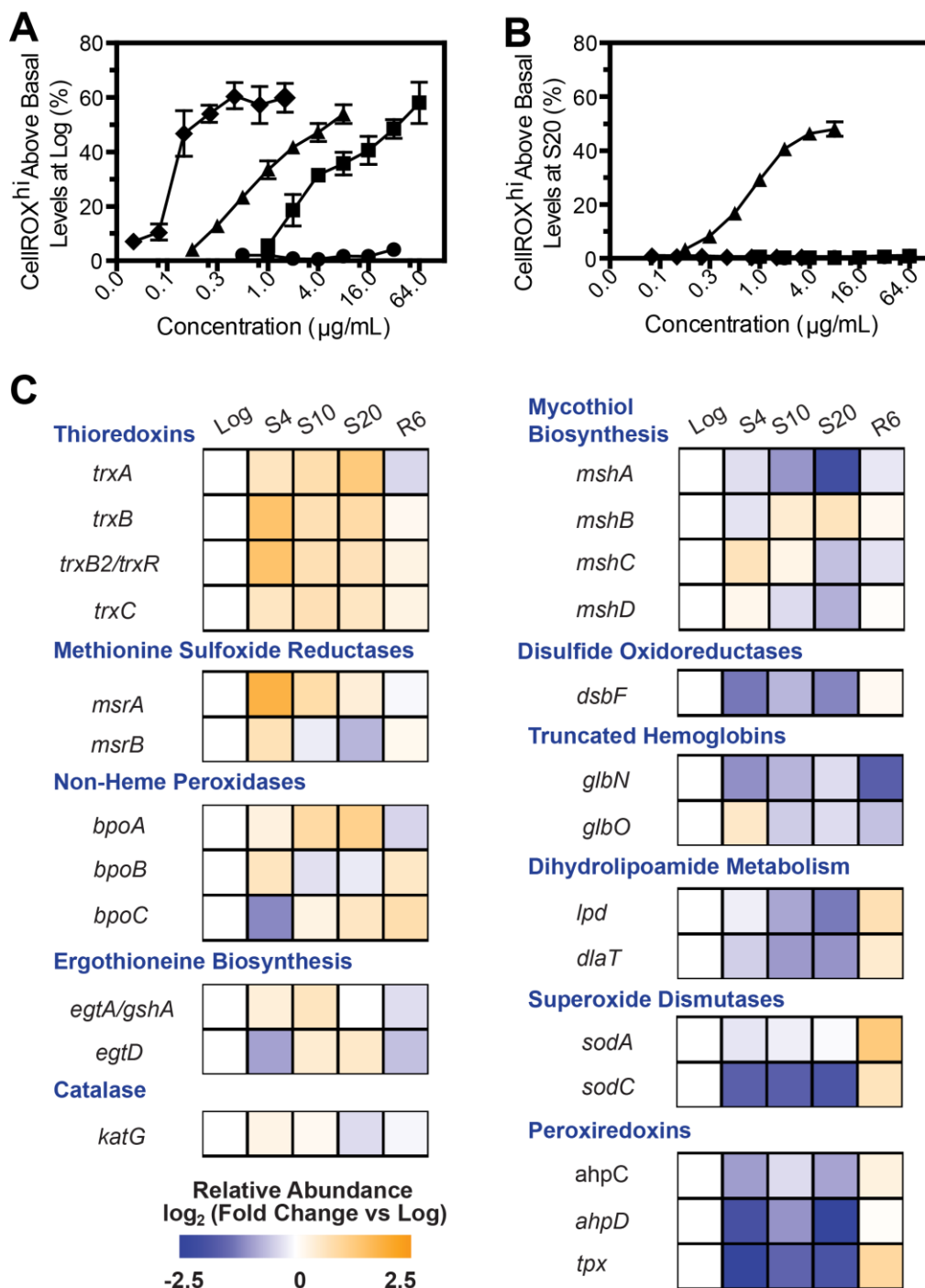
Figure 5.6. ROS production under antibiotic stress. (**A and B**) Percentage increments in cells with CellROX^{hi} phenotype after 48h of exposure to INH (◆), EMB (■), RIF (▲) or STM (●) at the indicated concentrations in Log (A) and S20 (B) cultures (n ≥ 8). (**C**) Transcript levels of genes involved in antioxidant defense across the starvation time course. Gene expression heat maps of antioxidant defense genes featured by parsimonious RDA (**Figure 5.5E**). Similar to **Figure 5.4**, colors of blocks represent extent of transcriptional regulation.

5.4.8. Biochemical and genetic validation of the CYP-mediated ketone body metabolism model of mycobacterial persistence

With only ~40% of the starvation-induced phenotype attributable to oxidative stress response pathways we initiated the confirmatory phase of our studies (**Figure 5.1A**) by testing the hypothesis that altered metabolic processes contribute to increases in basal ROS generation. RDA highlighted a strong correlation between 3HB metabolism and the number of cells with increased basal ROS production (CellROX^{hi}) (**Figure 5.5E**), which was validated by starvation-induced increases in 3HB levels (Fig. S1O). Since a ketotic state is also characterized by ketoacidosis, we measured the intracellular pH of BCG before, during and after starvation using flow cytometry with SNARF-5F. Starved cells in the persistent phase had a decrease in intracellular pH compared with cells in nutrient replete media (6.9 ± 0.1 vs 7.5 ± 0.1, **Figure 5.6A**). As a corollary, we mined the data from phenotype arrays for the sensitivity of Log, S4, S10, S20 and R6 cells to low pH. As expected, starved cultures were more resilient at pH 5.0 than Log and R6 cells (see **5.5.1. Extended results** and **Supplementary Figures 5.7A-C**). Thus, despite a >2.5-fold increase in intracellular 3HB levels, persistent BCG were better able to handle acidic pH stress.

The CYP-mediated ketone body metabolism model also entails breakdown of acetol to formaldehyde, a toxic alkylating agent, with the formaldehyde converted to formate and subsequently shunted into THF-mediated one carbon metabolic reactions, as discussed earlier. Consistent with this model, we observed up-regulation of several NAD^{+}-linked dehydrogenases that could oxidize formaldehyde to formate (**Figure 5.4**), as well as a coordinated up-regulation in the enzymes involved in B$_{12}$ biosynthesis for one-carbon (C$_{1}$) metabolism (**Supplementary Figure 5.4B**). This detoxification pathway would

render starved bacilli less sensitive to killing by formaldehyde. Indeed, S20 BCG were significantly more resistant to formaldehyde-induced cytotoxicity than Log BCG (**Figure 5.7B**).

With CYP activity as a well-established source of ROS [59, 62], we assessed the role of CYP up-regulation in ketone body metabolism and ROS generation. We could not use the azole-based CYP inhibitors clotrimazole, econazole or miconazole for these studies since the drugs elicited modest killing (<10%) at relevant doses (see **Supplementary Figures 5.7D-F**). However, exposure to millimolar concentrations of H_2O_2 proved to be bactericidal to S10 and S20, but not to Log or R6 BCG (**Figure 5.7C**). Hence, the cross-protection from acidification and formaldehyde did not extend to the biocidal activities of H_2O_2. Seemingly paradoxical, this idiosyncratic behavior can be explained as H_2O_2-facilitated killing through CYP inactivation [67]. While H_2O_2 specifically can be used as an electron donor to drive CYP activity by generating Cpd 0 directly from the ferric enzyme through the peroxidase shunt (**Figure 5.4**), the CYP heme group and catalytic cysteine residues are exquisitely susceptible to H_2O_2 inactivation by Fe-mediated reduction to generate locally damaging hydroxyl radicals or ferryl-oxo species (i.e., Fenton chemistry) [58]. This is especially plausible as all the preconditions for Haber-Weiss or Bray-Gorin catalysis were met within the confines of the reactome, including an acidic pH, availability of ferric and ferrous ions, steady-state generation of superoxide anion radical ($O_2^{\bullet-}$) from CYP activity, except one – the presence of excess H_2O_2 [44, 68]. Of relevance here, genes coding for H_2O_2-producing enzymes were down-regulated in starved bacilli (**Supplementary Figure 5.7G**). Furthermore, H_2O_2 treatment significantly lowered steady-state ROS levels in starved but not Log or R6 BCG (**Figure 5.7D**). More importantly, we found that 3HB accumulates within 4 h of H_2O_2

exposure, reaching over 600% of its initial levels 48 h post-treatment in S20 but not Log or R6 cells (**Figure 5.7E**). Correspondingly, greater killing was observed at 48 h than at 4 h following H_2O_2 treatment (**Supplementary Figure 5.7H**). These results directly link H_2O_2 cytotoxicity to the inhibition ketone body degradation through CYP inactivation.

Finally, to assess the relevance of our findings to *Mtb*, we quantified the expression of *cyp128, cyp135A1, relA, and fdxA* in Log, S4, S20 and R6 *Mtb* by qPCR. Similar to BCG, *cyp128* and *cyp135A1* expression levels were elevated during starvation compared with Log cells (**Figure 5.7G**). Elevated transcription of *relA* and *fdxA* were also confirmed in *Mtb* (**Supplementary Figure 5.7I**). Considering these findings, we propose a novel mechanism for the survival of persistent mycobacteria under nutrient deprivation, which implicates the accumulation and detoxification of ketone bodies by CYPs as a consequence of TAG utilization (summarized in **Figure 5.7H**). Signatures of these metabolic changes can be observed in the gene expression, metabolism and basal ROS production of these starved mycobacteria.

5.5. Discussion

The concept of persistence was presaged over 70 years ago when Bigger noted that prior exposure to physiological stresses in general (starvation, temperature, cytotoxic agents) "predestined" a larger bacterial subpopulation for enhanced survival [69]. For persisters, their antibiotic resistance profile is underscored by adaptive metabolic shifts required for stress survival [13, 30]. Here, we demonstrated that nutrient deprivation by culture in PBS predisposed mycobacterial populations to persister formation. And unlike previous studies
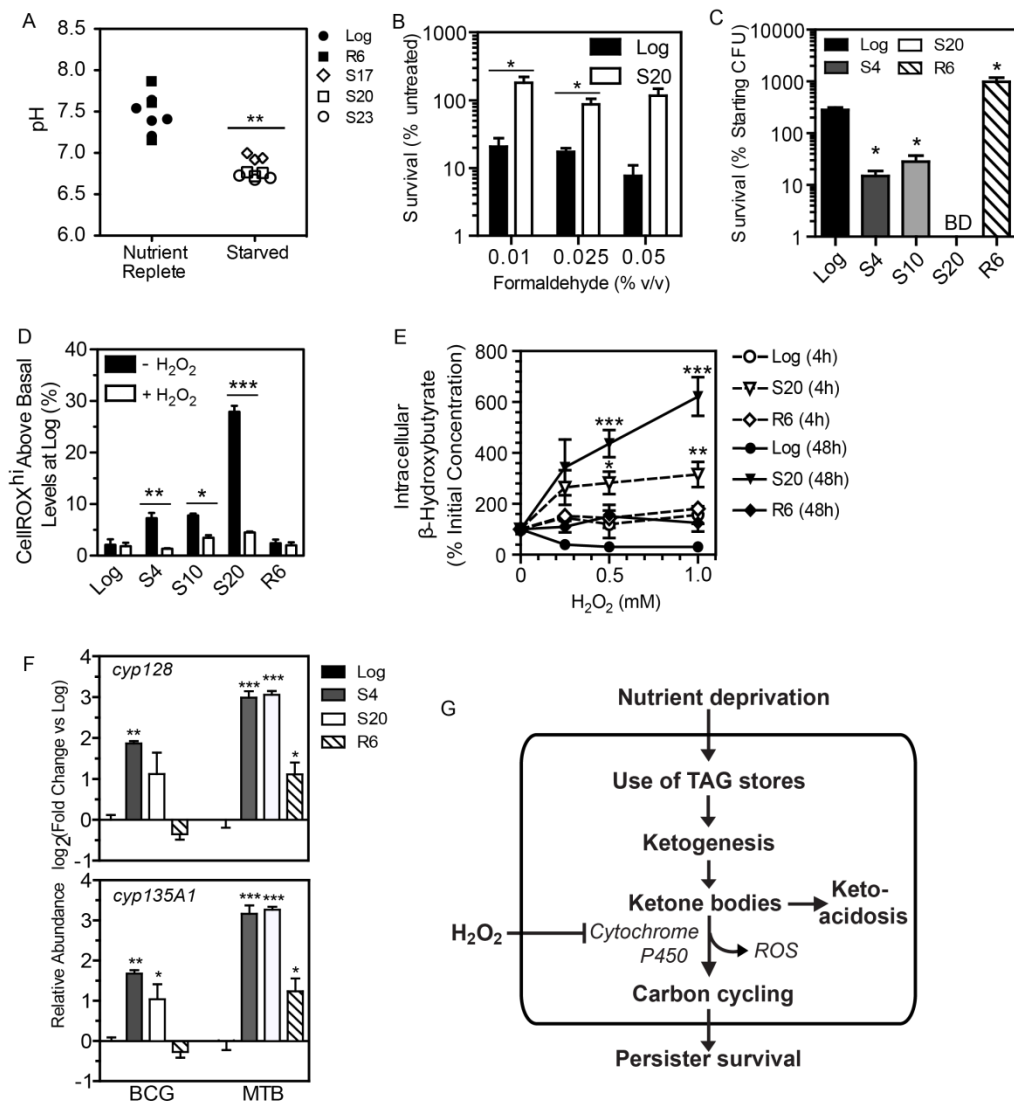
Figure 5.7. CYPs contributes to ROS production and play an essential role in ketone body metabolism during nutrient deprivation. (**A**) pH of Log, R6, S17, S20 and S23 cells determined by SNARF-5F flow cytometry. Each symbol denotes the median pH of a sample of 50,000 cells (**, unpaired, two-tailed t-test with Welch's correction). (**B**) Survival of Log and S20 cells to 4h formaldehyde exposures at indicated concentrations (*, unpaired, two-tailed t-test). (**C**) Survival of Log, S4, S10, S20 and R6 cells to 48h of H_2O_2 (0.5 mM) exposure. BD: below assay detection limit of 1000 CFU/mL (n = 6; one-way ANOVA with Bonferroni post-test). (**D**) Percentage of Log, S4, S10, S20 and R6 cells with elevated steady-state ROS production (CellROX^{hi}) after 48h of either 0.5 mM H_2O_2 or mock (PBS) exposure (n = 8; *,**, ***, unpaired, two-tailed t-test between H_2O_2 or mock treatment groups with Welch's correction). (**E**) Intracellular 3HB levels in Log, S20 and R6 cells after 4h and 48h of H_2O_2 exposures at the indicated concentrations. n ≥ 3; metabolic state of cells accounts for 32.3% and 54.4% of the total variance at 4h and 48h respectively (*,**, ***, two-way ANOVA with Bonferroni post-tests comparing cell state and H_2O_2 dose interactions). (**F**) Regulation of *cyp128* and *cyp135A1* in *Mtb* and BCG at Log, S4, S20 and R6 as determined by qPCR (n = 3; *,**, ***, one-way ANOVA with Dunnett's test (sample means vs Log)).(**G**) Model of starvation induced ketone body metabolism in mycobacterial persisters. Upon nutrient deprivation, mycobacteria increase

the utilization of their triacylglycerol stores. A portion of the acetyl-CoA from β-oxidation of fatty acids is converted to AcAc and 3HB (ketogenesis). These products lower the intracellular pH (acidosis). Over time, these ketone bodies are oxidized by CYPs and used to generate formate required for one-carbon cycling. The loss of CYP activity through H_2O_2 exposures causes persisters, otherwise able to survive for prolonged periods with little or no nutrients even in the presence of antibiotics, to lose viability.

in mycobacteria, these populations are generated according to a biphasic survival profile that follows well described double-exponential kinetics observed in other bacteria [70], and a starvation-induced antibiotic tolerance. However, these persisters are not transcriptionally or metabolically quiescent. Rather, we found signatures for a stringent response-initiated global transcriptional reprogramming associated with an altered metabolic state (**Figure 5.3**). By linking published disparate observations on lipid, lactate, 1,2-propandiol, acetol, methylglyoxal, formate and carbon dioxide metabolism [24, 56, 57, 63, 71-75] with our analysis of persister physiology, biochemistry and gene expression, we conclude that a central feature of the response of mycobacteria to starvation involves a reliance on CYPs for protection from cytotoxic ketosis. Correspondingly, CYP disruption by H_2O_2-based inactivation renders the cells susceptible to the cytotoxic effects of nutrient deprivation (**Figure 5.7**).

The transcriptional and metabolic remodeling underlying this starvation-induced ketosis provide a context for understanding several unique features of the mycobacterial persistent state revealed by our data, including drug tolerance, the metabolic phenotype, and the interplay between nutrient deprivation and endogeneous acidosis. Several mechanisms have been proposed to explain the antibiotic tolerance of persisters [76]. First, the complex mycolic acid-containing cell wall of mycobacteria serves as an intrinsic pharmacokinetic barrier to antibiotics under normal growth conditions [77], with the structure and permeability of the cell wall changing significantly

under starvation conditions [16, 17, 78]. We observed a similar reduction in ROS-sensitive dye uptake in starved compared to exponentially growing mycobacteria, which we accounted for in the flow cytometric analyses (**Supplementary Figure 5.2**) (McBee et al, unpublished data). A second mechanism of drug tolerance involves ATP-binding cassette (ABC) transporters and other drug efflux pumps [79, 80], several of which we observed to be up-regulated during starvation (**Supplementary Figures 5.3C and Figure 5.5E**). RIF tolerance is facilitated through BCG_1316c efflux pumps (*Mtb* Rv1258c homolog) [81], which was significantly upregulated during starvation and highlighted by redundancy analysis (**Figure 5.5E**). Incidentally, clotrimazole, miconazole and econazole, which pentacoordinate heme and block the CYP catalytic cycle [82, 83], have unexceptional killing activities on mycobacterial persisters (**Supplementary Figures 5.7D-F**) possibly due to the strong up-regulation of transmembrane transporters *mmpL5* and *mmpS5* [84]. Third, persistence may repress or down-regulate drug targets [85]. For instance, the efficacy of aminoglycosides, such as streptomycin, is limited in starved mycobacteria (**Figure 5.2C**) possibly due to reduced translational activities, evident the decrease in cellular rRNA and tRNA content (**Supplementary Figures 5.1H-K**), and the transcriptional down-regulation of genes coding for ribosomal proteins such as *rpsL* (**Supplementary Figure 5.6E**), together with a change in cell wall permeability (McBee et al, unpublished data). Furthermore, *GidB, embB, inhA* and *rpoB,* genes associated with antibiotic susceptibility [86] were highlighted in our redundancy analysis and repressed during starvation (**Figure 5.5E** and **Supplementary Figure 5.6E**). We propose a fourth mechanism, whereby exogenous stresses cause metabolic shifts that induce xenobiotic degradation pathways leading to antibiotic tolerance. Network mapping of our RNA-seq data revealed that the degradation pathways for bisphenol,

chloroalkane and chloroalkene, aminobenzoate and polycyclic aromatic hydrocarbons were enriched in starved BCG (**Supplementary Figure 5.3C**). With parallels in human drug metabolism, these degradation and detoxification pathways often involve oxidative biotransformation by CYPs [87]. Another parallel to human drug metabolism arises in the starvation-induced upregulation of mycobacterial conjugating enzymes, such as arylamine N-acetyltransferase (*nat;* 1.7-fold increase in S20 BCG, p = 1.4 x 10^{-6}), which uses INH as a substrate [88, 89]. Jointly, our results underscore the importance of bacterial pharmacodynamics and pharmacokinetics in antibiotic design and development.

Antibiotic tolerance in persisters strongly parallels the novel metabolic phenotype of increasing steady-state levels of ROS and dependence on ketone body degradation, with lipolysis and ketogenesis as metabolic signatures of the mycobacterial adaptive and persistent phases of starvation, respectively (**Figure 5.3**). These processes lead to an accumulation of AcAc and 3HB, which acidifies the cytoplasm (**Figure 5.7A**). In turn, cells progressively up-regulate CYP-mediated ketone body detoxification pathways (**Figures 5.5A-D**) that inherently generate ROS [60, 62, 90]. Our models predict *cyp135A1* and *cyp128* as vital enzymes in ketone body detoxification. Cyp128 activity is predicted, based on proximity to the *stf3* gene, to be the terminal oxidase of dihydromenaquinone-9 synthesis – the essential mycobacterial respiratory chain quinol electron carrier [91]. The precise reaction(s) catalyzed by Cyp135A1 is unclear. Based on Hausinger's (2007) speculations, Cyp135A1 would catalyze the conversion of acetone to acetol. The lack of significant up-regulation in AcAc decarboxylase (*BCG_3583*) expression (**Figure 5.4**) led us to consider the possibility that 3HB could be oxidized by CYP135A1 instead. This would result either in the formation of

1,2-propandiol or acetol through a second oxidation event (**Supplementary Figure 5.4A**). This alternate reaction pathway fits previous observations of lactate, 1,2-propandiol and acetol metabolism in mycobacteria [57, 71, 75]. As mycothiol-dependent formaldehyde dehydrogenase *mscR*/*adhE2* is down-regulated during starvation (**Figure 5.4**), subsequent alkylating damage from the accumulation of formaldehyde is likely mitigated by its conversion to formate, not *S*-formylmycothiol. However, the fate of formate is uncertain. Due to down-regulation in *fdhF*, it is unlikely that formate is oxidized to carbon dioxide, resulting in the generation of NADH [92]. Hence, we propose a model in which formate is used for acetogenesis through carbon fixation. Dissimilar to the Eastern Woods-Ljungdhal pathway [93], formate from ketone degradation (rather than carbon dioxide reduction) is used by FolD and MetH to methylate THF then fix carbon monoxide to generate acetyl-coA (**Supplementary Figure 5.4B**). Carbon fixation was previously demonstrated to occur in mycobacteria in glycerol-limited chemostat cultures, though Woods-Ljungdhal acetogenesis was not implicated due to the lack of an annotated formate-THF ligase [72]. BioCys' PathoLogic predicts that FolD possesses this activity [94]. Additionally, we observed coordinated up-regulation in the genes required for cobalamin (coenzyme B_{12}) synthesis, which is consistent with the generation of corrinoid proteins required for these methyl group transfers [93]. Though 5,10-methylenefolate could be siphoned off for amino acid and pyrimidine biosynthesis, we also noticed an up-regulation in acetyl-coA synthetase (*acs*) transcription during starvation, which is suggestive of acetate buildup from acetogenesis (**Supplementary Figure 5.4B**). Though speculative, formate could be a potential source of C1 for methylated THF in persisters. This warrants further investigation as it would not only expand the known metabolic plasticity of mycobacteria, but

also provide further insights into the underlying metabolic transitions required for the long-term survival of *Mtb* in hypoxic and nutrient-limiting granulomas.

One might conclude, given the ability of mycobacteria to fix carbon dioxide under nutrient limited conditions [72], that pathogens of the *Mtb* complex could potentially persist indefinitely in granulomas. By reevaluating the interplay between acid stress and starvation our results highlight an unassuming limiting factor to survival – the availability of magnesium cations (Mg^{2+}). *In vivo*, the macrophage compartment in which *Mtb* resides early in infection, ranges from pH 4.5-6.2, depending on the activation state of the macrophage [95]. *In vitro*, *Mtb* and BCG grow at pH 5.8-6.5 with growth retardation observed at ~pH 5.0 and below [96, 97]. Against a decreasing external pH, intracellular pH is maintained between pH 6.1 and 7.2 [52]. Survival, however, is contingent on population density, carbon sources in the test medium and the presence of Mg^{2+} ions [95, 98]. For instance, there is evidence that short starvation in phosphate-citrate buffer improves survival of mycobacteria at low pH [99, 100]. Conversely, when incubated in buffered media at the same pH with utilizable carbon sources, such as Tween 80, lactate or acetate, survival is compromised [95, 99, 101]. Our results confirmed that starved mycobacteria survive better during acidic pH stress (**Supplementary Figures 5.7A-C**). Starved cells accumulate acidic AcAc and 3HB, while sustaining their intracellular pH above 6.0 (**Figure 5.7A**); below which is lethal for BCG and SMG [52]. Reminiscent of the *phoP* independent metabolic remodeling required for acid stress survival [102], we noticed a down-regulation of *phoP* during starvation (**Figure 5.4**), suggesting that there are common features to the adaptive responses to starvation and acid stress. One commonality is the need for Mg^{2+}. Several actinomycetes have been shown to retain Mg^{2+} during starvation [38]. Furthermore, Mg^{2+}, but not Ca^{2+},

Mn^{2+} or Zn^{2+}, supplementation is required for growth of *Mtb* at pH 6 [98]. And veritably, we observed that more mycobacteria survived prolonged starvation in Mg^{2+}-containing DPBS than in PBS (**5.5.1. Extended results** and **Supplementary Figures 5.1B and C**). Altogether, these results support a role for Mg^{2+} ions in intracellular pH homeostasis and thus, starvation survival.

Just as Achilles possessed high-invulnerability to physical harm, persistent mycobacteria are extremely tolerant to physiological stressors, including prolonged starvation, antibiotics, acidic pH and formaldehyde. Yet, a deadly weakness lies in its formidable defenses – ketogenesis. The heavy reliance on cytochrome P450 for ketone body metabolism, while protective against ketosis, predisposes the cells to lethal oxidative challenges. These enzymes are widespread across bacterial pathogens, with *E. coli* being a notable exception, and utilized in a wide variety of biosynthetic and degradation pathways [58, 103]. Their essentiality to persister phenotype switching [104, 105], however, has, until now, been unexplored. Thus, the challenge remains to convert these new insights into drug development approaches and practical clinical applications.

5.6. Supplementary material

5.6.1. Extended results

5.6.1.1. Cannibalism is not a major source of nutrients for starved mycobacteria

We evaluated the possibility that starved BCG found sustenance from the non-viable cells (~95% of initial CFU) in culture. After 20-days, cultures were split, pelleted and either washed and reconstituted to its prior volume in fresh PBS or re-suspended in its own (native) supernatant then starved for another 20 days. The ratios of CFU recovered from cultures in fresh PBS against

those in their native supernatant did not differ (**Supplementary Figure 5.1A**).

Conceivably, cannibalism would cause a further reduction in the population of washed cultures, hence, no cannibalization was observed. This suggests either metabolic quiescence or cellular autophagy as a means of subsistence during extended starvation.

5.6.1.2. Divalent cations support survival during the adaptive phase of starvation

Two forms of PBS are commonly used in research, PBS and Dulbecco's PBS (DPBS; contains Mg^{2+} and Ca^{2+}). In addition to starving BCG in PBS, DPBS was also evaluated as a carbon and nutrient deprivation media. When starved in Mg^{2+}- and Ca^{2+}-ion containing DPBS significantly more BCG survived when compared to PBS (**Supplementary Figure 5.1B**). This phenomenon was reproducible in SMG whereby >99% of the inoculum was recovered under DPBS starvation compared with 30% in PBS (**Supplementary Figure 5.1C**). This survival profile of SMG under nutrient deprivation in PBS is similar to prior characterizations performed by Stallings et al [106].

5.6.1.3. Starvation adaptation alters antibiotic killing kinetics

When Log bacilli were treated with a minimal bactericidal dose of INH (MBC_{99}), 96% of killing occurs within 24h. By 48h, another 75% of surviving bacteria died. In contrast, at S10 and S20, only 39% of the viable cells were killed within 24 h followed by an additional 17% by 48 h. Cells at S4 displayed an intermediate phenotype. Similar trends were observed with RIF, STM and EMB (**Supplementary Figure 5.1D-G**). These findings indicate that the drug tolerant phenotype is linked to dynamic changes during adaption to starvation rather than as a response to the antibiotic exposure.

5.6.1.4. Acid tolerance in BCG persisters

The detection of acidosis in BCG persisters led us to investigate if starved bacilli are better adapted to survive lower pH environments. We retrieved data regarding pH sensitivity from the GENIII microplates metabolic phenotype screen (see **Discussion** in main text and **Figure 5.3**). Whereas no changes in cellular viability (recoverable CFU) were detected in pH 6.0 media, starved BCG were more tolerant to pH 5.0 than Log and R6 (**Supplementary Figure 5.7A**). This is also reflected in the optical densities and tetrazolium reduction capabilities (used as an indicator of cellular respiration [31, 32]) (**Supplementary Figure 5.7B and C**).

5.6.1.5. Features of interest in the starvation transcriptome

Our rRNA and tRNA depletion strategy greatly enriched for multi-gene operon cluster mRNAs for RNAseq. However, some non-coding RNA (ncRNA) still remained (**Supplementary Table S2**). Nonetheless, this provided a glimpse of how ncRNA contributes to the overall molecular phenotype. One of the most prominent and quantifiable ncRNA was the ribozyme *rnpB* (**Supplementary Figure 5.3A**)*.* Together with *rnpA,* they form the RNaseP complex, which catalyzes the removal of the 5'-leader sequence from pre-tRNAs in the presence of divalent cations (Mg^{2+} or Ca^{2+}) to produce the mature tRNAs [107, 108]. Continued *rnpB* transcription supports the observation of tRNA maintenance during early starvation (**Supplementary Figure 5.1K**), and provides evidence against translational quiescence in starved mycobacteria. Another prominent ncRNA found throughout the starvation time course is tmRNA, coded by *ssr* (**Supplementary Figure 5.3A**). tmRNA is involved in trans-translation by releasing mRNA from stalled ribosomes, thus ensuring protein fidelity [109]. Its maintenance during starvation, once again, suggests active protein translation in persisters.

In bacteria, RNA synthesis is initiated by sigma factors. Specialized sigma factors, such as SigF, are induced under stress, bind their target promoters and induce the transcription of stress related genes. During starvation the SigF regulon became significantly enriched at S10 despite peak *sigF* up-regulation at S4 (**Figure 4** and **Supplementary Figures 5.5E-H**). A plausible explanation for this phenomenon may be the spikes in *rsbW-usfY* transcription at S4 (**Supplementary Figure 5.3A**). *rsbW* codes for an anti-sigma factor protein that binds and negatively regulates SigF activity [110]. This suggests a layer of regulation governing the temporal expression of gene transcripts.

In addition to the SigF regulon, CYPs, antioxidant defense and antibiotic sensitivity genes, PCA also identified several genes involved with the TCA cycle or members of the Lsr2 regulon to co-vary with metabolic utilization phenotypes at various time points. These include *sucD*, *fum*, *acn*, *aceA* and *glcB* (TCA cycle), *sigF*, *canA*, *cyp128*, *BCG_2905* (SigF regulon), and *moaD1*, *PPE16*, *PE9*, *celA2b* (Lsr2 regulon) (**Supplementary Figure 5.5M**). GSEA further showed that in comparison to Log cultures, TCA cycle genes, generally, are significantly de-enriched at S4 but were indifferently expressed at R6 (**Supplementary Figures 5.5A-D**). One notable exception is *aceA* which is up-regulated at S4. In contrast, Lsr2 regulon genes are significantly enriched at S20 but became equally de-enriched at R6. It is important to consider that Lsr2 is a repressor of gene expression, thus, its targets become up-regulated only upon its suppression and down-regulated upon its restoration at R6 (**Figure 4** and **Supplementary Figures 5.I-L**). We also observed contributions from whiB-family transcription factors, cAMP signaling, PhoPR, TrcPR and KdpDE two-component signaling, which could facilitate global transcriptional reprogramming (**Figure 5.4**). Overall, RNA-seq provided

evidence for the global reprogramming of the transcriptome affected through active translation processes which enabled metabolic remodeling during prolonged starvation.

5.6.2. Supplementary figures

(Next page)

A

B

C

D Isoniazid (0.2 µg/mL)

E Rifampicin (0.25 µg/mL)

F Streptomycin (1.0 µg/mL)

G Ethambutol (8.0 µg/mL)

H 23S rRNA

I 16S rRNA

J 5S rRNA

K tRNA

L

M

N

O

(Legend – next page)
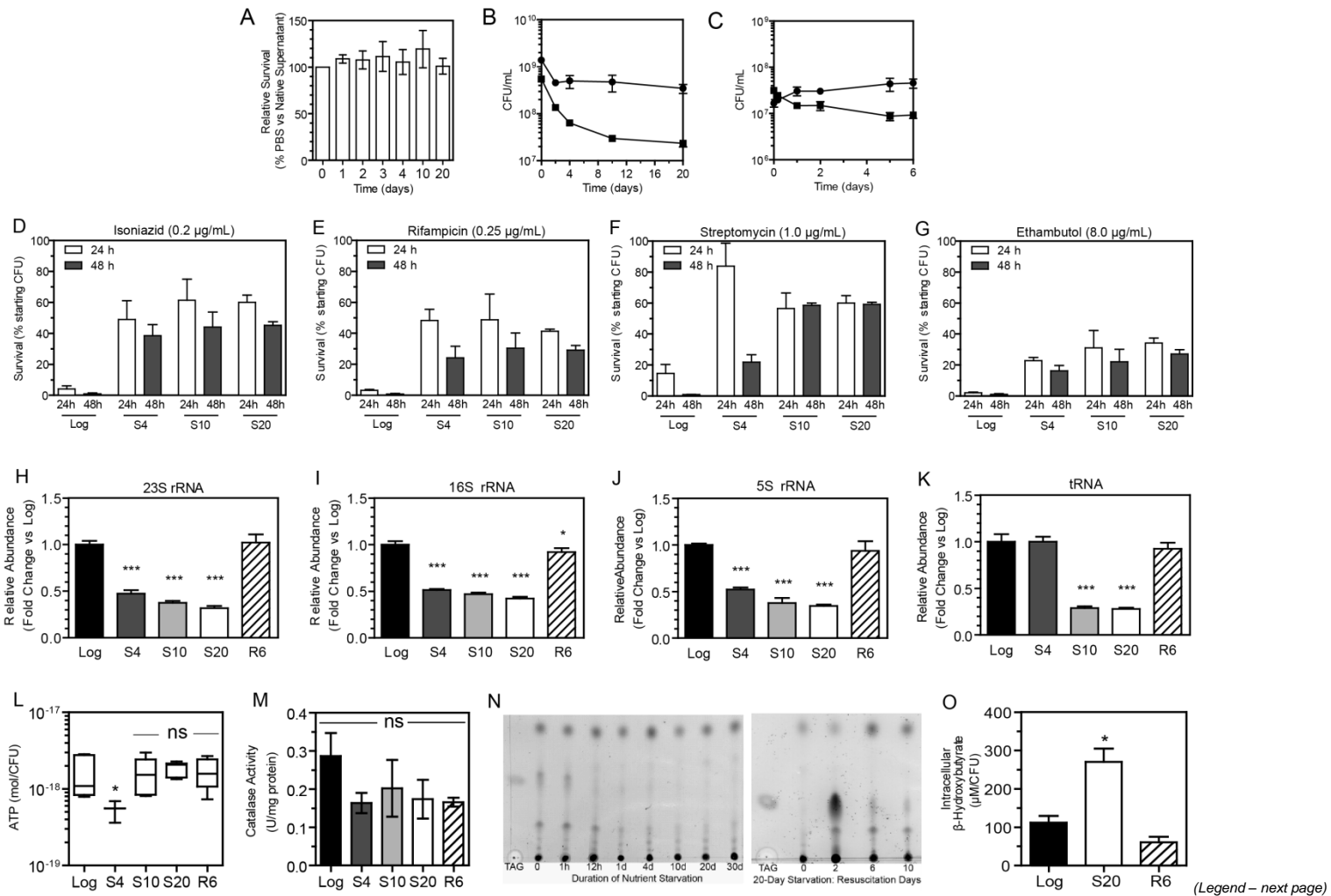
Supplementary Figure 5.1. Physiological and molecular features of mycobacterial starvation and persistence.

(**A**) CFU ratios of BCG recovered from 20-day cultures re-starved for 20 days in fresh PBS versus native supernatant. N = 6; p > 0.05; one-way ANOVA with Tukey's HSD.

(**B and C**) Survival curves of BCG (**B**) and SMG (**C**) starved in Mg^{2+} and Ca^{2+} ions containing DPBS (●) or PBS (■). For BCG, PBS data also presented in Figure 1; n = 6; p < 0.01; maximum cumulative D = 0.52 (two-sample Kolmogorov-Smirnov test). For SMG, n = 6; p < 0.001; maximum cumulative D = 0.77; two-sample Kolmogorov-Smirnov test.

(**D-G**) Starvation enriched for the proportion of persisters in BCG populations. Survival of BCG cultures at Log, S4, S10 and S20 to MBC_{99} concentrations; i.e. 99% killing after 5 days post-exposure, of INH (**D**), RIF (**E**), STM (**F**) and EMB (**G**). Save for STM at S4, antibiotic susceptible cells were mostly killed the first 24h while proportions of drug tolerant survivors at 48h became enriched over the starvation time course (from Log-S20).

(**H-K**) Intracellular levels of 23S (**H**), 16S (**I**), 5S (**J**) rRNA and tRNA (**K**) at Log, S4, S10, S20 and R6 (n = 3; * 0.01< p < 0.05, ***p < 0.001; one-way ANOVA with Dunnett's test vs Log).

(**L**) Intracellular ATP levels at Log, S4, S10, S20 and R6 (n ≥ 6; * 0.01< p < 0.05; one-way ANOVA with Dunnett's test vs Log).

(**M**) Catalase activity in Log, S4, S10, S20 and R6 cultures (n ≥ 6; p > 0.05; one-way ANOVA with Dunnett's test vs Log).

(**N**) Representative thin layer chromatograms of starved (left panel) and resuscitated (right panel) BCG at indicated time points (TAG: Triacylglycerol standard; d: days).

(**O**) Concentrations of intracellular 3HB for Log, S20 and R6 cultures (n = 6; * 0.01< p < 0.05; one-way ANOVA with Dunnett's test vs Log).
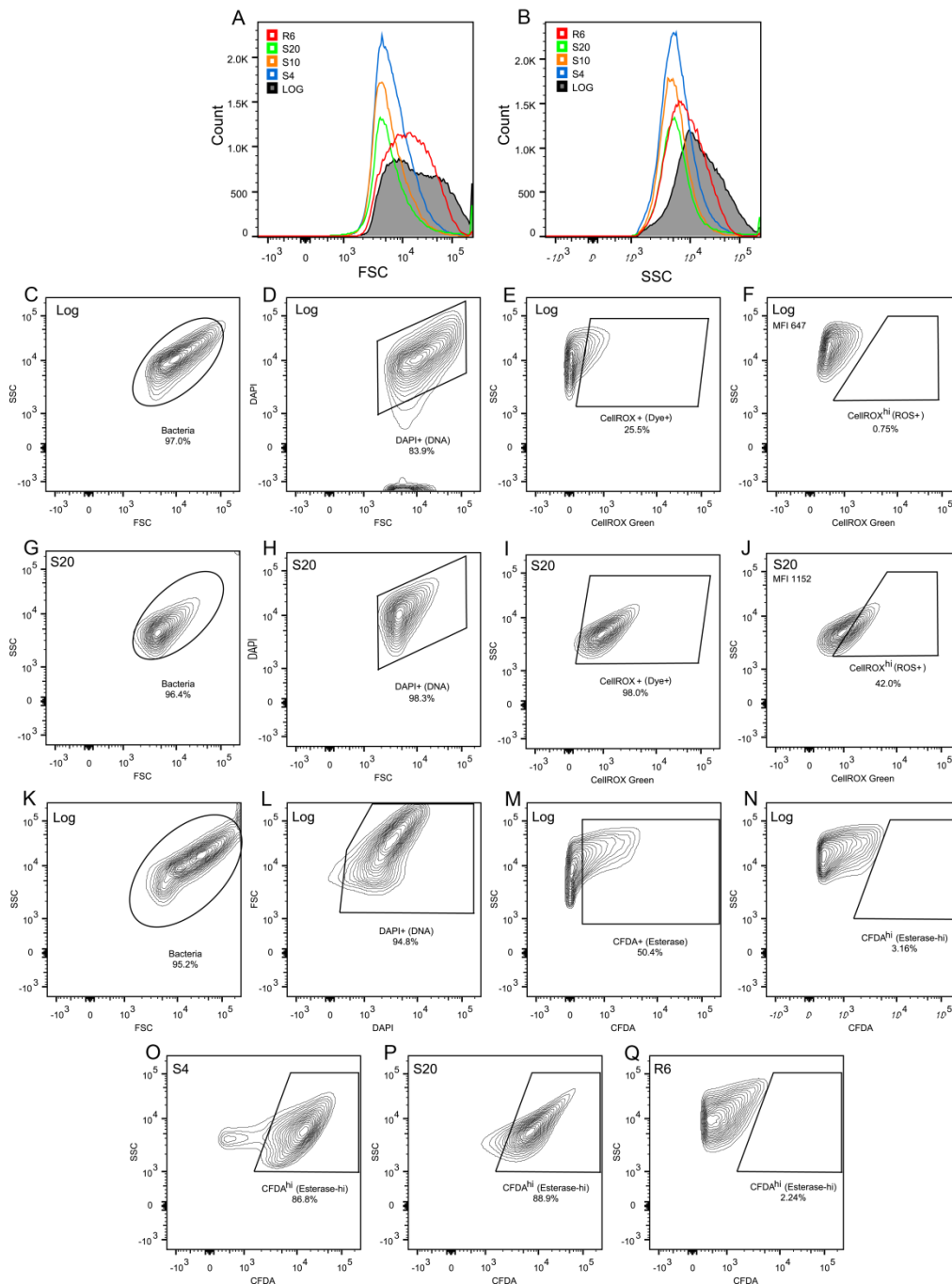
Supplementary Figure 5.2. Representative histograms and contour plots from flow cytometric analysis of BCG in Figures 5.2 and 5.3.

(**A and B**) Comparison of FSC and SSC histogram profiles of BCG during starvation, resuscitation, and log growth shows starvation in PBS decreases the heterogeneity of both FSC and SSC profiles indicating smaller overall cell size and reduced cellular complexity.

(**C-J**) Contour plots of Log and S20 BCG depicting gating scheme used to identify CellROX[hi] (ROS-producing) cells. Cells in the "size" gate (**C and G**) were gated for "DNA content" (**D and H**). Those with DNA were further identified as having intracellular CellROX® (**E and I**). CellROX[hi] cells were designated as those with fluorescence greater than basal levels (gate set at ~1%, F and J).

