Untangling topoisomerase inhibition by ParE toxins of Mycobacterium tuberculosis; a biochemical, structural, and phylogenetic approach

IZAAK NATHAN BECK

Untangling topoisomerase inhibition by ParE toxins of *Mycobacterium tuberculosis*; a biochemical, structural, and phylogenetic approach

Izaak N. Beck
BSc (Hons), MScR

Grey College

Thesis submitted for the degree of Doctor of Philosophy in Biological Sciences,
Department of Biosciences, Durham University, 2022

Abstract

Mycobacterium tuberculosis (Mtb), the causative agent of TB, remains a threat to global health with recent efforts towards its eradication being reversed in the wake of the COVID-19 pandemic. Increasing resistance to gyrase targeting second-line fluoroquinolone (FQ) antibiotics indicates the necessity to develop both novel therapeutics and our understanding of Mtb growth during infection. ParDE type toxin-antitoxin systems may lie at this intersection, with ParE toxins targeting gyrase as responsive elements to both host-associated and drug-induced stress during infection. Here we present biochemical and biophysical analyses exploring the ParE1 and ParE2 poisoning of Mtb gyrase, trapping cleavage complexes potentially via a contrasting mechanism to FQ antibiotics. We also propose a novel mechanism of post-translational ParE1 toxin release and system activation via complex remodelling, a potential first in the field. We show the ParE2 toxin to be closely related to the RelE toxin family via phylogenetic analyses and as part of a collaboration with the Genevaux group (Toulouse, France) present the crystal structure of the Mtb RelBE1 complex. We highlight differences in the RelE1 toxin which may contribute to alternate and novel mechanisms of toxicity employed by Mtb RelE toxins, all three of which are upregulated during infection. Altogether, this study combines a range of techniques to better our understanding of three TA systems important to the adaptability of Mtb. We lay the foundations for future work focussing on the molecular basis of Mtb ParE and RelE toxicity through biochemical, biophysical, and structural studies.

Declaration

This thesis is the outcome of my own work and is submitted solely for the degree of Doctor of Philosophy (PhD) in Biological Sciences, Durham University. Any results that arose from collaborative efforts are mentioned in-text.

All research presented herein, unless otherwise stated, were performed in the laboratory of Professor Tim R. Blower, Department of Biosciences, Durham University, during the period August 2018 to December 2022.

Izaak N. Beck

December 2022

Statement of Copyright

Contents

Chapter 2

Materials & Methods

List of Figures

List of Tables

List of Abbreviations

°C	Degrees centigrade
α	Alpha
β	Beta
γ	Gamma
Å	Ångstrom
aa	Amino acid
Alr	Alanine racemase
Ap	Ampicillin
ATP	Adenosine triphosphate
bp	Base pair
BPaLM	Bedaquiline, Pretomanid, Linezolid, Moxifloxacin
BRD	Breakage-reunion domain
BSA	Bovine serum albumin
C. crescentus	*Caulobacter crescentus*
CD	Circular dichroism
C-loop	*Corynebacteriales*-specific loop
Cm	Chloramphenicol
CoA	Coenzyme A
Cryo	Cryogenic
CSS	Complex significance score
CTD	C-terminal domain
cv	Column volume
Ddl	D-alanine:D-alanine ligase
Ddn	Nitroreducatase
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ds	Double-stranded
E. coli	*Escherichia coli*
EcGyr	*E. coli* gyrase
EDTA	Ethylenediaminetetraacetic acid
Ef-Tu	Elongation factor Tu
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium bromide
FQ	Fluoroquinolone
g	Gram(s)
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL
GFX	Gatifloxacin
GSK	GlaxoSmithKline