(**K-Q**) Contour plots depicting gating scheme used to identify CFDA[hi] (elevated esterase activity) cells. Cells within the "size" gate (**K**) were gated for "DNA content" (**L**). Cells were further gated for intracellular CFDA presence or esterase activity (M) followed by a threshold set for basal activity in Log (**N**). Representative CFDA contour plots for S4 (**O**), S20 (**P**), and R6 (**Q**) depict shifts in esterase activity during starvation.
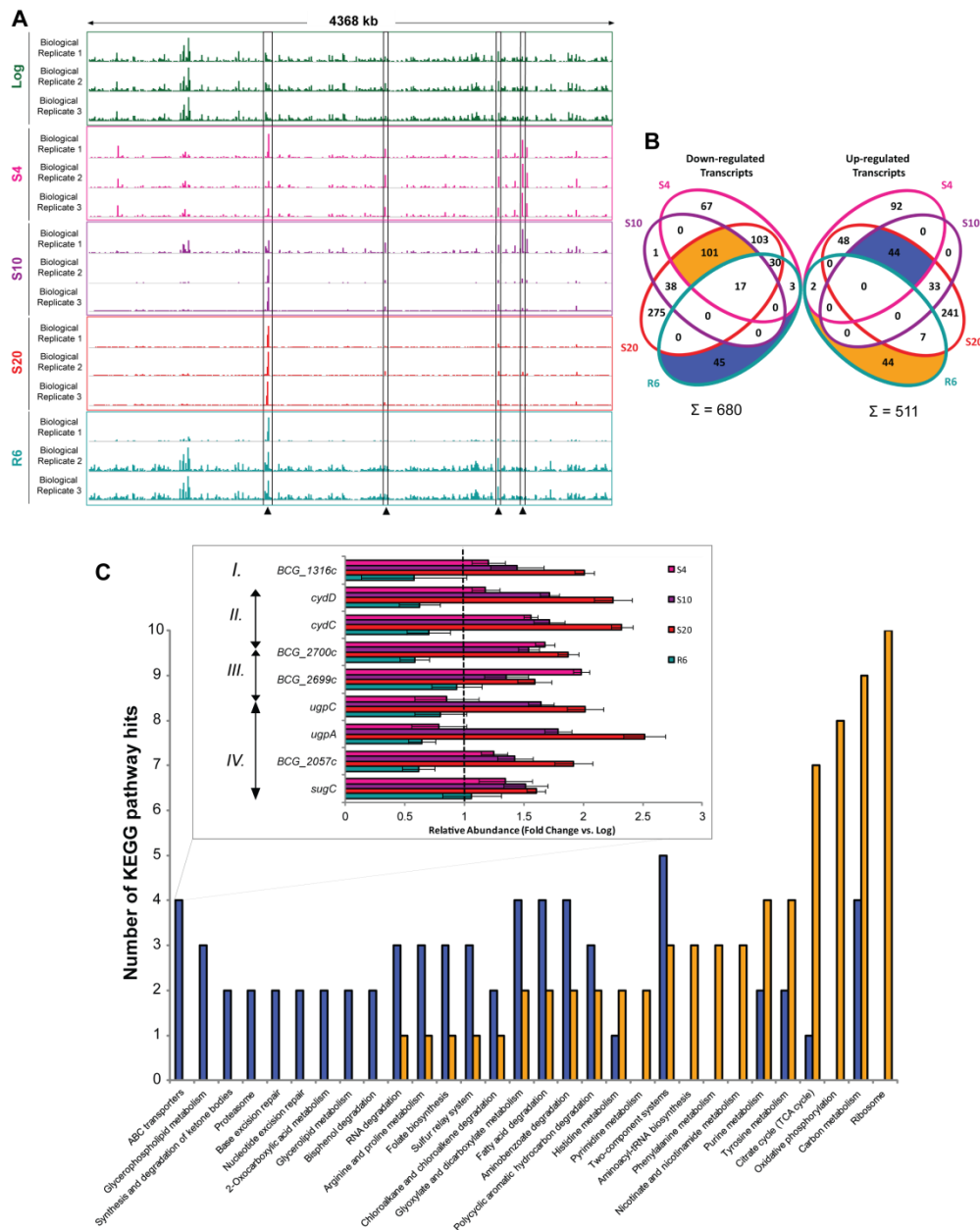
Supplementary Figure 5.3. RNA-seq and differential pathway analysis of the starvation transcriptome.

(**A**) Visualization of RNA-Seq coverage across the BCG genome. Each sequencing track represents the one biological replicate consisting of 2 appended technical replicate runs. Peaks heights indicate read coverage. Arrows (from left to right) highlight the prominent transcripts from 16S-23S-5S rRNA, *rnpB*, *ssr* and *rsbW-usfY* respectively. Time points are color coded consistently as Log (█), S4 (█), S10 (█), S20 (█) and R6 (█) throughout the figure.

(**B**) Venn diagram of significantly up- and down-regulated genes in S4, S10, S20 and R6 compared against Log cultures (>1.5 fold change; $p < 0.05$; De-seq multi-testing). A total of 680 down-regulated transcripts and 511 up-regulated transcripts fulfill these criteria. The individual and overlapping areas represent the number of uniquely or co-expressed transcripts at the indicated time points. Areas highlighted in blue (█) show genes down-regulated at

R6 but up-regulated during starvation (S4, S10 and S20) – core set 1. Conversely, areas highlighted in yellow (█) show genes up-regulated at R6 but down-regulated during starvation – core set 2. These core gene sets were subjected to pathway analysis (**C**).

(**C**) Pathway analysis showing canonical pathways (KEGG annotation) modulated by the genes whose expression differed among R6 and starved cultures (S4, S10 and S20). Specificity was determined by selecting for pathways that fit the condition: $\frac{core\ set\ 1\ hits}{core\ set\ 2\ hits} \geq 1.5$ $\boldsymbol{or}$ $\frac{core\ set\ 2\ hits}{core\ set\ 1\ hits} \geq 1.5$. Blue and yellow bars: core sets 1 and 2 respectively. *Insert*: Illustrated example of pathway enrichment with ABC transporters. 4 groups of hits were identified. *I.* multidrug-efflux transporter *BCG_1316c* (*Mtb Rv1258c* homolog). *II.* ATP-binding cassette, subfamily C (ABCC) cytochrome biosynthesis permeases cydC and cydD. *III.* Fluoroquinolones transport system permease proteins *BCG_2699c* and *BCG_2700c* (*Mtb Rv2686c* and *Rv2687c* homologs). *IV.* Multiple sugar transporters (*sugC* and *BCG_2057c*) and *sn*-glycerol 3-phosphate transporters (*ugpA* and *ugpC*). Transcripts from all aforementioned genes possess similar patterns of expression (bars graph with indicated time points).

Supplementary Figure 5.4. Proposed reaction schemes for the decomposition of 3-hydroxybutyrate (3HB) by cytochrome P450s (CYP) and the contributions ketone body degradation products to carbon cycling.

(**A**) Scheme 1: Terminal decarboxylation of 3HB generating vicinol diol (1,2-propanediol) products based on the proposed mechanisms of oxidative decarboxylation [111, 112]. 1,2-propanediol then feeds into pathway C (Figure S4B) or reaction Scheme 2.

Scheme 2: β-hydroxylation of 1,2-propanediol to acetol based on the proposed mechanism by Rude et al (2011) and observations of 1,2-propanediol and acetol metabolism in mycoabcteria [57, 113]. Acetol would be further oxidized by Bayer-Villager monooxygenases (BVMOs; see Figure 4). More detailed experimental work is needed to confirm these mechanisms.

(**B**) Proposed carbon cycling pathways in starved BCG. Formate [A], acetate [B] and 1,2-propanediol [C] are derived from the decomposition of hydroxymethylene acetate (Figure 4) and 3HB (Scheme 1). These feed into metabolic pathways mediated by *folD, acs* and 1,2-propandiol dehydrogenase respectively and lead to the anaplerotic node (See Discussion). Enzymes catalyzing each reaction are named in blue. Assignment of gene functions based on KEGG and BioCys mapping. Colors in heat maps denote up- and down-regulation of transcripts.
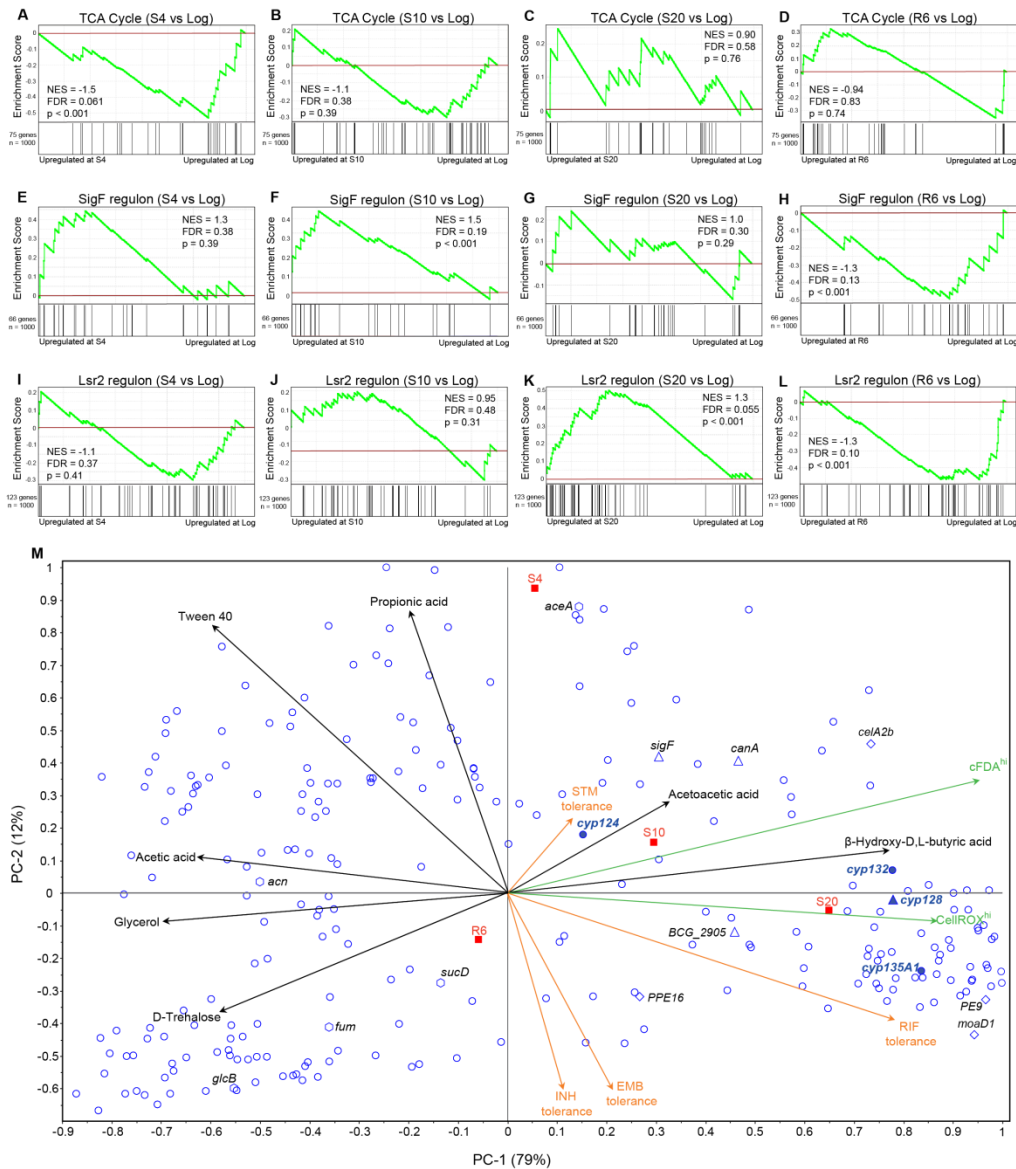
Supplementary Figure 5.5. Dynamic shifts in the expression of TCA cycle enzymes, SigF and Lsr2 regulon genes underline their covariance with observed phenotypes.

(**A-L**) GSEA determining the transcriptional enrichments of TCA cycle enzymes (**A-D**), SigF (**E-H**) and Lsr2 regulon (**I-L**) genes during starvation. GSEA plots indicates the degree to which members of each gene sets are overrepresented at the extreme left, up-regulated at S4 (**A, E and I**), S10 (**B, F and J**), S20 (**C, G and K**) or R6 (**D, H and L**), or on the extreme right (up-regulated at Log) of the entire gene set. Solid bars represent bound genes. NES: normalized enrichment score. FDR: false discovery rate.

(**M**) PCA triplot showing the covariance between time points (■), gene expression (○) and phenotypes ("environmental" variables). ◇, △ and ◇ symbols are used for TCA cycle, SigF and Lsr2 regulon genes respectively. Filled symbols are CYPs referred to in the main text. Arrows represent severity of phenotypes: orange (INH, EMB, RIF or EMB tolerance), green (CellROX[hi] or CFDA[hi]), and black (ability to utilize indicated metabolite for growth). 235 strong influencers of the phenotypes plotted. 79% and 12% of the total observed variances are accounted for by the principle components axes (PC-1 and PC-2).
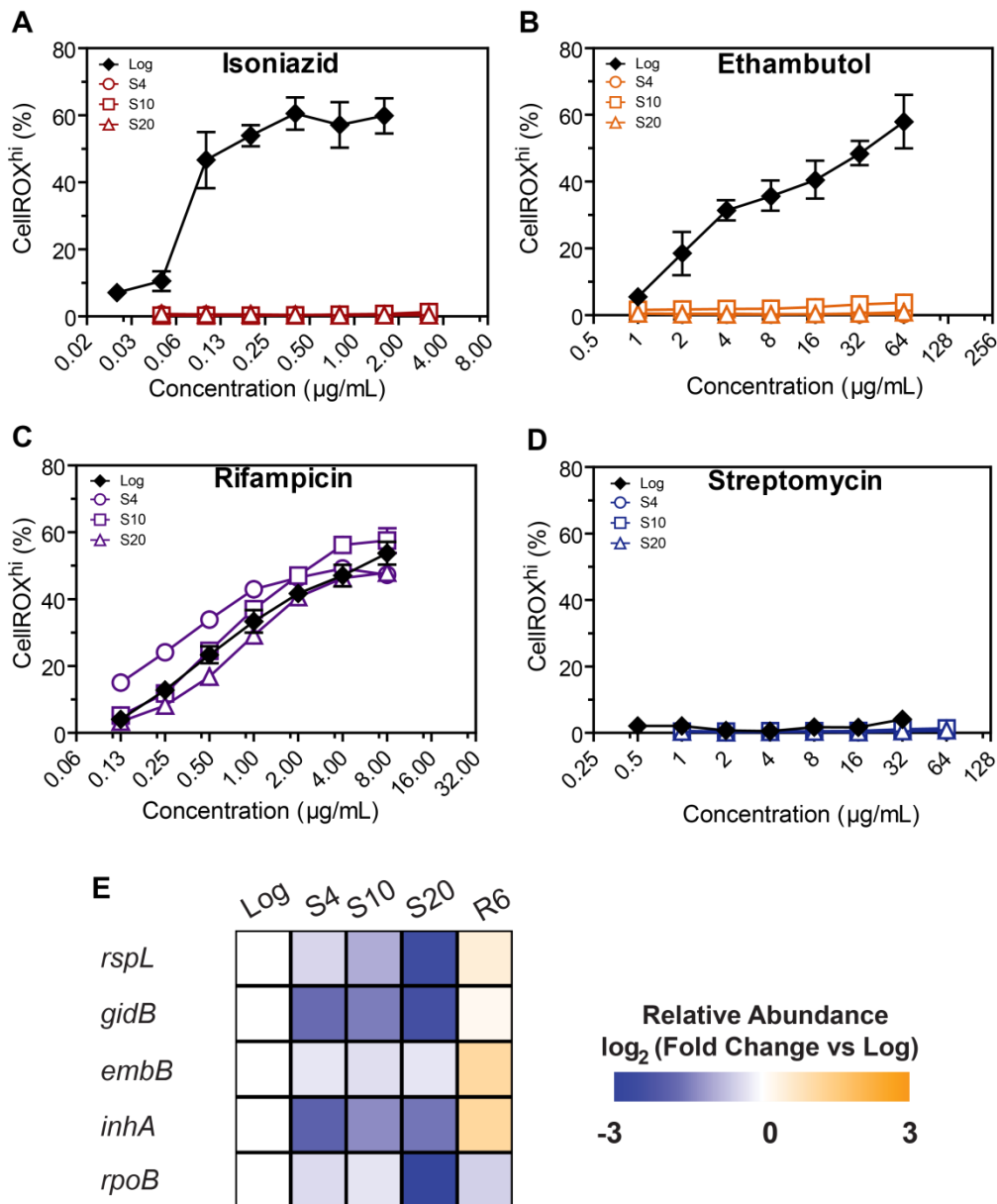
Supplementary Figure 5.6. Dependency of steady-state ROS production on antibiotic dose under nutrient replete and nutrient deprived conditions.

(**A-D**) Percentage of cells with the CellROX^{hi} phenotype after 48h of exposure to INH (**A**), EMB (**B**), RIF (**C**) or STM (**D**) at the indicated doses for Log, S4, S10 and S20 cultures. Data shown as arithmetic means ± SE; n ≥ 8.

(**C**) Transcript levels of genes associated with antibiotic sensitivity across the starvation time course. Gene expression heat maps of antibiotic sensitivity genes features by parsimonious RDA (**Figure 5.5E**). Similar to **Figure 5.4**, colors of blocks represent extent of transcriptional regulation. Orange: up-regulation. Blue: Down-regulation.
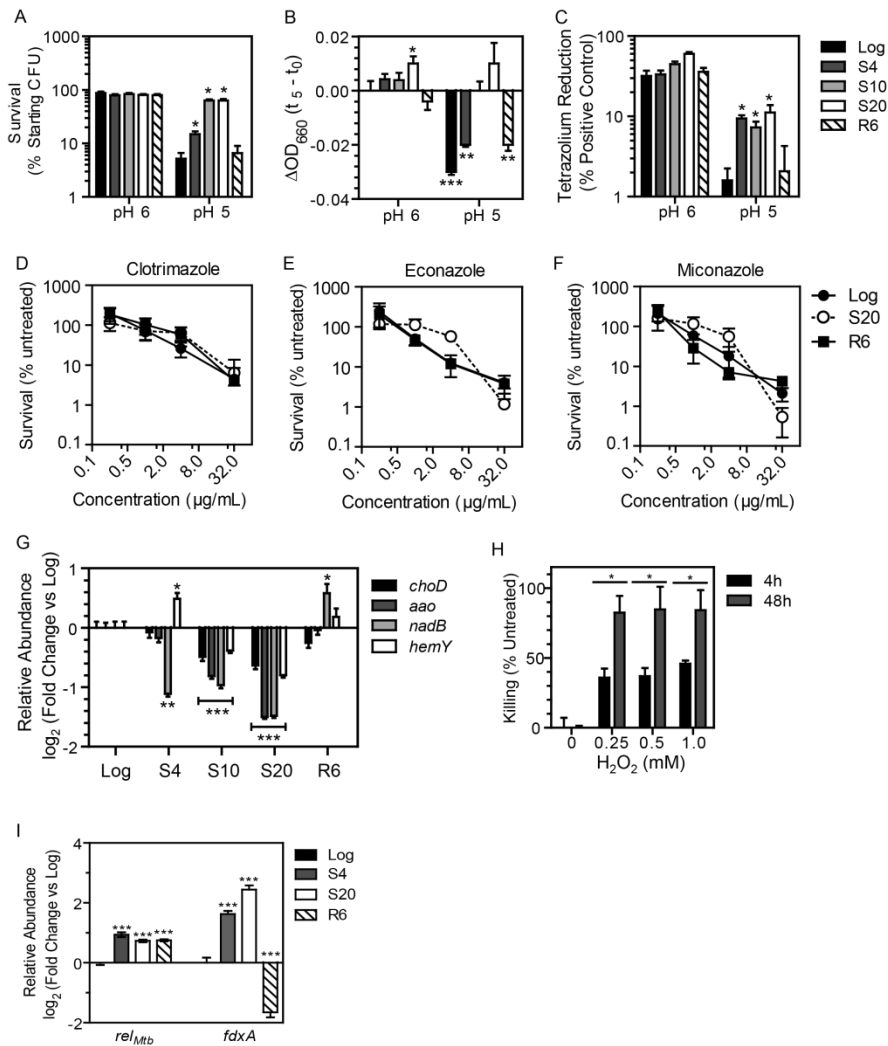
Supplementary Figure 5.7. Phenotypic acid resistance, azole indifference, hydrogen peroxide sensitivity and gene expression in *Mtb* supports the role of ketone body metabolism in mycobacterial persisters.

(**A-C**) Growth phenotypes and metabolic activities of Log, S4, S10, S20 and R6 BCG at pH 6 and pH 5. Viability – CFU (**A**), ΔOD (**B**) and reductive respiratory potential (**C**) – after 5 days of exposure to pH 6 and pH 5 media. n = 3, $0.01 < $ *p $ < 0.05$, $0.001 < $ **p $ < 0.01$, ***p $ < 0.001$ (one-way ANOVA with Dunnett's Multiple Comparison).

(**D-F**) Survival of Log, S20 and R6 BCG to azole-based CYP inhibitors. Cultures treated for 48h at indication clotrimazole (D), econazole (E) or miconazole (F) concentrations. n ≥ 3.

(**G**) Transcript levels of H_2O_2-producing enzymes during starvation. Gene expressions at indicated time points determined using RNA-seq. Enzymes catalyze the following reactions:

Cholesterol oxidase (*choD*): Cholesterol + O_2 ↔ Cholest-4-en-3-one + **H_2O_2**

D-amino acid oxidase (*aao*): D-Amino acid + H_2O + O_2 ↔ 2-Oxo acid + NH_3 + **H_2O_2**
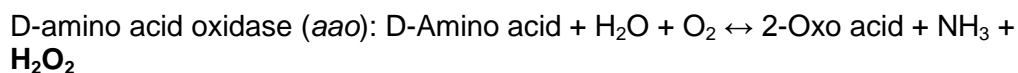
L-aspartate oxidase (*nadB*): L-Aspartate + $O_2 \leftrightarrow$ Iminoaspartate + **H_2O_2**

Protoporphyrinogen oxidase (*hemY*): Protoporphyrinogen IX + 3 $O_2 \leftrightarrow$ Protoporphyrin + 3 **H_2O_2**

n = 3, $0.01 < *p < 0.05$, $0.001 < **p < 0.01$, $***p < 0.001$ (one-way ANOVA with Dunnett's Multiple Comparison for each gene compared across time points).

(**H**) Killing of S20 BCG when exposed to H_2O_2. Unlike in **Figure 5.7C**, 25 mL of S20 cultures (starting with ~10^8 CFU/mL) were treated in 50 mL conical tubes at the indication H_2O_2 concentrations and proportion of the bacteria population killed was determined through CFU plating at neat - 10^3 dilutions. n = 3; $0.01 < *p < 0.05$ (unpaired, two-tailed t-test with Welch's correction).

(**G**) Regulation of *rel$_{MTB}$* and *fdxA* in *Mtb* at Log, S4, S20 and R6 as determined by qPCR (n = 3; one-way ANOVA with Dunnett's Multiple Comparison).

5.6.3. Supplementary tables

Supplementary Table 5.1. Variable Importance in Projection (VIP) scores and Kruskal-Wallis p-values from PLS-DA model of the metabolic phenotypes of Log, S4, S10, S20, and R6 cultures

[a] Induced abiotic reduction of tetrazolium dye

Metabolite	VIP	Kruskal-Wallis p-value
Acetoacetic acid	2.822	0.008
β-hydroxy-D,L-butyric acid	2.759	0.007
Glycerol	2.615	0.009
Tween 40	2.599	0.007
D-galactose[a]	2.260	0.007
Propionic acid	2.074	0.012
Acetic acid	1.489	0.013
D-trehalose	1.448	0.016
α-Ketoglutaric acid	1.403	0.009
Bromo-succinic acid[a]	1.399	0.010
D-fucose[a]	1.292	0.013
L-glutamic acid	1.246	0.010
Methyl pyruvate	1.113	0.005
Glucuronamide	1.103	0.015
D-arabitol	1.082	0.007
Citric acid	1.024	0.053
L-fucose	1.011	0.010
3-methyl-glucose[a]	1.009	0.012
L-aspartic acid	0.955	0.011
Gentiobiose	0.915	0.019
D-glucuronic acid	0.887	0.011
D-fructose-6-phosphate	0.834	0.011
D-lactic acid methyl ester	0.807	0.010
D-salicin[a]	0.778	0.046
L-lactic acid	0.778	0.021
L-histidine	0.752	0.005
D-melibiose	0.730	0.010
α-D-glucose	0.703	0.006
L-malic acid	0.671	0.008
α-keto-butyric acid	0.668	0.031
Pectin[a]	0.666	0.010
L-pyroglutamic acid	0.648	0.005
D-glucose-6-phosphate	0.644	0.007
D-mannose	0.636	0.007
L-rhamnose	0.620	0.007

Supplementary Table 5.2. Replicate numbers, sequencing depth and quality control parameters for RNA-seq of BCG transcriptome before, during and after starvation

Sample	Treatment	Biological Replicate	Technical Replicate 1	Technical Replicate 2	Coverage (% mapped)	rRNA (%)	3' to 5' ratio
			Number of Reads				
L1		1	1204645	1205127	99.4	0.2	1.07
L2	Exponential growth	2	1395678	1401660	99.5	0.5	1.05
L3		3	1985455	1987693	99.2	0.2	1.06
S4_1		1	1528656	1532367	87.3	2.6	1.05
S4_2	Starvation 4 days	2	1190809	1191191	92.9	0.6	1.02
S4_3		3	1577450	1583833	92.4	0.7	1.06
S10_1		1	1448874	1450220	95.0	1.4	0.92
S10_2	Starvation 10 days	2	1519684	1522886	93.4	11.6	0.98
S10_3		3	1002837	1003414	94.5	6.7	0.97
S20_1		1	1241005	1245825	88.5	11.2	0.96
S20_2	Starvation 20 days	2	1626951	1628710	94.6	7.5	0.97
S20_3		3	1304831	1307200	85.5	7.0	0.97
R6_1	Resuscitation	1	1203096	1205831	98.8	1.7	1.09
R6_2	6 days after 20 days	2	1143052	1146801	98.9	2.3	1.08
R6_3	starvation	3	1387610	1391396	90.7	20.5	1.11
Average			1384042	1386944	94.0	5.0	1.02

Supplementary Table 5.3. Primers used for qPCR

Gene	Direction	Sequence	Melting Temperature (°C)
relA	F	ATTGCCACCAGAAACACCGA	91.50
	R	GGTTCCGGGCGATGTGATTA	
fdxA	F	AGTGAGTGCGTGGATGTGATG	88.50
	R	TGGTGTTGATCGTCGGGTAGA	
hspX	F	GACGAGATGAAAGAGGGGCG	90.00
	R	GTCGTCCTCGTCAGCACCTA	
ethA	F	CGAGGCCGACGTTCTACTTAT	90.00
	R	GCGACTTCGACACTGGTTGC	
cyp135A1	F	GTGAAGCGGCGGAAAATCTC	90.50
	R	CACCCCAAACCCACAGAGTT	
cyp128	F	TCCCGAGAACAGTGCATTCC	90.50
	R	TCCGGGTTGATTTCGCTTGT	
sigA (Cappelli et al., 2006)	F	CGATGAGCCGGTAAAACGC	91.00
	R	GAGCCACTAGCGGACTTCGC	

5.7. References

1. Nathan, C., *Fresh approaches to anti-infective therapies.* Sci Transl Med, 2012. 4(140): p. 140sr2.
2. Grant, S.S. and D.T. Hung, *Persistent bacterial infections, antibiotic tolerance, and the oxidative stress response.* Virulence, 2013. 4(4): p. 273-83.
3. Lewis, K., *Persister cells.* Annu Rev Microbiol, 2010. 64: p. 357-72.
4. Lyczak, J.B., C.L. Cannon, and G.B. Pier, *Lung infections associated with cystic fibrosis.* Clin Microbiol Rev, 2002. 15(2): p. 194-222.
5. Blumberg, H.M., M.K. Leonard, Jr., and R.M. Jasmer, *Update on the treatment of tuberculosis and latent tuberculosis infection.* JAMA, 2005. 293(22): p. 2776-84.
6. Nguyen, D., et al., *Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria.* Science, 2011. 334(6058): p. 982-6.
7. Eoh, H. and K.Y. Rhee, *Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2013. 110(16): p. 6554-9.
8. Murray, C.J., et al., *Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014.
9. Kumar, A., et al., *Redox homeostasis in mycobacteria: the key to tuberculosis control?* Expert Rev Mol Med, 2011. 13: p. e39.
10. Amato, S.M., et al., *The role of metabolism in bacterial persistence.* Front Microbiol, 2014. 5: p. 70.
11. Dalebroux, Z.D. and M.S. Swanson, *ppGpp: magic beyond RNA polymerase.* Nat Rev Microbiol, 2012. 10(3): p. 203-12.
12. Balaban, N.Q., et al., *A problem of persistence: still more questions than answers?* Nat Rev Microbiol, 2013. 11(8): p. 587-91.
13. Maisonneuve, E. and K. Gerdes, *Molecular mechanisms underlying bacterial persisters.* Cell, 2014. 157(3): p. 539-48.
14. Betts, J.C., et al., *Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling.* Mol Microbiol, 2002. 43(3): p. 717-31.
15. Albrethsen, J., et al., *Proteomic profiling of Mycobacterium tuberculosis identifies nutrient-starvation-responsive toxin-antitoxin systems.* Mol Cell Proteomics, 2013. 12(5): p. 1180-91.
16. Yang, Y., et al., *A hydrolase of trehalose dimycolate induces nutrient influx and stress sensitivity to balance intracellular growth of Mycobacterium tuberculosis.* Cell Host Microbe, 2014. 15(2): p. 153-63.
17. Sarathy, J., et al., *Reduced drug uptake in phenotypically resistant nutrient-starved nonreplicating Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2013. 57(4): p. 1648-53.
18. Xie, Z., N. Siddiqi, and E.J. Rubin, *Differential antibiotic susceptibilities of starved Mycobacterium tuberculosis isolates.* Antimicrob Agents Chemother, 2005. 49(11): p. 4778-80.
19. Gengenbacher, M., et al., *Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability.* Microbiology, 2010. 156(Pt 1): p. 81-7.

20. Grant, S.S., et al., *Eradication of bacterial persisters with antibiotic-generated hydroxyl radicals.* Proc Natl Acad Sci U S A, 2012. 109(30): p. 12147-52.

21. Dwyer, D.J., et al., *Antibiotics induce redox-related physiological alterations as part of their lethality.* Proc Natl Acad Sci U S A, 2014. 111(20): p. E2100-9.

22. Liu, Y. and J.A. Imlay, *Cell death from antibiotics without the involvement of reactive oxygen species.* Science, 2013. 339(6124): p. 1210-3.

23. Low, K.L., et al., *Triacylglycerol utilization is required for regrowth of in vitro hypoxic nonreplicating Mycobacterium bovis bacillus Calmette-Guerin.* J Bacteriol, 2009. 191(16): p. 5037-43.

24. Khatri, B., et al., *High throughput phenotypic analysis of Mycobacterium tuberculosis and Mycobacterium bovis strains' metabolism using biolog phenotype microarrays.* PLoS One, 2013. 8(1): p. e52673.

25. Chen, J.W., J. Scaria, and Y.F. Chang, *Phenotypic and transcriptomic response of auxotrophic Mycobacterium avium subsp. paratuberculosis leuD mutant under environmental stress.* PLoS One, 2012. 7(6): p. e37884.

26. Galindo-Prieto, B., L. Eriksson, and J. Trygg, *Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS).* Journal of Chemometrics, 2014. 28(8): p. 623-632.

27. Trygg, J. and S. Wold, *Orthogonal projections to latent structures (O-PLS).* Journal of Chemometrics, 2002. 16(3): p. 119-128.

28. Meot, A., P. Legendre, and D. Borcard, *Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework.* Environmental and Ecological Statistics, 1998. 5(1): p. 1-27.

29. Valway, S.E., et al., *An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis.* N Engl J Med, 1998. 338(10): p. 633-9.

30. Olivares, J., et al., *The intrinsic resistome of bacterial pathogens.* Front Microbiol, 2013. 4: p. 103.

31. Bochner, B.R., *New technologies to assess genotype-phenotype relationships.* Nat Rev Genet, 2003. 4(4): p. 309-14.

32. Bochner, B.R., *Global phenotypic characterization of bacteria.* FEMS Microbiol Rev, 2009. 33(1): p. 191-205.

33. Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence.* Infect Immun, 1996. 64(6): p. 2062-9.

34. Galagan, J.E., et al., *The Mycobacterium tuberculosis regulatory network and hypoxia.* Nature, 2013. 499(7457): p. 178-83.

35. Hampshire, T., et al., *Stationary phase gene expression of Mycobacterium tuberculosis following a progressive nutrient depletion: a model for persistent organisms?* Tuberculosis (Edinb), 2004. 84(3-4): p. 228-38.

36. Steinhaus, E.A. and J.M. Birkeland, *Studies on the Life and Death of Bacteria: I. The Senescent Phase in Aging Cultures and the Probable Mechanisms Involved.* J Bacteriol, 1939. 38(3): p. 249-61.

37. Watson, S.P., M.O. Clements, and S.J. Foster, *Characterization of the starvation-survival response of Staphylococcus aureus.* J Bacteriol, 1998. 180(7): p. 1750-8.

38. Boylen, C.W. and M.H. Mulks, *The Survival of Coryneform Bacteria during Periods of Prolonged Nutrient Starvation.* Journal of General Microbiology, 1978. 105(2): p. 323-334.

39. Archuleta, R.J., P. Yvonne Hoppes, and T.P. Primm, *Mycobacterium avium enters a state of metabolic dormancy in response to starvation.* Tuberculosis (Edinb), 2005. 85(3): p. 147-58.

40. Gonzalez-Pastor, J.E., E.C. Hobbs, and R. Losick, *Cannibalism by sporulating bacteria.* Science, 2003. 301(5632): p. 510-3.

41. Chao, M.C. and E.J. Rubin, *Letting sleeping dos lie: does dormancy play a role in tuberculosis?* Annu Rev Microbiol, 2010. 64: p. 293-311.

42. Wu, J. and J. Xie, *Magic spot: (p) ppGpp.* J Cell Physiol, 2009. 220(2): p. 297-302.

43. Battesti, A. and E. Bouveret, *Bacteria possessing two RelA/SpoT-like proteins have evolved a specific stringent response involving the acyl carrier protein-SpoT interaction.* J Bacteriol, 2009. 191(2): p. 616-24.

44. Bataineh, H., O. Pestovsky, and A. Bakac, *pH-induced mechanistic changeover from hydroxyl radicals to iron(iv) in the Fenton reaction.* Chemical Science, 2012. 3(5): p. 1594-1599.

45. Daniel, J., et al., *Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in Mycobacterium tuberculosis as it goes into a dormancy-like state in culture.* J Bacteriol, 2004. 186(15): p. 5017-30.

46. Deb, C., et al., *A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in Mycobacterium tuberculosis.* J Biol Chem, 2006. 281(7): p. 3866-75.

47. Lee, W., et al., *Intracellular Mycobacterium tuberculosis exploits host-derived fatty acids to limit metabolic stress.* J Biol Chem, 2013. 288(10): p. 6788-800.

48. Rockabrand, D., et al., *Roles of DnaK and RpoS in starvation-induced thermotolerance of Escherichia coli.* J Bacteriol, 1998. 180(4): p. 846-54.

49. Weiss, L.A. and C.L. Stallings, *Essential roles for Mycobacterium tuberculosis Rel beyond the production of (p)ppGpp.* J Bacteriol, 2013. 195(24): p. 5629-38.

50. Aoki-Kinoshita, K.F. and M. Kanehisa, *Gene annotation and pathway mapping in KEGG.* Methods Mol Biol, 2007. 396: p. 71-91.

51. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. 4(1): p. 44-57.

52. Rao, M., et al., *Intracellular pH regulation by Mycobacterium smegmatis and Mycobacterium bovis BCG.* Microbiology, 2001. 147(Pt 4): p. 1017-24.

53. Putra, S.R., et al., *Distribution of mevalonate and glyceraldehyde 3-phosphate/pyruvate routes for isoprenoid biosynthesis in some gram-negative bacteria and mycobacteria.* FEMS Microbiol Lett, 1998. 164(1): p. 169-75.

54. Merilainen, G., et al., *The thiolase reaction mechanism: the importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation.* Biochemistry, 2009. 48(46): p. 11011-25.

55. Uchino, K., et al., *Isolated poly(3-hydroxybutyrate) (PHB) granules are complex bacterial organelles catalyzing formation of PHB from acetyl coenzyme A (CoA) and degradation of PHB to acetyl-CoA.* J Bacteriol, 2007. 189(22): p. 8250-6.

56.   Hausinger, R.P., *New insights into acetone metabolism.* J Bacteriol, 2007. 189(3): p. 671-3.

57.   Hartmans, S. and J.A.M. de Bont, *Acetol monooxygenase from Mycobacterium Py1 cleaves acetol into acetate and formaldehyde.* FEMS Microbiology Letters, 1986. 36(2-3): p. 155-158.

58.   Munro, A.W., et al., *What makes a P450 tick?* Trends Biochem Sci, 2013. 38(3): p. 140-50.

59.   Denisov, I.G., et al., *Structure and chemistry of cytochrome P450.* Chem Rev, 2005. 105(6): p. 2253-77.

60.   Koop, D.R., *Oxidative and reductive metabolism by cytochrome P450 2E1.* FASEB J, 1992. 6(2): p. 724-30.

61.   Leung, T.M. and N. Nieto, *CYP2E1 and oxidant stress in alcoholic and non-alcoholic fatty liver disease.* J Hepatol, 2013. 58(2): p. 395-8.

62.   Puntarulo, S. and A.I. Cederbaum, *Production of reactive oxygen species by microsomes enriched in specific human cytochrome P450 enzymes.* Free Radic Biol Med, 1998. 24(7-8): p. 1324-30.

63.   Pethe, K., et al., *A chemical genetic screen in Mycobacterium tuberculosis identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy.* Nat Commun, 2010. 1: p. 57.

64.   Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. 102(43): p. 15545-50.

65.   Legendre, P. and E.D. Gallagher, *Ecologically meaningful transformations for ordination of species data.* Oecologia, 2001. 129(2): p. 271-280.

66.   Albesa, I., et al., *Oxidative stress involved in the antibacterial action of different antibiotics.* Biochem Biophys Res Commun, 2004. 317(2): p. 605-9.

67.   Karuzina, II, et al., *Heme and apoprotein modification of cytochrome P450 2B4 during its oxidative inactivation in monooxygenase reconstituted system.* Free Radic Biol Med, 1999. 26(5-6): p. 620-32.

68.   Enami, S., Y. Sakamoto, and A.J. Colussi, *Fenton chemistry at aqueous interfaces.* Proc Natl Acad Sci U S A, 2014. 111(2): p. 623-8.

69.   Bigger, J., *Treatment of Staphylococcal Infections with Penicillin by Intermittent Sterisation.* The Lancet, 1944. 244(6320): p. 497-500.

70.   Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch.* Science, 2004. 305(5690): p. 1622-5.

71.   Loebel, R.O., E. Shorr, and H.B. Richardson, *The Influence of Foodstuffs upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli.* J Bacteriol, 1933. 26(2): p. 139-66.

72.   Beste, D.J., et al., *(1)(3)C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation.* PLoS Pathog, 2011. 7(7): p. e1002091.

73.   Baek, S.H., A.H. Li, and C.M. Sassetti, *Metabolic regulation of mycobacterial growth and antibiotic sensitivity.* PLoS Biol, 2011. 9(5): p. e1001065.

74.   Park, S.W., et al., *Growth of mycobacteria on carbon monoxide and methanol.* J Bacteriol, 2003. 185(1): p. 142-7.

75.   Cousins, F.B., *The lactic acid oxidase of the mycobacteria.* Biochem J, 1956. 64(2): p. 297-304.

76.   Walsh, C., *Molecular mechanisms that confer antibacterial drug resistance.* Nature, 2000. 406(6797): p. 775-81.

77. Jarlier, V. and H. Nikaido, *Mycobacterial cell wall: structure and role in natural resistance to antibiotics.* FEMS Microbiol Lett, 1994. 123(1-2): p. 11-8.

78. Nyka, W., *Studies on the effect of starvation on mycobacteria.* Infect Immun, 1974. 9(5): p. 843-50.

79. Gengenbacher, M. and S.H. Kaufmann, *Mycobacterium tuberculosis: success through dormancy.* FEMS Microbiol Rev, 2012. 36(3): p. 514-32.

80. da Silva, P.E., et al., *Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis.* FEMS Immunol Med Microbiol, 2011. 63(1): p. 1-9.

81. Adams, K.N., et al., *Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism.* Cell, 2011. 145(1): p. 39-53.

82. Balding, P.R., et al., *How do azoles inhibit cytochrome P450 enzymes? A density functional study.* J Phys Chem A, 2008. 112(50): p. 12911-8.

83. McLean, K.J., et al., *The Mycobacterium tuberculosis cytochromes P450: physiology, biochemistry & molecular intervention.* Future Med Chem, 2010. 2(8): p. 1339-53.

84. Milano, A., et al., *Azole resistance in Mycobacterium tuberculosis is mediated by the MmpS5-MmpL5 efflux system.* Tuberculosis (Edinb), 2009. 89(1): p. 84-90.

85. Ando, H., et al., *Downregulation of katG expression is associated with isoniazid resistance in Mycobacterium tuberculosis.* Mol Microbiol, 2011. 79(6): p. 1615-28.

86. da Silva, P.E. and J.C. Palomino, *Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs.* J Antimicrob Chemother, 2011. 66(7): p. 1417-30.

87. Anzenbacher, P. and E. Anzenbacherova, *Cytochromes P450 and metabolism of xenobiotics.* Cell Mol Life Sci, 2001. 58(5-6): p. 737-47.

88. Sim, E., et al., *Arylamine N-acetyltransferases in mycobacteria.* Curr Drug Metab, 2008. 9(6): p. 510-9.

89. Sholto-Douglas-Vernon, C., et al., *Mutational and expression analysis of tbnat and its response to isoniazid.* J Med Microbiol, 2005. 54(Pt 12): p. 1189-97.

90. Linhart, K., H. Bartsch, and H.K. Seitz, *The role of reactive oxygen species (ROS) and cytochrome P-450 2E1 in the generation of carcinogenic etheno-DNA adducts.* Redox Biology, (In Press, Accepted Manuscript).

91. Holsclaw, C.M., et al., *Structural characterization of a novel sulfated menaquinone produced by stf3 from Mycobacterium tuberculosis.* ACS Chem Biol, 2008. 3(10): p. 619-24.

92. Fan, J., et al., *Quantitative flux analysis reveals folate-dependent NADPH production.* Nature, 2014. 510(7504): p. 298-302.

93. Ragsdale, S.W. and E. Pierce, *Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation.* Biochim Biophys Acta, 2008. 1784(12): p. 1873-98.

94. Caspi, R., et al., *The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.* Nucleic Acids Res, 2014. 42(Database issue): p. D459-71.

95. Vandal, O.H., C.F. Nathan, and S. Ehrt, *Acid resistance in Mycobacterium tuberculosis.* J Bacteriol, 2009. 191(15): p. 4714-21.

96.   Abramovitch, R.B., et al., *aprABC: a Mycobacterium tuberculosis complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome.* Mol Microbiol, 2011. 80(3): p. 678-94.

97.   Portaels, F. and S.R. Pattyn, *Growth of mycobacteria in relation to the pH of the medium.* Ann Microbiol (Paris), 1982. 133(2): p. 213-21.

98.   Piddington, D.L., A. Kashkouli, and N.A. Buchmeier, *Growth of Mycobacterium tuberculosis in a defined medium is very restricted by acid pH and Mg(2+) levels.* Infect Immun, 2000. 68(8): p. 4518-22.

99.   Jackett, P.S., V.R. Aber, and D.B. Lowrie, *Virulence and resistance to superoxide, low pH and hydrogen peroxide among strains of Mycobacterium tuberculosis.* J Gen Microbiol, 1978. 104(1): p. 37-45.

100.   Vandal, O.H., et al., *A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis.* Nat Med, 2008. 14(8): p. 849-54.

101.   Loebel, R.O., E. Shorr, and H.B. Richardson, *The Influence of Adverse Conditions upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli.* J Bacteriol, 1933. 26(2): p. 167-200.

102.   Baker, J.J., B.K. Johnson, and R.B. Abramovitch, *Slow growth of Mycobacterium tuberculosis at acidic pH is regulated by phoPR and host-associated carbon sources.* Mol Microbiol, 2014. 94(1): p. 56-69.

103.   Lewis, D.F.V. and A. Wiseman, *A selective review of bacterial forms of cytochrome P450 enzymes.* Enzyme and Microbial Technology, 2005. 36(4): p. 377-384.

104.   Wakamoto, Y., et al., *Dynamic persistence of antibiotic-stressed mycobacteria.* Science, 2013. 339(6115): p. 91-5.

105.   Tsuru, S., et al., *Adaptation by stochastic switching of a monostable genetic circuit in Escherichia coli.* Mol Syst Biol, 2011. 7: p. 493.

106.   Stallings, C.L., et al., *CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence.* Cell, 2009. 138(1): p. 146-59.

107.   DiChiara, J.M., et al., *Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis.* Nucleic Acids Res, 2010. 38(12): p. 4067-78.

108.   Wegscheid, B. and R.K. Hartmann, *In vivo and in vitro investigation of bacterial type B RNase P interaction with tRNA 3'-CCA.* Nucleic Acids Res, 2007. 35(6): p. 2060-73.

109.   Keiler, K.C., *Biology of trans-translation.* Annu Rev Microbiol, 2008. 62: p. 133-51.

110.   Hartkoorn, R.C., et al., *Genome-wide definition of the SigF regulon in Mycobacterium tuberculosis.* J Bacteriol, 2012. 194(8): p. 2001-9.

111.   Ortiz de Montellano, P. and J. De Voss, *Substrate Oxidation by Cytochrome P450 Enzymes*, in *Cytochrome P450*, P. Ortiz de Montellano, Editor. 2005, Springer US. p. 183-245.

112.   Rude, M.A., et al., *Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from Jeotgalicoccus species.* Appl Environ Microbiol, 2011. 77(5): p. 1718-27.

113.   Cocks, G.T., T. Aguilar, and E.C. Lin, *Evolution of L-1, 2-propanediol catabolism in Escherichia coli by recruitment of enzymes for L-fucose and L-lactate metabolism.* J Bacteriol, 1974. 118(1): p. 83-8.

6. Reprogrammed tRNAs read a code of codons to regulate mycobacterial dormancy

6.1. Abstract

Codon usage bias and RNA modifications are ubiquitous features of life; yet, the manner in which these genetic and epigenetic traits interact to influence gene expression remains elusive. For members of the *Mycobacterium tuberculosis* Complex especially, persistence in the presence of cell-mediated immunity is dependent upon the timely regulation of dormancy programs within hypoxic granulomas. Seeking to understand the mechanism behind the selective translation of stress response genes when *M. bovis* BCG is exposed to hypoxia, we examined the molecular factors contributing to translation efficiency. Of the 40 modified ribonucleosides in BCG tRNA, levels of wobble 5-oxyacetyl-uridine (cmo^5U) in $tRNA^{Thr(\underline{U}GU)}$ herald changes in DosR abundance – the transcription factor critical to the establishment of hypoxic dormancy. Translation of *dosR* and other transcripts biased in $Thr^{AC\underline{G}}$ usage are synchronously enhanced during hypoxia. Recombinant BCG with *dosR* coded by other threonine codons showed varied fitness during and after hypoxic exposures in accordance with the availability of modified tRNAs. Our results connect previous observations on mycobacterial dormancy with mechanistic insights as to how synonymous codon substitutions affect protein abundances and cellular phenotype.

6.2. Introduction

The consistent and meticulous manner in which organisms choose one of several synonymous codons to code for an amino acid is unlikely a quirk of biology. Instead, this suggests that a 'code of codons' exists [1]. This code

impacts protein translation and could serve to regulate gene expression by prioritizing the translation of specific transcripts over others [2-4]. This secondary feature of genetic coding is frequently obscured in piece-wise measurements of gene transcript and protein abundance and in species-to-species genome-wide comparisons, but easily observed by gene-specific codon usage of the transcriptome [5-7]. Similarly, the role tRNA modifications play in biology cannot be fully appreciated without considering that it a dynamic system that responds to cellular stress [5, 8, 9]. These post-transcriptional modifications to the four canonical ribonucleosides are capable of altering base-pairing interactions between codons and anticodons and thus, expand or limit tRNA decoding capabilities [10-12]. How do these genetic and epigenetic traits come together to affect gene expression? One possibility is that stress reprograms tRNAs by transiently altering their modifications. Changes in wobble modifications on anticodons, particularly, would facilitate increased or restricted decoding of cognate and near-cognate codons which are over- or under-used in transcripts. Consequently, translational elongation of transcripts bearing preferentially read codons are enhanced, in rate, accuracy, or both, leading to the selective up-regulation of these genes [1, 3].

In Chapters 2-4, we developed an analytical platform for the isolation, purification, characterization and quantification of cellular RNA and its biochemical modifications. In Chapter 5, we dissected the phenomenon of persistence in nutrient starved mycobacteria and revealed how stress responses could cause metabolic shifts that determine phenotypic antibiotic resistance. Here, we move from transcriptional to translational control of stress responses, wherein our knowledge of RNA biology and mycobacteriology come together to help us understand how mycobacterial

pathogens use tRNA modifications to control gene expression by the selective translation of gene transcripts necessary for surviving hypoxic stress.

6.3. Material and methods

6.3.1. Reagents

Unless otherwise stated, chemical reagents were purchased from Sigma-Aldrich, bacterial culture reagents from BD Biosciences, Purelink RNA extraction kits were purchased from Life Technologies, LC-MS grade solvents from Thermo-Fisher, RNA and DNA oligonucleotides from Integrated DNA Technologies, restriction enzymes from New England Biolabs and peptide standards from Genescript.

6.3.2. Bacterial cultures

For mycobacteria pre-cultures, *M. bovis* Bacille Calmette-Guérin (str. Pasteur 1173P2; BCG) were grown at 37 ^{o}C in a shaking incubator in Middlebrook 7H9 broth supplemented with 0.5% (w/v) albumin, 0.2% (w/v) glucose, 0.085% (w/v) NaCl, 0.2% (v/v) glycerol and 0.05% (v/v) Tween 80 to an OD_{600} of 0.6 – 0.8. For exponential growth, BCG pre-cultures were innoculated into Dubos broth (supplemented with 10% Dubos medium albumin, and 0.03% Tween 80) and passaged in roller bottles for balanced growth. A non-replicating state was induced in BCG by subjecting them to the slow withdrawal of oxygen as described by Wayne and Hayes [13] with modifications made by Ravindran et al. [14] for a ~40 fold expansion of batch cultures. Briefly, a culture of exponentially growing BCG in Dubos broth was diluted to an OD_{600} of 0.005 and placed in a tightly-sealed (latex-lined cap) 1 L glass flask (Duran) with stirring at 80 rpm and air headspace ratio of 0.5. For resuscitation of dormant cultures by re-aeration, hypoxic cultures were

transferred to 1.5 L vent-capped Erlenmeyer flasks (Corning) and shaken at 180 rpm. Levels of oxygen were tracked by methylene blue decolorization, oxygen indicator strips (BD) and dissolved oxygen probe sensor (Vernier). Cultures with added methylene blue were not used in further experiments. At indicated time points (Fig. S1), flasks were sacrificially opened; and bacteria with pelleted, washed, and flash frozen in liquid nitrogen. For CFU determinations, serial dilutions were plated on 7H11 agar supplemented with 10% (v/v) oleic acid-albumin-dextrose-catalase (OADC). Cultures were routinely assessed for contamination by microscopy (Gram, Ziehl-Neelsen, and auramine-rhodamine staining) and streaking on blood agar. When appropriate, hygromycin and kanamycin, were added at 80 and 30 µg/mL respectively to broth and agar.

One Shot® TOP10 *Escherichia coli* (F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *araD*139 Δ(*araleu*)7697 *galU galK rpsL* (StrR) *endA1 nupG*; Life Technologies) were grown in Luria-Bertani (LB) broth and agar. When appropriate, hygromycin and kanamycin were added at 150 and 50 µg/mL respectively.

6.3.3. RNA extraction and purification

Total RNA was extracted from BCG pellets as described previously [15]. To prevent the formation of Tris-RNA adducts [16], TRIzol (Life Technologies) instead of Tris-EDTA buffer saturated phenol:chloroform:isoamyl alcohol was used with bead-beating for cell lysis. Additionally, on-column DNAse I (Qiagen) digest was performed in the presence of 10 mM ammonium acetate (pH 7.0) instead of Tris-buffer. Multi-dimensional HPLC for the separation and isolation and quantification of RNA species was performed as previously described [17]. RNA composition and integrity was determined using the

appropriate Agilent Bioanalyzer RNA chips. Total RNA samples with a RNA Integrity Number (RIN) of 8.0 or greater were used for qPCR experiments. tRNA samples that were >99.9% pure by size exclusion chromatography were used for LC-MS/MS analysis.

6.3.4. Identification and quantification of tRNA modifications

Purified BCG tRNA (0.5 µg per sample) was hydrolyzed enzymatically as detailed in [18]. To prevent the formation of Tris-RNA adducts, HEPES buffer (pH 8.0) was used instead of Tris-HCl buffer (pH 8.0). Furthermore, all enzymes were dialyzed against HEPES buffer (pH 8.0) immediately prior to use. Reverse phase-LC of the hydrolyzed tRNA was performed as preciously described [18, 19]. Neutral loss scan (NLS), molecular feature extraction (MFE) and targeted ion fragmentation (targeted MS[n]) were used to identify 40 modifications in BCG tRNA. Neutral loss scan was performed on an Agilent 6460 LC-QQQ spectrometer with ESI Jetstream ionization, searching for compounds with loss of ribose (-136 m/z) or 2'-O-methyl-ribose (-146 m/z) upon fragmentation. High accuracy masses for molecular ions and CID fragment were obtained using either Agilent 6520 LC-QTOF spectrometer with an ESI ionization or LTQ Orbitrap XL (Thermo-Scientific). Untargeted feature finding was performed using molecular feature extraction (Agilent Workstation Qualitative Analysis vB05.06). Molecules reproducibly observed (by retention time, molecular mass, features of MS[2] fragmentation) in all biological replicates at one time point were validated by comparisons with commercially available standards (See Table S1), comparisons with theoretical molecular masses of ribonucleosides found in ChemSpider (http://www.chemspider.com/), Modomics (http://modomics.genesilico.pl/) and the RNA modifications database (http://mods.rna.albany.edu/), targeted MS[n], isotopic envelope analysis and salt adduct analysis. Dynamic multiple

reaction monitoring (dMRM) on an Agilent 6460 LC-QQQ spectrometer was used to quantify these modifications [18]. Quantitative comparisons between biological replicates from various time points were made possible by correcting for biological variation in total tRNA quantities by dividing raw peak area for the ribonucleoside by the UV absorbance (in-line detector) peak areas for the 4 canonical ribonucleosides and normalizing spectra signals against that of the spiked internal standard ($[^{15}N]_5$-deoxyadenosine) to adjust for day-to-day fluctuation in MS sensitivity. All mass spectrometers were operated in positive ion mode.

6.3.5. Sequencing and quantification of tRNA-specific oligonucleotides

We developed a LC-MS/MS-based platform to map tRNA modifications, perform label-free absolute quantification of tRNA copy numbers and determine the extent of modification on those tRNA copies by combining the principles of bottom-up shotgun proteomics, amide-HILIC oligonucleotide liquid chromatography, and response factor calibration outlined in previous studies [20-22]. However, the following alterations were made:

The tRNA^{Thr} pool was characterized and quantified using Agilent 6520 QTOF and Agilent 6460 QQQ spectrometers coupled with an Agilent 1290 infinity LC system with online diode array for UV-Vis spectrometry, operated in negative ion mode. BCG tRNA sequences were downloaded from the Genomic tRNA database (http://gtrnadb.ucsc.edu/) and RNase U2, T1 and A digestion products were predicted using Mongo Oligo Mass Calculator (http://mods.rna.albany.edu/masspec/Mongo-Oligo). >6 mer digestion products were evaluated for sequence uniqueness (by BLASTn against genomic tRNA sequences) and RNAse U2 selected for further experiments as it generates unique products from the anticodon stem-loop of all 3 tRNA^{Thr}.

We eliminated sequences with positional isomers that could be generated from RNAse U2 digests with or without missed cleavages, and identified a unique oligonucleotide tag for every RNA[Thr]. RNA and DNA oligomers with the same sequences (with the exception of U>T for DNA) were purchased and used to optimize LC and MS parameters for maximal chromatographic separation (by retention times), sequence coverage and signal strength of the fragments, unique transitions with high signal to noise ratios (S/N > 10) and minimal source fragmentation.

RNAse U2 (Thermo-Scientific) digestion of BCG tRNA was performed followed by the removal of 5' and 3' phosphates by bacterial alkaline phosphatase (Life Technologies). Reaction performed in 10 mM ammonium acetate pH7.0 at 37°C for 4 h (2.5h of U2 digest and 1.5h of dephosphorylation) in the presence of deaminase inhibitors (0.5 µg/mL coformycin, 5 µg/mL tetrahydrouridine) and antioxidants (50 µM desferrioxamine, 50 µM butylated hydroxytoluene). Enzymes were dialyzed against 10 mM ammonium acetate prior to use to prevent the formation of Tris-RNA adducts. Proteins were removed by filtration (Microcon YM-10), desalted by ZipTip$_{C18}$ (Millipore) and concentrated by vacuum centrifuge. The extent of digestion and size of products were assessed by small Bioanalyzer chips (Agilent) and by MALDI-MS (Voyager DE, AB SCIEX) using 2,4,6-trihydroxyacetophenone (THAP) as matrix. RNA fragments were reconstituted in 70% acetonitrile (v/v) for LC-MS/MS.

Amide-HILIC LC separation was performed on a TSK-gel Amide-80 column (2.0 mm ID x 150 mm, 3 µm particle size) using a binary solvent system consisting of 8 mM ammonium acetate in ultrapure water (solvent A) and acetonitrile (solvent B). HPLC was performed at a flow rate of 0.1 mL/min. The gradient of solvent A was as follows: 0-2 min, held at 10% (v/v); 2-3 min,

10%-15%; 3-5.5 min, 15%-30%; 5.5-20.5 min, 30%-60%; 20.5-25 min, 60%-70%; 25-28 min, 70%-10%. The HPLC column was maintained at 50 ºC. QTOF mass spectrometer was operated at gas temperature 325 °C, gas flow 8 L/min, nebulizer 30 psi, Vcap 3800 V, Fragmentor 250 V, Skimmer 1 65 V, octopole RF peak of 750. Targeted MS^2 performed every 5V from collision energies from 25V to 45V and products scanned from 100 m/z to 1500 m/z. QQQ mass spectrometer operated at gas temperature 325 °C, gas flow 10 L/min, nebulizer 32 psi, sheath gas temperature 300 °C, sheath gas flow 11 L/min, capillary 4000V, Vcharging 500. For MRM, MS1 acquisition was performed at wide resolution and MS2 acquisition at unit resolution. Dwell time per transition at 150 ms, fragmentor between 130-160 V, collision energy between 30-40V and cell accelerator voltage between 5-7V.

As the positions of BCG tRNA modifications had not been mapped, we generated a list of potential modified oligomers by taking into account the 40 modifications identified in BCG tRNA (**Figure. 1**), predicted modifications by tRNAmod (http://crdd.osdd.net/raghava/trnamod/) and comparisons with tRNA sequences catalogued in Modomics (http://modomics.genesilico.pl/). The exact masses of these potential modified oligomers and their c1 and y1 ion fragments were calculated. Selected reaction monitoring (SRM) worklists were created to screen the tRNA pool of Log, H18 and R6 cells for oligomers with these predicted c1 and y1 transitions. SRM screen for c1 and y1 transitions were performed separately and in technical duplicate on the LC-QQQ. Oligomers with matching the retention times for their predicted c1 and y1 transitions were sequenced by targeted MS^2 on the LC-QTOF at collision energies of 0, 15, 30 and 45V. Fragment analysis, aided by SOS [23] and RoboOligo (https://u.osu.edu/paulsample/robooligo/), enabled us to map the exact location of each modification based on their a-B, c, w, and y ions.

Analysis of free bases and internal fragments (mainly matching ion fragments with w type cuts at the 3' end or a-B type cut from the 5' end, generating a pNpf or NpNf fragment) further validated our structural assignments [24]. The two strongest signals (Fig S4A, table S2) were selected to quantify and identify each modified oligomer by MRM in the same run (**Supplementary Figure 6.4A**).

Response factors for RNA oligomers against DNA oligomers of the same sequence were computed by external calibration (**Figure S4B**). With the assumption that these response factors are applicable to modified oligomers of the same length and sequence, we quantified the amounts of each tRNA species by the ratios of peak areas between the quantifier transitions of the modified RNA fragment to 1 pmol of spiked DNA oligomers of the same length and sequence (**Supplementary Figure 6.4B**).

6.3.6. Protein extraction and processing

Six replicate samples of BCG pellets at 8 experimental time points (Log, H4, H6, H9, H14, H18, R3 and R6) were extracted with 8 M urea, 1 M thiourea with 2% SDS supplemented with protease inhibitor complete cocktail (Roche) by bead beating (50 Hz for 3 five-minute cycles, TissueLyzer II (Qiagen)) with 0.2 mm silica beads. Bead beating chambers were chilled at -20 °C before and between cycles. Extracts were centrifuged at 14 000 g for 15min at 4 °C and the supernatant collected. Supernatant were diluted with equal volumes of ultra-pure water and supplemented with HALT protease inhibitor cocktail (Pierce BioSciences), $MgCl_2$ (2mM final) and Tris-HCl (pH8.0, 10 mM final) for Benzonase digestion (25 Units per mL lysate) at 20 °C for 3 h. Samples were chilled on ice and sonicated at (20% maximum amplitude, five 10 second cycles, Fisher-Scientific FB15061) to break up any aggregates that might had

formed. Overnight trichloroacetic acid/acetone precipitation was performed with 2D Clean-Up kits (GE Healthcare) as instructed by the manufacturer. Air-dried protein pellets were resuspended in 10 mM triethylammonium bicarbonate (TEAB) buffer (pH 8.5) and protein concentrations determined by BCA assay (Pierce BioSciences). Protein quality and quantities was checked by SDS-PAGE electrophoresis (12% polyacrylamide gels) and purity by UV spectrometry (Nanodrop, Thermo Scientific). Protein aliquots (100 µg) were collected, lyophilized, and stored at –80 °C prior to further processing.

Lyophilized proteins were reconstituted in 100 mM TEAB and 10% acetonitrile (v/v) by bath sonication. Samples were reduced and alkylated using the iTRAQ kit (Applied Biosystems) according to the manufacturer's instructions. Protein concentrations were estimated using Bradford assays (Bio-Rad). 50 µg of protein was precipitated by cold acetone, redissolved in 6 M urea, diluted with five volumes of 100 mM TEAB with 10% (v/v) acetonitrile and digested with trypsin in a 1:30 (w/w) ratio overnight at 37 °C. Samples were dried by vacuum centrifugation and stored at –20 °C prior to analysis.

6.3.7. iTRAQ labeling and peptide fractionation

Aliquots of digested protein (from 50 µg of total protein) were split into 3 portions as technical replicates and labeled with 8-plex iTRAQ reagents according to the manufacturer's instructions. To avoid bias from any one tag during analysis, labels for samples from each time point were randomized and unblinded post-analysis. After iTRAQ labeling, samples were desalted with Sep-Pak Plus C18 cartridges (Waters), dried by vacuum centrifugation and reconstituted in IPG buffer (Agilent) without glycerol. Isoelectric focusing was performed from pH 3-10 over 24 wells on an Agilent 3100 OFFGEL

fractionator according to the manufacturer's protocol (OG24PE00). All 24

fractions were collected and analyzed by nano-LC-MS/MS.

6.3.8. LC-MS/MS analysis of the BCG proteome

iTRAQ proteomics experiments were performed on an Agilent 1200 nano-LC-

Chip/MS interfaced to an Agilent 6510 QTOF LC/MS. The LC system

consisted of a capillary pump for sample loading, a nanoflow pump and a

temperature-controlled microwell-plate autosampler. The HPLC-Chip

configuration consisted of a 160 nL enrichment column and a 150 mm x 75

µm analytical column (G4240-62001 Zorbax 300SB-C18). Mobile phases

employed were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in

acetonitrile (solvent B). A 120 min long gradient LC separation was used with

10 minutes for column wash and equilibration between runs. Samples were

loaded unto the enrichment column at 1% (v/v) B at flow rates of 3 µL/min. On

the nano-flow pump, the gradient of solvent B was as follows: 0-1 min, held at

1% (v/v), flow rate from 0.4-0.2 µL/min; 1-101 min, 1%-45%, flow rate held at

0.2 µL/min; 101-121 min, 45%-75%, flow rate held at 0.2 µL/min; 121-122

min, 75%-98%, flow rate from 0.2-0.4 µL/min; 122-126 min, held at 98%, flow

rate held at 0.4 µL/min; 126-127 min, 98%-1%, flow rate held at 0.4 µL/min;

127-130 min, held at 1%, flow rate held at 0.4 µL/min. LC-QTOF was

operated at high resolution (4 GHz) in positive ion mode with the following

source conditions: gas temperature 325 °C, drying gas 5 L/min, fragmentor

225 V. Capillary voltage was adjusted between 1500-2100 V manually to

achieve a steady spray. Data was acquired from 200–1700 m/z with an

acquisition rate of 4 spectra/s in MS mode and from 50–2200 m/z with an

acquisition rate of 2 spectra/s in MS/MS mode.

LC/MS data were extracted and evaluated using the molecular feature extractor (MFE) algorithm in MassHunter Qualitative Analysis software (B04.00). Test injections (3-4) from each fraction of the first technical replicate were made to optimize injection volumes for the second and third biological replicates for maximal extracted molecules with peptide-like features. For each fraction, the MFE list of molecular ions was exported and used to exclude the acquisition of spectra from these ions in subsequent runs. As such, every fraction from each technical replicate was run twice, first without and later with the exclusion list. Data from MassHunter Qualitative Analysis was exported to Mass Profiler Professional (version B02.02) for analysis of technical reproducibility. This process was repeated for all 3 biological replicates.

6.3.9. Proteomics data processing and database searching

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by Spectra Mill (Agilent; v B.04.00.127). All MS/MS samples were analyzed using Spectrum Mill and X!Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). Spectrum Mill and X!Tandem were set up to search the SwissProt.BCG.Pasteur.1173P2 database (selected for all curated and non-curated proteins. Retrieved Apr 27 2014, 3891 entries) with tryptic digest fragments with an ion mass tolerance of 50 PPM and a parent ion tolerance of 20 PPM. Carbamidomethyl of cysteine and iTRAQ 8plex of lysine and the n-terminus were specified in Spectrum Mill and X!Tandem as fixed modifications. Ammonia-loss of the n-terminus, deamidated of asparagine and oxidation of methionine were specified in Spectrum Mill as variable modifications. Glu->pyro-Glu of the n-terminus, ammonia-loss of the n-terminus, gln->pyro-Glu of the n-terminus, deamidated of asparagine, oxidation of methionine, acetyl of lysine, carbamidomethyl of cysteine,

phosphorylation of serine, threonine and tyrosine and iTRAQ8plex of lysine and the n-terminus were specified in X!Tandem as variable modifications.

6.3.10. Criteria for protein identification

Scaffold (version Scaffold_4.3.0, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [25]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Annotated with GO terms from 25994.M_bovis_Pasteur_1173P2.goa (downloaded from NCBI May 5, 2014) was performed [26].

6.3.11. Relative protein quantification by iTRAQ

Scaffold Q+ (version Scaffold_4.3.4, Proteome Software Inc.) was used to quantify iTRAQ Label Based Quantitation peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [25]. Proteins sharing significant peptide evidence were grouped into clusters. Channels were corrected by the matrix as:

[0.000,0.000,0.929,0.0689,0.00220];

[0.000,0.00940,0.930,0.0590,0.00160];

[0.000,0.0188,0.931,0.0490,0.001000];

[0.000,0.0282,0.932,0.0390,0.000700];

[0.000600,0.0377,0.933,0.0288,0.000];

[0.000900,0.0471,0.933,0.0188,0.000];

[0.00140,0.0566,0.933,0.00870,0.000];

[0.000,0.000,0.000,0.000,0.000]; and

[0.00270,0.0744,0.921,0.00180,0.000]

in all samples according to the i-Tracker algorithm [27]. Acquired intensities in the experiment were globally normalized across all acquisition runs. Individual quantitative samples were normalized within each acquisition run. Intensities for each peptide identified were normalized within the assigned protein. The reference channels were normalized to produce a 1:1 fold change. All quantitative calculations were performed using means of the multiplicatively normalized data. Differentially expressed proteins were determined using one-way ANOVA using Bonferroni multiple testing correction.

6.3.12. Strain construction

Strains, derivatives of BCG (str. Pasteur 1173P2), used in this study are listed in **Supplementary Table 6.1**. *dosSR* knockout and complementation are constructed by the methods of Lin et al., Bartek et al., Bardarov et al. and Stover et al. [28-31]. Briefly, primers *dosR$_{hr}$*FR, *dosS$_{hr}$*FR and *dosRreg*FR (sequences in **Supplementary Table 6.1**) were used to amplify the *dosR* 3' flanking region, *dosS* 5' flanking region and *dosSR* promoter regions respectively (**Supplementary Figure 6.7A**). The promoter regions encompass all identified promoter and transcription factor binding elements

identified [32-34]. These regions were sub-cloned into TOPO$^{®}$ vector by TA

cloning (TOPO® TA cloning kit, Life Technologies), propagated in TOP10

chemically competent *E. coli*, and sequenced to select for vectors that

correctly amplified the PCR fragment. Inserts were excised by the appropriate

restriction enzymes and ligated into pYUB854 (for *dosS$_{hr}$*FR and *dosRreg*FR

inserts) or pMV306 (for *dosR$_{reg}$* FR insert) vectors using the DNA Ligation Kit

Mighty Mix (Takara) according to the manufacturer's instructions.

Transfection of WT BCG with pYUB854 containing the *dosR$_{hr}$*FR and

*dosS$_{hr}$*FR inserts generated the Δ*dosSR* strain. Δ*dosSR* was in turn

complemented with pMV306 containing the *dosR$_{reg}$*FR insert and 1 of 5

possible *dosSR* constructs (synthesized by g-blocks (IDT) and altered by site-

directed mutagenesis (Genescript). This generated the Δ*dosSR*::*dosSR(WT)*,

Δ*dosSR*::*dosSR(ACA)*, Δ*dosSR*::*dosSR(ACC)*, Δ*dosSR*::*dosSR(ACG)* and

Δ*dosSR*::*dosSR(ACT)* strains. Deletion at the correct locus was verified by

hygromycin resistance, gel electrophoresis of PCR products and by qPCR.

Successful complementation was determined by kanamycin resistance,

restriction mapping, sequencing and qPCR.

6.3.13. Reverse transcription-qPCR

Quantification of targeted mRNA sequences was performed as described in

[15], using primers in **Supplementary Figure 6.7A**. *sigA* served as internal

loading control.

6.3.14. Data processing and statistical analysis

Data from mass spectrometric measurements and gene expression were

deflated as fold changes against Log values. Logarithmic transformations

were used for datasets with skewed or wide distributions and indicated in their

respective figures legends. Data from bacterial growth were curve-fitted to an

order 5 polynomial curve using Excel's LINEST function (Microsoft).
Comparisons between 2 samples were made using the appropriate two-tailed
t tests after equality of variances were tested using F-tests. Comparisons
between multiple samples were determined using One-way or Two-way
ANOVA when comparing 1 or 2 factors respectively. Bonferroni's or Dunnett's
Multiple Comparison or Tukey's HSD post-hoc tests were used where it best
reflects how sample means were compared and stated in their respective
figure legends. These statistical tests were performed using Prism 5
(Graphpad). Unless otherwise stated, all data are represented as arithmetic
means ± SE. To aid interpretation on statistical significances, $p < 0.05$, $p < 0.01$ and $p < 0.001$ are denoted as *, ** and *** respectively.

Differential abundance of tRNAs, tRNA modifications and proteins was
analyzed by a random effects Bayes model using the BETR algorithm in MeV
(http://www.tm4.org/mev.html). Percentages of modified tRNA species are
determined by peak area normalization with response factors. Cluster
analysis was performed using two-way hierarchical clustering with average
distances and complete linkages and visualized using MeV. Shifts in
proteome of BCG were analyzed by principle component analysis (PCA)
using the NIPALS algorithm. Interpretations of the relationships between
codon usage predictors (codon frequency) and protein up-or down- regulation
($\log_2$ median fold change) was analyzed by partial least squares regression
(PLS). Alternate start and stop codons were treated as categorical variables.
Outliers which could cause over-fitting were removed by inspection of variable
residuals and leverages. Validation was performed using cross-validation and
the significance of variables determined by Marten's uncertainty test. The
Root Mean Square Error of Prediction (RMSEP), slope and correlation
coefficient of predicted versus measured correlation line was used to evaluate

the efficiency of applied regression model. Eigenvector based multivariate statistics performed using UnscramblerX® (v10.3, Camo).

6.4. Results

6.4.1. A systems-level analysis to characterize translational control of mycobacterial dormancy responses

Undoubtedly, coordinating gene expression to facilitate stress responses is essential for human pathogens like *M. tuberculosis* to survive the host immunity and persist in hypoxic lung granulomas [35, 36]. Hence, we sought to determine if tRNA modifications and codon usage bias, previously shown to be essential in the stress response of *S. cerevisiae* [5, 8, 19], play significant roles in the hypoxia-induced dormancy response of mycobacteria [37, 38]. To do this, we pursued a systems-level analysis of the mycobacterial dormancy response. This systems-level approach is divided into two parts (**Figure 6.1**). The first involves the exposure of *M. bovis* BCG to hypoxia to induce dormancy, extraction, purification and quantification of cellular RNA, qualitative and quantitative study of tRNA modifications and the mapping of these modifications. The second involves the global analysis of hypoxia-induced changes in the proteome, codon usage across all protein coding genes, targeted quantification of the tRNA pool and mRNA levels of key regulatory genes, such as *dosR*, followed by integrating both datasets to build quantitative models of genetic decoding and translation efficiency; and finally test our hypothesis of tRNA modification-mediated selective translation of codon biased, stress response genes using synthetic biology approaches.

6.4.2. Hypoxia induces a systemic reprogramming of tRNA modifications in BCG

For selective translation of codon-biased genes to occur, there needs to be
stress-induced shifts in either tRNA abundance or modification. These
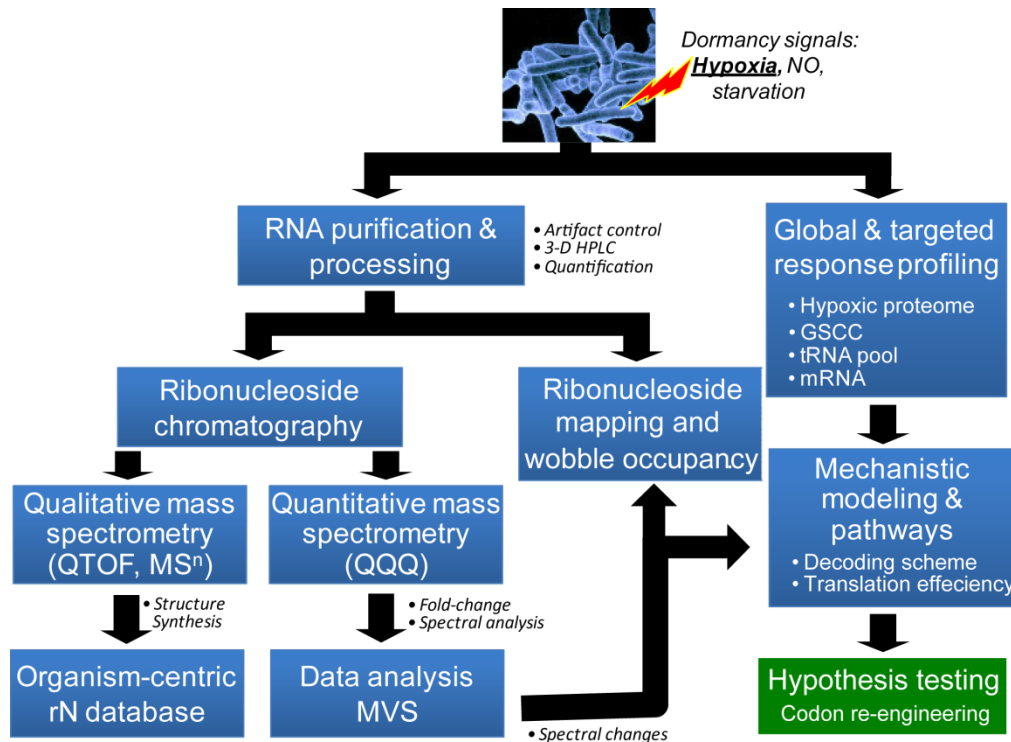changes would select groups of genes which over-or under uses synonymous

Figure 6.1. Experimental workflow for the systems-level analysis of translational control of hypoxia-induced dormancy responses. This analysis involves the induction of mycobacterial dormancy through hypoxia exposures [13, 39]; extraction, purification and quantification of cellular RNA [15, 17]; identification and characterization of all RNA modification in BCG building an organism-centric ribonucleoside (rN) database [40]; quantitative analysis of the dynamic changes of these modification under hypoxic stress by multivariate statistics (MVS) [18]; mapping key modification and determining wobble occupancy on tRNA species; global profiling of the hypoxic proteome and codon usage over time [3]; target analysis of the transcript levels of critical stress response regulators and the tRNAs that decode them; building mechanistic models of genetic decoding and selective translation and finally testing the predictions of this model through codon re-engineering experiments.

codons. Therefore, we began by characterizing the full repertoire of tRNA
modifications in BCG. Combining high mass accuracy with fast scan speed
liquid chromatography tandem mass spectrometry (LC-MS/MS) techniques
[40], we identified 40 distinct chemical modifications spanning a 1000-fold

dynamic range (**Figure 6.2A** and Appendix I). Next, we quantified the changes in these modifications as BCG entered and exited a non-replicating state by dynamic multiple reaction monitoring (dMRM) [18]. We found that each stage of hypoxic dormancy in a classical Wayne model experimental setup can be defined by a signature set of tRNA modifications [13]. In addition to exponentially dividing cells in aerated cultures (aerated growth -AG), dormant BCG (non-replicating persistence stages 1 and 2 – NRP1 and 2) possessing unique patterns of tRNA modifications (Log and R6, H6 and H9, H14, 18 and 21; representing days 0, 24, 6, 9, 14 and 21 of the experimental time course), hierarchical clustering (HCL) also distinguished two transitional phases at the entry, H4 (day 4), and exit ,R1-R3 (days 19 and 21), of hypoxia-induced dormancy. These phases we termed hypoxic transition (HT) and early resuscitation (ER) respectively (**Figure 6.2B** and **Supplementary Figure 6.1A**). Therefore, hypoxia induced tRNA reprogramming; and this could affect gene expression through selective translation of stress response transcripts.