Gyr	Gyrase
GyrA	Gyrase subunit A
GyrB	Gyrase subunit B
HCl	Hydrogen chloride
hr	Hour
HR	High resolution
hSENP2	Human Sentrin-specific protease 2
HTH	Helix-turn-helix
InhA	Enoyl-acyl carrier protein reductase
IPTG	Isopropyl β – D – thiogalactopyranoside
IR	Inverted repeat
Kav	Partition coefficient
kcal	Kilocalorie
KCl	Potassium chloride
Kd	Dissociation constant
kDa	Kilodalton
Km	Kanamycin
L	Linear
LB	Luria broth
LIC	Ligation independent cloning
LFX	Levofloxacin
MAG	Methyladenine DNA glycosylase
MALDI	Matrix-assisted laser desorption/ionisation
mAU	Milli arbitrary units
MccB17	Microcin B17
MDR-TB	Multidrug resistant tuberculosis
mg	Milligram
MgCl	Magnesium chloride
MGI	*Mycobacterium tuberculosis* gyrase inhibitor
M. jannaschii	*Methanocaldococcus jannaschii*
mM	Millimolar
M. opportunistum	*Mesorhizobium opportunistum*
Mr	Molecular Weight
mRNA	Messenger ribonucleic acid
M. smegmatis	*Mycobacterium smegmatis*
MS	Mass spectrometry
MSA	Multiple sequence alignment
M. tuberculosis / Mtb	*Mycobacterium tuberculosis*
MXF	Moxifloxacin
N	Nicked
NA	Not available
NaCl	Sodium chloride
NBTI	Novel bacterial topoisomerase inhibitor

nM	nanomolar
Ni-NTA	Nickel nitrilotriacetic acid
NTD	N-terminal domain
OD	Optical density
OmpA	Outer membrane protein A
PAGE	Polyacrylamide gel electrophoresis
P. aeruginosa	*Pseudomonas aeruginosa*
PanD	Aspartate decarboxylase
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PDB	Protein Data Bank
P. horikoshii	*Pyrococcus horikoshii*
pI	Isoelectric point
PISA	Proteins Interfaces Surfaces Assemblies
Pre-XDR-TB	Pre-extensively resistant tuberculosis
Qnr	Quinolone resistance protein
R	Relaxed
Rpf	Resuscitation promoting factor
RR-TB	Rifampicin resistant tuberculosis
RHH	Ribbon-helix-helix
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
ROI	Reactive oxygen intermediate
rpm	Revolutions per minute
rpf	Resuscitation promoting factor
Rst	Stokes Radius
S	Supercoiled
S. agalactiae	*Streptococcus agalactiae*
S. aureus	*Staphylococcus aureus*
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SEM	Standard error of mean
sRNA	Small ribonucleic acid
ss	Single-stranded
SUMO	Small Ubiquitin-like Modifier
TA	Toxin-Antitoxin
TAE	Tris Acetic acid EDTA
TAG	Triacylglycerol
TB	Tuberculosis
TBE	Tris Boric acid EDTA
Tc	Tetracycline
TEMED	*N,N,N',N'*-tetramethylethylenediamine

TEV	Tobacco Etch Virus
Tm	Melting temperature
tmRNA	Transfer messenger ribonucleic acid
TOF	Time of flight
TopA	Topoisomerase A
Topo	Topoisomerase
TOPRIM	Topoisomerase-primase
tRNA	Transfer ribonucleic acid
Tyr	Tyrosine
V	Volts
vol/vol	Volume/volume
Vc	Geometric column volume
V. cholerae	*Vibrio chloerae*
Ve	Elution volume
Vo	Void volume
Vol	Volume
WHO	World Health Organisation
wHTH	Winged helix-turn-helix
WT	Wild-type
wt	Weight
UDP	Uridine diphosphate
µL	Microlitre
µm	Micrometre
µM	Micromolar
UNAG	Uridine diphosphate-N-acetylglucosamine
XDR-TB	Extensively drug-resistant tuberculosis

Acknowledgements

I never thought I would do a PhD, in fact I never thought I would continue studying beyond my undergraduate. The last five years of my education have been a steep learning curve and one that I may not have kept up with were it not for the support and guidance of Professor Tim Blower. I couldn't be more thankful for his supervision, mentorship, and friendship across both my MScR and PhD. It has been an incredibly enjoyable journey and I have benefitted immensely from the opportunity to learn from him. There are few things I am genuinely proud of so far in life but alongside being from Yorkshire, being part of his group is one of them. If only I had attended his lectures during my BSc.

I thank the past and present members of the TRB lab for always being there for chats of all kinds, whether it be in the lab as I avoid an experiment, in the office over a coffee and a crossword, or at the pub to drown the sorrows of science. I also thank those outside of the TRB group with whom we share both the lab, office, and department for making working at Durham considerably better than its facilities might suggest. I hope we all keep in touch.