6.4.3. *dosR*, the master regulator of the initial hypoxic response, is biased in Thr^{ACG} usage – a feature shared by Group I genes

One potential modification tunable transcript is *dosR*, a transcription factor that activates a 48 member regulon involved in development of the dormant, non-replicating state (**Supplementary Figure 6.1B**) [41, 42]. Gene-specific codon counting (GSCC) analysis showed that this gene is significantly enriched in Thr^{ACG} codons and impoverished in Thr^{ACC} codons (**Supplementaty Figure 6.3**). G-ending codons in 4-fold degenerate codon boxes (for Leu. Val, Ser, Pro, Thr and Ala) could be read by modified wobble uridines (**Supplementary Figure 6.2**), and interestingly, we observed that levels of 5-oxyacetyl-uridine (cmo^{5}U), noted for its ability to decode G-ending

codons [43, 44], increase by >350% at H4, 6 and 9 (Fig 1B). Expanding

GSCC to all 3951 protein-coding genes, we observed that *dosR* lies at the

periphery of a set of ~90 genes (Group 1 genes, **Supplementary Figure 6.3**),

sharing several codon biases including an over-usage of Thr^{ACG} and an

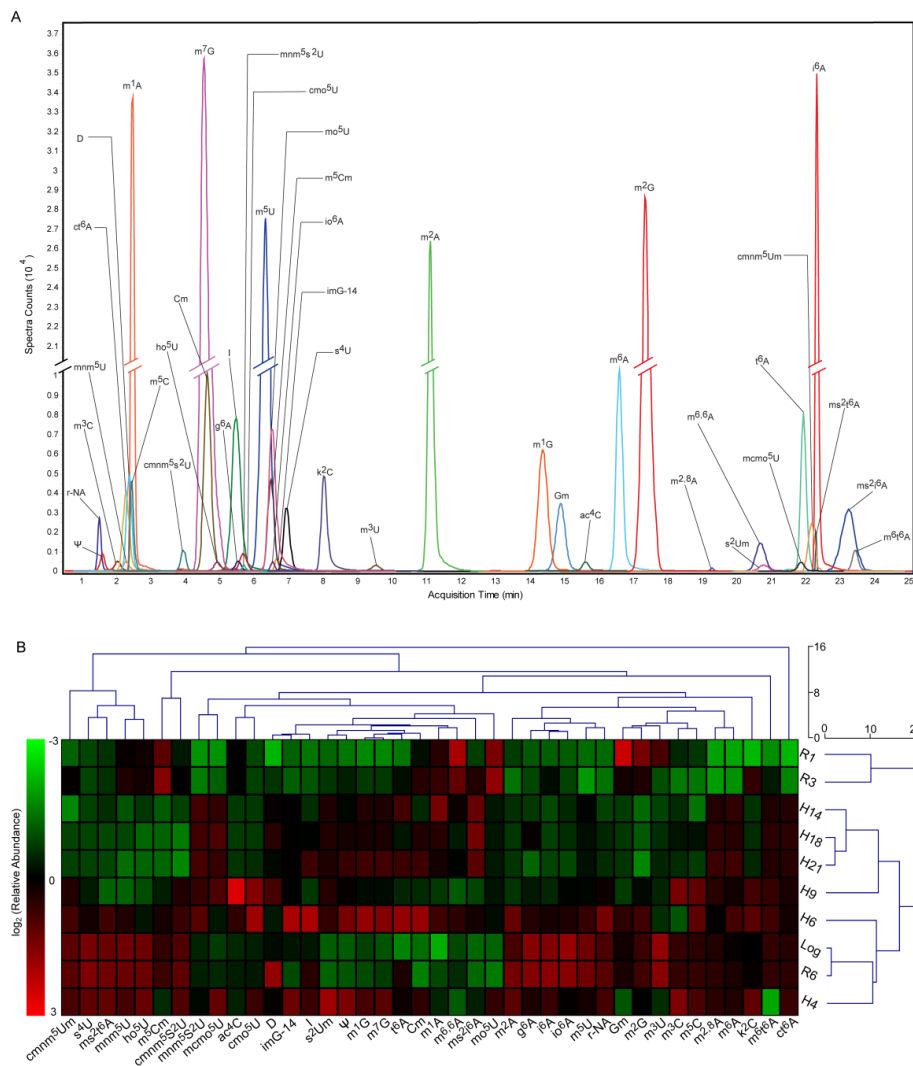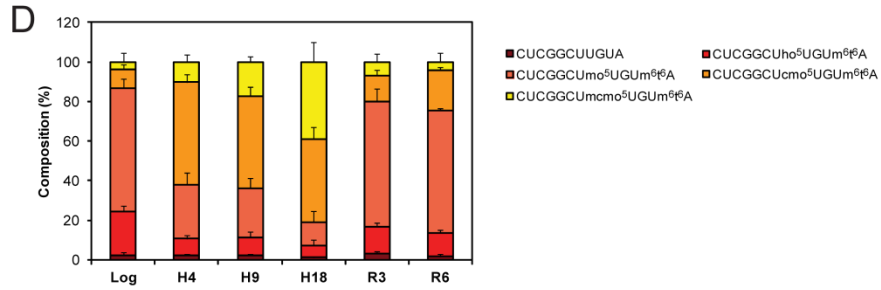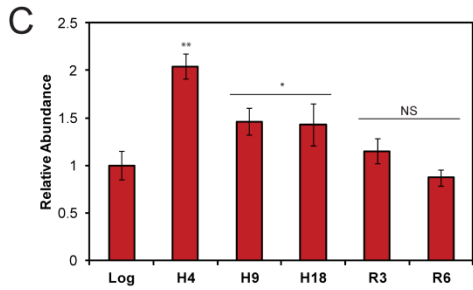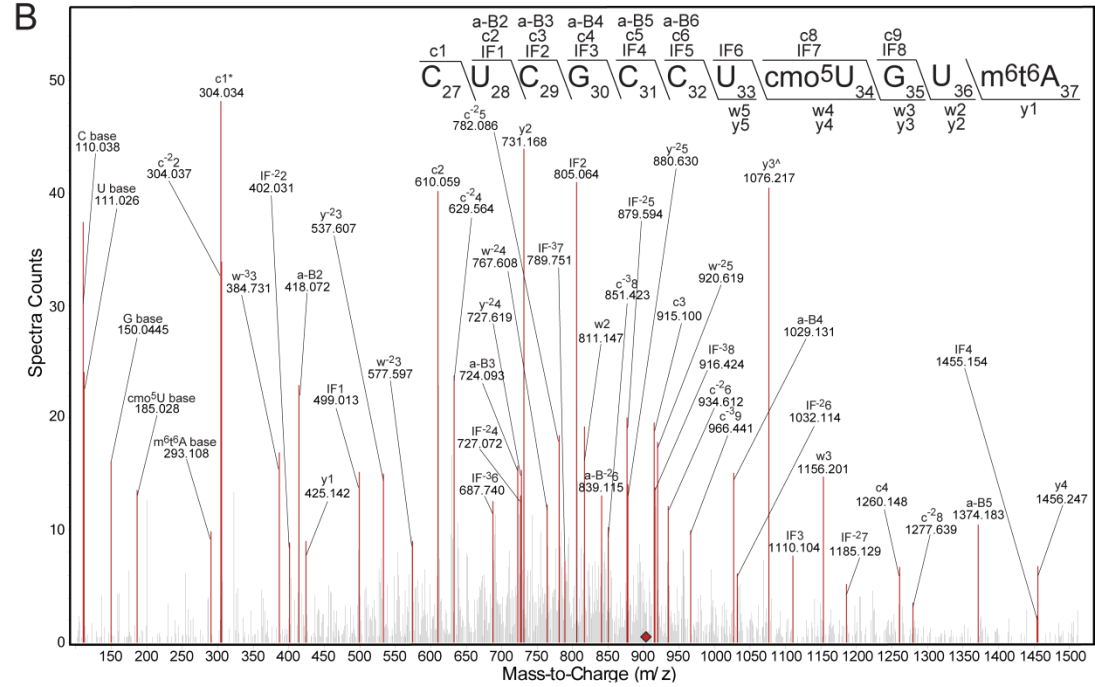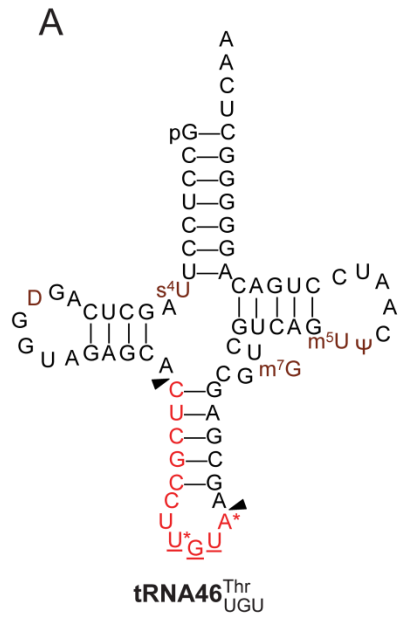under-usage of the optimal codon Thr^{ACC} (for Thr^{ACG} Cohen's d = 0.70; for

Figure 6.2. **Dynamics of tRNA modifications as BCG enter and exit hypoxic dormancy**. (**A**) Composite extracted ion chromatogram of 40 modified ribonucleosides in BCG tRNA. Full names, structures and LC-MS/MS characterizations found in Appendix I. (**B**) HCL analysis of changes in the relative levels of BCG tRNA modification induced by hypoxia and re-aeration on days 0, 4, 6, 9, 14, 18, 21 and 24 of the Wayne Model (**Supplementary Figure 6.1A**). HCL was performed on mean-centered data (n = 6) and visualized as a heatmap of log$_2$ fold-changes relative to Log cultures.

Thr^{ACC} Cohen's d = -1.19; z-test: p < 0.05). Thus, the BCG genome contains sets of genes with biases in codon usage which could be acted upon by differentially modified tRNAs.
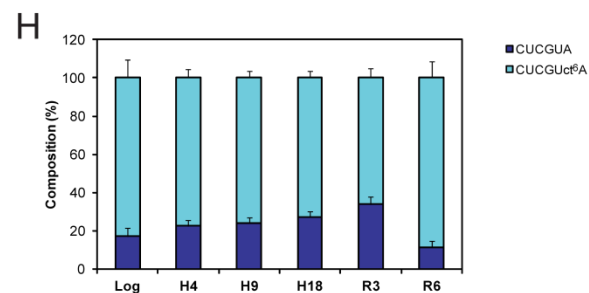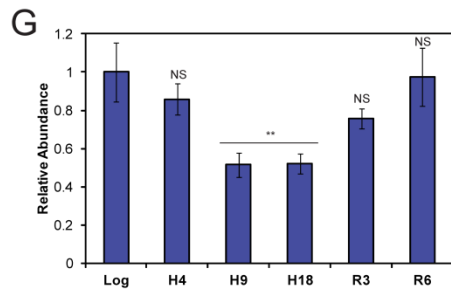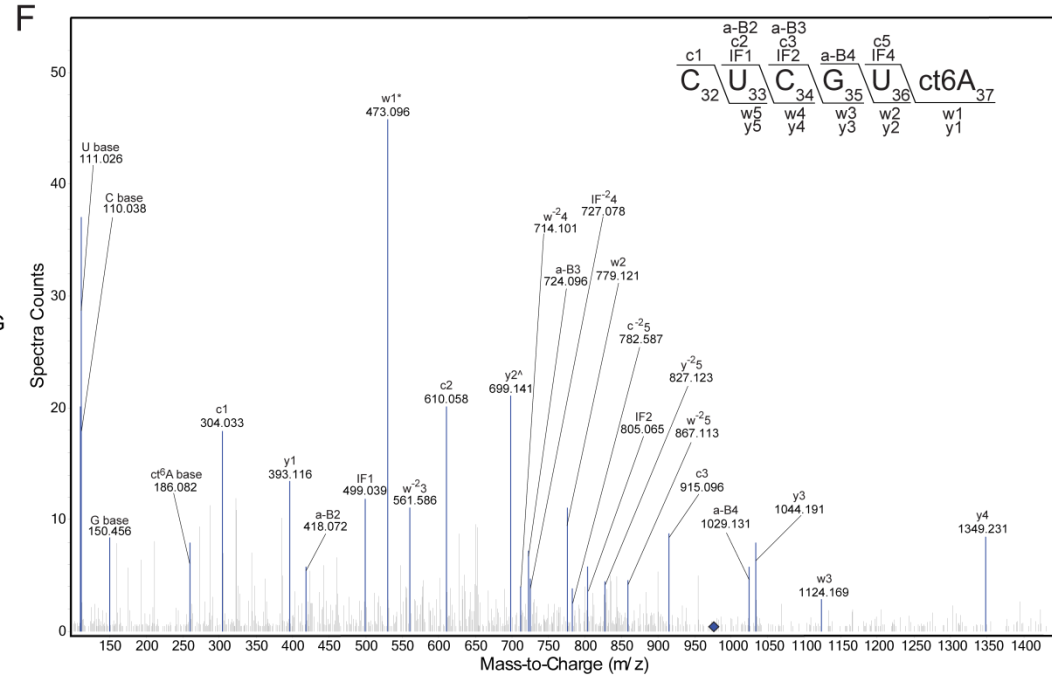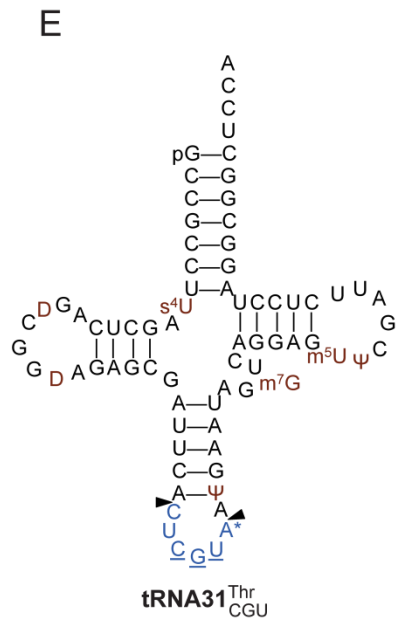
6.4.4. Remodeling of the tRNA^{Thr} pool during hypoxia

Abundances of tRNA modifications, however, are neither the sole determinants of translation elongation efficiency, nor are 5-oxyacetyled uridines the only wobble base capable of decoding G-ending codons. Conventionally, the effects of codon usage bias on translation have been attributed to of the copy numbers of available tRNAs [45, 46], and, contrary to experimental evidence [47], tRNA modifications are thought to be all-or-none events. This excludes the possibility of populations of tRNA isoacceptors possessing several different modification states. Furthermore, 5-methoxy-uridine (mo^{5}U) had also been identified in BCG and this modification has been found on the wobble position of tRNA$^{Thr(\underline{U}GU)}$ in *Bacillus subtilis* [8, 48]. To directly assess the tRNA^{Thr} pool, and determine the wobble occupancy of tRNA$^{Thr(\underline{U}GU)}$, we applied the principles of shotgun proteomics to sequence and quantify RNAse U2 digestion products which uniquely identify a tRNA isoacceptor (see **Materials and methods**). We appraised unique oligonucleotide fragments from the 3 tRNA^{Thr} isoacceptors in the BCG genome (**Figures 6.3A, E and I**) and found modest changes (from 0.5- 2.1 folds against Log levels) in tRNA abundances across the experimental time course (**Figures 6.3C, G and K**). The magnitude of these changes are in line with previous quantification of stress induced perturbations in tRNA copy number by tRNA-seq [49].

Moreover, we detected time-dependent changes in the extent of cmo^{5}U (**Figure 6.3B**), mo^{5}U and 5-hydroxy-uridine (ho^{5}U) modifications on position

E

A
C
C
U
pG—C
C—G
G—C
C—G
C—G
C—G
U—A   UCCUC U U A
s⁴U   A | | | | | G
C   D G A C U C G A   AGGAG   C
G | | | | | C U
G D A G A G C G   U A G m⁷G
A—U   m⁵U Ψ
U—A
U—A
C—G
A   Ψ
◄ C   A ►
U   A*
C U
G

tRNA31$_{CGU}^{Thr}$

F

G

H

I

tRNA7$^{Thr}_{GGU}$

J

K
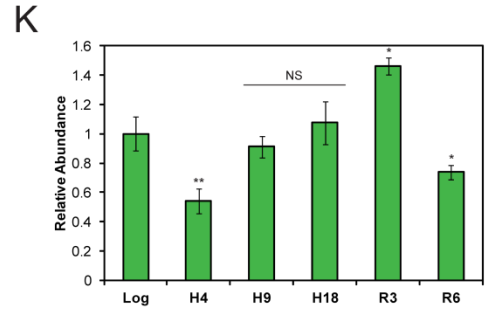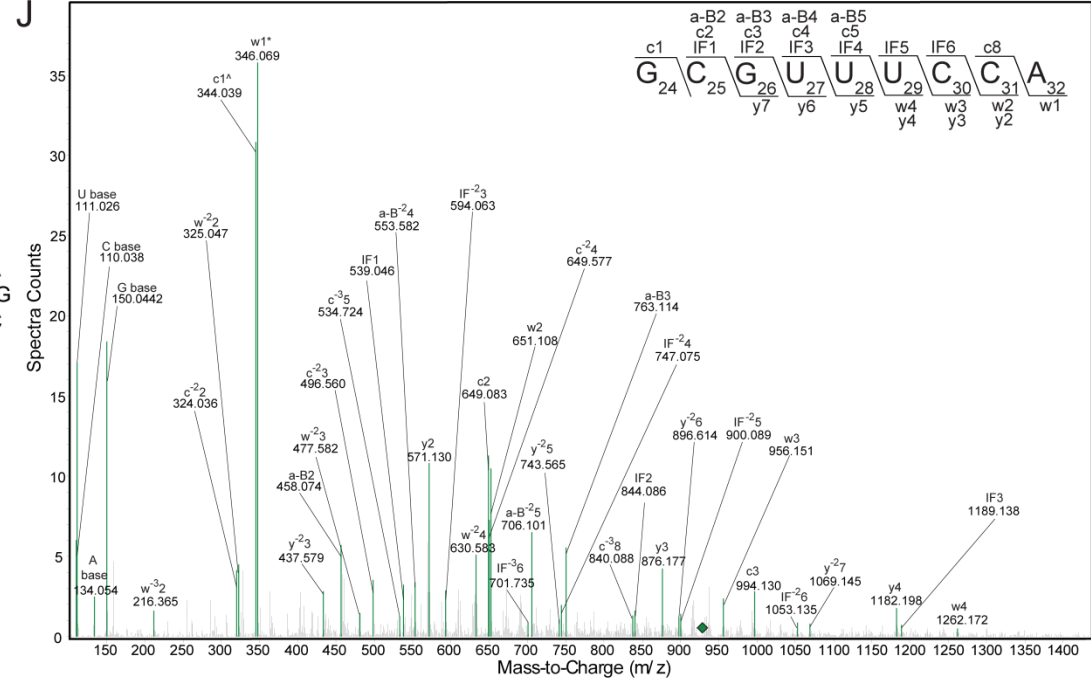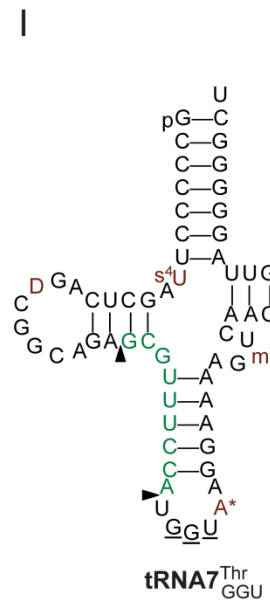
Figure 6.3. Hypoxia induces remodeling of the tRNA^{Thr} pool. Total tRNA was digested with RNAse U2 generating oligoribonucleotides containing unique fragments (arrows) from (**A**) tRNA46$^{Thr(\underline{U}GU)}$, (**E**) tRNA31$^{Thr(\underline{U}GU)}$ and (**I**) tRNA7$^{Thr(\underline{U}GU)}$ which were analyzed by LC-MS/MS. Positions of conserved 4-thiouridine (s^{4}U), dihydrouridine (D), pseudouridine (Ψ), 7-methylguanosine (m^{7}G), 5-methyluridine (m^{5}U), wobble (U*) and adenosine37 (A*) are shown in brown. (**B**) CID spectra of CUCGCCUcmo^{5}UGUmt^{6}t^{6}A (m/z 907.613, z = -4,◆), (**F**) CUCGUct^{6}A (m/z 979.644, z = -2,◆) and (**J**) GCGUUUCCA (m/z 929.121, z = -3,◆). a-B, c, w and y ions series presented as defined by nomenclature. IF: internal fragments. *: Quantifier ion. ^ Identifier ion. Abundances of (**C**) tRNA46$^{Thr(\underline{U}GU)}$, (**G**) tRNA31$^{Thr(\underline{U}GU)}$ and (**K**) tRNA7$^{Thr(\underline{U}GU)}$ as determined by targeted MRM. Values are fold changes relative to quantities at Log (mean ± SEM, n = 4; NS: not significant, p< 0.05, p < 0.01 and p < 0.001 are denoted as *, ** and *** respectively determined by One-way ANOVA with Dunnett's test vs Log). Ratios of (**D**) Modified oligonucleotides from the anticodon stem-loop of tRNA46$^{Thr(\underline{U}GU)}$ and (**H**) tRNA46$^{Thr(\underline{C}GU)}$ at indicated time points expressed as percentages of their sum total.

34 (wobble) of tRNA$^{Thr(\underline{U}GU)}$, and *N*6-methyl-*N*6-threonylcarbamoyladenosine (m^{6}t^{6}A) and cyclic *N*6-threonylcarbamoyladenosine (ct^{6}A) on position 37 of tRNA$^{Thr(\underline{U}GU)}$ and tRNA$^{Thr(\underline{C}GU)}$, respectively (**Figures 6.3D, F and H**). These results are reaffirmed by previous tRNA mapping studies and together indicate a widespread remodeling of the tRNA^{Thr} pool as cells enter and exit hypoxic dormancy [16, 48, 50].

6.4.5. Gene transcripts overusing Thr$^{AC\underline{G}}$ but not Thr$^{AC\underline{C}}$ are selectively translated during the hypoxia transition

Are the effects of tRNA^{Thr} pool remodeling reflected in the proteome? If so, transcripts, like *dosR*, which over-uses Thr$^{AC\underline{G}}$ while under-using Thr$^{AC\underline{C}}$ would be differentially translated before, during and after hypoxic exposures. When selective translation occurs, protein abundances would change independent of mRNA levels but in accordance with the availabilities of tRNA^{Thr} species capable of decoding them [1]. We performed 8-plex iTRAQ proteomics to compare the relative quantities of proteins across the experimental time course at Log, H4, H6, H9, H14, H18, R3 and R6. Cross-validating protein identifications between Spectrum Mill and X!tandem, we matched 122,546

spectra to 36,751 unique peptides which mapped onto 2,455 proteins (62%
proteome coverage at 4.9% peptide FDR) (**Supplementary Figures 6.5A
and B**). Of these, 965 could be consistently quantified by 2 or more peptides
in all samples at >99% confidence (protein FDR < 0.01%; no missing values)
(**Figure 6.4A** and **Supplementary Figure 6.5C**). These proteins are
distributed across all major gene ontology categories and principle
component analysis (PCA) of their relative abundances recapitulated all 5
stages of hypoxia-induced dormancy and resuscitations upon re-aeration
visualized by HCL of tRNA modification (**Figures 6.2B, 6.4A, 6.4B and
Supplementary Figures 6.5E-G**). Changes in the level of DosR protein
significantly diverge from its mRNA counterpart: whereas levels of *dosR*
transcripts drop rapidly after H4, DosR reached its maximal levels at H9 and
was maintained until re-aeration (post-H18). In contrast, changes in
tRNA$^{Thr(\underline{cmo5U}GU)}$ levels more closely matched those of DosR (**Supplementary
Figure 6.5H**). Moreover, *dosR* is not alone. We applied partial least squares
(PLS) regression to determine if there was a fundamental relationship
between the most significantly up- or down-regulated proteins (> 2 fold-
change and with $p < 0.05$, **Supplementary Figures 6.6A, C, D, G, J, M and
P**) and their codon usages. Interestingly at H4, not only could up- or down-
regulated proteins be distinguished based on their codon usages, up-
regulated proteins tended to use Thr$^{AC\underline{G}}$ over Thr$^{AC\underline{C}}$ and down-regulated
proteins the reverse ($0.83 < R^2 < 0.91$ for Factor-1 which contributes to 38%
of overall variance) (**Figures 6.4C,D** and **Supplementary Figure 6.6B**). At
other time points, codon choices between Asn^{AAT}/Asn^{AAC}, Asp^{GAT}/Asp^{GAC},
Cys^{TGT}/Cys^{TGC}, His^{CAT}/His^{CAC}, Lys^{AAA}/Lys^{AAG} and Tyr^{TAT}/Tyr^{TAC} contributed to
the separation between up- or down-regulated proteins (**Supplementary
Figure 6.6**). These results highlight the subtle but pervasive influence of
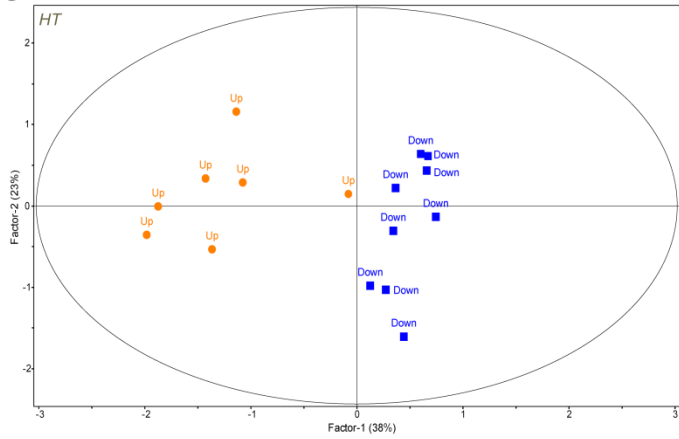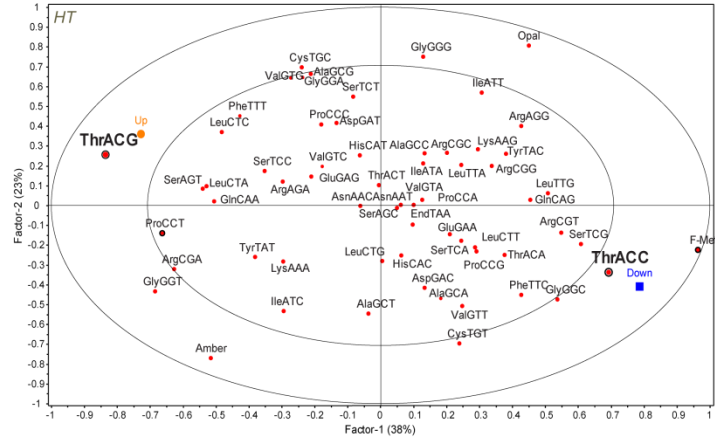codon usage on gene expression.
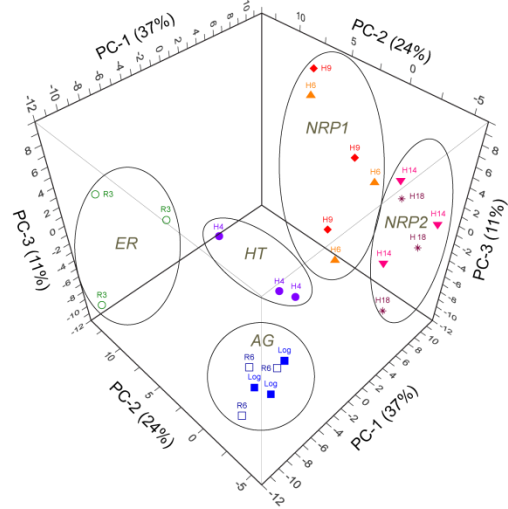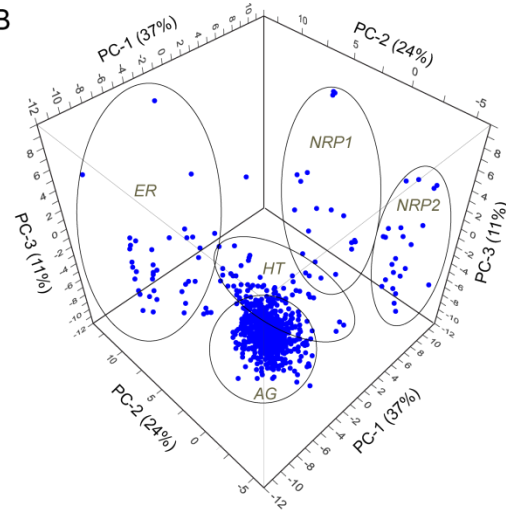
Figure 6.4. Choice between ThrACG amd ThrACC influences protein up- or down-regulation. PCA (**A**) scores and (**B**) loadings plots showing the clustering of samples (scores) based on fold changes in the 965 most quantifiable proteins (loadings). Groupings of sample eigenvectors (Log ■, H4●, H6▲, H9◆, H14▼, H18∗, R3○ and R6□) and protein variables (●) highlighted using data eclipses. These reflect growth phenotypes (**Supplementary Figure 6.1A**) and HCL sample clusters (**Figure 6.2B**). 72% of observed variance can be explained by 3 principle components (n = 3, PC-1: 37%, PC-2: 24%, PC-3: 11%). Acronyms for growth stage explained in text. PLS regression analysis of significantly up- (●) or down- (■) regulated proteins and their codon usages (●) at H4 visualized by (**C**) scores and (**D**) X,Y correlation loadings – representing codon usage (X) against extent of protein up- or down-regulation (Y). Protein selected show >2 fold-change and p < 0.05 (n = 3; unpaired, two-tailed t-test). Eclipse in (**C**) represents Hoteling T^2 limit at p-value of 0.05 (F-test) while those in (**D**) indicate the explained variance. Outer and inner ellipses indicate 100% and 50% explained variance respectively. Codons contributing significantly to the regression (cross validation, by Marten's uncertainty test) are circled in black. 61% of observed variance can be explained by 2 latent factors (n = 3, Factor1: 38%, Factor-2: 23%).

6.4.6. Choice between synonymous threonine codons affects *dosR* expression and hypoxia survival

To determine if the choice between Thr codons affects translation efficiency and the dormancy response, we generated a series of recombinant BCG with *dosR* using solely 1 of 4 possible codons (ACA, ACC, ACG and ACU) to encode threonine (See **Material and methods**, **Supplementary Figure 6.7A**). These are termed *ΔdosSR*::*dosSR(ACA) ΔdosSR*::*dosSR(ACC), ΔdosSR*::*dosSR(ACG)* and *ΔdosSR*::*dosSR(ACT)* respectively. *dosS* and *dosR* belong to the same operon and our previous RNAseq analysis confirmed that they are co-transcribed on the same mRNA (submitted to NCBI GEO). Knocking out both genes (*ΔdosSR*) is deleterious for mycobacteria during hypoxia and upon re-aeration, but its restoration through complementation under transcriptional control of its native promoters and transcription factors (*ΔdosSR*::*dosSR(WT)*) restored fitness (**Figure 6.5D** and **Supplementary Figure 6.7B**) [51]. Conversely, overexpression of *dosR* retarded growth [52]. In exponential phase growth, under aerobic conditions,

these synonymous Thr codon mutations were silent. We observed neither discrepancy in growth rate nor altered levels of *dosR* expression (**Figures 6.5A and B**). These mutants also possessed similar transcription efficiency for *dosR* relative to *dosS* (**Figure 6.5C**). However, at points in time when we observed shifts in translation efficiency for DosR relative to DosS – that is, at the hypoxic transition (from H4-H6) and at early resuscitation (From H18 - R3) (**Figure 6.5C**), the mutants diverged in fitness. More *ΔdosSR::dosSR(ACC)* and *ΔdosSR::dosSR(ACT)* bacilli were recovered at H6, suggesting that growth arrest at NRP1 was retarded. This momentary growth advantage was lost by H18 where a loss of viability brought the bacteria numbers (CFU/mL) back to wild type levels. Upon re-aeration, however, the ability to resuscitate *ΔdosSR::dosSR(ACC)* and *ΔdosSR::dosSR(ACG)* mutants was diminished and cultures had reduced viability at R6 (**Figure 6.5D**). *dosR* expression was higher than wild-type for *ΔdosSR::dosSR(ACA)* and *ΔdosSR::dosSR(ACG)* at H6 and H18, while *ΔdosSR::dosSR(ACC)* and *ΔdosSR::dosSR(ACT)* possessed lower than wild type *dosR* expression at H6. Intriguingly, while *dosR* expression was decreased for all mutants upon re-aeration, higher than wild type levels of *dosR* were observed for *ΔdosSR::dosSR(ACC)* and *ΔdosSR::dosSR(ACG)* (**Figure 6.5E**). Simultaneously, we tracked DosR activity by measuring *hspX* expression. *hspX* is a prominent member of the *dosR* regulon, and its induction by *dosR* slows growth [53]. We found that *hspX* expression mirrors that of *dosR* mRNA levels implying that all 4 synonymous Thr mutants produced functional DosR protein that was translated in a timely manner (**Supplementary Figure 6.7C**).

6.5. Discussion

The seemingly paradoxical phenotypes of these *dosR* synonymous threonine mutants under hypoxic stress could be explained by taking
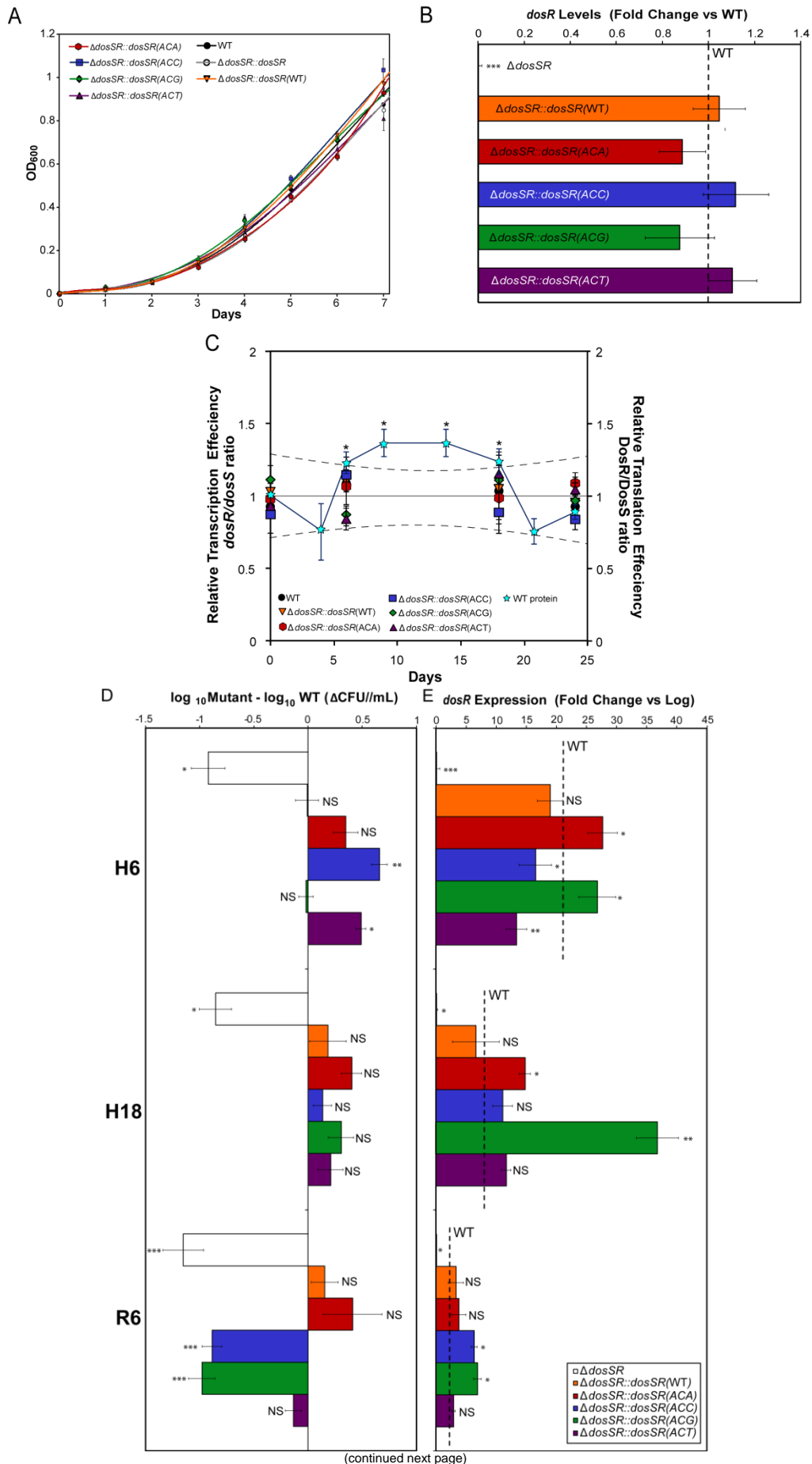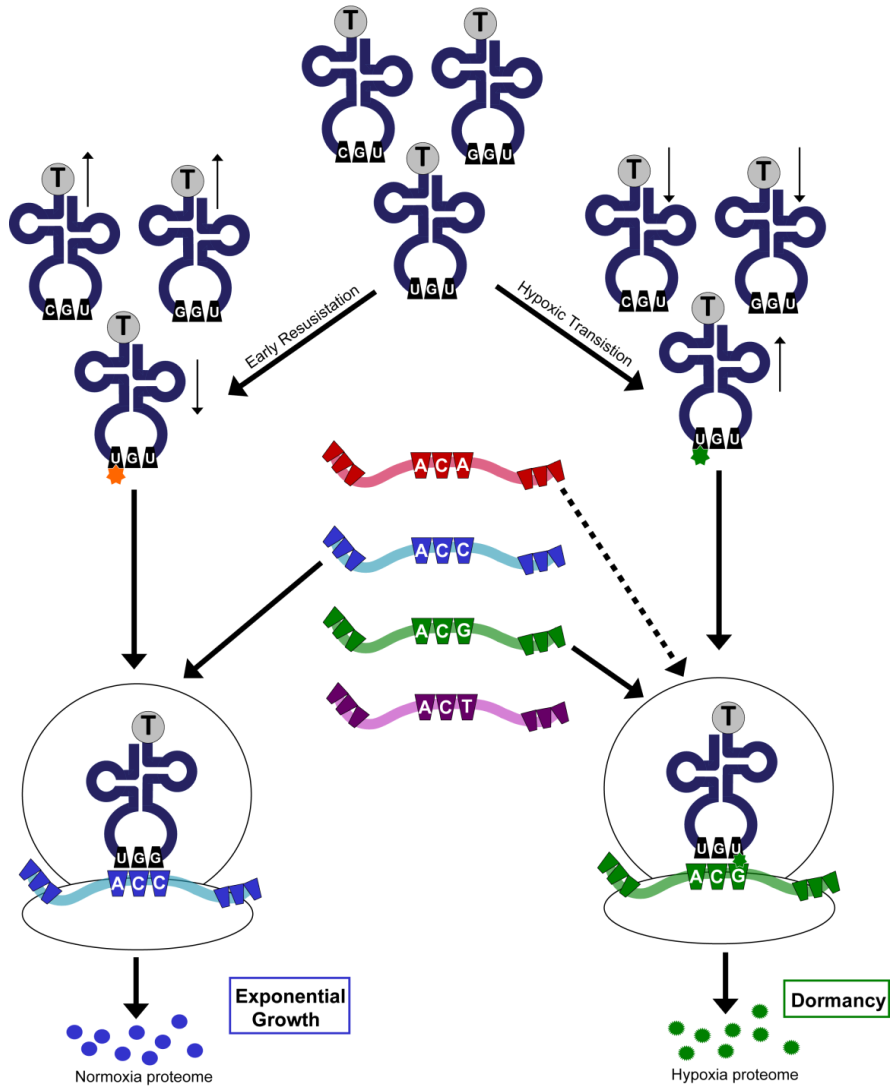
Figure 6.5. *dosR* **mutants re-engineered to used synonymous Thr codons showed altered growth phenotypes and *dosR* expression**. Notations used for each strain explained in text, figure S7A and table S2. (**A**) Strains had similar growth profiles and (**B**) *dosR* mRNA levels (save for Δ*dosSR* – negative control; One-way ANOVA with Tukey's HSD). Dashed line shows mean *dosR* expression in WT (**C**) Relative efficiencies of *dosR* transcription relative to *dosS* (black trace; left y-axis) and DosR translation relative to DosS (blue trace, right y-axis). Dashed lines show the 95% confidence boundary for *dosR*/*doss* ratios being y = 1 (by ANCOVA),thus DosR/DosS ratios beyond this boundary are diverging significantly from 1 (indicated by *). (**D**) Difference in fitness and (**E**) *dosR* expression of mutant strains relative to WT at H6, H18 and R6.(n ≥ 5; mean ± SEM; NS: not significant, $p < 0.05$, $p < 0.01$ and $p < 0.001$ are denoted as *, ** and *** respectively determined by two-way ANOVA with Bonferroni post-tests (vs. Log) considering interactions between mutations and hypoxia.) (**F**) Schematic for selective translation of codon biased transcripts where changes in tRNA abundance and modifications regulate protein abundance.

into account the changes induced by hypoxia in the tRNA^{Thr} pool and that DosR activation, through phosphorylation by DosS or DosT (**Supplementary Figure 6.1A**), is necessary and sufficient for *dosR* transcriptional up-regulation [32, 33]. At the hypoxic transition, (H4) tRNA$^{Thr(\underline{U}GU)}$ levels increased while levels of tRNA$^{Thr(\underline{C}GU)}$ and tRNA$^{Thr(\underline{C}GU)}$ decreased (**Figure 6.3**). tRNA$^{Thr(\underline{U}GU)}$, in turn, was reprogrammed by cmo^{5}U wobble modifications to preferably decode Thr^{ACG}. Thus, selective translation of transcripts over-using Thr^{ACG} was achieved. Reading of the cognate, but rare, Thr^{ACA} was permitted, but the ability to decode Thr^{ACC} and Thr^{ACU} was reduced with decreasing tRNA$^{Thr(\underline{G}GU)}$ levels. Hence, translation of ACG- and ACA-enriched but not ACC- or ACU-enriched *dosR* transcripts was enhanced. Since DosR positively regulates its own transcription [32, 41], we observed increased *dosR* levels for *ΔdosSR*::*dosSR(ACG)* and *ΔdosSR*::*dosSR(ACA)* mutants which induced *hspX* expression. This contributed to growth arrest at NRP1 (H6). Over-production of *dosR* in *ΔdosSR*::*dosSR(ACG)* mutants, however, impeded the resuscitation of cultures at R6. In early resuscitation (R3-R6), tRNA$^{Thr(\underline{U}GU)}$ and tRNA$^{Thr(\underline{C}GU)}$ returned to their Log-phase levels while tRNA$^{Thr(\underline{G}GU)}$ momentarily increased. Wobble uridines were methoxylated to mo^{5}U instead. These changes enhanced the reading of Thr^{ACC} codons while transcripts over-using Thr^{ACG} suffer from negative selection. *dosR* and *hspX* expression thus reverted to their Log-phase levels. In *dosSR*::*dosSR(ACC)* mutants, translation of ACC-enriched *dosR* was diminished at H6 but enhanced at R6, thereby, selective translation of *dosR* was mistimed, which led to reduced resuscitability (**Figure 6.5F**). Hence, we teased apart the effects synonymous mutations on gene expression and elucidated the molecular mechanism for selective translation of codon-biased genes through tRNA pool modulations.

The studies presented here also provide insights into the biosynthesis of mo^{5}U and cmo^{5}U in mycobacteria. In prokaryotes, cmo^{5}U is one of the most prevalent wobble base modifications involved in decoding 6 of the 8 fully degenerate 4-codon boxes – or up to 24 of the 61 amino acid coding codons (**Supplementary Figure 6.2)** [12, 44]. Our results indicate that wobble uridine modifications in the tRNA^{Thr} pool are in a state of flux and that the steady-state of wobble cmo^{5}U increases under hypoxia, which is in agreement with early studies in *E. coli*, *S. typhimurium*, and *B. subtilis* showing that cmo^{5}U is preferentially formed under anaerobic conditions (**Figure 6.3**) [54]. The precise biochemical pathway required for its formation remains a mystery. Chorismic acid is a key metabolite in its formation and the loss of its biosynthetic pathway (AroA-E) prevented the formation of wobble cmo5U [54]. Indeed, iTRAQ proteomics detected a spike in chorismate synthase (AroC) levels at the hypoxic transition (between H4-H9) – reaching over ~1.5 folds of its levels at Log (Appendix II). How chorismate used in the synthesis of cmo^{5}U is unknown. It was assumed based on the phenotype of *cmoB* mutants which possessed wobble ho^{5}U but not cmo^{5}U or mo^{5}U that the enzyme CmoB methylates ho^{5}U to mo^{5}U which is then carboxylated by a second enzyme, CmoA, thought to be a *S*-adenosyl-methionine (SAM) dependent methyltransferase, to cmo^{5}U (**Supplementary Figure 6.8A**) [55, 56]. In agreement, null mutation of CmoA results in the accumulation of tRNA containing mo^{5}U34 and ho^{5}U34 [55]. However, recent structural studies on CmoA and CmoB revealed that chorismate is presumably converted to prephenate which is used by CmoA to generate the ylide carboxy-*S*-adenosyl-*L*-methionine (Cx-SAM) [56, 57]. CmoB then uses the sulfonium heteroatom to directly acetylate ho^{5}U in a single step reaction without the formation of mo^{5}U intermediate [57]. This finding together with our observation that levels of mo^{5}U and cmo^{5}U move in opposite trends strongly

indicates that either CmoB itself in the absence of Cx-SAM, or a third hitherto undiscovered enzyme, can catalyze the methylation of ho^{5}U to mo^{5}U in competition to the tRNA 5-oxyacetyluridine(34) synthase activity of CmoB (**Supplementary Figure 8B**). Hypoxia shifts the equilibrium between these two competing pathways towards the formation of cmo^{5}U by modulating CmoA activity. Though these had not been annotated in BCG, however, we noted that their closest homologs BCG_0612[1] (55% coverage of CmoA, E-value = 4e^{-4} by blastp) and BCG_2975c (51% coverage of CmoB, E-value = 2 e^{-5} by blastp) increased by ~6.4 and ~1.5 folds respectively in the hypoxic transition (Appendix II).