I must thank several PIs at Durham for sharing their passion and wisdom with me and for their mentorship and guidance over this project, notably Prof. Ehmke Pohl, Prof. Nigel Robinson, Dr Peter Chivers and Dr Gary Sharples. I also thank collaborating labs for their support throughout, including the Maxwell Group (JIC), Genevaux Group (Toulouse), and Fineran Group (Otago), alongside my co-supervisor at Newcastle University, Dr Jon Marles-Wright.

A special thanks must also go to the BBSRC DTP for funding this work and supporting me throughout my PhD, including their support of my PIPS placement; the Three Brothers Brewing team were amazing hosts and I believe I made some great friends through the experience.

Beyond academia, I owe a great deal to the Grey College community for making over 8 years at Durham University some very tough years to beat. I have made a number of incredible friends for life through Grey and I extend special thanks to its staff for making the college a second home to me since I started here in 2014. My Mum never worried, supposedly.

Finally, I thank my family and friends. Without you I simply wouldn't have had the confidence to pursue this, nor the stamina to make it through.

Publications

Work completed across this funding period has contributed to three publications to date, although, none of the data published has contributed to this thesis.

- Beck, I.N. *et al.* (2020) 'Antitoxin autoregulation of M. tuberculosis toxin-antitoxin expression through negative cooperativity arising from multiple inverted repeat sequences', *Biochemical Journal*, 477(12), pp. 2401–2419. Available at: https://doi.org/10.1042/BCJ20200368.

- Usher, B. *et al.* (2021) 'Crystal structure of the anti-CRISPR repressor Aca2', *Journal of Structural Biology*, 213(3). Available at: https://doi.org/10.1016/j.jsb.2021.107752.

- Beck I.N., Picton D.M., Blower T.R. (2022) 'Crystal structure of the BREX phage defence protein BrxA', *Current Research in Structural Biology*, pp. 211-219. Available at: https://doi: 10.1016/j.crstbi.2022.06.001.

Chapter 1. Introduction

1.1 Tuberculosis epidemiology and global burden

Despite concerted efforts by the WHO over the past two decades, tuberculosis (TB), a largely curable disease, remains a major cause of morbidity and a leading cause of mortality worldwide (WHO, 2022). As a communicable disease, it is estimated that over a quarter of the world's population would demonstrate an immunological response to *Mycobacterium tuberculosis* (Mtb), the causative agent of TB (WHO, 2022). Until the COVID-19 pandemic, TB was the leading cause of death by a single infectious agent, ranking above HIV since 2014 (WHO, 2014).

The Global TB report 2022 (WHO, 2022) highlights that all major targets of the post-2015 'end TB strategy' are now significantly off-track, and that 2035 is an unlikely date by which deaths, incidence, and catastrophic costs associated with this illness are reduced by over 90 % or eliminated entirely.

Targets for the end of 2022 included the treatment of 40 million persons infection with Mtb; only 66 % of this target was reached by the end of 2021, and only 54 % of target number of cases in children being treated. These figures become even more worrying when considering the treatment of drug-resistant forms of TB, where 43 % of adult cases and only 15 % of child cases targets were being treated by the end of 2021 (WHO, 2022).

With regards to global spending on universal access to TB prevention, diagnosis, treatment, and care, the WHO set a target of $ 30 billion USD (bn) by 2022; only $ 5.4 bn was spent in 2021 (WHO, 2022). Further to this, the $ 2 bn target for research spending by 2022 is also considerably behind schedule with only $ 915 million spent in 2020 (WHO, 2022). Altogether, these metrics assessing the global performance in the fight against TB indicate that the current strategy is not adequate in reaching the 2035 goals.

1.1.1 A disease back on the rise

Unfortunately, the COVID-19 pandemic had a devastating impact on the global situation surrounding TB. The reduction in access to diagnostics and treatment has undeniably contributed to the reversal of positive trends and inroads made from 2005 to 2019 (Zimmer *et al.*, 2022). For example, new diagnoses dropped to the level last seen in 2012 (WHO, 2022), therefore the true extent of TB globally is currently unknown.