At any one point in time, there are between 3000-8000 mRNA molecules in a bacterial cell [58, 59]. With copy numbers as low as 0.6 - 2 per gene, finding the appropriate transcripts (such as those for *dosR* and *cmoA*) for translation under the pressure of a rapidly changing external environment could prove to be a monumental task. In this chapter, we showed that BCG elegantly scheduled transcripts for translation by reprogramming tRNAs to better read specific codons over their synonymous counterparts. Genes, such as *dosR*, using a set of complementary codons are transiently prioritized and their translation enhanced. Hence, while the genetic code is universal, genetic coding and decoding are not; and choices like those between codons like Thr^{ACG} and Thr^{ACC} help coordinate stress responses through selective translation. The challenge, once again, is to convert these newfound insights into drug development approaches and practical clinical applications.

6.6. Supplementary material

6.6.1. Supplementary figures

---

[1] We also note that BCG_0612 is heavily biased in ACA usage (Z = 2.75, p < 0.01) – a rare codon decoded solely by tRNA$^{Thr(UGU)}$ and its modified derivatives (**Supplementary Figure 6.2**) and preferably read during the hypoxic transition (**Figure 6.5F**).
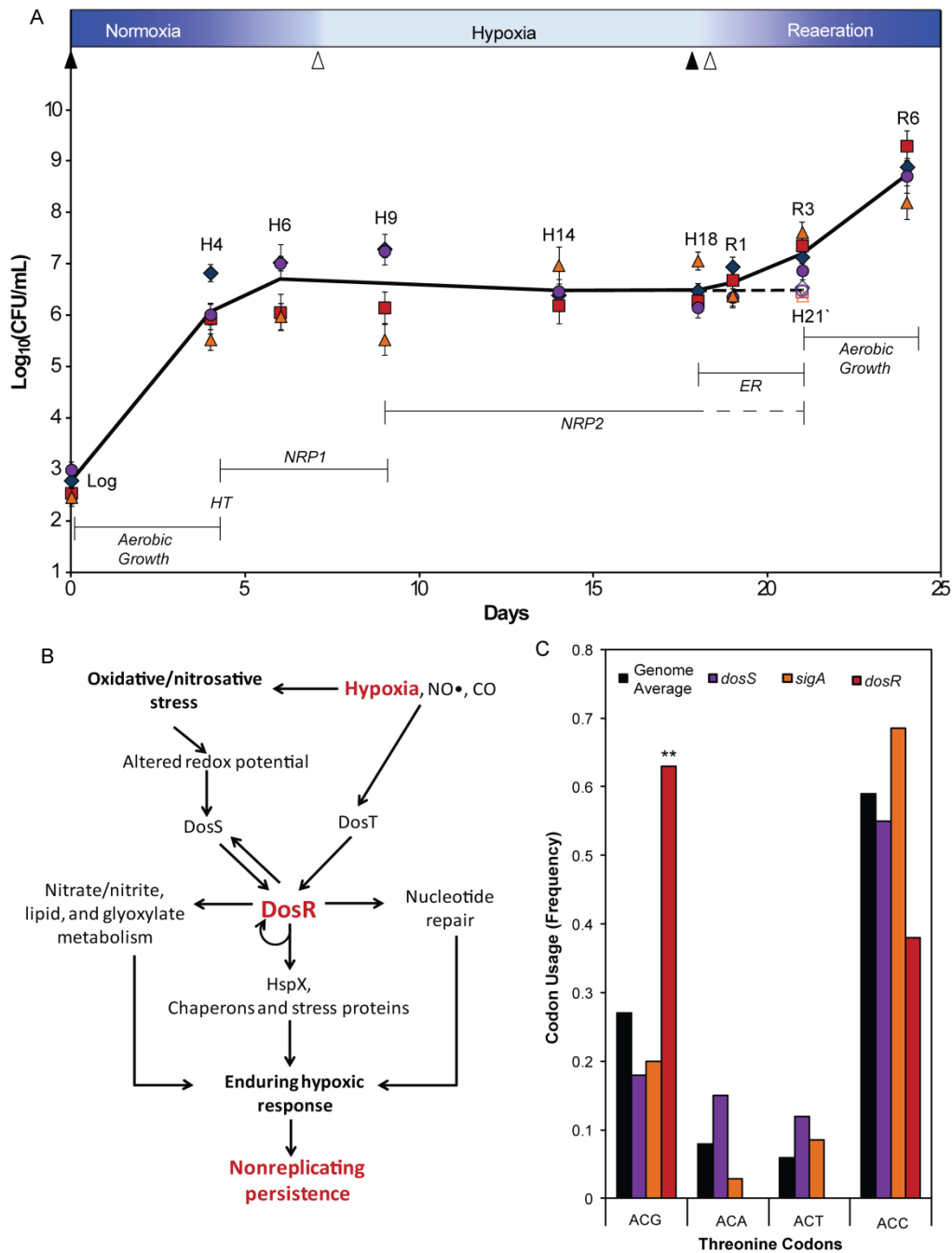
Supplementary Figure 6.1. Hypoxia-induced dormancy in *M. bovis* BCG.

(**A**) Growth of BCG in the Wayne culture system adapted by Ravindran et al. [14]. Diamonds, circles, squares and triangles each represent the means of ≥3 biological replicate cultures grown in the same batch. Error bars represent relative error of means. Growth curve plotted by joining geometric means of all samples across time points. Top panel: Intensity plot of methylene blue decolorization upon reduction. The formation of leucomethylene blue was tracked by absorption at 665 nm and correlated to color intensity between blue (100% - initial A_{665}) and white (0% - A_{665} = 0). Flasks were sealed and opened at days 1 and 18 (black arrows), complete methylene blue decolorization and recolorized was observed at days 7 and 18.5 (white

arrows). To determine if cultures remained dormant and viable if flasks remained sealed, one flask from each batch remained sealed for another 3 days till day 21 (denoted H21`). Time points are denoted as Log (day 0), H4 (day 4), H6 (day 6), H9 (day 9), H14 (day 14), H21` (day 21- dashed lines), R3 (day 21) and R6 (day 24). Stage of growth are denoted in *italics* aerobic growth, *HT* (hypoxic transition), *NRP2 (*non-replicating persistence stages 1), *NRP2* (non-replicating persistence stages 2) and *ER* (early resuscitation). No statistical changes in CFU/mLwere detected for H4, H6, H9, H14, H18, R1, R3 and H21` samples by one-way ANOVA with Tukey's HSD.

(**B**) Model for the regulation of hypoxia-induced non-replicating persistence by DosR. Under hypoxic conditions, DosS and DosT are sensor histidine kinases that activate DosR. Induction of DosR up-regulates the dos regulon (which *dosR* and *dosS* are members of), leading to changes in cellular metabolism, protein homeostasis and nucleotide repair. A subsequent enduring hypoxic response is induced, enabling the cell to enter non-replicating persistence [41, 42, 60-62]. Adapted from **Figure 1.2**.

(**C**) Usage frequencies of threonine codons, ACG, ACA, ACT and ACC, across all protein-coding genes in the genome and in genes *dosS*, *sigA* and *dosR*.

1st Codon Base (5' end) / 2nd Codon Base	U	C	A	G	3rd Codon Base (3' end)
U	Phe ■G(m)AA Leu ■Xnm5(s2)U(m)AA □CAA	Ser ■GGA ■Xo5UGA ■CGA	Tyr ■(G?)UA Ochre Amber	Cys ■GCA Opal Trp ■C(m)UC	U C A G
C	Leu ■GAG ■Xo5UAG □CAG	Pro ■GGG ■Xo5UGG □CGG	His ■(G?)UG Gln ■Xnm5(s2)U(m)UG □CUG	Arg ICG■ ■CCG	U C A G
A	Ile ■GAU ●k2CAU Met ■ac4CAU ■CAU^imet	Thr ■GGU ■Xo5UGU □CGU	Asn ■(G?)UU Lys ■Xnm5(s2)U(m)UU ■CUU	Ser ■GGA Arg ■Xnm5(s2)U(m)CU ■CCU	U C A G
G	Val □GAC ■Xo5UAC □CAC	Ala ■GGC ■Xo5UGC □CGC	Asp ■(G?)UC Glu ■Xnm5(s2)U(m)UC ■CUC	Gly ■GCC ■Xnm5(s2)U(m)CC ■CCC	U C A G

(Legends- next page)

Supplementary Figure 6.2. Presumptive genetic decoding in BCG.

In this coding chart, symbols connected with lines indicate codons which are read by the same tRNA. Stacked symbols in AUH and GCN boxes represent isodecoder pairs for tRNA$^{Ile(\underline{G}AU)}$ and tRNA$^{Ala(\underline{U}CC)}$ respectively. Squares denote cognate Watson-Crick base pairings while circles denote non-cognate pairings. Filled symbols denote productive pairings between codon and anticodons that had been validated experimentally in prokaryotes [44, 63-67]. Codon-anticodon pairings assigned based on wobble rules of Crick [68], and observations of Yokoyama and Nishimura [69], Agris et al. [70], Grosjean et al.[12] and Roth [71]. Xnmn^5(s^2)U$_m$: 5-iminomethyl-U$_{34}$ family of hypermodified uridines in which mnm^{5}U, mnm^{5}s^{2}U, cmnm^{5}s^{2}U and cmnm^{5}U$_m$ had been found in BCG tRNA. Xo^{5}U: 5-oxyU$_{34}$ family of hypermodified uridines in which ho^{5}U, mo^{5}U, cmo^{5}U and mcmo^{5}U had been found in BCG tRNA. (G?): Quenosine (G$_{34}$) family of hypermodified guanosines; however no members had been detected in this study. Full names, chemical structures and LC-MS/MS characteristics of all modifications can be found in table S1. Amber, Ochre and Opal are stop codons. Only F-met^{ATG}, the cannonical start codon, is featured in this scheme.
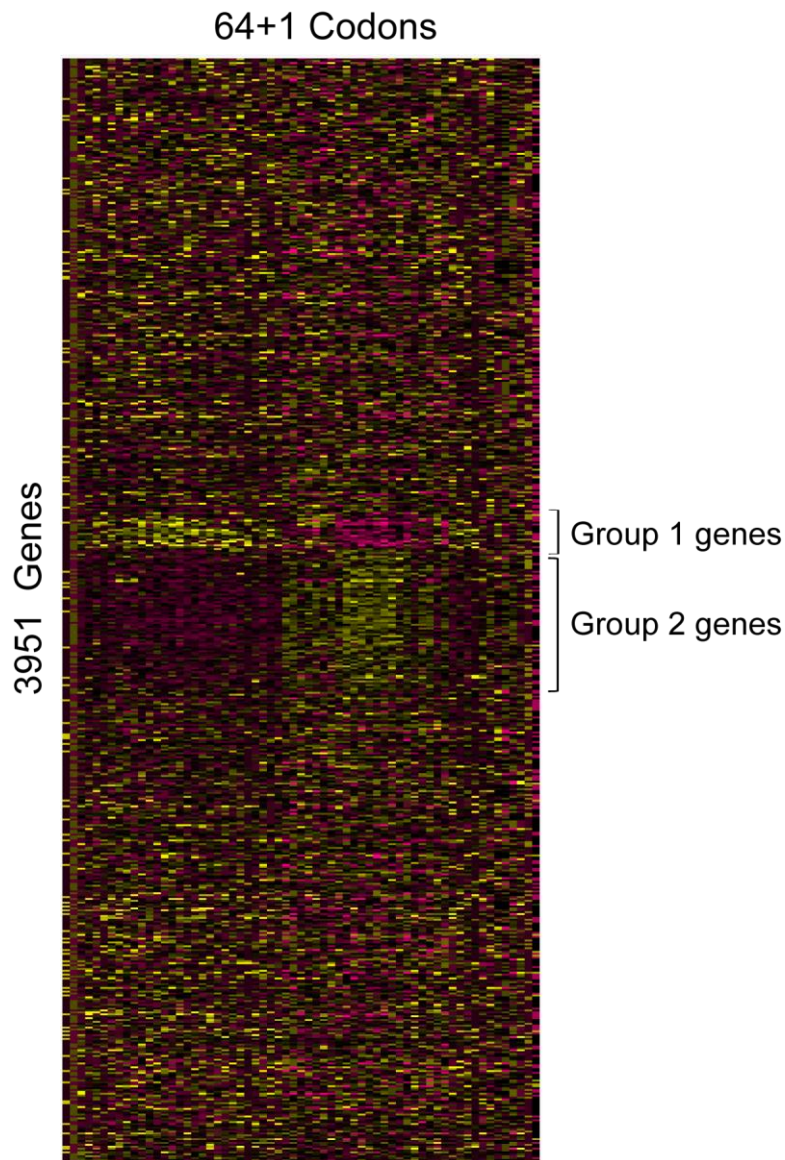
Supplementary Figure 6.3. Heat map of codon usage patterns across the BCG transcriptome.

The gene-specific codon counting (GSCC) algorithm was used to analyze all 3951 protein-coding genes in BCG for all codons choices (including alternate start '+1' and stop codons) [72]. Visualization by heat map of over-(yellow) and under-(purple) used codons revealed 580 open reading frames that possessed codon usages that significantly deviated (Cohen's d > 0.5) from the genome average. Group 1 genes overuse Ala^{GCT}, Asn^{AAT}, Asp^{GAT}, Cys^{TGT}, Gly^{GGT}, His^{CAT}, Ile^{ATA}, Ile^{ATT}, Leu^{TTA}, Leu^{TTG}, Phe^{TTT}, Pro^{CCT}, Thr^{ACG}, Tyr^{TAT}, Val^{GTT} while under-using Ala^{GCC}, Arg^{CGC}, Asn^{AAC}, Asp^{GAC}, Cys^{TGC}, Gly^{GGC}, His^{CAC}, Ile^{ATC}, Leu^{CTC}, Leu^{CTG}, Phe^{TTC}, Pro^{CCC}, Thr^{ACC}, Tyr^{TAC}, Val^{GTC}. Group 2 genes overuse Ala^{GCC}, Arg^{CGC}, Asn^{AAC}, Asp^{GAC}, Gly^{GGC}, His^{CAC}, Ile^{ATC}, Leu^{CTG}, Phe^{TTC}, Thr^{ACC} while under-using Ala^{GCA}, Ala^{GCT}, Arg^{CGA}, Asn^{AAT}, Asp^{GAT}, Gly^{GGT}, His^{CAT}, Ile^{ATA}, Ile^{ATT}, Leu^{TTG}, Leu^{CTT}, Phe^{TTT}, Pro^{CCA}, Pro^{CCT}, Ser^{AGT}, Thr^{ACA}, Thr^{ACG}, Thr^{ACT}, Val^{GTA} and Val^{GTT}. 12 of these codon choices feature as determinants of protein up- or down-regulation during hypoxic and normoxic shifts (**Figure 6.4C and Supplementary Figure 6.5**).
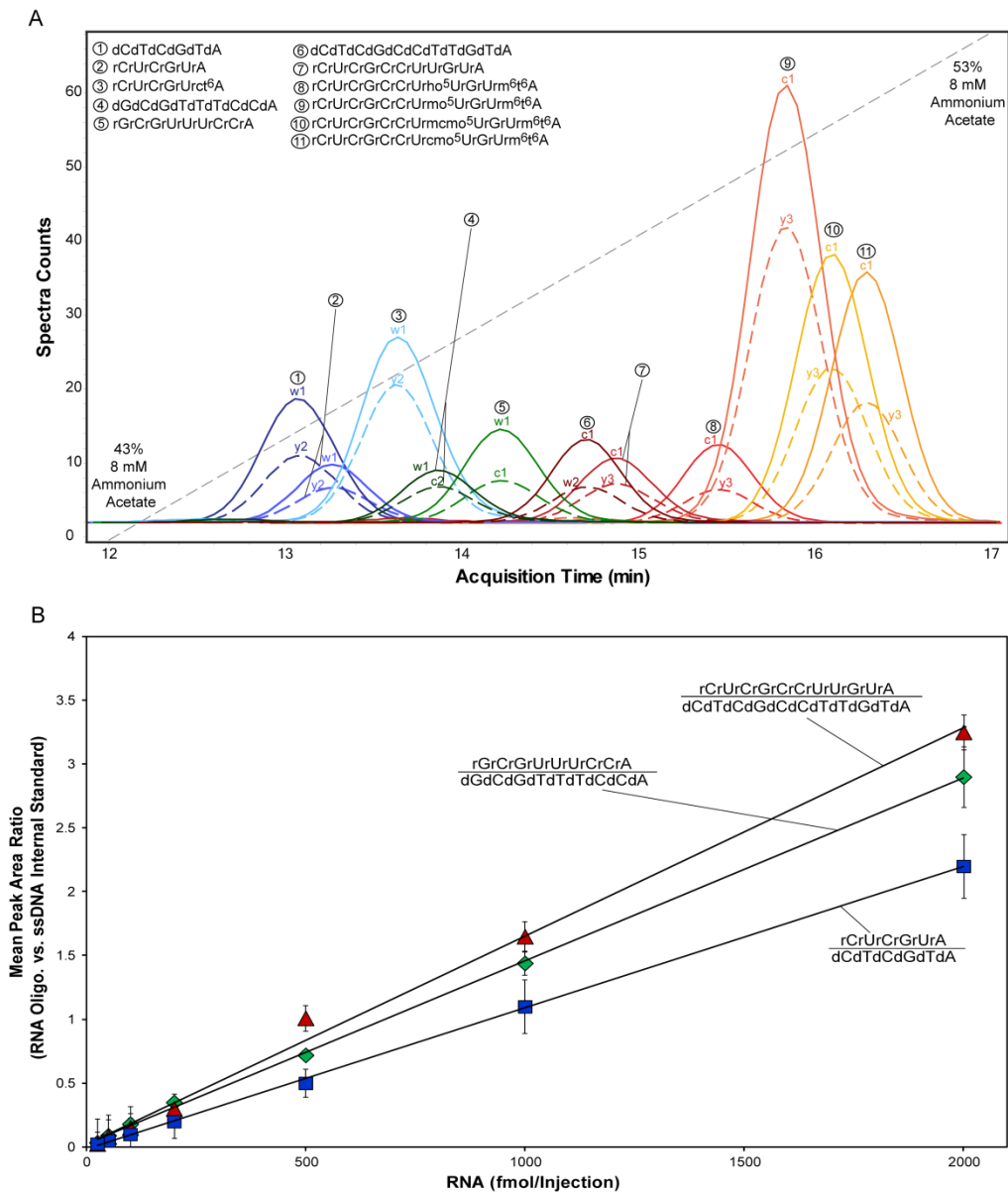
A

① dCdTdCdGdTdA
② rCrUrCrGrUrA
③ rCrUrCrGrUrct^{6}A
④ dGdCdGdTdTdCdCdA
⑤ rGrCrGrUrUrUrCrCrA

⑥ dCdTdCdGdCdCdTdTdGdTdA
⑦ rCrUrCrGrCrCrUrUrGrUrA
⑧ rCrUrCrGrCrCrUrho^{5}UrGrUrm^{6}t^{6}A
⑨ rCrUrCrGrCrCrUrmo^{5}UrGrUrm^{6}t^{6}A
⑩ rCrUrCrGrCrCrUrmcmo^{5}UrGrUrm^{6}t^{6}A
⑪ rCrUrCrGrCrCrUrcmo^{5}UrGrUrm^{6}t^{6}A

53%
8 mM
Ammonium
Acetate

43%
8 mM
Ammonium
Acetate

B

Figure 6.4. Label-free absolute quantification of the tRNA^{Thr} pool in BCG.

(**A**) Composite extracted ion chromatograms of the MRM transitions used to quantify unique RNAse U2 fragments which define each tRNA^{Thr} species. Oligonucleotides labeled 1-11 by sequence identity (Precursor ion). Each peak is labeled with the identity of its product ion monitored. Dashed line shows the elusion gradient of ammonium acetate.

(**B**) Calibration curves for the response of RNA oligonucleotides against 1 pmol of their DNA counterparts of the same sequence. Calibration was performed across 2 orders of magnitude (25-2000 fmol RNA). The mean peak area ratios were fitted to linear equations where correlation coefficient (R^2) was determined to >0.995 for all 3 curves. The reproducibility of the instrument response was studied by triplicate injections, showing RSD < 11 % for all concentrations. Response factors were calculated from the gradient of these curves. MRM transitions used are found in **Supplementary Table 6.1**.
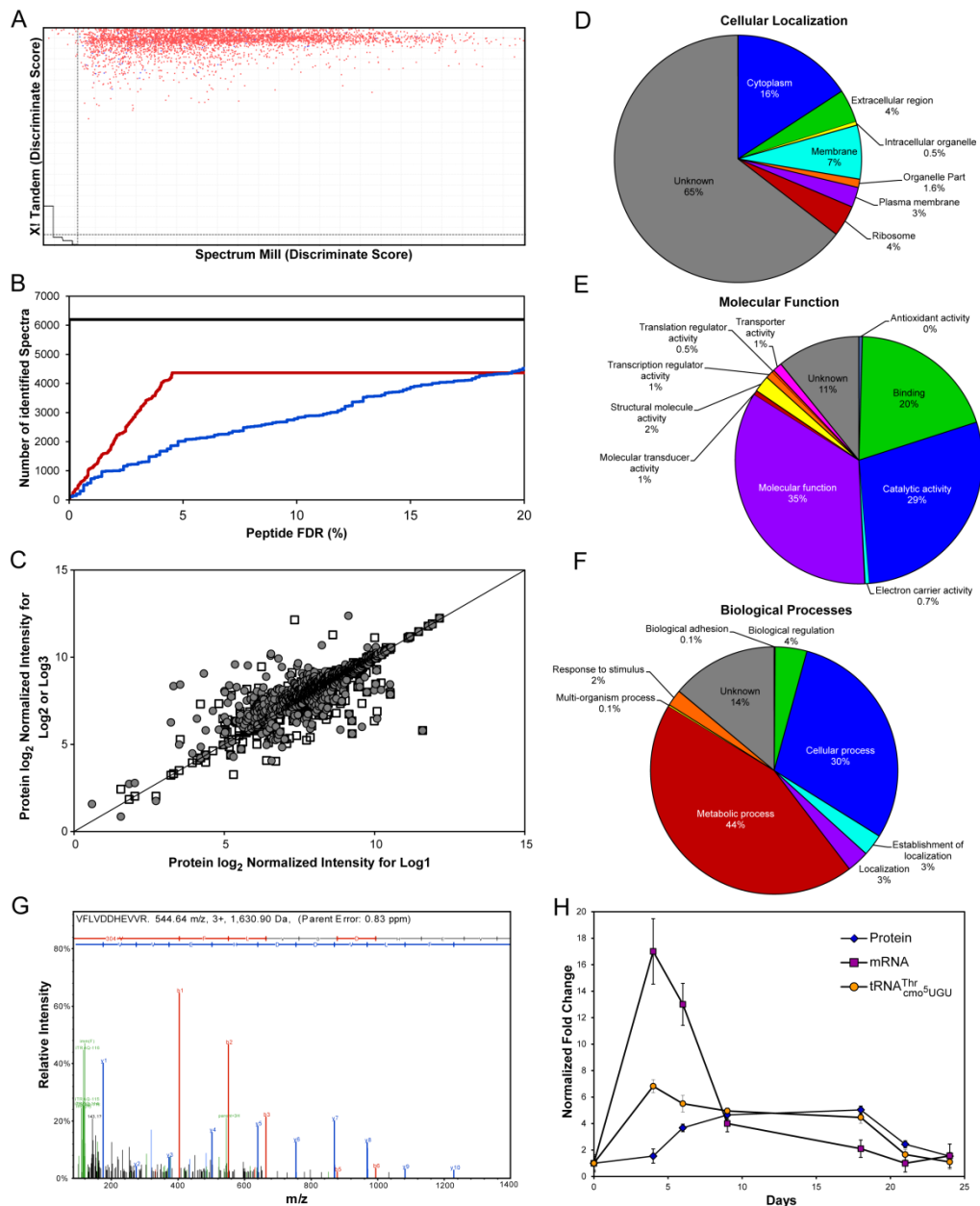
Supplementary Figure 6.5. iTRAQ-based proteomic analysis of BCG entering and emerging from hypoxia-induced dormancy.

(**A**) Scatterplot of spectrum to peptide matches between X!Tandem and Spectrum Mill. The x and y axes of the scatter plot are the search engines' discriminant scores determined by Scaffold. Vertical and horizontal dashed lines define the 95% peptide probability filer. Correct peptide to protein assignments are in red; incorrect in blue.

(**B**) Peptide receiver operator curve (ROC) plot for the selection of global peptide false discovery rate (FDR). Sensitivity is defined as the number of correctly identified spectra (y-axis; true positives); fall-out is defined as the peptide false discovery rate (FDR) (x-axis, false positives). Blue, red and black traces are the ROCs for Spectrum Mill, X!Tandem and combined Spectrum Mill and X!Tandem scores respectively. At 4.9% FDR (plateau in X!Tandem scores), 2455 proteins were identified.

(**C**) Scatter plot for intensities of the iTRAQ labels (randomized for each biological replicate) for all proteins in Log samples. The x-axis plots the quantitative values for the first biological replicate; y-axis plots the quantitative values for the second (□) and third (●) biological replicate. The 965 most quantifiable proteins fall within the 95% confidence interval of the optimal 45° reporter ion correlation line which represents a 1:1 ratio between the selected quantitative samples.

(**D-F**) Gene ontology analysis of the 965 most quantifiable BCG proteins. The classification of the protein set was performed according to the 1st level gene ontology terms: (**D**) "Cellular component", (**E**) "Molecular function" and (**F**) "Biological process". Pie charts show the breakdown according to 2nd level classifications. Molecular function in (**E**) could be further parsed into categories for binding, electron carrier activity, enzyme regulator activity, guanyl-nucleotide exchange factor activity, metallochaperone activity, nutrient reservoir activity, receptor activity, receptor regulator activity.

(**G**) MS^2 spectra of VFLVDDHEVVR, 1 of 11 peptides which uniquely identifies DosR. *b* and *y* ion series shown in red and blue respectively. Relative quantities of DosR across each experimental time point were defined by the iTRAQ intensities in green.

(**H**) Changes in the abundances of DosR protein, *dosR* mRNA and tRNA$^{Thr(\underline{cmo5U}GU)}$ across the experimental time course (Fig. S1A). All abundances normalized to their values at day 0 (Log) as fold changes.
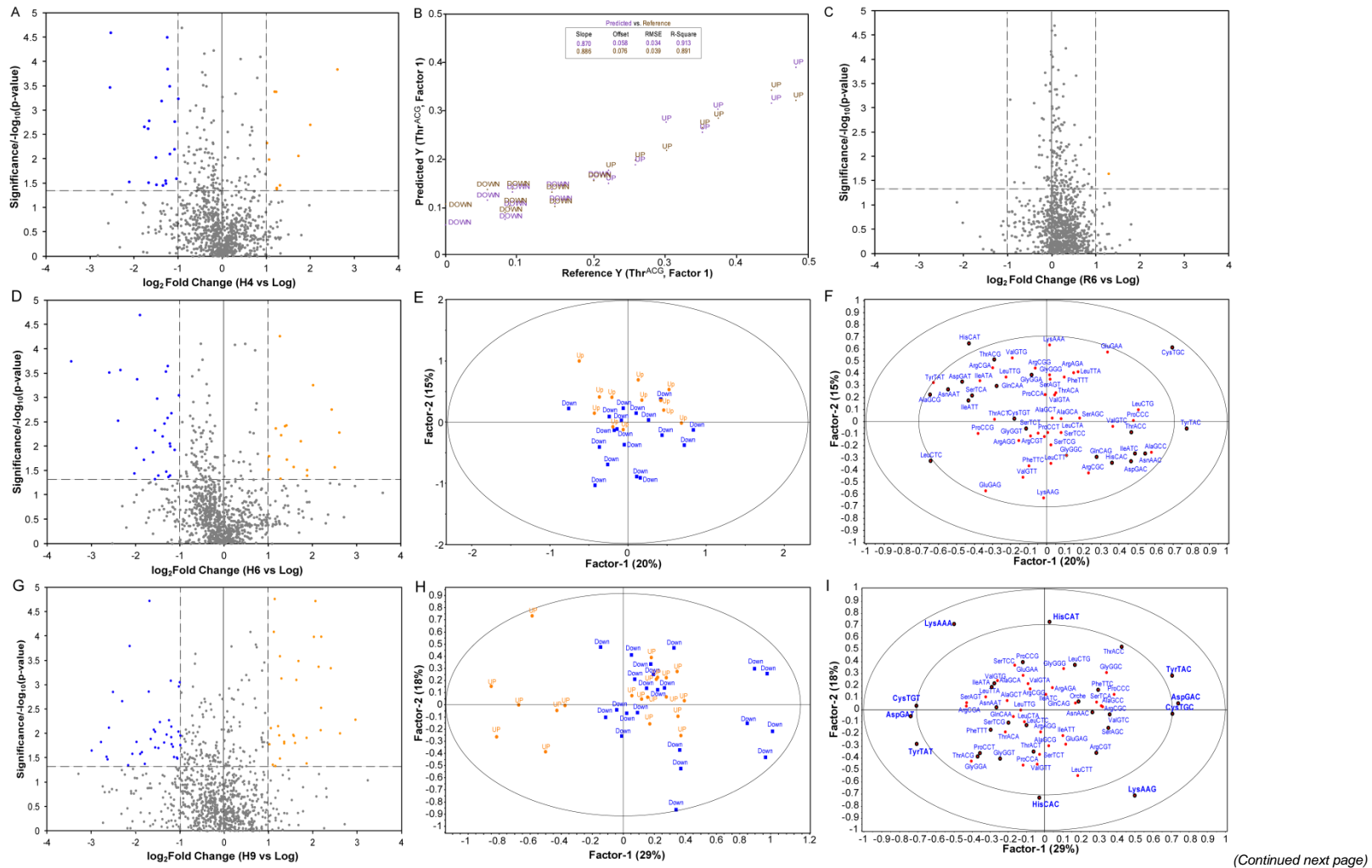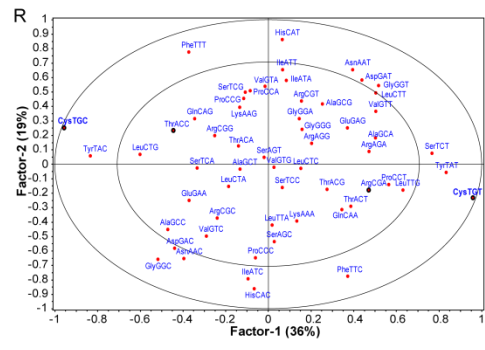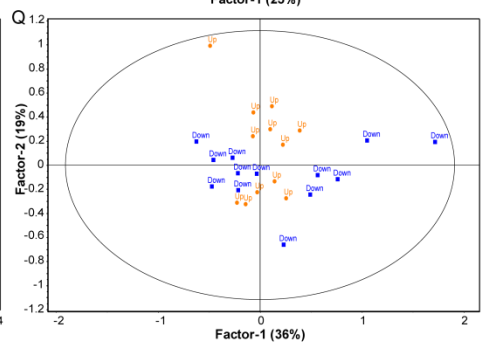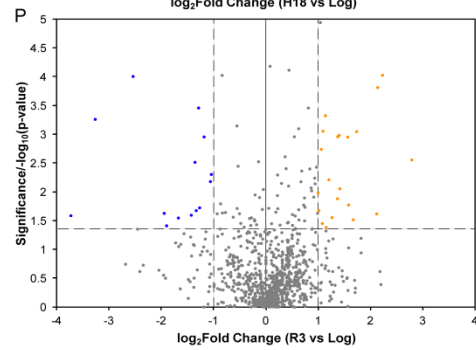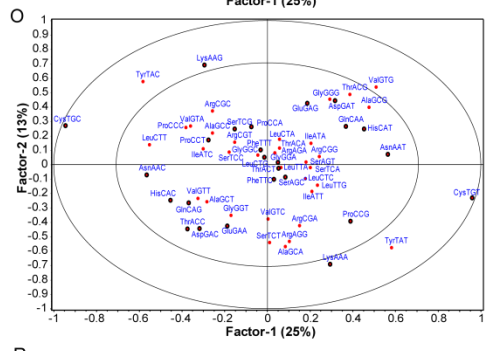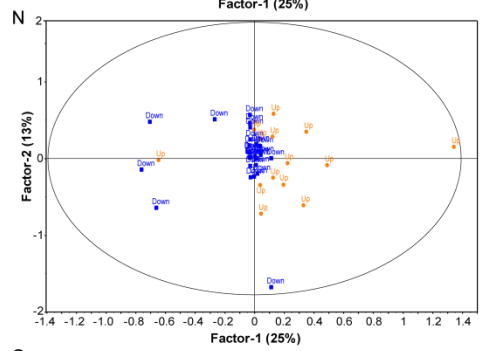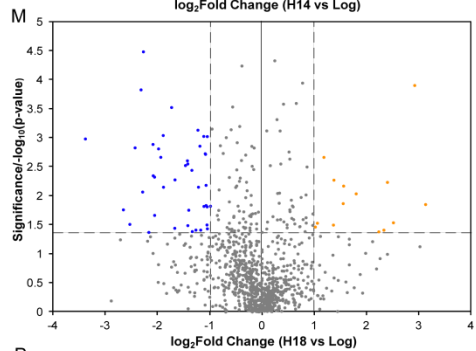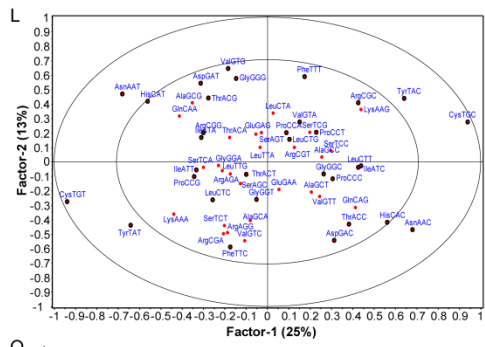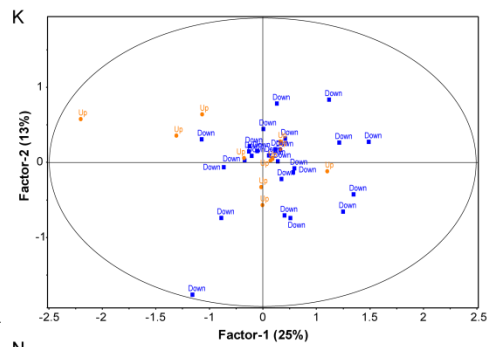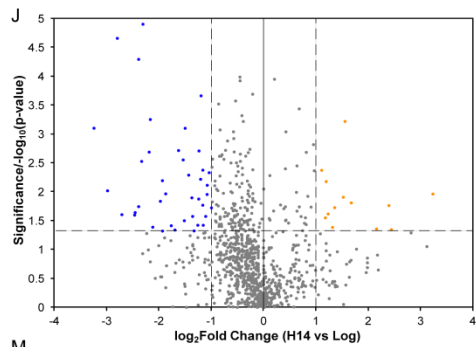
(Legends – next page)

Supplementary Figure 6.6. Entry and exit from a hypoxia-induced dormancy causes the up- and down regulation of proteins with distinct codon usage signatures.

Volcano plots comparison of proteins up- or down-regulation between (**A**) H4 and Log, (**D**) H6 and Log, (**G**) H9 and Log, (**J**) H14 and Log, (**M**) H18 and Log, (**P**) R3 and Log, and (**C**) R6 and Log. The x axis indicates the differential protein expression profiles, plotting the fold-induction ratios in a $\log_2$ scale. The y axis indicates the statistical significance of the difference in expression (p-value from a t-test) in a $\log_{10}$ scale. Proteins >2 fold up-regulated with $p < 0.05$ are shown in blue; proteins >2 fold down-regulated with $p < 0.05$ in orange. These proteins are further analyzed for codon usage by PLS.

In R6 vs. Log (**C**), no proteins made the cut-off, hence, PLS regression analysis was not performed.

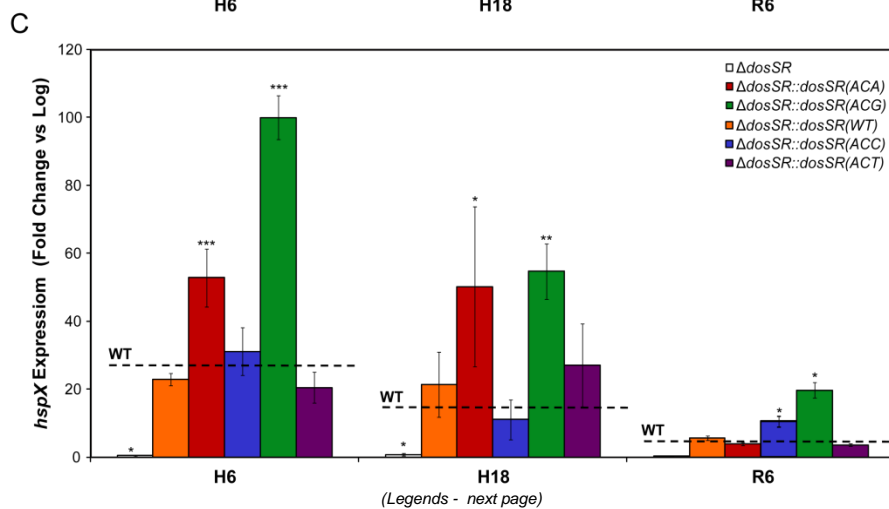(**B**) RMSE plot for Thr^{ACG} (Y-variable) against the strongest latent factor (Factor 1). The predicted Thr^{ACG} value (purple) from the PLS model is plotted against the measured Thr^{ACG} value (brown). The goodness of this prediction can be validated by the differences between the slopes, offset, RMSE and R^2 between the prediction and calibration reference. UP and DOWN refer to up- and down-regulated proteins at H4 relative to Log (Fig. 3C and S6A).

PLS regression analysis of the most significantly up- (●) or down- (■) regulated proteins and their codon usages (●) visualized by scores (**E** – at H6; **H** – at H9; **K** – at H14; **N** – at H18; **Q** – at R3) and loadings plots (**F** – at H6; **I** – at H9; **L** – at H14; **O** – at H18; **R** – at R3). Eclipse in scores plots represents Hoteling T^2 limit at p-value of 0.05 (F-test) while those in loadings plots indicate the explained variance. Outer and inner ellipses indicate 100% and 50% explained variance respectively. Codons contributing significantly to the regression (cross validation, by applying Marten's Uncertainty Test) are circled in black. Percentages of observed variances explained by the Factor 1 and 2 are indicated in parentheses along their respective axes.

A

BCG_3154

BCG_3157c

$dosR_{reg}$-F

$dosR_{reg}$-R

WT

5' homologous region

dosS

dosR

TF binding sites

3' homologous region

$dosS_{hr}$-F

$dosS_{hr}$-R

$dosR_{hr}$-F

$dosR_{hr}$-R

pYUB845-$dosSR_{FR}$

5' homologous region

hyg^r

3' homologous region

Homologous recombination

ΔdosSR

BCG_3154

hyR

BCG_3157c

Integration at attB site

dosS

dosR*

TF binding sites

Complementation with one of 5 dosR constructs

WT Thr

All Thr^{ACA}

All Thr^{ACC}

All Thr^{ACG}

All Thr^{ACT}

ΔdosSR::dosSR(WT)

ΔdosSR::dosSR(ACA)

ΔdosSR::dosSR(ACC)

ΔdosSR::dosSR(ACG)

ΔdosSR::dosSR(ACT)

B

OD_{600}

□ ΔdosSR
■ ΔdosSR::dosSR(WT)
■ ΔdosSR::dosSR(ACA)
■ ΔdosSR::dosSR(ACC)
■ ΔdosSR::dosSR(ACG)
■ ΔdosSR::dosSR(ACT)

WT

WT

WT

H6

H18

R6

***

**

***

*

***

C

hspX Expressiom (Fold Change vs Log)

□ ΔdosSR
■ ΔdosSR::dosSR(ACA)
■ ΔdosSR::dosSR(ACG)
■ ΔdosSR::dosSR(WT)
■ ΔdosSR::dosSR(ACC)
■ ΔdosSR::dosSR(ACT)

WT

WT

WT

H6

H18

R6

***

***

*

*

*

**

*

*

*

(Legends - next page)

Supplementary Figure 6.7. *dosR* mutants, reengineered with altered threonine codon usage, possess varied fitness and mistimed DosR activity.

(**A**) Construction of mutant BCG strains containing *dosR* with altered threonine codon usage. dosSR were first deleted. Subsequent complementation re-introduced reengineered *dosSR* contructs with its native promoter generating strains with genotypes described in table S2. See Materials and Methods for full description of the process.

(**B**) Turbidity (optical density at 600 nm, OD_{600}) of cultures of recombinant BCG at H6, H18 and R6. Dashed line denotes the mean OD_{600} for wild type cultures.

(**C**) Fold change in *hspX* expression of mutant strains relative to WT at H6, H18 and R6.