It is estimated that there were 10.6 million new cases of TB in 2021 (56.5 % adult men, 32.5 % adult women), an overall increase of 4.5 % from 2019 (WHO, 2022). The incidence rate, that is cases per 100 000, increased by around 3.6 % between 2020 and 2021, reversing almost two decades of 2 % year-on-year decline for a final net reduction of around only 10 % for the 2015 to 2021 period (WHO, 2022). Between 2005 and 2019 there had been a year-on-year reduction in

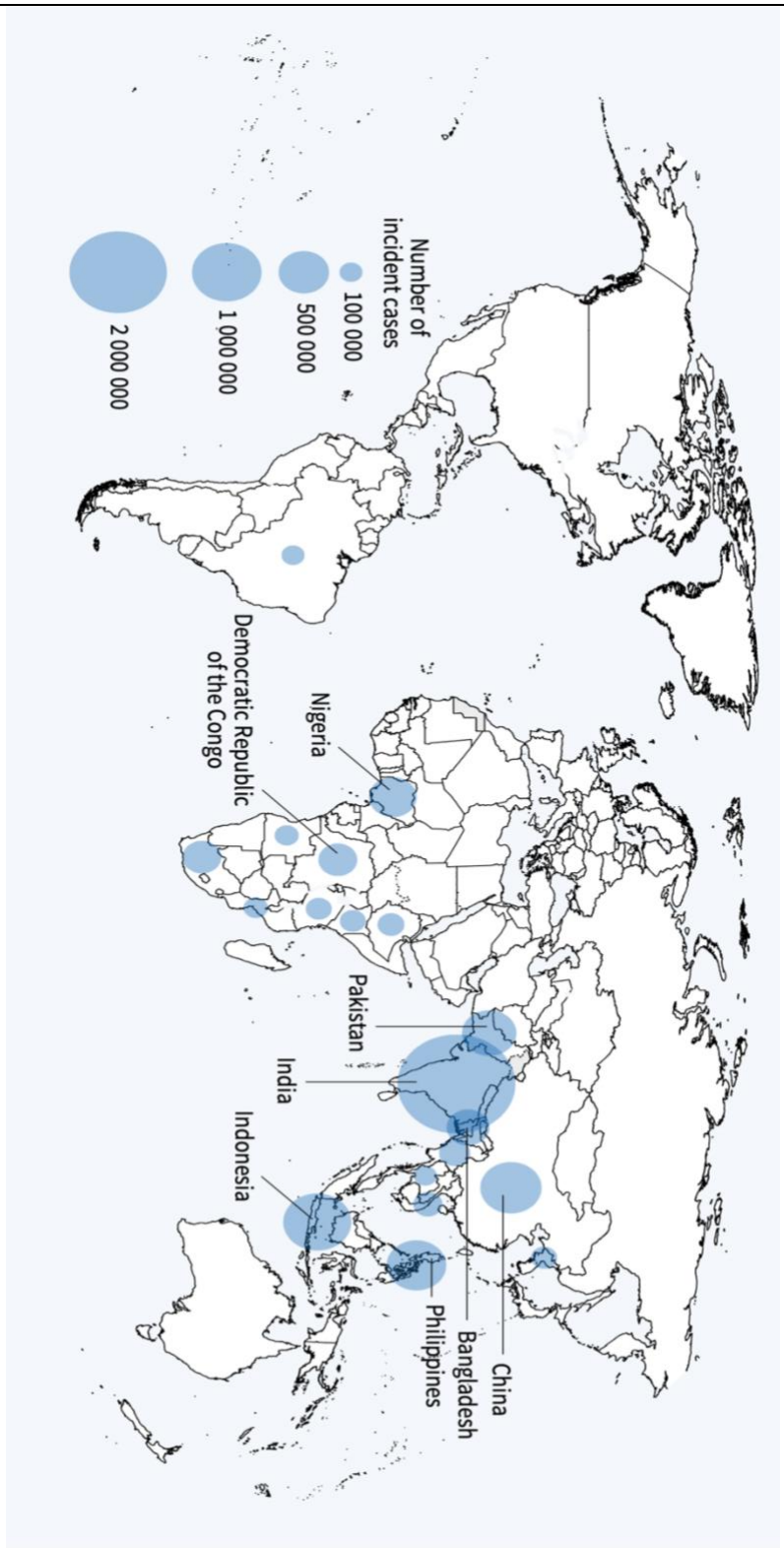
the number of deaths from TB alone from 1.6 million (WHO, 2005) to around 1.4 million (WHO, 2019). In 2021, there was an estimated 1.4 million deaths among HIV-negative people and 187 000 deaths among those with HIV for a combined total of approximately 1.6 million deaths. Unfortunately, this is back to the level last seen in 2017 (WHO, 2017).

The burden of drug-resistant TB is also estimated to have increased with 450 000 new cases of rifampicin-resistant TB (RR-TB) in 2021. Between 2019 and 2020 it is also estimated that the provision of treatment to individuals with RR-TB and multi-drug resistant TB (MDR-TB, TB resistant to rifampicin and isoniazid treatments) fell by around 17 %, although this has made a partial recovery by around 7.5 % in 2021 (WHO, 2022).

Around 5 – 10 % of individuals infected with Mtb go on to develop the disease (Millet *et al.*, 2013) and without treatment, the mortality rate of TB is around 45 % in HIV-negative individuals, and nearly 100% in HIV-positive individuals (WHO, 2021). Treatment of drug-sensitive TB with currently recommended regimes is estimated to cure around 85 % of patients (Torres *et al.*, 2019; Izudi, Tamwesigire and Bajunirwe, 2020). Several countries have had great success in reducing incidence rates to fewer than 10 per 100 000; unsurprisingly these are the more economically developed countries, while incidence rates as high as 500 per 100 000 are observed in countries with high rates of poverty (WHO, 2022). A staggering 87 % of all TB cases can be attributed to only 30 countries, termed 'high-burden' countries (WHO, 2022). Concerningly, only 8 of these countries account for over two thirds of the global total (**Figure 1.1**). The incidence rate, and in turn rate of deaths, can be reduced by action to address TB determinants such as poverty, malnourishment, HIV infection, smoking and diabetes.

The current global situation regarding TB is in a state of desperate need for the development of research breakthroughs, such as novel vaccines to rapidly reduce the global incidence rate worldwide to levels achieved in low-burden countries. Better diagnosis, treatment regimens and medicines are required to more rapidly cure those infected (Fogel, 2015; Tang *et al.*, 2015; Nguyen *et al.*, 2019). Achieving these goals will require better understanding of *M. tuberculosis* as a pathogen and investigation of potentially novel mechanisms to inhibit the Mtb life cycle.

Figure 1.1 Number of incident cases of TB; the 8 highest-burden countries



The 8 highest burden countries are labelled with circular representations of numbers of incident cases. Figure is from the WHO Global TB Report 2022.

1.2 The life cycle of *M. tuberculosis* and progression of TB

M. tuberculosis is a notoriously slow growing highly aerobic bacillus and is primarily a pathogen of the mammalian respiratory system, infecting the lungs (Gordon and Parish, 2018). In some cases, the bacterium can also infect the kidneys, spine, or brain (Maltezou, Spyridis and Kafetzis, 2000; Gordon and Parish, 2018). The infection generally progresses through three phases (ultimately ending in death for many if untreated), starting with exposure whereby Mtb is transmitted from person-to-person through the air in droplets expelled when an infected person coughs or sneezes (Mathema *et al.*, 2017). After exposure comes a prolonged period of latent infection where the individual remains asymptomatic, and indeed some individuals may clear the infection themselves (Emery *et al.*, 2021). Finally, the individual progresses to the disease state whereby they present with signs and symptoms of infection, a phase in which infectious aerosols are generated to facilitate transmission (Smith, 2003).

1.2.1 Exposure to *M. tuberculosis* and early infection

Exposure occurs when a person has been in contact with another person who has active, disease state, TB and has inhaled droplets containing numerous Mtb bacteria (Pfyffer, 2015). The infectious dose, (ID_{50}, number of organisms required to establish infection in 50 % of normal adult humans by a given route), is estimated to be as low as 10 for Mtb making it one of the most infectious bacterial species, and diseases, known (Fauci and Morens, 2012).

At this point, the exposed individual will present with a negative skin test, normal chest X-ray and have no disease symptoms. The innate immune system will begin to attempt to eradicate the infection, and in some cases, this is successful (Emery *et al.*, 2021). Most individuals, however, only have partially effective initial immune responses against Mtb as the bacterium inhabits professional phagocytic cells such as macrophages, neutrophils, and later dendritic cells (Wallgren, 1948; Wolf *et al.*, 2007; Kang *et al.*, 2011). In most infections, recruitment of these phagocytic cells would restrict and even eliminate the pathogen; this early innate immune response contrastingly benefits the survival of Mtb in most cases as additional cellular niches are provided for bacterial population expansion (Davis and Ramakrishnan, 2009; Zhai *et al.*, 2019).