For (**B**) and (**C**) n ≥ 5; mean ± SEM; $p < 0.05$, $p < 0.01$ and $p < 0.001$ are denoted as *, ** and *** respectively determined by two-way ANOVA with Bonferroni post-tests (vs. Log) considering interactions between mutations and hypoxia. Together with changes in CFU (Fig. 4D) and dosR expression (Fig S4E), choice of threonine codon contributes 29.9-42.7% to the total variance ($p < 0.0001$); hypoxic exposures 20.3-38.4% ($p < 0.0001$)and interactions between the two 21.6-22.7% ($p < 0.0001$).
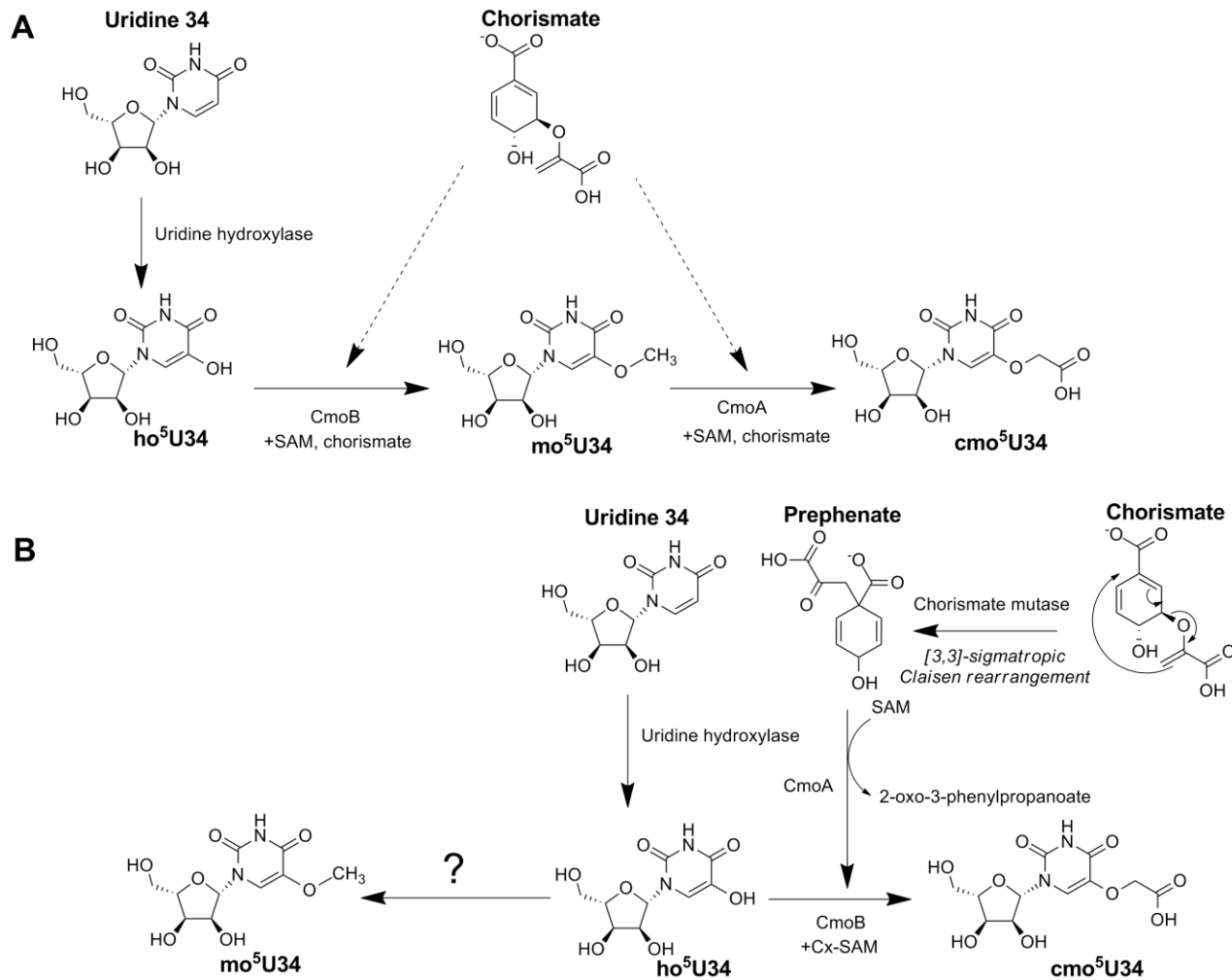
Supplementary Figure 6.8. Proposed biosynthetic pathways for the synthesis of cmo⁵U and mcmo⁵U.

(**A**) Biosynthetic pathway proposed by Nasvall et al [55] based on the phenotypes of cmoA and cmoB null mutants. cmo^{5}U34 is thought to be formed in 2 steps. First, ho^{5}U34 is methylated to mo^{5}U34 by CmoB. And subsequently, mo^{5}U34 is carboxylated to cmo^{5}U34 by CmoA. Both enzymes require SAM and chorismate or one of its derivatives as cofactors (dashed arrow). The identity of uridine hydrolase is currently unknown.

(**B**) Alternate biosynthetic pathway integrating the observations of Kim et al. [57], Byrne et al. [56]; Schneider et al. [73]; Nasvall et al. [44, 55]; Hagervall et al. [54]; Murao et al. [50] and Chionh et al. (this study). Chorismate is converted to prephenate by chroismate mutases (BCG_1002c – *Mtb* Rv0948c homolog and BCG_1922c – Rv1885c homolog). Prephenate is used as a cofactor by CmoA to generate Cx-SAM from SAM, releasing 2-oxo-phenylpropanoate as a by-product. Cx-SAM is a co-factor used by CmoB to directly acetylate ho^{5}U34 to cmo^{5}U34 without generating mo^{5}U34 intermediates. (**?**) mo^{5}U34 is made from ho^{5}U34 by a competing reaction catalyzed by either CmoB in the absence of CmoA induction or by another unidentified methyltransferase.

ho^{5}U34: wobble 5-hydroxyuridine; mo^{5}U34: wobble 5-methoxyuridine; cmo^{5}U: wobble uridine-5-oxyacetic acid; SAM: S-adenosyl-methionine; Cx-SAM: carboxy-*S*-adenosyl-*L*-methionine (reactive ylide form)

6.6.2. Supplementary tables

Supplementary Table 6.1. Primer, oligoribonucleotide, peptide sequences and strains used in this study.

(**A**) **Primers for cloning of 5' and 3' homologous regions**

Name	Direction	Sequence (5'-3')$^{\#}$	T_m (^{o}C)
dosR$_{hr}$	F	TTTCTAGAGCTGGTATGCACCGCACAAT	66.1
	R	TTCTTAAGCCAGTAACGTACCGCTGAA	
dosS$_{hr}$	F	TTACTAGTCTCAGCTGTGGTGTGGCATT	65.4
	R	TTCTCGAGCCGTTGTCGCAGTAGCTCTT	

(**B**) **Primers for cloning *dosSR* promoter and TF binding region**

Name	Direction	Sequence (5'-3')$^{\#}$	T_m (^{o}C)
dosR$_{reg}$	F	TTGTCGACCGGCGGATCGGAGATTGAT	65.1
	R	TTAAGCTTCGGCGGATCGGAGATTGATC	

(**C**) **Primers for screening of mutants**

Gene	Direction	Sequence (5'-3')	T_m (^{o}C)
dosR	F	GTGCCCTGGTGGTAAAGGTCT	61.0
	R	GAGGCCCCCTGTTGTCATGG	
dosS	F	CGGTTCGGGTGCGCGATGAA	61.0
	R	GGTGAAAGCCCCTGGACCGC	
ureC *	F	TCGACCCCGGCGGCCATC	66.0
	R	GGTGGACGGTCGGATCCG	
hyg^{r} *	F	ACTGCTTGTCCGATATCTGAT	61.0
	R	GAACTGGCGCAGTTCCTCT	

(**D**) **Primers for qPCR**

Gene	Direction	Sequence (5'-3')	Experimental denaturing temperature (^{o}C)
dosR	F	AGGCGATGGCCAGGGTTCCT	90.0
	R	GCGCGCGCCAACTCCATTCCCT	
dosS	F	CGGTTCGGGTGCGCGATGAA	92.0
	R	ACAAGCCGGAACACCGTCGC	
hspX	F	GACGAGATGAAAGAGGGGCG	90.0
	R	GTCGTCCTCGTCAGCACCTA	
sigA$^{\wedge}$	F	CGATGAGCCGGTAAAACGC	91.0
	R	GAGCCACTAGCGGACTTCGC	

(E) Oligonucleotides for the quantitation of tRNA species

tRNA	Role`	Sequence (5'-3')	MRM Transition[#]
tRNA$_{CGU}$	Cal	rCrUrCrGrUrA	Quan: 916.1 → 346.1 (w1)
			Qual: 916.1 → 572.1 (y2)
	IStd	CTCGUA	Quan: 882.2 → 330.1 (w1)
			Qual: 882.2 → 634.1 (w2)
tRNA$_{GGU}$	Cal	rCrUrCrGrCrCrUrUrGrUrA	Quan: 929.1 → 346.1 (w1)
			Qual: 929.1 → 344.0 (c1)
	IStd	CTCGCCTTGTA	Quan: 895.2 → 330.1 (w1)
			Qual: 895.2 → 619.1 (c2)
tRNA$_{UGU}$	Cal	rGrCrGrUrUrUrCrCrA	Quan: 849.4 → 304.0 (c1)
			Qual: 849.4 → 917.2 (y3)
	IStd	GCGTTTCCA	Quan: 819.4 → 303.0 (c1)
			Qual: 819.4 → 634.1 (w2)

(F) Peptides for the targeted quantification of proteins

Protein	Role˜	Sequence (N- to C- terminus)	MRM Transition[#]
DosR	Quan	QDPLSGLTDQER	679.8 → 818.4 (y7)
	Qual	TLLGLLSEGLTNK	679.9 → 861.6 (y8)
DosS	Quan	TIPVAGAVLR	498.8 → 782.5 (y8)
	Qual	DESFGTLYLTDK	694.8 → 910.5 (y9)
SigA	Quan	SVKPASAPQDTTTSTIPK	686.15 → 548.4 (y2)
	Qual	KDAELTASADSVR	657.7 → 361.2 (y3)

(**G**) **Strains and characteristics**

Strain	Genotype
BCG Pasteur 1173P2	Wild type (WT)
ΔdosSR	As 1173P2 plus Δ(dosS-dosR)::hyR
ΔdosSR::dosSR(WT)	As ΔdosSR plus attB::[pMV306::dosS-dosR]::kan
ΔdosSR::dosSR(ACA)	As ΔdosSR plus attB::[pMV306::dosS-dosR(40C>T + 61C>T + 115C>T +157G>T + 187C>T + 202G>T + 400G>T + 409C>T]::kan
ΔdosSR::dosSR(ACC)	As ΔdosSR plus attB::[pMV306::dosS-dosR(40C>G + 61C>G + 115C>G + 187C>G + 409C>G]::kan
ΔdosSR::dosSR(ACG)	As ΔdosSR plus attB::[pMV306::dosS-dosR(157G>C + 202G>C+ 400G>C]::kan
ΔdosSR::dosSR(ACT)	As ΔdosSR plus attB::[pMV306::dosS-dosR(40C>A + 61C>A+ 115C>A +157G>A + 187C>A + 202G>A + 400G>A + 409C>A]::kan

[#] Restriction endonuclease sites underlined.

* Positive controls. Sequences obtained from [29].

^Sequences Obtained from [74].

`Cal: External calibration standards. IStd: Spiked internal standards used for relative quantitation.

Quan: Quantifier ion. Qual: Qualifier ion. Fragment identity indicated in parentheses.

~ Quan: Quantifier peptide. Qual: Qualifier peptide. Unique peptide sequences from tryptic digests of target proteins.

6.7. References

1. Dedon, P.C. and T.J. Begley, *A system of RNA modifications and biased codon use controls cellular stress response at the level of translation.* Chem Res Toxicol, 2014. **27**(3): p. 330-7.
2. Plotkin, J.B. and G. Kudla, *Synonymous but not the same: the causes and consequences of codon bias.* Nat Rev Genet, 2011. **12**(1): p. 32-42.
3. Gu, C., T.J. Begley, and P.C. Dedon, *tRNA modifications regulate translation during cellular stress.* FEBS Lett, 2014. **588**(23): p. 4287-4296.
4. Subramaniam, A.R., T. Pan, and P. Cluzel, *Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria.* Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2419-24.
5. Begley, U., et al., *Trm9-catalyzed tRNA modifications link translation to the DNA damage response.* Mol Cell, 2007. **28**(5): p. 860-70.
6. Gingold, H., O. Dahan, and Y. Pilpel, *Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome.* Nucleic Acids Res, 2012. **40**(20): p. 10053-63.
7. Shah, P. and M.A. Gilchrist, *Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10231-6.
8. Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.
9. Laxman, S., et al., *Sulfur amino acids regulate translational capacity and metabolic homeostasis through modulation of tRNA thiolation.* Cell, 2013. **154**(2): p. 416-29.
10. Agris, P.F., *Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications.* EMBO Rep, 2008. **9**(7): p. 629-35.
11. Persson, B.C., *Modification of tRNA as a regulatory device.* Mol Microbiol, 1993. **8**(6): p. 1011-6.
12. Grosjean, H., V. de Crecy-Lagard, and C. Marck, *Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.* FEBS Lett, 2010. **584**(2): p. 252-64.
13. Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence.* Infect Immun, 1996. **64**(6): p. 2062-9.
14. Ravindran, M.S., et al., *Targeting lipid esterases in mycobacteria grown under different physiological conditions using activity-based profiling with tetrahydrolipstatin (THL).* Mol Cell Proteomics, 2014. **13**(2): p. 435-48.
15. Hia, F., et al., *Mycobacterial RNA isolation optimized for non-coding RNA: high fidelity isolation of 5S rRNA from Mycobacterium bovis BCG reveals novel post-transcriptional processing and a complete spectrum of modified ribonucleosides.* Nucleic Acids Res, 2014.
16. Miyauchi, K., S. Kimura, and T. Suzuki, *A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification.* Nat Chem Biol, 2013. **9**(2): p. 105-11.
17. Chionh, Y.H., et al., *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Res, 2013. **41**(17): p. e168.

18. Su, D., et al., *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry.* Nat Protoc, 2014. **9**(4): p. 828-41.

19. Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.

20. Wetzel, C. and P.A. Limbach, *The global identification of tRNA isoacceptors by targeted tandem mass spectrometry.* Analyst, 2013. **138**(20): p. 6063-72.

21. Easter, R.N., et al., *Separation and identification of oligonucleotides by hydrophilic interaction liquid chromatography (HILIC)-inductively coupled plasma mass spectrometry (ICPMS).* Analyst, 2010. **135**(10): p. 2560-5.

22. Hong, Q., et al., *Label-free absolute quantitation of oligosaccharides using multiple reaction monitoring.* Anal Chem, 2014. **86**(5): p. 2640-7.

23. Rozenski, J. and J.A. McCloskey, *SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry.* J Am Soc Mass Spectrom, 2002. **13**(3): p. 200-3.

24. Schurch, S., *Characterization of nucleic acids by tandem mass spectrometry - The second decade (2004-2013): From DNA to RNA and modified sequences.* Mass Spectrom Rev, 2014.

25. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

26. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

27. Shadforth, I.P., et al., *i-Tracker: for quantitative proteomics using iTRAQ.* BMC Genomics, 2005. **6**: p. 145.

28. Bardarov, S., et al., *Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis.* Microbiology, 2002. **148**(Pt 10): p. 3007-17.

29. Lin, W., et al., *Urease activity represents an alternative pathway for Mycobacterium tuberculosis nitrogen metabolism.* Infect Immun, 2012. **80**(8): p. 2771-9.

30. Stover, C.K., et al., *New use of BCG for recombinant vaccines.* Nature, 1991. **351**(6326): p. 456-60.

31. Bartek, I.L., et al., *The DosR regulon of M. tuberculosis and antibacterial tolerance.* Tuberculosis (Edinb), 2009. **89**(4): p. 310-6.

32. Bagchi, G., et al., *Transcription and autoregulation of the Rv3134c-devR-devS operon of Mycobacterium tuberculosis.* Microbiology, 2005. **151**(Pt 12): p. 4045-53.

33. Chauhan, S. and J.S. Tyagi, *Cooperative binding of phosphorylated DevR to upstream sites is necessary and sufficient for activation of the Rv3134c-devRS operon in Mycobacterium tuberculosis: implication in the induction of DevR target genes.* J Bacteriol, 2008. **190**(12): p. 4301-12.

34. Rustad, T.R., et al., *Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network.* Genome Biol, 2014. **15**(11): p. 502.

35. Kumar, A., et al., *Redox homeostasis in mycobacteria: the key to tuberculosis control?* Expert Rev Mol Med, 2011. **13**: p. e39.

36. Via, L.E., et al., *Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates.* Infect Immun, 2008. **76**(6): p. 2333-40.

37.   Chao, M.C. and E.J. Rubin, *Letting sleeping dos lie: does dormancy play a role in tuberculosis?* Annu Rev Microbiol, 2010. **64**: p. 293-311.
38.   Rustad, T.R., et al., *Hypoxia: a window into Mycobacterium tuberculosis latency.* Cell Microbiol, 2009. **11**(8): p. 1151-9.
39.   Low, K.L., et al., *Triacylglycerol utilization is required for regrowth of in vitro hypoxic nonreplicating Mycobacterium bovis bacillus Calmette-Guerin.* J Bacteriol, 2009. **191**(16): p. 5037-43.
40.   Chan, C.T., et al., *Identification of N6,N6-dimethyladenosine in transfer RNA from Mycobacterium bovis Bacille Calmette-Guerin.* Molecules, 2011. **16**(6): p. 5168-81.
41.   Galagan, J.E., et al., *The Mycobacterium tuberculosis regulatory network and hypoxia.* Nature, 2013. **499**(7457): p. 178-83.
42.   Boon, C. and T. Dick, *How Mycobacterium tuberculosis goes to sleep: the dormancy survival regulator DosR a decade later.* Future Microbiol, 2012. **7**(4): p. 513-8.
43.   Weixlbaumer, A., et al., *Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines.* Nat Struct Mol Biol, 2007. **14**(6): p. 498-502.
44.   Nasvall, S.J., P. Chen, and G.R. Bjork, *The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons.* RNA, 2007. **13**(12): p. 2151-64.
45.   Dong, H., L. Nilsson, and C.G. Kurland, *Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.* J Mol Biol, 1996. **260**(5): p. 649-63.
46.   Stoecklin, G. and S. Diederichs, *tRNAs: new tricks from old dogs.* EMBO J, 2014. **33**(18): p. 1981-3.
47.   Moukadiri, I., et al., *The output of the tRNA modification pathways controlled by the Escherichia coli MnmEG and MnmC enzymes depends on the growth conditions and the tRNA species.* Nucleic Acids Res, 2014. **42**(4): p. 2602-23.
48.   Murao, K., T. Hasegawa, and H. Ishikura, *5-methoxyuridine: a new minor constituent located in the first position of the anticodon of tRNAAla, tRNAThr, and tRNAVal from Bacillus subtilis.* Nucleic Acids Res, 1976. **3**(10): p. 2851-60.
49.   Pang, Y.L., et al., *Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number.* Nucleic Acids Res, 2014. **42**(22): p. e170.
50.   Murao, K., et al., *On the biosynthesis of 5-methoxyuridine and uridine-5-oxyacetic acid in specific procaryotic transfer RNAs.* Nucleic Acids Res, 1978. **5**(4): p. 1273-81.
51.   Leistikow, R.L., et al., *The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy.* J Bacteriol, 2010. **192**(6): p. 1662-70.
52.   Flores Valdez, M.A. and G.K. Schoolnik, *DosR-regulon genes induction in Mycobacterium bovis BCG under aerobic conditions.* Tuberculosis (Edinb), 2010. **90**(3): p. 197-200.
53.   Hu, Y., et al., *Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo.* Infect Immun, 2006. **74**(2): p. 861-8.
54.   Hagervall, T.G., et al., *Chorismic acid, a key metabolite in modification of tRNA.* J Bacteriol, 1990. **172**(1): p. 252-9.
55.   Nasvall, S.J., P. Chen, and G.R. Bjork, *The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo.* RNA, 2004. **10**(10): p. 1662-73.

56.   Byrne, R.T., et al., *S-Adenosyl-S-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 6): p. 1090-8.

57.   Kim, J., et al., *Structure-guided discovery of the metabolite carboxy-SAM that modulates tRNA function.* Nature, 2013. **498**(7452): p. 123-6.

58.   Bernstein, J.A., et al., *Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays.* Proc Natl Acad Sci U S A, 2002. **99**(15): p. 9697-702.

59.   Milo, R. and R. Phillips, *Concentrations and Absolute Numbers: How Many mRNAs are in a Cell?*, in *Cell Biology by the Numbers*2014, Garland Science: New York, NY. p. 157-158.60.   Kumar, A., et al., *Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor.* Proc Natl Acad Sci U S A, 2007. **104**(28): p. 11568-73.

61.   Rustad, T.R., et al., *The enduring hypoxic response of Mycobacterium tuberculosis.* PLoS One, 2008. **3**(1): p. e1502.

62.   Selvaraj, S., et al., *In silico analysis of DosR regulon proteins of Mycobacterium tuberculosis.* Gene, 2012. **506**(1): p. 233-41.

63.   Sorensen, M.A., et al., *Over expression of a tRNA(Leu) isoacceptor changes charging pattern of leucine tRNAs and reveals new codon reading.* J Mol Biol, 2005. **354**(1): p. 16-24.

64.   Takai, K., et al., *Recognition of UUN codons by two leucine tRNA species from Escherichia coli.* FEBS Lett, 1994. **344**(1): p. 31-4.

65.   Murphy, F.V.t., et al., *The role of modifications in codon discrimination by tRNA(Lys)UUU.* Nat Struct Mol Biol, 2004. **11**(12): p. 1186-91.

66.   Rodriguez-Hernandez, A., et al., *Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position.* J Mol Biol, 2013. **425**(20): p. 3888-906.

67.   Yarian, C., et al., *Accurate translation of the genetic code depends on tRNA modified nucleosides.* J Biol Chem, 2002. **277**(19): p. 16391-5.

68.   Crick, F.H., *Codon--anticodon pairing: the wobble hypothesis.* J Mol Biol, 1966. **19**(2): p. 548-55.

69.   Yokoyama, S. and S. Nishimura, *Modified nucleotides and codon recognition*, in *tRNA: Structure, Biosynthesis and Function*, D. Soll and U.L. Rajbhandary, Editors. 1995, American Society for Microbiology: Washington, D.C. p. 207-223.

70.   Agris, P.F., F.A. Vendeix, and W.D. Graham, *tRNA's wobble decoding of the genome: 40 years of modification.* J Mol Biol, 2007. **366**(1): p. 1-13.

71.   Roth, A.C., *Decoding properties of tRNA leave a detectable signal in codon usage bias.* Bioinformatics, 2012. **28**(18): p. i340-i348.

72.   Tumu, S., et al., *The gene-specific codon counting database: a genome-based catalog of one-, two-, three-, four- and five-codon combinations present in Saccharomyces cerevisiae genes.* Database (Oxford), 2012. **2012**: p. bas002.

73.   Schneider, C.Z., et al., *The two chorismate mutases from both Mycobacterium tuberculosis and Mycobacterium smegmatis: biochemical analysis and limited regulation of promoter activity by aromatic amino acids.* J Bacteriol, 2008. **190**(1): p. 122-34.

74.   Volpe, E., et al., *Gene expression profiling of human macrophages at late time of infection with Mycobacterium tuberculosis.* Immunology, 2006. **118**(4): p. 449-60.

7. Targeting mycobacterial stress responses for biomarker and drug discovery: A perspective

7.1. Summary

The purpose of my research, as presented in this thesis, is to advance our understanding of how pathogens of the *Mycobacterium tuberculosis* (*Mtb*) complex regulate their gene expression in response to physiological stresses. Specifically, much of the work was focused on studying the effects of nutrient deprivation and hypoxia – stresses associated with the non-replicating, persistent and drug tolerant state of mycobacteria within pulmonary granulomas. To achieve this, I developed an analytical platform for the isolation, purification, characterization and quantification of intracellular RNA and its secondary modifications (Chapters 2-4). When combined with whole cell phenotype and metabolic screens, global analysis of mRNA levels revealed a metabolic shift towards ketogenesis in nutrient deprived mycobacterial persisters (Chapter 5). Separately, the dynamic reprogramming of tRNA by enzymatic modifications was quantified during hypoxia and, when this was matched with simultaneous changes in the proteome, we discovered that selective translation of codon-biased transcripts regulate the hypoxia-induced dormancy response (Chapter 6). Yet, as more and more information on the stress response of mycobacteria becomes available, it is increasingly difficult not to be amazed – and worried – about the sophisticated strategies evolved by these pathogens to circumvent their host's defenses and survive in seemingly hostile environments. Thus, in this closing chapter, I identify shortcomings in our current approaches to studying myobacteria, build new models that link our findings to clinical observations, and propose ways in

which our newfound knowledge could be employed to develop biomarkers of disease state and strategies for drug discovery.

7.2. Clinical significance: Persister reactivation as a consequence of diabetic ketoacidosis

The growing "co-epidemic" of TB and diabetes mellitus (DM) is a global public health concern [1, 2]. At present, 15% of TB cases worldwide are attributable to DM. Furthermore, having diabetes triples the risk of acquiring active TB and patients with inherent DM and latent TB are 4-times more likely to relapse [3-6]. More worrying is the 2- to 9-fold increased risk of MDR TB among diabetics [7-12]. Data from one longitudinal study even suggests that diabetic patients frequently relapsed with resistant strains [64]. Although the definite pathophysiological mechanism of the effect of DM as a predisposing risk factor for TB is unknown, several hypotheses had been suggested. They include depressed cellular immunity, dysfunction of alveolar macrophages, altered levels of pro-inflammatory cytokines including IFN-γ, IL-8 and TNF-α, pulmonary microangiopathy, and micronutrient deficiency [2, 13-18]. However, there is no explanation for the impact of DM on *Mtb* drug resistance [2, 13]. Our findings related to ketone body metabolism in starvation-induced mycobacterial persisters (detailed in Chapter 5) may shed light on this phenomenon.

We propose a model wherein diabetic ketoacidosis induces the reactivation of antibiotic tolerant persisters. Diabetic ketoacidosis occurs in patients with both type 1 and type 2 DM [19, 20]. In both cases, the lack of proper dietary control coupled with exogeneous stress and insulin deficiency or resistance leads to the release of free fatty acids from adipose tissue (lipolysis), which are

converted, again in the liver, first into acetyl-coA then to the ketone bodies, acetoacetate (AA) and β-hydroxybutyrate (βHB) [21]. Part of the acetyl-coA is also shunted through the TCA cycle into glucose by gluconeogenesis (**Figure 7.1**) [22]. The release of large amounts (>50-times normal blood concentrations) of AA, βHB and glucose into the bloodstream cause metabolic acidosis and hyperglycermia [21]. Our data suggest that in diabetics with latent TB, non-replicating persisters could be reactivated by the ketoacidotic conditions. Due to their nutrient deprivation-induced ketogenic state within lung granulomas, mycobacterial persisters would readily take up and utilize host AA, βHB and glucose (**Figures 5.3C**, **5.3E**, **5.7G** and **7.1**). Glucose removes the block on replication, while the continued presence of ketone bodies maintains the induction of cytochrome P450s. Both glucose and ketone bodies are used as carbon sources for persister reactivation. The latter, in particular, induces an altered pharmacodynamic state through the facilitates antibiotic tolerance and the development of genetic drug resistance
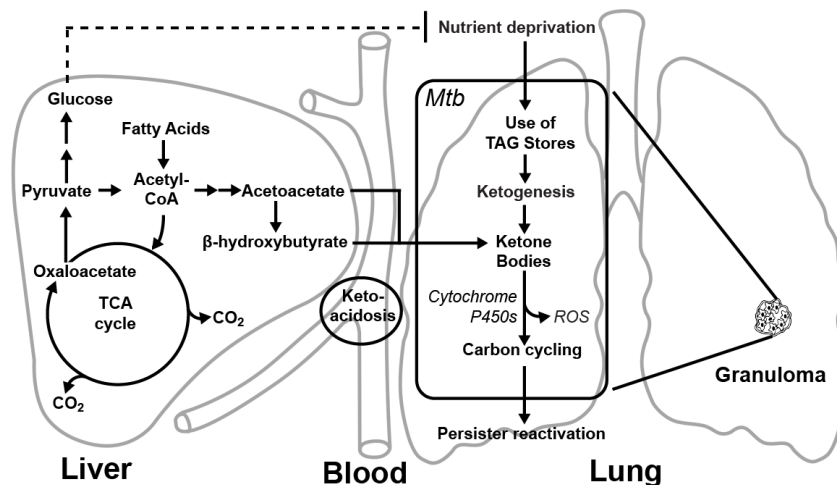

Figure 7.1. Proposed host-pathogen metabolic interactions leading to *Mtb* persister reactivation in diabetics experiencing ketoacidosis. During diabetic ketoacidosis, the metabolism of fatty acids occurs at high rates in the liver producing ketone bodies (AA and βHB) that enter the bloodstream and reduce its pH. These ketone bodies are transported to other organs, including the lung where nonreplicating *Mtb* resides. This provides a carbon source for persister reactivation. Concurrent gluconeogenesis in the liver releases glucose that helps to remove stringency signals blocking regrowth.

induction of bacterial cytochrome P450s in these reactivated persisters, which (**Chapter 5**; **Figure 7.1**).

We could directly test this hypothesis in two steps. First, starvation-induced BCG and *Mtb* persisters could be exposed to AA, βHB and glucose in batch cultures and their drug susceptibility tested. Our phenotypic screens indicate that these persisters readily utilize these metabolites as carbon sources, and if our hypothesis is true, we would expect persisters to be reactivated by ketone bodies, in both the presence and absence of glucose, to possess higher antibiotic tolerances than those resuscitated by nutrient replete 7H9 media (**Figure 5.2A-D**). The levels of bacterial cytochrome p450s and their activities would be determined using qPCR and chemical probes respectively [23, 24]; thereby establishing an experimental link between carbon utilization and antibiotic tolerance. Second, further testing of this host-pathogen metabolic interaction could be performed in animal models [15]. For instance, a Wistar rat model for *Mtb* could be used [25, 26]. In this model, the treatment of *Mtb*-infected rats with dexamethasone induced the reactivation of bacilli otherwise suppressed by the host immunity [25]. Glucagon, a key hormone required for the initiation of diabetic ketoacidosis [20], is known to be induced by corticosteorids such as dexamethasone [27]. Hence, we could treat *Mtb* infected rats with glucagon to determine if this hormone acts downstream of dexamethasone to reactivate *Mtb* growth. Furthermore, ketosis or insulin insensitivity could be induced in these animals by switching them over to ketogenic or high-fat hypercaloric diets respectively [28, 29]. If our hypothesis is true, we would expect to observe an increase in bacterial burden in these rats that is closely associated with the levels of ketone bodies in the blood. Levels of bacterial cytochrome P450s could be measured by qPCR, and ideally, tRNA modification-based biomarkers could be used to ascertain drug

tolerance (see following section). Together, these studies could help aid clinical assessment of MDR-TB risks in ketosis-prone type 1 and type 2 diabetes patients.

7.3. Disease diagnosis: tRNA modifications as biomarkers of TB pathogenesis, stress exposure and antibiotic susceptibility

Despite my identification of 40 tRNA modifications in BCG – the most in any microorganism to date (*E.coli*: 27 modifications; *S. cerevisiae*: 24 modifications. **Table 1.4**) – and the fact that each stage of hypoxia- induced dormancy possessed distinctive tRNA modification patterns (**Figure 6.2**) - some of which (e.g., cmo^5U) are highly correlated with the hypoxic stress response (**Supplementary Figure 6.5H**) - several immediate questions remain unanswered. Has the full spectrum of tRNA modifications truly been determined? Are the same modifications found in *Mtb*? Most importantly, how relevant are the changes that we detect to the actual survival of mycobacterial pathogens in their hosts?

Ongoing research in our laboratory has partially addressed the first two questions. 2'-*O*-Methyladenosine (Am) and queuosine (Q, and its precursors or its hypermodified derivatives) are two prominent modifications thought to be conserved in prokaryotes that we did not found in our analysis of BCG tRNA modifications. We have since acquired chemically synthesized Am standards (Berry & Associates, Inc.), which led to identification and quantification of Am in the tRNA from *M. smegmatis* and *Pseudomonas aeruginosa*, 16S and 23S rRNA from BCG (unpublished), and vRNA from Dengue virus serotype 2 [30]. Preliminary analysis of tRNA from *Mtb* strain H37Rv by Molecular Feature Extraction of high accuracy LC-QToF mass spectra [31], conversely, failed to find Am (Lim et al., unpublished). These

results suggest that members of the *Mtb* complex might not possess Am in their tRNA. The failure to detect Q in BCG tRNA, similarly, might be reflective of its biology. The absence from BCG of *tgt* (clusters of orthologous groups (COG)-0343) and *queA* (COG0809) homologs required for the biosynthesis of Q suggests that this modified base is not naturally found in mycobacteria [32, 33]. Nonetheless, rare, uncharacterized modifications, such as the recently discovered wobble 2-geranylthiouridine and its derivatives, which are found in a small fraction (~ 2.8-6.7%) of *E. coli* tRNA^{Lys}UUU and tRNA^{Gln}UUG [34], could be missed due to insufficient analytical sensitivity. In addition, our current electron spray ionization (ESI) and ion transmission (fragmentor voltage) or collision induced dissociation (CID) settings might be too harsh to preserve *N*- and *O*-linked RNA glycosylations [35]. Repeating our analyses in negative ion mode might also yield 2'-*O*-phosphoribosyl modifications (2'-O-ribosyladenosine phosphate (Ar(p)) and 2'-O-ribosylguanosine phosphate (Gr(p)) that had been described previously in *S. cerevisiae* [36, 37], as well as phosphate modifications such as phosphorothioations [38, 39]. Finally, *Mtb* might contain unique tRNA modifications not found in BCG. For instance, lipidomic comparisons between *Mtb* and BCG recently identified 1-tuberculosinyladenosine (1-TbAd) as a unique terpene-adducted member of the *Mtb* metabolite pool [40]. It is quite possible that similar modifications could be found in *Mtb* tRNA.

The longstanding knowledge that modified ribonucleosides are excreted rather than recycled by nucleoside salvage pathways [41-43], and observations that *Mtb* secretes unique nucleosides like 1-TbAd [40, 44] raises the possibility that these molecules could be used as biomarkers of infection. More importantly, each stage of hypoxia survival induced signature changes in the abundance of tRNA modifications (Chapter 6). The potential of tRNA

modifications as biochemical indicators of stress (including antibiotic) exposure and disease state is highlighted by these signature changes, as well as our previous results from *S. cerevisiae*, which indicate that exposures to different cytotoxic agents, specifically, methylmethanesulfonate (MMS), hydrogen peroxide (H_2O_2), sodium arsenite ($NaAsO_2$) and sodium hypochlorite (NaOCl), give unique patterns of modifications closely associated with the mechanism of toxicity [45]. Similarly, the bactericidal effects of antibiotics can be thought of as a form of cytotoxic stress. Hence, I hypothesize that not only would classes of antibiotics induce changes in tRNA modification signatures based on their killing mechanisms, but this signature would differ significantly in antibiotic susceptible and resistant (both phenotypic or genetic) mycobacteria. This is an easily testable hypothesis, which, if true, could be a potentially exploitable property of mycobacterial stress responses for the development of biomarkers of cell state and drug susceptibility.

Worldwide, effective management of TB is hamstrung, in part, by the failure to obtain timely and accurate diagnosis of the disease [46]. Commonly used diagnostic techniques, such as the Tuberculin skin test (Mantoux test) that screens for the presence of antibodies against mycobacterial proteins, will return false positive results in persons who have had the BCG vaccine [47, 48], making it unsuitable in many parts of the world – including Singapore. Traditional sputum smear microscopy can only detect active infections in patients with a relatively high disease burden and is unable to distinguish *Mtb* from other non-tuberculous mycobacterial pathogens (NTMs). In addition, it is also difficult to rapidly diagnose MDR-TB. Definitive diagnosis requires culturing that could take more than a month, which introduces significant delays to appropriate treatment [49]. New molecular diagnostic methods in

particular the Xpert® MTB/RIF tests show promise [50, 51], but, being a PCR-based assay, is not always practical for field work in developing countries with high disease burdens [52]. Thus, there is an unmet need in the development for rapid, noninvasive diagnostics for disease state and antibiotic susceptibility in both the clinical and laboratory setting.

Given their metabolic stability, distinctiveness and biological roles in stress responses, tRNA modifications would make attractive candidates for biomarker development. Thus, I propose to examine the utility of these ubiquitous natural products as biochemical indicators of mycobacterial dormancy and antibiotic susceptibility in a series of steps:

(1) Compare the spectrum of modified ribonucleosides in BCG with that of *Mtb*. The latter might contain unique RNA modifications that are uncharacterized and could thus be useful in distinguishing *Mtb* infections from those caused by other mycobacteria.

(2) Perform systems-level analysis on changes in tRNA modifications under cytotoxic (with antibiotic, nitrosative and oxidative agents, et cetera.) and physiological stresses (nutrient deprivation, acid pH, *etc.*) for both *Mtb* and BCG. The stress-induced changes in tRNA modifications from persisters and drug resistant mutant strains would be analyzed, laying the groundwork for further studies to identify critical tRNA modification signatures that define each stress response.

(3) Compare these modification profiles using multivariate statistics and machine learning algorithms to determine sets of modifications that discriminates each group. These would serve as potential biomarker candidates of stress exposure and antibiotic resistance.

(4) Examine the "secretome" of *Mtb* in *ex vivo* macrophage infection models [44, 53-55] and in the urine, serum and lung of infected animals [49, 56] for the presence of these modified ribonucleosides biomarkers.

Several possible scenarios could emerge from the proposed studies. The ideal one is the discovery of a group of ribonucleosides that is always present in the infected animals with active disease, another for latent disease and neither are present in the uninfected animals. However, it is likely that no single group of ribonucleosides can distinguish infection status. Nonetheless, groups of ribonucleosides among infected animals might be sufficiently unique and reproducible that they could serve as fingerprints of disease state and progression. Just as modification signatures serve as fingerprints of stress exposure *in vitro*, panels of ribonucleoside markers might be predictive of increased lethality or antibiotic resistance *in vivo*. Failing that, the complex host immunity-induced stress response signatures represented by the tRNA modification profiles of mycobacteria recovered from these infection models could be parsed using multivariate breakdown and discriminant function analysis into its essential components. These components would then be compared to the tRNA signatures of stress exposures previously acquired in step 2 to associate infection outcomes and antibiotic susceptibility status with their corresponding stress responses.

Preliminary studies in *M. smegmatis* show that cytotoxic exposure to antibiotics and H_2O_2 induces dose-dependent changes in the tRNA modifications. Some of these profiles possess components that closely resemble those of antibiotic tolerant persisters produced by nutrient starvation (Cai *et al.*, unpublished). Furthermore, members of our laboratory had also managed to recover modified ribonucleosides from the urine of BCG infected Wistar rats [57]. These early results will hopefully lead into proof-of-concept

studies in *Mtb* infection models, which if successful, could be extended into the analysis of the urine from TB patients. Identification and validation of urinary biomarkers could lead to the development of a rapid, non-invasive test for the disease state and drug susceptibility of *Mtb* infections, which would be a major boon to global health.

7.4. Drug discovery: targeting tRNA modifications to disrupt mycobacterial dormancy and antibiotic tolerance

Our studies of tRNA modifications in mycobacteria do more than just inform us of their potential as functional biomarkers, they also lead us to evaluate the mechanism of selective translation wherein reprogrammed tRNA regulate the translation priority and abundance of codon biased genes as a target for TB drug discovery. Though it is now generally acknowledged that the choice between synonymous codons can significantly influence protein abundance by changes in translation efficiency [58-62], there is a disparity in our understanding of the biophysical mechanisms of translational selection and it biological function and significance. Current models of translational section is based on a yeast–*E.coli* paradigm, stating that highly expressed genes use a subset of "optimal" codons in accordance with their respective major isoacceptor tRNA levels [63-66]. Recent computational and experimental models share three common themes: (1) codon bias affects translational rate at the elongation step [8, 60-62, 65, 67-70]; (2) associations between local mRNA secondary structures to translational rates and protein expression and folding [58, 71-74]; (3) availability of tRNAs influence overall translational efficiency and ribosomal usage [63, 65, 69, 72, 75-77]. The models also share three limitations: (1) maximal growth rate is considered the ideal outcome,

even for cells under stress [62, 67, 70, 75, 76, 78-80]; (2) arbitrary designation of mRNA secondary structures as beneficial or detrimental to translational elongation, cell survival and growth, leading to contradictory findings and conclusions [58, 68, 71-73]; and (3) usage of the number of tRNA isoacceptor and isodecoder genes as proxy estimations of tRNA copy numbers which fails to appreciate changes in tRNA abundance and modification status in the tRNA pool [62, 63, 68, 69, 72, 75-77]. Nonetheless, these studies all support the premise that codons choice contains more information than merely amino acids identity [64, 74, 81]. The biological nature of this secondary message, however, remains elusive.

Together with recent progress with ribosome profiling [65, 82, 83], our recent advances in deep tRNA sequencing (tRNA-seq) [84] and quantification of tRNA by targeted mass spectrometric (Chapter 6) directly resolved several technical issues associated with the aforementioned models. However, the current narrow focus on elongation rates and secondary structures as determinants of translational control makes our current understanding relatively "static", in that, while these features are undoubtedly necessary for proper cellular functioning, they neither account for adaptive changes in cellular composition nor participate in dynamic processes [59, 62, 65, 70, 78, 80]. For instance, our codon re-engineered *dosR* mutants all possess identical 3' and 5' structures (no changes were made in the flanking *dosS* and *BCG_3157c* sequences, **Supplementary Figure 6.7**) and modest changes to local *dosR* secondary structures (ΔG^0 were stable at 28-32 kCal/mol for the top 5 structures for all *dosR* sequences – wild type and mutants – as predicted by mfold: http://mfold.rna.albany.edu). Yet, codon re-reengineered *dosR* mutants that showed no differences in growth phenotype at maximal growth rates in aerobic conditions, showed altered fitness under hypoxic

stress (**Figure 6.5**). As alluded to in Chapter 6, one possible explanation for this phenomenon, from the standpoint of network biology, is that codon usage and tRNA modifications work together to ensure an orderly scheduling of transcripts for translation.

At any point in time, cells need to integrate and run multiple genetic programs comprised of gene transcripts. However, these programs use the same system resources. Two such resources are amino acids and ribosomes. Naturally, given sufficient time and resources all transcripts can be translated into proteins. The amount of resources each genetic program requires would be largely determined by the abundance of their representative transcripts. However, during stress when specific stress response programs would need to be prioritized ahead of others, a mechanism would be required to allocate system resources accordingly. A system monitor, a programming construct from computer science, which controls access, by multiple processes, to a common resource in a parallel programming environment, could be used to allocate priority to specific genetic programs [85-87]. tRNA could play this resource monitoring and allocating role. To do so, the "value" of the tRNA is re-written by enzymatic modification to designate its decoding partner, unlocking translation for specific codon using transcripts but excluding others.

To illustrate, we consider three tasks that are to be performed under stress, one of high priority – essential stress response genes such as *dosR*, a second of medium priority – housekeeping genes, and a third of low priority – genes whose expression is detrimental to survival under stress, such as those for the rapid synthesis of cellular material in preparation for division. Normally, high priority tasks would be allocated resources first, followed by medium priority ones, and finally low priority tasks (**Figure 7.2A**). At the hypoxic transition (**Supplementary Figure 6.1**), we observed that

tRNA$^{Thr(\underline{mo5U}GU)}$ is re-written to tRNA$^{Thr(\underline{cmo5U}GU)}$ which shifts the priority of translation to ACG (and ACA) biased transcripts from ACC biased ones. Decreases in tRNA$^{Thr(\underline{C}UG)}$ levels and steady levels of tRNA$^{Thr(\underline{G}UG)}$ ensures the primacy of stress response gene transcripts over housekeeping ones while excluding the processing of those using rare ACT codons (**Figure 7.2B**). By swapping synonymous codons, we inverted the priority of transcripts for translation. In *ΔdosSR::dosSR(ACC)* mutants for example (**Figure 6.5**), initial exposure to hypoxia failed to halt bacterial growth, eventually to the loss of viability and resuscitability. When *dosR* lost its ACG bias for the "optimal" ACC codon shared by other housekeeping genes (such as *sigA*), its priority in the translational queue was lost. tRNA$^{Thr(cmo5U)}$ normally used to grant access to threonines fails to recognize the re-engineered *dosR* transcripts, forcing it to compete with other ACC-rich housekeeping genes for threonine (**Figure 7.2D**). We note while *dosR* is best known for the inducing the transcription of dos regulon genes, it also binds to and suppress the expression of dozens of other genes [88, 89], and its complete loss in *dosSR* knockouts is deleterious to survival under hypoxia (**Figure 6.5D**). Thus, priority inversion occurs when one synonymous threonine codon is swapped for another, limiting the capacity of the system to execute high priority tasks (**Figure 7.2C and D**).

Obviously, this priority assignment model of selective translation is highly speculative. The validity of this model, however, could be tested by performing quantitative proteomics and transcriptomics analysis on the *dosR* re-engineered mutants. Rigourous mathemtical descriptions would be required to define its parameters and generate predictions of system behavior. Its general applicability would also need to be ascertained using synthetic biology approaches to engineer gene circuits in other organisms. Nonetheless, based on our understanding of tRNA modification mediated
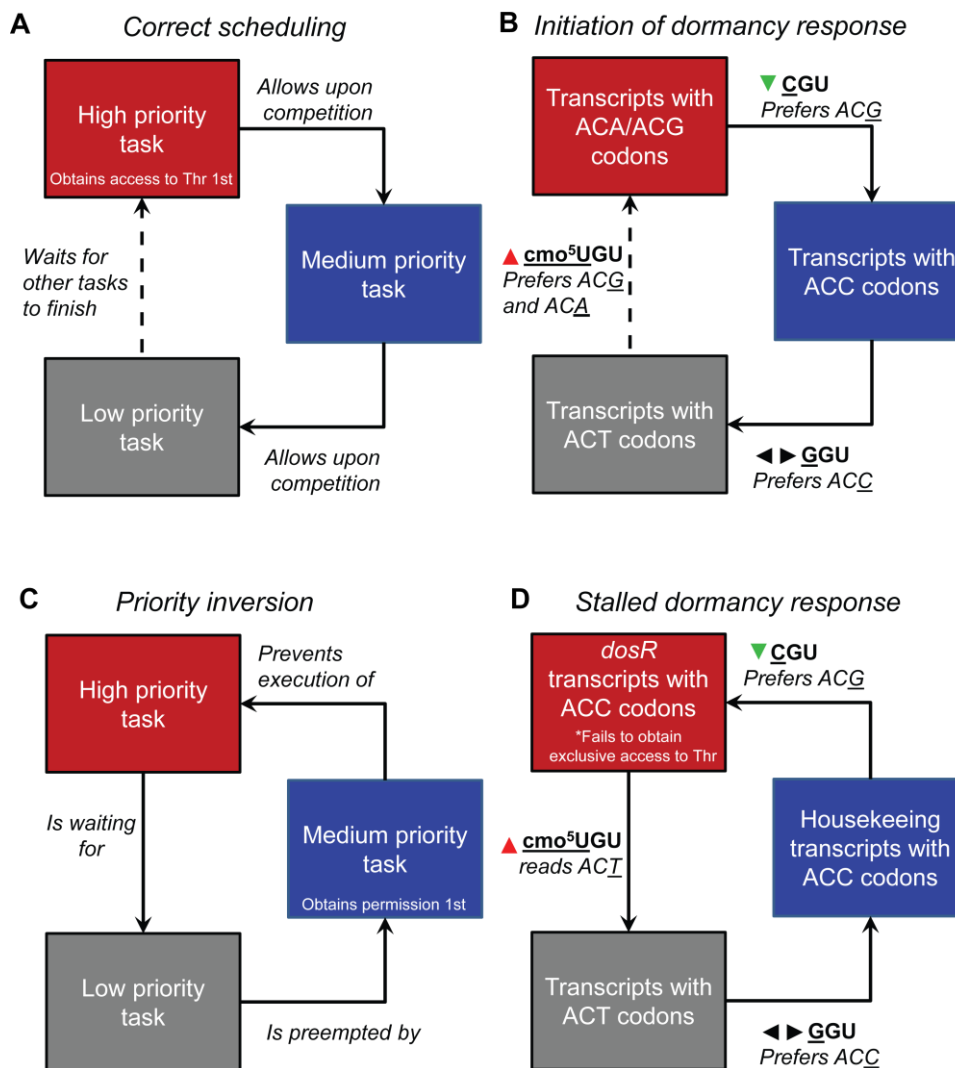
A *Correct scheduling*

High priority task
Obtains access to Thr 1st

Allows upon competition

Medium priority task

Waits for other tasks to finish

Low priority task

Allows upon competition

B *Initiation of dormancy response*

Transcripts with ACA/ACG codons

▼ **CGU**
Prefers ACG

Transcripts with ACC codons

▲ **cmo⁵UGU**
Prefers ACG and ACA

Transcripts with ACT codons

◄ ► **GGU**
Prefers ACC

C *Priority inversion*

High priority task

Prevents execution of

Medium priority task
Obtains permission 1st

Is waiting for

Low priority task

Is preempted by

D *Stalled dormancy response*

dosR transcripts with ACC codons
Fails to obtain exclusive access to Thr

▼ **CGU**
Prefers ACG

Housekeeing transcripts with ACC codons

▲ **cmo⁵UGU**
reads ACT

Transcripts with ACT codons

◄ ► **GGU**
Prefers ACC

Figure 7.2. tRNAs act as system monitors to schedule mRNA for translation. (**A**) Correct scheduling of translational tasks from high to medium to low priority. Access to the amino acid threonine is used as an example of a common system resource. (**B**) To initiate the dormancy response, levels of tRNA^{Thrcmo5UGU} increase allocating Thr to transcripts with ACG or ACA first. Though tRNA^{ThrCGU} decrease, it too favours transcripts using Thr^{ACG}. Levels of tRNA^{ThrGGU} which favors Thr^{ACC} over Thr^{ACT} codons remains relatively unchanged, thus, giving rise to the priority inheritance order of Thr^{ACG}≥Thr^{ACA}>Thr^{ACC}>Thr^{ACT}. Thr^{ACA} is a rare codon and used to a limited extent in a small number of transcripts. (**C**) A problematic scenario in scheduling occurs when a high priority task is indirectly preempted by a medium priority task effectively "inverting" the relative priorities of the two tasks. (**D**) This occurred for *dosR* transcripts when we swapped its Thr^{ACG} codons for Thr^{ACC}. While the remodeling events in the tRNA^{Thr} pool remains unchanged, *dosR* transcripts lost their priority. Instead, it now competes with other Thr^{ACC}-using transcripts for threonines. Thr^{ACC} is the most common Thr codon used in *Mtb* and BCG, hence, the translation of *dosR* becomes a purely stochastic event. While low levels of DosR aid cell survival, the dormancy response (represented by the dos regulon) fails to trigger, cells continue to replicate under hypoxia. This effect is exacerbated by the limited ability of tRNA^{Thrcmo5UGU} to read transcripts using Thr^{ACT}, further pushing mutant dosR transcripts down the translation queue.

selective translation, I propose that it is feasible to design a hybrid whole-cell target-based drug screen for theraputics which would block the dormancy response of mycobacteria.

Contemporary drug discovery efforts are based on two broad types of screens: phenotypic or whole-cell screens and target-based screens. The former looks at the effects, or phenotypes, that compounds induce in cells, tissues or whole organisms, whereas the latter measures the effect of compounds on a purified target protein via *in vitro* assays [90-92]. Each approach has its own drawbacks. While phenotypic screening can potentially lead to the identification of a molecule that modifies cellular phenotypes by acting on a previously undescribed target or by acting simultaneously acting on multiple targets, much effort and chemoproteomic resources are required to elucidate the target or targets affected by the lead compound [91, 93]. Target-based screening, on the other hand, uses our understanding of genetics, structural biology and biochemistry to provide specific and high throughput testing of large compound libraries for molecules that interact with a validated drug target. However, this reductionist approach fails to capture the pharmacology of small molecules in complex biological systems and as a result many candidate drug compounds fail to illicit their anticipated responses *in vivo* [94, 95]. Hence, several efforts had been made to develop hybrid screens which incorporate the strengths of both classical phenotypic screens and target-based screens without their associated limitations [95, 96].

Most antibiotics, including first line TB drugs, are more effective against replicating than non-replicating bacteria [97-99]. However, a major need in global health is to eradicate persistent or non-replicating subpopulations of *Mtb* [100, 101]. Mycobacterial persisters, however, they take days or weeks to generate and are difficult to maintain. Furthermore, it is difficult to assess cell

killing activities of compounds in persister cultures rapidly. Hence, it is difficult to develop phenotypic screens based on mycobacterial persister. Development of target-based screens is also complicated as mycobacterial persisters likely possess altered pharmacokinetic and pharmacodynamic properties which could invalidate *in vitro* findings (Chapter 5).

One solution is to prevent *Mtb* from entering dormancy altogether and thus, sensitize the pathogen to antibiotics that kill replicating bacteria. Instead of targeting just a single target such as *dosR* [102-104], the translation of multiple stress response proteins, including *dosR*, could be inhibited by targeting cmo^{5}U biosynthesis. This biosynthetic pathway, however, is not annotated in mycobacteria. Nonetheless, comparative genomic tools such as Integrated Microbial Genomes system (http://img.jgi.doe.gov/cgi-bin/w/main.cgi/), Clusters of Orthologous Groups (http://www.ncbi.nlm.nih.gov/COG/) and BioCys (http://biocyc.org/) could be used to identify the enzymes, CmoA and CmoB, responsible for cmo^{5}U biosynthesis. Already, we had identified several candidate genes including *BCG_0612* and *BCG_2975c* with a relatively high protein sequence homology to *E.coli cmoA* and *cmoB* respectively (see Chapter 6). Furthermore, we can evaluate the molecular phenotypes of transposon mutants from the TraSH library for the absence of cmo^{5}U [105, 106]. The identification of *cmoA* and *cmoB* allows us to construct a strain over-expressing these two enzymes. *cmoAB* over-expression would drive mycobacteria into a dormant-like state by increasing the tRNA$^{cmo5U}_{UGU}$ levels and driving selective translation of Thr^{ACG} biased proteins without the need for extensive and sustained stress exposures (**Figure 7.2**).

To attain a measurable signal for on-target therapeutic activates for high throughput screening, we would then introduce a dual fluorescence
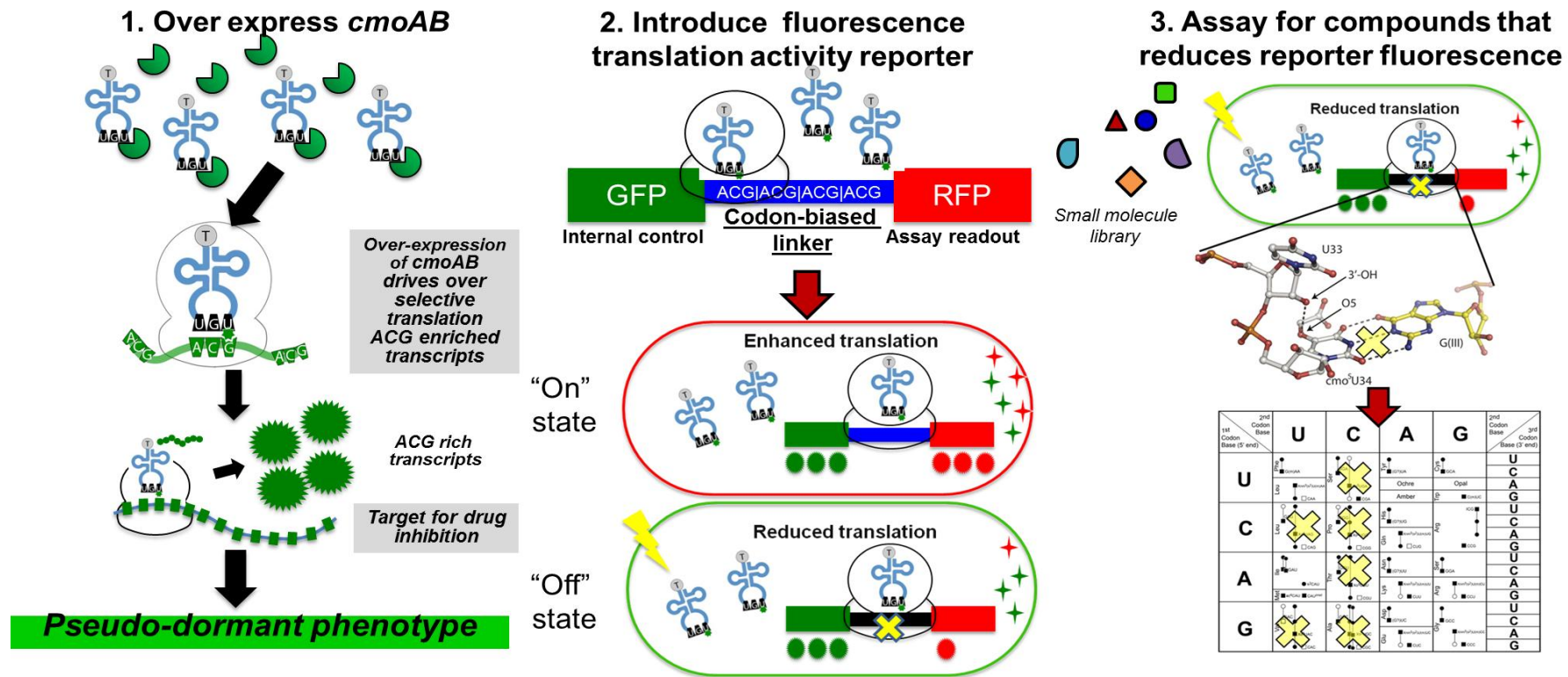
Figure 7.3. Proposed hybrid target-based phenotypic screen for small molecules that inhibit mycobacterial dormancy. 1. Over-expression of *cmoAB* leads to a "pseudo-dormant' phenotype presenting a target for drug inhibition without the need for extended hypoxic exposures. 2. A dual fluorescent reporter is introduced. When small molecules successfully enter and turn "off" enhanced translation, cells would appear more green than red. 3. High throughput screens against chemical libraries would reveal lead compound with the ability to disrupt multiple codon-anticodon interactions. Adapted from Protein Databank structure 2UU9 and Cai et al. (unpublished)

translation activity reporter into the genome of the *cmoAB* overexpression strain. This reporter codes for a fusion protein consisting of an internal control fluorescent domain (represented by green fluorescent protein (GFP) in **Figure 7.2**), a codon-biased linker made solely out of Thr^{ACG} codons, and a selective translation signaling fluorescent domain (represented by red fluorescent protein (RFP) in **Figure 7.2**). Baseline translational activities of cells exposed to different compounds could be tracked using GFP fluorescence while high levels of RFP fluorescence would only be detected when selective translation by tRNA$^{\underline{cmo5U}GU}$ took place. Increased tRNA$^{\underline{cmo5U}GU}$ levels would lead to the efficient read-through of the Thr^{ACG} biased linker allowing the RFP domain to be translated readily. Hence, we now have an indicator for successful on-target activity. This comprises of an "on" signal made out of both red and green fluorescence and an "off" signal made mainly of green fluorescence. The former indicates the failure of the compound to inhibit the enhanced translation of the Thr^{ACG} biased linker while the latter indicates successful on-target inhibition leading to reduced translation of the RFP domain (**Figure 7.2**). Cultures of these cells could be grown and dispensed into 384 well or 96 well pates for screens against drug compound libraries based on GFP and RFP fluorescence. Unlike traditional targeted-based screens against CmoA or CmoB, this approach also possess the added advantage of the whole cell, including the mycobacterial cell wall – a considerable pharmacological barrier [107] – remaining intact. Thus, hits are more likely to work *in vivo*. Secondary screens for target specificity could be performed by targeted mass spectrometry measurements of cmo^{5}U levels from drug candidate treated cells. An inhibitor of cmo^{5}U biosynthesis would be extremely potent as it would directly affect the decoding of 21 codons in 6 of the 8 four-codon boxes (**Figure 7.2** and **Supplementary Figure 6.2**); and could become an efficacious first-in-class drug against dormant *Mtb*.

7.5. Conclusion

The success of mycobacteria pathogens can be attributed to its extraordinary stealth and capacity to adapt to environmental changes throughout the course of infection. These changes include nutrient starvation and hypoxia, various exogenous stress conditions and the intra-phagosomal environment. In this thesis, the stress responses induced by nutrient starvation and hypoxia were studied in detail. To facilitate these studies we developed of an analytical platform for the isolation, purification and quantification of RNA and its modifications. This platform is generally applicable to other biological systems and will be useful to RNA scientists.

More importantly, my work furthers our understanding of mycobacterial stress responses and how it contributes to phenotypic drug resistance. The study of nutrient starvation led us to understand how a shift towards ketogenic metabolism and radical formation offer mechanistic insights into starvation-induced antibiotic tolerance and persistence. Key findings include:

1.   Nutrient deprivation generates viable but reversible persisters.

2.   Starvation survival is modulated in a biphasic manner – an early adaptive phase and an extended persistent phase.

3.   Lipolysis and ketogenesis are metabolic signatures of the adaptive and persistent phases respectively.

4.   Ketoacidosis is prevented by cytochrome P450s which oxidizes ketone bodies generating reactive oxygen species.

5.   The cytochrome P450 reactome bridges carbon cycling and xenobiotic detoxification contributing to antibiotic tolerance.

6.   Targeted killing of persisters achieved through disruption of the P450 reactome with Fenton chemistry.

Though mycobacterial stress responses have been an active field of research in the past two decades, knowledge has been mostly limited to adaptations at the level of transcription. Knowledge of the translational adaptions to physiological stresses had been hitherto limited due to the lack of a platform to quantify tRNA, its modifications and their dynamics. Just as previous work performed in the Dedon laboratory on the cytotoxic stress response of *S. cerevisiae* indicates strongly wobble modifications on tRNAs control gene expression through selective translation [108], I have provided evidence to suggest that the survival of pathogenic mycobacteria under hypoxic conditions is governed by a similar process. This process mechanistically involves key chemical modifications at the wobble base of tRNA to select for specific codons that are innately biased in all protein coding mRNA transcripts which, in turn, are essential to mount an appropriate response to external stresses. Key findings include:

1.   tRNA modifications are dynamic and change as a function of exposure to hypoxic stress.

2.   The wobble occupancy of cmo^5U and mo^5U on $tRNA^{Thr\underline{U}GU}$ are hallmarks of the dormant and replicating states.

3.   Signatures of tRNA reprogramming are prevalent in the hypoxic proteome.

4.   Up- and down-regulated proteins possess codon enrichments which define their expression during translation.

5.   Disruption of this "Code of Codons" adversely affects survival under hypoxia.

These studies show that metabolic shifts and tRNA reprogramming play essential roles in mycobacterial pathogenesis; and that their modulation could become promising targets for the development of new clinical and laboratory

tools that can aid TB management, improve diagnosis, and reverse antibiotic

tolerance.

7.6. References

1.   Sullivan, T. and Y. Ben Amor, *The co-management of tuberculosis and diabetes: challenges and opportunities in the developing world.* PLoS Med, 2012. **9**(7): p. e1001269.
2.   Dooley, K.E. and R.E. Chaisson, *Tuberculosis and diabetes mellitus: convergence of two epidemics.* Lancet Infect Dis, 2009. **9**(12): p. 737-46.
3.   The Lancet, *Diabetes and tuberculosis—a wake-up call.* The Lancet Diabetes & Endocrinology. **2**(9): p. 677.
4.   Jeon, C.Y. and M.B. Murray, *Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies.* PLoS Med, 2008. **5**(7): p. e152.
5.   Lonnroth, K., G. Roglic, and A.D. Harries, *Improving tuberculosis prevention and care through addressing the global diabetes epidemic: from evidence to policy and practice.* Lancet Diabetes Endocrinol, 2014. **2**(9): p. 730-9.
6.   Weixlbaumer, A., et al., *Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines.* Nat Struct Mol Biol, 2007. **14**(6): p. 498-502.
7.   Chang, J.T., et al., *Effect of type 2 diabetes mellitus on the clinical severity and treatment outcome in patients with pulmonary tuberculosis: a potential role in the emergence of multidrug-resistance.* J Formos Med Assoc, 2011. **110**(6): p. 372-81.
8.   Zhang, Q., H. Xiao, and I. Sugawara, *Tuberculosis complicated by diabetes mellitus at shanghai pulmonary hospital, china.* Jpn J Infect Dis, 2009. **62**(5): p. 390-1.
9.   Bashar, M., et al., *Increased incidence of multidrug-resistant tuberculosis in diabetic patients on the Bellevue Chest Service, 1987 to 1997.* Chest, 2001. **120**(5): p. 1514-9.
10.   Perez-Navarro, L.M., et al., *[Factors associated to pulmonary tuberculosis in patients with diabetes mellitus from Veracruz, Mexico].* Gac Med Mex, 2011. **147**(3): p. 219-25.
11.   Fisher-Hoch, S.P., et al., *Type 2 diabetes and multidrug-resistant tuberculosis.* Scand J Infect Dis, 2008. **40**(11-12): p. 888-93.
12.   Gu, C., T.J. Begley, and P.C. Dedon, *tRNA modifications regulate translation during cellular stress.* FEBS Lett, 2014. **588**(23): p. 4287-96.
13.   Restrepo, B.I. and L.S. Schlesinger, *Host-pathogen interactions in tuberculosis patients with type 2 diabetes mellitus.* Tuberculosis (Edinb), 2013. **93 Suppl**: p. S10-4.
14.   Odone, A., et al., *The effect of diabetes and undernutrition trends on reaching 2035 global tuberculosis targets.* Lancet Diabetes Endocrinol, 2014. **2**(9): p. 754-64.
15.   Podell, B.K., et al., *Increased severity of tuberculosis in Guinea pigs with type 2 diabetes: a model of diabetes-tuberculosis comorbidity.* Am J Pathol, 2014. **184**(4): p. 1104-18.

16. Webb, E.A., et al., *High prevalence of Mycobacterium tuberculosis infection and disease in children and adolescents with type 1 diabetes mellitus.* Int J Tuberc Lung Dis, 2009. **13**(7): p. 868-74.

17. Martens, G.W., et al., *Tuberculosis susceptibility of diabetic mice.* Am J Respir Cell Mol Biol, 2007. **37**(5): p. 518-24.

18. Cegielski, J.P. and D.N. McMurray, *The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals.* Int J Tuberc Lung Dis, 2004. **8**(3): p. 286-98.

19. Misra, S., N. Oliver, and A. Dornhorst, *Diabetic ketoacidosis: not always due to type 1 diabetes.* BMJ, 2013. **346**: p. f3501.

20. Welch, B.J. and I. Zib, *Case Study: Diabetic Ketoacidosis in Type 2 Diabetes: "Look Under the Sheets".* Clinical Diabetes, 2004. **22**(4): p. 198-200.

21. Hall, S.E., et al., *Ketone body kinetics in humans: the effects of insulin-dependent diabetes, obesity, and starvation.* J Lipid Res, 1984. **25**(11): p. 1184-94.

22. Beech, J.S., et al., *Gluconeogenesis and the protection of hepatic intracellular pH during diabetic ketoacidosis in rats.* Biochem J, 1989. **263**(3): p. 737-44.

23. Hickman, D., et al., *Evaluation of the selectivity of In vitro probes and suitability of organic solvents for the measurement of human cytochrome P450 monooxygenase activities.* Drug Metab Dispos, 1998. **26**(3): p. 207-15.

24. Donato, M.T. and J.V. Castell, *Strategies and molecular probes to investigate the role of cytochrome P450 in drug metabolism: focus on in vitro studies.* Clin Pharmacokinet, 2003. **42**(2): p. 153-78.

25. Singhal, A., et al., *Experimental tuberculosis in the Wistar rat: a model for protective immunity and control of infection.* PLoS One, 2011. **6**(4): p. e18632.

26. Heng, Y., et al., *Mycobacterium tuberculosis infection induces hypoxic lung lesions in the rat.* Tuberculosis (Edinb), 2011. **91**(4): p. 339-41.

27. Wise, J.K., R. Hendler, and P. Felig, *Influence of glucocorticoids on glucagon secretion and plasma amino acid concentrations in man.* J Clin Invest, 1973. **52**(11): p. 2774-82.

28. Akiyama, T., et al., *High-fat hypercaloric diet induces obesity, glucose intolerance and hyperlipidemia in normal adult male Wistar rat.* Diabetes Res Clin Pract, 1996. **31**(1-3): p. 27-35.

29. Likhodii, S.S., et al., *Dietary fat, ketosis, and seizure resistance in rats on the ketogenic diet.* Epilepsia, 2000. **41**(11): p. 1400-10.

30. Dong, H., et al., *2'-O methylation of internal adenosine by flavivirus NS5 methyltransferase.* PLoS Pathog, 2012. **8**(4): p. e1002642.

31. Katajamaa, M. and M. Oresic, *Data processing for mass spectrometry-based metabolomics.* J Chromatogr A, 2007. **1158**(1-2): p. 318-28.

32. Reader, J.S., et al., *Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine.* J Biol Chem, 2004. **279**(8): p. 6280-5.

33. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes.* BMC Bioinformatics, 2003. **4**: p. 41.

34. Dumelin, C.E., et al., *Discovery and biological characterization of geranylated RNA in bacteria.* Nat Chem Biol, 2012. **8**(11): p. 913-9.

35. Morelle, W. and J.C. Michalski, *Analysis of protein glycosylation by mass spectrometry.* Nat Protoc, 2007. **2**(7): p. 1585-602.
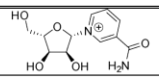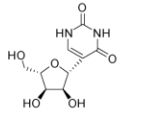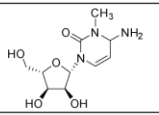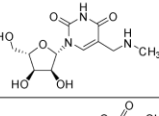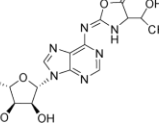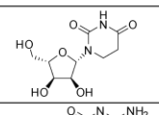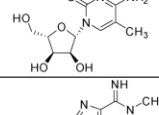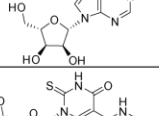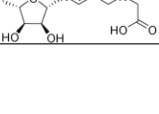
36. Desgres, J., et al., *Presence of phosphorylated O-ribosyl-adenosine in T-psi-stem of yeast methionine initiator tRNA.* Nucleic Acids Res, 1989. **17**(3): p. 865-82.

37. Keith, G., et al., *Identification and structural characterization of O-beta-ribosyl-(1"----2')-adenosine-5"-phosphate in yeast methionine initiator tRNA.* Nucleic Acids Res, 1990. **18**(20): p. 5989-93.

38. Zhou, X., et al., *A novel DNA modification by sulphur.* Mol Microbiol, 2005. **57**(5): p. 1428-38.

39. Wang, L., et al., *Phosphorothioation of DNA in bacteria by dnd genes.* Nat Chem Biol, 2007. **3**(11): p. 709-10.

40. Layre, E., et al., *Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c.* Proc Natl Acad Sci U S A, 2014. **111**(8): p. 2978-83.

41. Teichert, F., et al., *Evaluation of urinary ribonucleoside profiling for clinical biomarker discovery using constant neutral loss scanning liquid chromatography/tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2011. **25**(14): p. 2071-82.

42. Dudley, E., et al., *Urinary modified nucleosides as tumor markers.* Nucleosides Nucleotides Nucleic Acids, 2003. **22**(5-8): p. 987-9.

43. Gehrke, C.W., et al., *Patterns of urinary excretion of modified nucleosides.* Cancer Res, 1979. **39**(4): p. 1150-3.

44. Lau, S.K.P., et al., *Identification of specific metabolites in culture supernatant of Mycobacterium tuberculosis using metabolomics: exploration of potential biomarkers.* Emerg Microbes Infect, 2015. **4**: p. e6.

45. Chan, C.T., et al., *A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress.* PLoS Genet, 2010. **6**(12): p. e1001247.

46. Zumla, A., et al., *Tuberculosis.* N Engl J Med, 2013. **368**(8): p. 745-55.

47. Rabalais, G., G. Adams, and B. Stover, *PPD skin test conversion in health-care workers after exposure to Mycobacterium tuberculosis infection in infants.* Lancet, 1991. **338**(8770): p. 826.

48. Joncas, J.H., R. Robitaille, and T. Gauthier, *Interpretation of the PPD skin test in BCG-vaccinated children.* Can Med Assoc J, 1975. **113**(2): p. 127-8.

49. O'Garra, A., et al., *The immune response in tuberculosis.* Annu Rev Immunol, 2013. **31**: p. 475-527.

50. Boehme, C.C., et al., *Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study.* Lancet, 2011. **377**(9776): p. 1495-505.

51. Theron, G., et al., *Accuracy and impact of Xpert MTB/RIF for the diagnosis of smear-negative or sputum-scarce tuberculosis using bronchoalveolar lavage fluid.* Thorax, 2013. **68**(11): p. 1043-51.

52. Sohn, H., et al., *Xpert MTB/RIF testing in a low tuberculosis incidence, high-resource setting: limitations in accuracy and clinical impact.* Clin Infect Dis, 2014. **58**(7): p. 970-6.

53. Lappann, M., et al., *Comparative proteome analysis of spontaneous outer membrane vesicles and purified outer membranes of Neisseria meningitidis.* J Bacteriol, 2013. **195**(19): p. 4425-35.

54. Pethe, K., et al., *Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation.* Proc Natl Acad Sci U S A, 2004. **101**(37): p. 13642-7.
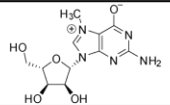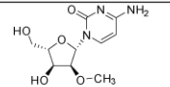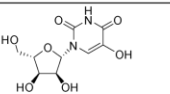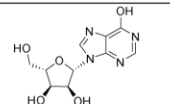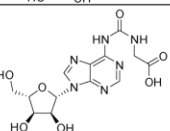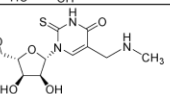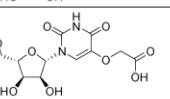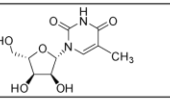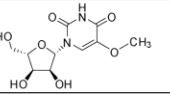
55. Welin, A., et al., *Importance of phagosomal functionality for growth restriction of Mycobacterium tuberculosis in primary human macrophages.* J Innate Immun, 2011. **3**(5): p. 508-18.

56. Gupta, U.D. and V.M. Katoch, *Animal models of tuberculosis.* Tuberculosis (Edinb), 2005. **85**(5-6): p. 277-93.

57. Russell, B.S., *Nucleic acid modifications in bacterial pathogens - impact on pathogenesis, diagnosis, and therapy*, in *Department of Biological Engineering*. 2014, Massachusetts Institute of Technology: Cambridge. p. 129.

58. Tuller, T., et al., *Translation efficiency is determined by both codon bias and folding energy.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3645-50.

59. Sauna, Z.E. and C. Kimchi-Sarfaty, *Understanding the contribution of synonymous mutations to human disease.* Nat Rev Genet, 2011. **12**(10): p. 683-91.

60. Guimaraes, J.C., M. Rocha, and A.P. Arkin, *Transcript level and sequence determinants of protein abundance and noise in Escherichia coli.* Nucleic Acids Res, 2014. **42**(8): p. 4791-9.

61. Novoa, E.M., et al., *A role for tRNA modifications in genome structure and codon usage.* Cell, 2012. **149**(1): p. 202-13.

62. Cannarozzi, G., et al., *A role for codon order in translation dynamics.* Cell, 2010. **141**(2): p. 355-67.

63. dos Reis, M., R. Savva, and L. Wernisch, *Solving the riddle of codon usage preferences: a test for translational selection.* Nucleic Acids Res, 2004. **32**(17): p. 5036-44.

64. Saunders, R. and C.M. Deane, *Synonymous codon usage influences the local protein structure observed.* Nucleic Acids Res, 2010. **38**(19): p. 6719-28.

65. Stadler, M. and A. Fire, *Wobble base-pairing slows in vivo translation elongation in metazoans.* RNA, 2011. **17**(12): p. 2063-73.

66. Bulmer, M., *Coevolution of codon usage and transfer RNA abundance.* Nature, 1987. **325**(6106): p. 728-30.

67. Carbone, A., A. Zinovyev, and F. Kepes, *Codon adaptation index as a measure of dominating codon bias.* Bioinformatics, 2003. **19**(16): p. 2005-15.

68. Tuller, T., et al., *An evolutionarily conserved mechanism for controlling the efficiency of protein translation.* Cell, 2010. **141**(2): p. 344-54.

69. Wohlgemuth, S.E., T.E. Gorochowski, and J.A. Roubos, *Translational sensitivity of the Escherichia coli genome to fluctuating tRNA availability.* Nucleic Acids Res, 2013. **41**(17): p. 8021-33.

70. Shah, P. and M.A. Gilchrist, *Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10231-6.

71. Mao, Y., et al., *Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae.* Nucleic Acids Res, 2014. **42**(8): p. 4813-22.

72. Trotta, E., *Selection on codon bias in yeast: a transcriptional hypothesis.* Nucleic Acids Res, 2013. **41**(20): p. 9382-95.

73. Shabalina, S.A., N.A. Spiridonov, and A. Kashina, *Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity.* Nucleic Acids Res, 2013. **41**(4): p. 2073-94.

74. Zhou, M., et al., *Non-optimal codon usage affects expression, structure and function of clock protein FRQ.* Nature, 2013. **495**(7439): p. 111-5.
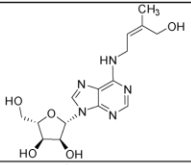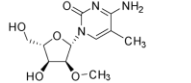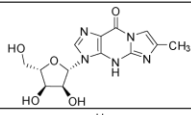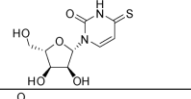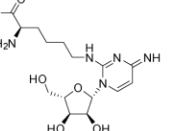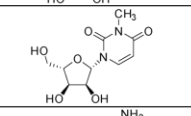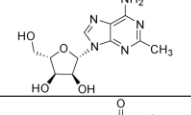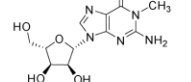
75. Gingold, H., O. Dahan, and Y. Pilpel, *Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome.* Nucleic Acids Res, 2012. **40**(20): p. 10053-63.

76. Iben, J.R. and R.J. Maraia, *tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote.* RNA, 2012. **18**(7): p. 1358-72.

77. Limor-Waisberg, K., et al., *Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies.* Nucleic Acids Res, 2011. **39**(14): p. 6016-28.

78. Andersson, G.E. and P.M. Sharp, *Codon usage in the Mycobacterium tuberculosis complex.* Microbiology, 1996. **142 (Pt 4)**: p. 915-25.

79. Subramaniam, A.R., T. Pan, and P. Cluzel, *Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria.* Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2419-24.

80. Vieira-Silva, S., et al., *Investment in rapid growth shapes the evolutionary rates of essential proteins.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20030-5.

81. Hershberg, R. and D.A. Petrov, *Selection on codon bias.* Annu Rev Genet, 2008. **42**: p. 287-99.

82. Ingolia, N.T., *Ribosome profiling: new views of translation, from single codons to genome scale.* Nat Rev Genet, 2014. **15**(3): p. 205-13.

83. Ingolia, N.T., et al., *Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes.* Cell Rep, 2014. **8**(5): p. 1365-79.

84. Pang, Y.L., et al., *Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number.* Nucleic Acids Res, 2014. **42**(22): p. e170.

85. Hoare, C.A.R., *Monitors: an operating system structuring concept.* Commun. ACM, 1974. **17**(10): p. 549-557.

86. Lamport, L., *A new solution of Dijkstra's concurrent programming problem.* Commun. ACM, 1974. **17**(8): p. 453-455.

87. Lampson, B.W. and D.D. Redell, *Experience with processes and monitors in Mesa.* Commun. ACM, 1980. **23**(2): p. 105-117.

88. Galagan, J.E., et al., *The Mycobacterium tuberculosis regulatory network and hypoxia.* Nature, 2013. **499**(7457): p. 178-83.

89. Peterson, E.J., et al., *A high-resolution network model for global gene regulation in Mycobacterium tuberculosis.* Nucleic Acids Res, 2014. **42**(18): p. 11291-303.

90. An, W.F. and N. Tolliday, *Cell-based assays for high-throughput screening.* Mol Biotechnol, 2010. **45**(2): p. 180-6.

91. Eder, J., R. Sedrani, and C. Wiesmann, *The discovery of first-in-class drugs: origins and evolution.* Nat Rev Drug Discov, 2014. **13**(8): p. 577-87.

92. Sundberg, S.A., *High-throughput and ultra-high-throughput screening: solution- and cell-based approaches.* Curr Opin Biotechnol, 2000. **11**(1): p. 47-53.

93. Hall, S.E., *Chemoproteomics-driven drug discovery: addressing high attrition rates.* Drug Discov Today, 2006. **11**(11-12): p. 495-502.

94. Swinney, D.C. and J. Anthony, *How were new medicines discovered?* Nat Rev Drug Discov, 2011. **10**(7): p. 507-519.

95. Moffat, J.G., J. Rudolph, and D. Bailey, *Phenotypic screening in cancer drug discovery - past, present and future.* Nat Rev Drug Discov, 2014. **13**(8): p. 588-602.
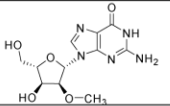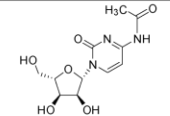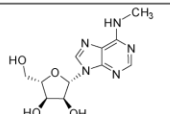
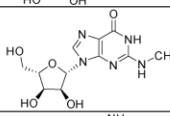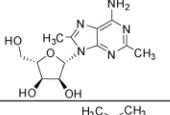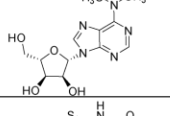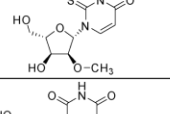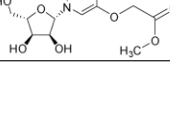96.   Lee, J.A., et al., *Modern phenotypic drug discovery is a viable, neoclassic pharma strategy.* J Med Chem, 2012. **55**(10): p. 4527-38.

97.   Walsh, C., *Where will new antibiotics come from?* Nat Rev Microbiol, 2003. **1**(1): p. 65-70.

98.   Levin, B.R. and D.E. Rozen, *Non-inherited antibiotic resistance.* Nat Rev Microbiol, 2006. **4**(7): p. 556-62.

99.   Gengenbacher, M., et al., *Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability.* Microbiology, 2010. **156**(Pt 1): p. 81-7.

100.   Munoz-Elias, E.J. and J.D. McKinney, *Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence.* Nat Med, 2005. **11**(6): p. 638-44.

101.   Rogerson, B.J., et al., *Expression levels of Mycobacterium tuberculosis antigen-encoding genes versus production levels of antigen-specific T cells during stationary level lung infection in mice.* Immunology, 2006. **118**(2): p. 195-201.

102.   Murphy, D.J. and J.R. Brown, *Identification of gene targets against dormant phase Mycobacterium tuberculosis infections.* BMC Infect Dis, 2007. **7**: p. 84.

103.   Zhang, Y., *The magic bullets and tuberculosis drug targets.* Annu Rev Pharmacol Toxicol, 2005. **45**: p. 529-64.

104.   Hasan, S., et al., *Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis.* PLoS Comput Biol, 2006. **2**(6): p. e61.

105.   Long, J.E., et al., *Identifying Essential Genes in Mycobacterium tuberculosis by Global Phenotypic Profiling.* Methods Mol Biol, 2015. **1279**: p. 79-95.

106.   Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Mol Microbiol, 2003. **48**(1): p. 77-84.

107.   Nikaido, H., *Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria.* Semin Cell Dev Biol, 2001. **12**(3): p. 215-23.

108.   Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.
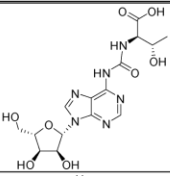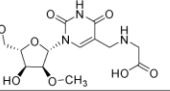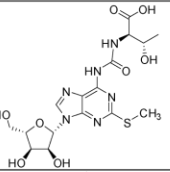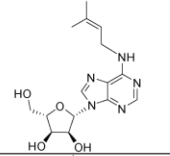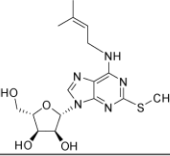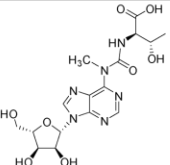
Appendix I: Modified ribonucleosides in Mycobacterium bovis BCG tRNA

Identity	Symbol	Structure	Retention Time (min)	Monoisotopic Mass (amu)	Observed Molecular ion (m/z)	Major Fragment (m/z)	Neutral Loss (amu)	Quantifier SRM Transistion (Precursor ion → Product ion)	Standards
N-ribosylnicotinamide	r-NA		1.51	255.098	255.101 ± 0.005	123.055 ± 0.002	132.046	255.1 → 123.1	Alkaline hydrolysis of NAD+[a]
Pseudouridine	Ψ		1.63	244.070	245.077 ± 0.003	209.052 ± 0.004[b]	36.025	245.1 → 209.1	Carbosynth Ltd. USA
3-Methylcytidine	m3C		2.01	257.101	258.102 ± 0.004	126.061 ± 0.002	132.041	258.1 → 126.1	MP Biochemicals LLC. USA
5-Methylaminomethyluridine	mnm5U		2.17	287.112	288.116 ± 0.006	156.074 ± 0.004	132.042	288.1 → 156.1	Armengod ME and Navarro-Gonzalez C. Laboratorio de Genética Molecular, Centro de Investigación Príncipe Felipe, Spain
Cyclic *N*6-threonylcarbamoyladenosine	ct6A		2.21	394.124	395.128 ± 0.004	263.089 ± 0.005	132.039	395.1 → 236.1	N/A[c]
Dihydrouridine	D		2.25	246.085	247.092 ± 0.005	115.050 ± 0.002	132.042	247.1 → 115.1	Advanced Technology & Industrial Co., Ltd. Hong Kong, China
5-Methylcytidine	m5C		2.53	257.101	258.107 ± 0.004	126.064 ± 0.002	132.043	258.1 → 126.1	Sigma-Aldrich. USA
1-Methyladenosine	m1A		2.62	281.112	282.115 ± 0.007	150.073 ± 0.003	132.042	282.1 → 150.1	Barry and Associates Inc. USA
5-Carboxymethylaminomethyl-2-thiouridine	cmnm5s2U		3.96	347.079	348.082 ± 0.004	216.047 ± 0.009	132.042	348.1 → 216.0	Armengod ME and Navarro-Gonzalez C. Laboratorio de Genética Molecular, Centro de Investigación Príncipe Felipe, Spain[d]

Identity	Symbol	Structure	Retention Time (min)	Monoisotopic Mass (amu)	Observed Molecular ion (m/z)	Major Fragment (m/z)	Neutral Loss (amu)	Quantifier SRM Transistion (Precursor ion → Product ion)	Standards
7-Methylguanosine	m^{7}G		4.78	297.107	298.113 ± 0.005	166.071 ± 0.002	132.042	298.1 → 166.1	Sigma-Aldrich. USA
2'-O-methylcytidine	Cm		4.91	257.101	258.103 ± 0.006	112.050 ± 0.002	146.053	258.1 → 112.1	Carbosynth Ltd. USA
5-Hydroxyuridine	ho^{5}U		4.96	260.064	261.68 ± 0.008	129.026 ± 0.009	132.042	261.1 → 129.0	Carbosynth Ltd. USA
Inosine	I		5.53	268.081	269.088 ± 0.007	137.047 ± 0.002	132.042	268.1 → 137.0	Sigma-Aldrich. USA
N^6-glycinylcarbamoyladenosine	g^{6}A		5.55	368.108	369.113 ± 0.005	136.062 ± 0.003	132.042	369.1 → 136.1	N/A
5-Methylaminomethyl-2-thiouridine	mnm^{5}s^{2}U		5.79	303.089	304.098 ± 0.012	172.057 ± 0.012	132.041	304.1 → 172.1	N/A^e
Uridine 5-oxyacetic acid	cmo^{5}U		5.82	318.070	319.078 ± 0.010	187.035 ± 0.011	132.043	319.1 → 187.0	Dziergowska A and Malkiewicz A. Institute of Organic Chemistry, Technical University of Lodz. Poland
5-Methyluridine	m^{5}U		6.42	258.085	259.089 ± 0.004	127.047 ± 0.003	132.042	259.1 → 127.0	Sigma-Aldrich. USA
5-Methoxyuridine	mo^{5}U		6.54	274.080	275.084 ± 0.009	143.042 ± 0.007	132.042	275.1 → 143.0	BOC Sciences-Creative Dynamics Inc. USA

Identity	Symbol	Structure	Retention Time (min)	Monoisotopic Mass (amu)	Observed Molecular ion (m/z)	Major Fragment (m/z)	Neutral Loss (amu)	Quantifier SRM Transistion (Precursor ion → Product ion)	Standards
N^6-(cis-hydroxyisopentenyl)adenosine	io^{6}A		6.69	351.154	352.156 ± 0.006	220.115 ± 0.005	132.041	352.2 → 220.1	N/A
5,2'-O-dimethylcytidine	m^{5}Cm		6.71	271.117	272.122 ± 0.009	126.065 ± 0.002	146.057	272.1 → 126.1	Barry and Associates Inc. USA
4-Demethylwyosine	imG-14		6.75	321.107	322.115 ± 0.008	190.074 ± 0.007	132.041	322.1 → 190.1	N/A^f
4-Thiouridine	s^{4}U		6.94	260.047	261.051 ± 0.006	129.009 ± 0.008	132.042	261.0 → 129.0	Sigma-Aldrich. USA
2-Lysidine	k^{2}C		8.06	371.180	372.182 ± 0.006	240.140 ± 0.007	132.042	372.2 → 240.1	N/A
3-Methyluridine	m^{3}U		9.56	258.085	259.089 ± 0.011	127.043 ± 0.011	132.046	259.1 → 127.0	Barry and Associates Inc. USA
2-Methyladenosine	m^{2}A		11.27	281.112	282.116 ± 0.005	150.074 ± 0.008	132.042	282.1 → 150.1	Carbosynth Ltd. USA
1-Methylguanosine	m^{1}G		14.36	297.107	298.111 ± 0.006	166.071 ± 0.002	132.040	298.1 → 166.1	N/A^g

Identity	Symbol	Structure	Retention Time (min)	Monoisotopic Mass (amu)	Observed Molecular ion (m/z)	Major Fragment (m/z)	Neutral Loss (amu)	Quantifier SRM Transistion (Precursor ion → Product ion)	Standards
2'-O-methylguanosine	Gm		14.98	297.107	298.109 ± 0.005	152.056 ± 0.009	146.053	298.1 →152.1	Barry and Associates Inc. USA
N^4-Acetylcytidine	ac^4C		15.5	285.096	286.101 ± 0.005	154.056 ± 0.005	132.044	286.1 → 154.1	Santa Cruz Biotechnology, Inc. USA
N^6-methyladenosine	m^6A		16.85	281.112	282.114 ± 0.008	150.071 ± 0.005	132.043	282.1 → 150.1	Carbosynth Ltd. USA
N^2-methylguanosine	m^2G		17.31	297.107	298.107 ± 0.007	166.064 ± 0.004	132.043	298.1 → 166.1	Sigma-Aldrich. USA
2,8-Dimethyladenosine[d]	$m^{2,8}A$		19.29	295.128	296.133 ± 0.012	164.090 ± 0.011	132.043	296.1 → 164.1	N/A[h]
N^6,N^6-dimethyladenosine	$m^{6,6}A$		20.82	295.128	296.131 ± 0.007	164.088 ± 0.008	132.043	296.1 → 164.1	Chembridge Corp. USA
2-Thio-2'-O-methyluridine	s^2Um		20.87	274.062	275.067 ± 0.009	129.008 ± 0.011	146.059	275.1 → 129.0	N/A[i]
Uridine 5-oxyacetic acid methyl ester	$mcmo^5U$		21.12	332.086	333.085 ± 0.008	201.041± 0.005	132.044	333.1 → 201.0	N/A[j]

Identity	Symbol	Structure	Retention Time (min)	Monoisotopic Mass (amu)	Observed Molecular ion (m/z)	Major Fragment (m/z)	Neutral Loss (amu)	Quantifier SRM Transistion (Precursor ion → Product ion)	Standards
N^6-threonylcarbamoyladenosine	t^6A		21.99	412.134	413.136 ± 0.005	136.062 ± 0.004	132.043	413.1 → 136.1	BioLog Life Science Institute. Germany
5-Carboxymethylaminomethyl- 2'-O-methyluridine	$cmnm^5Um$		22.15	345.117	346.117 ± 0.008	200.058 ± 0.010	146.059	346.1 → 200.1	N/A[k]
2-Methylthio-N^6-threonyl carbamoyladenosine	ms^2t^6A		22.20	458.112	459.117 ± 0.012	182.049 ± 0.009	132.043	459.1 → 182.0	N/A[l]
N^6-isopentenyladenosine	i^6A		22.56	335.159	336.162 ± 0.008	204.119 ± 0.006	132.043	336.2 →+204.1	Carbosynth Ltd. USA
2-Methylthio-N^6-isopentenyladenosine	ms^2i^6A		23.31	381.147	382.151 ± 0.007	250.108 ± 0.005	132.043	382.2 → 250.1	Carbosynth Ltd. USA
N^6-methyl-N^6-threonylcarbamoyladenosine	m^6t^6A		23.5	426.150	427.151 ± 0.011	295.111 ± 0.007	132.040	427.2 → 295.1	N/A[m]

Notes:

a: Method of S. Kellner et al., Chem Commun (Camb) 50, 3516 (Apr 4, 2014).

b: Previously reported by E. Dudley et al., Rapid Commun Mass Spectrom 19, 3075 (2005).

c: MS2 fragmention at collision energies of 8eV, 15eV and 30eV matched with previously reported CID spectra (K. Miyauchi, S. Kimura, T. Suzuki, Nat Chem Biol 9, 105 (Feb, 2013).).

d: Isolation and purified by I. Moukadiri, M. J. Garzon, G. R. Bjork, M. E. Armengod, Nucleic Acids Res 42, 2602 (Feb, 2014).

e: Further confirmed by MS2 fragmentation of cmnm^{5}s^{2}U standard at collision energies of 8eV, 15eV and 30eV.

f: Previously identified by Y. H. Chionh et al., Nucleic Acids Res 41, e168 (Sep, 2013). and C. T. Chan et al., Molecules 16, 5168 (2011).

g. Further confirmed by MS2 fragmentation of m^{2}G and m^{7}G standards at collision energies of 8eV, 20eV and 45eV.

h: MS2 and MS3 fragments match with previous reported CID spectra (A. M. Giessing et al., RNA 15, 327 (Feb, 2009).

i: Further confirmed by MS2 fragmentation of s^{2}U standard at collision energies of 8eV, 25eV and 45eV.

j: Further confirmed by MS2 fragmentation of cmo^{5}U standard at collision energies of 8eV, 15eV and 30eV.

k: Further confirmed by MS2 fragmentation of cmnm^{5}U standard at collision energies of 8eV, 15eV and 30eV.

l: Further confirmed by MS2 and MS3 fragmentation of t^{6}A and ms^{2}i^{6}A standards at collision energies of 8eV, 15eV, 25eV and 45eV.

m: Further confirmed by MS2 of t^{6}A at collision energies of 8eV, 15eV and 30 eV.

Appendix II: Sequences inserted at *attnB* site of Δ*dosSR* complements[1]

ΔdosSR::dosSR(WT)[2,3]:

cggcggatcggagattgatcgacgccgtgccggccgcatcgagatgcgctctcacagcgcggctatgcccaa
atattctgtgggtcaagcgatatgcagccgatggacggccgctggttcggcagctgtcggcaactgtaagccattt
ctgggactttgctgtgaaaagctgggcgatggttgtggacctggacgagccacccgtgcgataggtgagattcat
tctcgccctgacgggttgcgtctgtcatcggtcgataaggactaacggccctcaggtggggaccaacgcccctg
ggagatagcggtccccgccagtaacgtaccgctgaaccgacgggatgtatccgccccagcgaaggagacg
gcg*atg*agcgatcctcggccagctcgggcagtggtcgttggtatcgacgggtcaagggcggcaacgcatgcg
gcgttgtgggcggtcgatgaggcggtgaaccgagacattccgctgcgactggtgtacgtcatcgatccgtccca
actgtccgccgccggcgagggcggtgggcaatcagcggcccgagcggcgctgcacgacgcctctcggaag
gtcgaggccaccgggcaaccggtcaagatcgaaacggaggttctgtgcggcaggccgctcaccaagctgat
gcaggagtccaggtccgcggcgatgctgtgcgtcggttcggtggggcttgatcatgtccgcggtcgccggggttc
ggtcgcggcgaccctggctgggtcggccttatgccccgtggcggtgattcacccgtcgccggccgagccagcg
acaacctcccaggtcagcgcggttgtcgcggaggtggacaatggtgtggtgctgcggcacgcattcgaggagg
ccaggctgcgcggagttccgctgcgggccgtggctgtccacgctgctgaaacacccgatgacgtcgaacagg
gcagccggttggcgcatgtacacctgagccgtcggctcgcccactggacccggctctaccccgaggtgcgggt
ggatcgggccatcgccggcggcagtgcgtgccgtcatctggccgccaacgcaaagccgggtcagctgttcgtc
gcggactcacactccgcgcacgaattgtgcggtgcataccagcccggatgcgccgtacttacggtacgcagtg
ccaacttgtag***AAGCTT***gatcttgggagtggtgccctg*gtg*gtaaaggtcttcttggtcgatgaccacgaggtg
gtgcgtcgtggtctggttgacttgcttggggccgatcccgagcttgacgtcgtaggtgaggcgggttcggtcgccg
aggcgatggccagggttcctgccgcgcgcccagatgtcgcggtgctggatgtccggttgcccgatggcaacgg
cattgaactgtgccgcgatctgttgtcccgcatgcccgatctgcgctgtctgatcctcacgtcctacacctctgacga
ggccatgctagatgcgattctcgccggtgccagcggatatgtcgtcaaagacatcaagggaatggagttggcg
cgcgccgtcaaagatgtgggcgctggacggtcgctgctggacaatcgggccgcggccgcgctgatggccaa
gctgcgcggtgccgccgagaagcaggacccgctatcaggccttaccgaccaggagcggacgctactgggcc
tgcttagcgagggcctgaccaacaagcagatcgccgaccgaatgttcctagccgaaaagacggtgaagaact
acgtgtcgcggttgctggccaagctgggcatggaacgtcggacgcaagccgcggtattcgcgacggagttgaa
gcgctcgcggccacccggtgatggacc*atga*caacagggggcctcgtcgacgaaaacgacggcgccgcaa
tgcgtccactgcgtcacacgctctcccaactacgcctgcacgagctgctggtcgaggtgcaggaccgggtcga
gcagatcgtcgagggccgggaccgcctcgatggtctggtggaggccatgctcgtggtcacagcgggcctgga
cctggaggcaaccctacgcgctatcgtgcattcagcgaccagccttgtcgatgcgcgctatggcgctatggaggt
gcacgaccggcagcatcgggtattgcactttgtctatgaaggcatcgacgaggagaccgttcggcggatcggc
cacctaccgaaaggcctaggcgtcatcgggctgctcatcgaagatcccaaaccgttacggctggacgatgtttct
gcgcacccggcctcgattggttttccgccgtatcatccgccgatgcgtaccttcctcgggggtaccggttcgggtgc
gcgatgaatcgttcggcactctgtacctgactgacaagaccaacgggcaaccgttcagcgacgacgacgagg
ttctggtccaggcgctggcggccgccgcgggtatcgcagtcgcgaatgcccggctctaccagcaggctaaggc
gcgtcagtcgtggatcgaggccacccgtgacatcgccaccgagttgttgtccggcaccgaacccgcgacggtg
ttccggcttgtcgccgcggaggcgctcaagctgacggcggctgacgctgccctggtagccgttcccgtcgacga
ggacatgcctgccgctgacgtggggggagctgctggtgattgaaacagtcggcagcgctgtggcttccactgttgg

---

[1] Confirmed by Sanger sequencing (performed by GeneScript USA Inc.).
[2] *BCG_3157c* (blue); *dosR* (yellow); *dosS* (green). Start codons underlined.
[3] ***AAGCCT***: HindIII site in non-coding region used for cloning.

gcgaacgattccggtggcgggcgcggtgctgcgggaggtcttcgtcaacggcattccgcgacgggtcgaccg
ggtcgatttggaaggcctggacgaactggccgacgcaggtccggcgctgctgttgccgctgcgggccagaggt
accgtagcgggtgtcgttgttgtgctgagtcaaggcggtccaggggctttcaccgacgaacaactcgagatgat
ggccgcgttcgccgaccaggccgcgctggcttggcaattggccacttcgcaacgtcggatgcgcgaactcgac
gtactgaccgaccgggatcgtatcgcccgtgacctccatgaccatgtcatccagcggctcttcgcgattggcctg
gctttgcagggtgctgtcccgcacgaacgtaatcctgaagtgcagcaacgactctcggacgtggtagacgatct
gcaagacgttatacaggaaatccggaccaccatttatgacctgcacggagcatcgcagggtatcactcggctc
cggcagcgaatcgatgcggccgtagcccaatttgccgactcggggttgcgcaccagcgttcaattcgtgggtcc
attgtcggtggtcgacagcgcgctcgccgatcaggccgaggcggtggttcgggaagcggtcagcaacgcggtt
cgccatgcgaaggccagcacgttgaccgtccgggtcaaagtcgacgacgacttgtgcatcgaggtgaccgac
aacggccgcgggctgccccgacgagttcaccggaagcggcttaacgaacctgcggcagcgggcagagcag
gccggcggcgaattcaccctcgcgagcgtaccgggcgcgagcggaacagtgctgcgatggtcagcaccgttg
tcgcagtag*GTCGAC*

ΔdosSR::dosSR(ACG):

cggcggatcggagattgatcgacgccgtgccggccgcatcgagatgcgctctcacagcgcggctatgcccaa
atattctgtgggtcaagcgatatgcagccgatggacggccgctggttcggcagctgtcggcaactgtaagccattt
ctgggactttgctgtgaaaagctgggcgatggttgtggacctggacgagccacccgtgcgataggtgagattcat
tctcgccctgacgggttgcgtctgtcatcggtcgataaggactaacggccctcaggtggggaccaacgcccctg
ggagatagcggtccccgccagtaacgtaccgctgaaccgacgggatgtatccgccccagcgaaggagacg
gcg*atg*agcgatcctcggccagctcgggcagtggtcgttggtatcgacgggtcaagggcggcaacgcatgcg
gcgttgtgggcggtcgatgaggcggtgaaccgagacattccgctgcgactggtgtacgtcatcgatccgtccca
actgtccgccgccggcgagggcggtgggcaatcagcggcccgagcggcgctgcacgacgcctctcggaag
gtcgaggccaccgggcaaccggtcaagatcgaaacggaggttctgtgcggcaggccgctcaccaagctgat
gcaggagtccaggtccgcggcgatgctgtgcgtcggttcggtggggcttgatcatgtccgcggtcgccggggttc
ggtcgcggcgaccctggctgggtcggccttatgccccgtggcggtgattcacccgtcgccggccgagccagcg
acaacctcccaggtcagcgcggttgtcgcggaggtggacaatggtgtggtgctgcggcacgcattcgaggagg
ccaggctgcgcggagttccgctgcgggccgtggctgtccacgctgctgaaacacccgatgacgtcgaacagg
gcagccggttggcgcatgtacacctgagccgtcggctcgcccactggacccggctctaccccgaggtgcgggt
ggatcgggccatcgccggcggcagtgcgtgccgtcatctggccgccaacgcaaagccgggtcagctgttcgtc
gcggactcacactccgcgcacgaattgtgcggtgcataccagcccggatgcgccgtacttacggtacgcagtg
ccaacttgtag***AAGCTT***gatcttgggagtggtgccctg*gtg*gtaaaggtcttcttggtcgatgaccacgaggtg
gtgcgtcgtggtctggttgacttgcttgggggccgatcccgagcttgacgtcgtaggtgaggcgggttcggtcgccg
aggcgatggcccagggttcctgccgcgcgcccagatgtcgcggtgctggatgtccggttgcccgatggcaacg
cattgaactgtgccgcgatctgttgtcccgcatgcccgatctgcgctgtctgatcctcacgtcctacacgtctgacg
aggccatgctagatgcgattctcgccggtgccagcggatatgtcgtcaaagacatcaagggaatggagttggc
gcgcgccgtcaaagatgtgggcgctggacggtcgctgctggacaatcgggccgcggccgcgctgatggcca
agctgcgcggtgccgccgagaagcaggacccgctatcaggccttacggaccaggagcggacgctactgggc
ctgcttagcgagggcctgacgaacaagcagatcgccgaccgaatgttcctagccgaaaagacggtgaagaa
ctacgtgtcgcggttgctggccaagctgggcatggaacgtcggacgcaagccgcggtattcgcgacggagttg
aagcgctcgcggccacccggtgatggacc*atga*caacaggggcctcgtcgacgaaaacgacggcgccgc
aatgcgtccactgcgtcacacgctctcccaactacgcctgcacgagctgctggtcgaggtgcaggaccgggtc
gagcagatcgtcgagggccgggaccgcctcgatggtctggtggaggccatgctcgtggtcacagcgggcctg
gacctggaggcaaccctacgcgctatcgtgcattcagcgaccagccttgtcgatgcgcgctatggcgctatgga
ggtgcacgaccggcagcatcgggtattgcactttgtctatgaaggcatcgacgaggagaccgttcggcggatcg
gccacctaccgaaaggcctaggcgtcatcgggctgctcatcgaagatcccaaaccgttacggctggacgatgtt
tctgcgcacccggcctcgattggttttccgccgtatcatccgccgatgcgtaccttcctcgggggtaccggttcgggt
gcgcgatgaatcgttcggcactctgtacctgactgacaagaccaacgggcaaccgttcagcgacgacgacga
ggttctggtccaggcgctggcggccgccgcgggtatcgcagtcgcgaatgcccggctctaccagcaggctaag
gcgcgtcagtcgtggatcgaggccacccgtgacatcgccaccgagttgttgtccggcaccgaacccgcgacg
gtgttccggcttgtcgccgcggaggcgctcaagctgacggcggctgacgctgccctggtagccgttcccgtcga
cgaggacatgcctgccgctgacgtgggggagctgctggtgattgaaacagtcggcagcgctgtggcttccactg
ttgggcgaacgattccggtggcgggcgcggtgctgcgggaggtcttcgtcaacggcattccgcgacgggtcga
ccgggtcgatttggaaggcctggacgaactggccgacgcaggtccggcgctgctgttgccgctgcgggccag
aggtaccgtagcgggtgtcgttgttgtgctgagtcaaggcggtccaggggctttcaccgacgaacaactcgaga
tgatggccgcgttcgccgaccaggccgcgctggcttggcaattggccacttcgcaacgtcggatgcgcgaactc
gacgtactgaccgaccgggatcgtatcgcccgtgacctccatgaccatgtcatccagcggctcttcgcgattggc

ctggctttgcagggtgctgtcccgcacgaacgtaatcctgaagtgcagcaacgactctcggacgtggtagacga
tctgcaagacgttatacaggaaatccggaccaccatttatgacctgcacggagcatcgcagggtatcactcggc
tccggcagcgaatcgatgcggccgtagcccaatttgccgactcggggttgcgcaccagcgttcaattcgtgggtc
cattgtcggtggtcgacagcgcgctcgccgatcaggccgaggcggtggttcgggaagcggtcagcaacgcgg
ttcgccatgcgaaggccagcacgttgaccgtccgggtcaaagtcgacgacgacttgtgcatcgaggtgaccga
caacggccgcgggctgccccgacgagttcaccggaagcggcttaacgaacctgcggcagcgggcagagcag
gccggcggcgaattcaccctcgcgagcgtaccgggcgcgagcggaacagtgctgcgatggtcagcaccgttg
tcgcagtag*GTCGAC*

ΔdosSR::dosSR(ACC):

cggcggatcggagattgatcgacgccgtgccggccgcatcgagatgcgctctcacagcgcggctatgcccaa
atattctgtgggtcaagcgatatgcagccgatggacggccgctggttcggcagctgtcggcaactgtaagccattt
ctgggactttgctgtgaaaagctgggcgatggttgtggacctggacgagccacccgtgcgataggtgagattcat
tctcgccctgacgggttgcgtctgtcatcggtcgataaggactaacggccctcaggtggggaccaacgcccctg
ggagatagcggtccccgccagtaacgtaccgctgaaccgacgggatgtatccgccccagcgaaggagacg
gcg*atg*agcgatcctcggccagctcgggcagtggtcgttggtatcgacgggtcaagggcggcaacgcatgcg
gcgttgtgggcggtcgatgaggcggtgaaccgagacattccgctgcgactggtgtacgtcatcgatccgtccca
actgtccgccgccggcgagggcggtgggcaatcagcggcccgagcggcgctgcacgacgcctctcggaag
gtcgaggccaccgggcaaccggtcaagatcgaaacggaggttctgtgcggcaggccgctcaccaagctgat
gcaggagtccaggtccgcggcgatgctgtgcgtcggttcggtggggcttgatcatgtccgcggtcgccggggttc
ggtcgcggcgaccctggctgggtcggccttatgccccgtggcggtgattcacccgtcgccggccgagccagcg
acaacctcccaggtcagcgcggttgtcgcggaggtggacaatggtgtggtgctgcggcacgcattcgaggagg
ccaggctgcgcggagttccgctgcgggccgtggctgtccacgctgctgaaacacccgatgacgtcgaacagg
gcagccggttggcgcatgtacacctgagccgtcggctcgcccactggacccggctctaccccgaggtgcgggt
ggatcgggccatcgccggcggcagtgcgtgccgtcatctggccgccaacgcaaagccgggtcagctgttcgtc
gcggactcacactccgcgcacgaattgtgcggtgcataccagcccggatgcgccgtacttacggtacgcagtg
ccaacttgtag***AAGCTT***gatcttgggagtggtgccctg*gtg*gtaaaggtcttcttggtcgatgaccacgaggtg
gtgcgtcgtggtctggttgacttgcttggggccgatcccgagcttgacgtcgtaggtgaggcgggttcggtcgccg
aggcgatggcccagggttcctgccgcgcgcccagatgtcgcggtgctggatgtccggttgcccgatggcaacgg
cattgaactgtgccgcgatctgttgtcccgcatgcccgatctgcgctgtctgatcctcacctcctacacctctgacga
ggccatgctagatgcgattctcgccggtgccagcggatatgtcgtcaaagacatcaagggaatggagttggcg
cgcgccgtcaaagatgtgggcgctggacggtcgctgctggacaatcgggccgcggccgcgctgatggccaa
gctgcgcggtgccgccgagaagcaggacccgctatcaggccttaccgaccaggagcggaccctactgggcc
tgcttagcgagggcctgaccaacaagcagatcgccgaccgaatgttcctagccgaaaagaccgtgaagaact
acgtgtcgcggttgctggccaagctgggcatggaacgtcggacccaagccgcggtattcgcgaccgagttgaa
gcgctcgcggccacccggtgatggacc*atga*caacaggggcctcgtcgacgaaaacgacggcgccgcaa
tgcgtccactgcgtcacacgctctcccaactacgcctgcacgagctgctggtcgaggtgcaggaccgggtcga
gcagatcgtcgagggccgggaccgcctcgatggtctggtggaggccatgctcgtggtcacagcgggcctgga
cctggaggcaaccctacgcgctatcgtgcattcagcgaccagccttgtcgatgcgcgctatggcgctatggaggt
gcacgaccggcagcatcgggtattgcactttgtctatgaaggcatcgacgaggagaccgttcggcggatcggc
cacctaccgaaaggcctaggcgtcatcgggctgctcatcgaagatcccaaaccgttacggctggacgatgtttct
gcgcacccggcctcgattggttttccgccgtatcatccgccgatgcgtaccttcctcgggggtaccggttcgggtgc
gcgatgaatcgttcggcactctgtacctgactgacaagaccaacgggcaaccgttcagcgacgacgacgagg
ttctggtccaggcgctggcggccgccgcgggtatcgcagtcgcgaatgcccggctctaccagcaggctaaggc
gcgtcagtcgtggatcgaggccacccgtgacatcgccaccgagttgttgtccggcaccgaacccgcgacggtg
ttccggcttgtcgccgcggaggcgctcaagctgacggcggctgacgctgccctggtagccgttcccgtcgacga
ggacatgcctgccgctgacgtggggggagctgctggtgattgaaacagtcggcagcgctgtggcttccactgttgg
gcgaacgattccggtggcgggcgcggtgctgcgggaggtcttcgtcaacggcattccgcgacgggtcgaccg
ggtcgatttggaaggcctggacgaactggccgacgcaggtccggcgctgctgttgccgctgcgggccagaggt
accgtagcgggtgtcgttgttgtgctgagtcaaggcggtccaggggctttcaccgacgaacaactcgagatgat
ggccgcgttcgccgaccaggccgcgctggcttggcaattggccacttcgcaacgtcggatgcgcgaactcgac
gtactgaccgaccgggatcgtatcgcccgtgacctccatgaccatgtcatccagcggctcttcgcgattggcctg

gctttgcagggtgctgtcccgcacgaacgtaatcctgaagtgcagcaacgactctcggacgtggtagacgatct
gcaagacgttatacaggaaatccggaccaccatttatgacctgcacggagcatcgcagggtatcactcggctc
cggcagcgaatcgatgcggccgtagcccaatttgccgactcggggttgcgcaccagcgttcaattcgtgggtcc
attgtcggtggtcgacagcgcgctcgccgatcaggccgaggcggtggttcgggaagcggtcagcaacgcggtt
cgccatgcgaaggccagcacgttgaccgtccgggtcaaagtcgacgacgacttgtgcatcgaggtgaccgac
aacggccgcgggctgcccgacgagttcaccggaagcggcttaacgaacctgcggcagcgggcagagcag
gccggcggcgaattcaccctcgcgagcgtaccgggcgcgagcggaacagtgctgcgatggtcagcaccgttg
tcgcagtag***GTCGAC***

ΔdosSR::dosSR(ACA):

cggcggatcggagattgatcgacgccgtgccggccgcatcgagatgcgctctcacagcgcggctatgcccaa
atattctgtgggtcaagcgatatgcagccgatggacggccgctggttcggcagctgtcggcaactgtaagccattt
ctgggactttgctgtgaaaagctgggcgatggttgtggacctggacgagccacccgtgcgataggtgagattcat
tctcgccctgacgggttgcgtctgtcatcggtcgataaggactaacggccctcaggtggggaccaacgcccctg
ggagatagcggtccccgccagtaacgtaccgctgaaccgacgggatgtatccgccccagcgaaggagacg
gcg*atg*agcgatcctcggccagctcgggcagtggtcgttggtatcgacgggtcaagggcggcaacgcatgcg
gcgttgtgggcggtcgatgaggcggtgaaccgagacattccgctgcgactggtgtacgtcatcgatccgtccca
actgtccgccgccggcgagggcggtgggcaatcagcggcccgagcggcgctgcacgacgcctctcggaag
gtcgaggccaccgggcaaccggtcaagatcgaaacggaggttctgtgcggcaggccgctcaccaagctgat
gcaggagtccaggtccgcggcgatgctgtgcgtcggttcggtggggcttgatcatgtccgcggtcgccggggttc
ggtcgcggcgaccctggctgggtcggccttatgccccgtggcggtgattcacccgtcgccggccgagccagcg
acaacctcccaggtcagcgcggttgtcgcggaggtggacaatggtgtggtgctgcggcacgcattcgaggagg
ccaggctgcgcggagttccgctgcgggccgtggctgtccacgctgctgaaacacccgatgacgtcgaacagg
gcagccggttggcgcatgtacacctgagccgtcggctcgcccactggacccggctctacccccgaggtgcgggt
ggatcgggccatcgccggcggcagtgcgtgccgtcatctggccgccaacgcaaagccgggtcagctgttcgtc
gcggactcacactccgcgcacgaattgtgcggtgcataccagcccggatgcgccgtacttacggtacgcagtg
ccaacttgtag***AAGCTT***gatcttgggagtggtgccctg*gtg*gtaaaggtcttcttggtcgatgaccacgaggtg
gtgcgtcgtggtctggttgacttgcttgggggccgatcccgagcttgacgtcgtaggtgaggcgggttcggtcgccg
aggcgatggcccaggggttcctgccgcgcgcccagatgtcgcggtgctggatgtccggttgcccgatggcaacg
cattgaactgtgccgcgatctgttgtcccgcatgcccgatctgcgctgtctgatcctcacatcctacacatctgacg
aggccatgctagatgcgattctcgccggtgccagcggatatgtcgtcaaagacatcaagggaatggagttggc
gcgcgccgtcaaagatgtgggcgctggacggtcgctgctggacaatcgggccgcggccgcgctgatggcca
agctgcgcggtgccgccgagaagcaggacccgctatcaggccttacagaccaggagcggacactactgggc
ctgcttagcgagggcctgacaaacaagcagatcgccgaccgaatgttcctagccgaaaagacagtgaagaa
ctacgtgtcgcggttgctggccaagctgggcatggaacgtcggacacaagccgcggtattcgcgacagagttg
aagcgctcgcggccacccggtgatggacc*atga*caacaggggggcctcgtcgacgaaaacgacggcgccgc
aatgcgtccactgcgtcacacgctctcccaactacgcctgcacgagctgctggtcgaggtgcaggaccgggtc
gagcagatcgtcgagggccgggaccgcctcgatggtctggtggaggccatgctcgtggtcacagcgggcctg
gacctggaggcaaccctacgcgctatcgtgcattcagcgaccagccttgtcgatgcgcgctatggcgctatgga
ggtgcacgaccggcagcatcgggtattgcactttgtctatgaaggcatcgacgaggagaccgttcggcggatcg
gccacctaccgaaaggcctaggcgtcatcgggctgctcatcgaagatcccaaaccgttacggctggacgatgtt
tctgcgcacccggcctcgattggtttttccgccgtatcatccgccgatgcgtaccttcctcggggtaccggttcgggt
gcgcgatgaatcgttcggcactctgtacctgactgacaagaccaacgggcaaccgttcagcgacgacgacga
ggttctggtccaggcgctggcggccgccgcgggtatcgcagtcgcgaatgcccggctctaccagcaggctaag
gcgcgtcagtcgtggatcgaggccacccgtgacatcgccaccgagttgttgtccggcaccgaacccgcgacg
gtgttccggcttgtcgccgcggaggcgctcaagctgacggcggctgacgctgccctggtagccgttcccgtcga
cgaggacatgcctgccgctgacgtgggggagctgctggtgattgaaacagtcggcagcgctgtggcttccactg
ttgggcgaacgattccggtggcgggcgcggtgctgcgggaggtcttcgtcaacggcattccgcgacgggtcga
ccgggtcgatttggaaggcctggacgaactggccgacgcaggtccggcgctgctgttgccgctgcgggccag
aggtaccgtagcgggtgtcgttgttgtgctgagtcaaggcggtccaggggctttcaccgacgaacaactcgaga
tgatggccgcgttcgccgaccaggccgcgctggcttggcaattggccacttcgcaacgtcggatgcgcgaactc
gacgtactgaccgaccgggatcgtatcgcccgtgacctccatgaccatgtcatccagcggctcttcgcgattggc

ctggctttgcagggtgctgtcccgcacgaacgtaatcctgaagtgcagcaacgactctcggacgtggtagacga
tctgcaagacgttatacaggaaatccggaccaccatttatgacctgcacggagcatcgcagggtatcactcggc
tccggcagcgaatcgatgcggccgtagcccaatttgccgactcggggttgcgcaccagcgttcaattcgtgggtc
cattgtcggtggtcgacagcgcgctcgccgatcaggccgaggcggtggttcgggaagcggtcagcaacgcgg
ttcgccatgcgaaggccagcacgttgaccgtccgggtcaaagtcgacgacgacttgtgcatcgaggtgaccga
caacggccgcgggctgcccgacgagttcaccggaagcggcttaacgaacctgcggcagcgggcagagcag
gccggcggcgaattcaccctcgcgagcgtaccgggcgcgagcggaacagtgctgcgatggtcagcaccgttg
tcgcagtag*GTCGAC*

ΔdosSR::dosSR(ACT):

cggcggatcggagattgatcgacgccgtgccggccgcatcgagatgcgctctcacagcgcggctatgcccaa
atattctgtgggtcaagcgatatgcagccgatggacggccgctggttcggcagctgtcggcaactgtaagccattt
ctgggactttgctgtgaaaagctgggcgatggttgtggacctggacgagccacccgtgcgataggtgagattcat
tctcgccctgacgggttgcgtctgtcatcggtcgataaggactaacggccctcaggtggggaccaacgcccctg
ggagatagcggtccccgccagtaacgtaccgctgaaccgacgggatgtatccgccccagcgaaggagacg
gcg*atg*agcgatcctcggccagctcgggcagtggtcgttggtatcgacgggtcaagggcggcaacgcatgcg
gcgttgtgggcggtcgatgaggcggtgaaccgagacattccgctgcgactggtgtacgtcatcgatccgtcccа
actgtccgccgccggcgagggcggtgggcaatcagcggcccgagcggcgctgcacgacgcctctcggaag
gtcgaggccaccgggcaaccggtcaagatcgaaacggaggttctgtgcggcaggccgctcaccaagctgat
gcaggagtccaggtccgcggcgatgctgtgcgtcggttcggtggggcttgatcatgtccgcggtcgccgggggttc
ggtcgcggcgaccctggctgggtcggccttatgccccgtggcggtgattcacccgtcgccggccgagccagcg
acaacctcccaggtcagcgcggttgtcgcggaggtggacaatggtgtggtgctgcggcacgcattcgaggagg
ccaggctgcgcggagttccgctgcgggccgtggctgtccacgctgctgaaacacccgatgacgtcgaacagg
gcagccggttggcgcatgtacacctgagccgtcggctcgcccactggacccggctctaccccgaggtgcgggt
ggatcgggccatcgccggcggcagtgcgtgccgtcatctggccgccaacgcaaagccgggtcagctgttcgtc
gcggactcacactccgcgcacgaattgtgcggtgcataccagcccggatgcgccgtacttacggtacgcagtg
ccaacttgtag***AAGCTT***gatcttgggagtggtgccctg*gtg*gtaaaggtcttcttggtcgatgaccacgaggtg
gtgcgtcgtggtctggttgacttgcttggggccgatcccgagcttgacgtcgtaggtgaggcgggttcggtcgccg
aggcgatggcccaggggttcctgccgcgcgcccagatgtcgcggtgctggatgtccggttgcccgatggcaacg
cattgaactgtgccgcgatctgttgtcccgcatgcccgatctgcgctgtctgatcctcacttcctacacttctgacga
ggccatgctagatgcgattctcgccggtgccagcggatatgtcgtcaaagacatcaagggaatggagttggcg
cgcgccgtcaaagatgtgggcgctggacggtcgctgctggacaatcgggccgcggccgcgctgatggccaa
gctgcgcggtgccgccgagaagcaggacccgctatcaggccttactgaccaggagcggactctactgggcct
gcttagcgagggcctgactaacaagcagatcgccgaccgaatgttcctagccgaaaagactgtgaagaacta
cgtgtcgcggttgctggccaagctgggcatggaacgtcggactcaagccgcggtattcgcgactgagttgaagc
gctcgcggccacccggtgatggacc*atga*caacaggggggcctcgtcgacgaaaacgacggcgccgcaatg
cgtccactgcgtcacacgctctcccaactacgcctgcacgagctgctggtcgaggtgcaggaccgggtcgagc
agatcgtcgagggccgggaccgcctcgatggtctggtggaggccatgctcgtggtcacagcgggcctggacct
ggaggcaaccctacgcgctatcgtgcattcagcgaccagccttgtcgatgcgcgctatggcgctatggaggtgc
acgaccggcagcatcgggtattgcactttgtctatgaaggcatcgacgaggagaccgttcggcggatcggcca
cctaccgaaaggcctaggcgtcatcgggctgctcatcgaagatcccaaaccgttacggctggacgatgtttctgc
gcacccggcctcgattggtttccgccgtatcatccgccgatgcgtaccttcctcgggggtaccggttcgggtgcgc
gatgaatcgttcggcactctgtacctgactgacaagaccaacgggcaaccgttcagcgacgacgacgaggttc
tggtccaggcgctggcggccgccgcgggtatcgcagtcgcgaatgcccggctctaccagcaggctaaggcgc
gtcagtcgtggatcgaggccacccgtgacatcgccaccgagttgttgtccggcaccgaacccgcgacggtgttc
cggcttgtcgccgcggaggcgctcaagctgacggcggctgacgctgccctggtagccgttcccgtcgacgagg
acatgcctgccgctgacgtgggggagctgctggtgattgaaacagtcggcagcgctgtggcttccactgttgggc
gaacgattccggtggcgggcgcggtgctgcgggaggtcttcgtcaacggcattccgcgacgggtcgaccgggt
cgatttggaaggcctggacgaactggccgacgcaggtccggcgctgctgttgccgctgcgggccagaggtac
cgtagcgggtgtcgttgttgtgctgagtcaaggcggtccaggggctttcaccgacgaacaactcgagatgatgg
ccgcgttcgccgaccaggccgcgctggcttggcaattggccacttcgcaacgtcggatgcgcgaactcgacgt
actgaccgaccgggatcgtatcgcccgtgacctccatgaccatgtcatccagcggctcttcgcgattggcctggc

tttgcagggtgctgtcccgcacgaacgtaatcctgaagtgcagcaacgactctcggacgtggtagacgatctgc
aagacgttatacaggaaatccggaccaccatttatgacctgcacggagcatcgcagggtatcactcggctccg
gcagcgaatcgatgcggccgtagcccaatttgccgactcggggttgcgcaccagcgttcaattcgtgggtccatt
gtcggtggtcgacagcgcgctcgccgatcaggccgaggcggtggttcgggaagcggtcagcaacgcggttcg
ccatgcgaaggccagcacgttgaccgtccgggtcaaagtcgacgacgacttgtgcatcgaggtgaccgacaa
cggccgcgggctgcccgacgagttcaccggaagcggcttaacgaacctgcggcagcgggcagagcaggcc
ggcggcgaattcaccctcgcgagcgtaccgggcgcgagcggaacagtgctgcgatggtcagcaccgttgtcg
cagtag*GTCGAC*