Several bacterial mechanisms and virulence factors assist in bacterial survival within the phagosome (Schnappinger *et al.*, 2003; Ehrt and Schnappinger, 2009; Zhai *et al.*, 2019). Upon phagocytosis of Mtb bacilli, immune responses ultimately produce reactive oxygen intermediates (ROI) alongside a mildly acidic environment (Robinson, 2008); Mtb may resist this through detoxification of ROI via catalase-peroxidase (Rouse *et al.*, 1995) and a Cu/Zn superoxide dismutase (Piddington *et al.*, 2001). Processes such as phagolysosome maturation and killing by acidic degradation can also be inhibited, or at least delayed, through lipoarabinomannan

secretion (Chackerian *et al.*, 2002; Vergne, Chua and Deretic, 2003); this toxin subsequently blocks macrophage cytosolic Ca^{2+} increase, a process required for success of the calmodulin-dependent production of phosphatidylinositol 3 kinase (PI3P) required for fusion (Vergne, Chua and Deretic, 2003; Vergne, Gilleron and Nigou, 2014). Even when lysosome fusion is successful, Mtb can resist acidification through mechanisms of pH homeostasis (Vandal *et al.*, 2008). Several mechanisms are also deployed to prevent host-cell apoptotic death, allowing for prolonged survival and accumulation before release (Briken and Miller, 2008; Blomgran *et al.*, 2012). The mycobacterial gene, *nuoG*, encoding a subunit of the type I NADH dehydrogenase (NDH-1), alongside the SecA2 protein (a superoxide dismutase secretion system protein) and protein kinase E (PknE, expression of which is induced in nitric oxide stress) have been implicated in preventing host cell death by apoptosis (Hinchey *et al.*, 2007; Jayakumar, Jacobs and Narayanan, 2008). Beyond survival within cells, the Mtb type VII secretion system ESX1 can promote necrotic cell death of the infected cell, subsequently recruiting more macrophages and permitting uptake into fresh phagocytes (Garces *et al.*, 2010; Lee *et al.*, 2011; Mittal *et al.*, 2018).

Secondary to, and in part a result of, these Mtb mechanisms initially driving the expansion of the bacterial population within phagocytic cells, there is a significantly delayed adaptive immune response; current estimates suggest this takes about 42 days in humans (Wallgren, 1948). The rate limiting step in this is thought to be in the transport of bacteria from the lungs to lung-draining lymph nodes for antigen presentation and activation of naïve CD4+ T cells (Wolf *et al.*, 2008; Shafiani *et al.*, 2010; Blomgran and Ernst, 2011). Once this stage is complete, proliferation, differentiation, and trafficking of effector CD4+ T cells occurs in a kinetically similar manner to those seen for soluble protein antigens (Wolf *et al.*, 2008).

The onset of adaptive immune responses in humans ultimately results in the arrest of bacterial population expansion and can result in some cases with transient disease symptoms (Ernst, 2012). However, most individuals become asymptomatic and importantly cannot transmit infection; at this point a person is considered to have latent TB as the bacteria are kept in check, walled-off in granulomas (Davis and Ramakrishnan, 2009).

1.2.2 Latent/Persistent TB

In this stage of the infection, as previously mentioned, individuals are asymptomatic, do not test positive by a skin test, and are unable to transmit the infection (Lee, 2016). The precise nature and dynamics of the infection throughout this stage are regularly debated (van Crevel, Ottenhoff and van der Meer, 2002; Ernst, 2012; Ramakrishnan, 2012). Mtb persist in the host through this period, demonstrating slowed growth and metabolism, potentially in an equilibrium with the host immune system (Ernst, 2012), or simply walled-off within developing granulomas consisting of accumulated immune cells (Davis and Ramakrishnan, 2009).

Overall, the progressive growth of Mtb is arrested and coincides with the accumulation of CD4+ and CD8+ T cells in the lungs (Mogues *et al.*, 2001). A multitude of factors contribute to the ineffective clearing of the infection (Vergne, Chua and Deretic, 2003; Wolf *et al.*, 2008; Shafiani *et al.*, 2010), including impaired antigen presentation (Noss *et al.*, 2001), induction of anti-inflammatory mediators (Divangahi *et al.*, 2010), variable regulation of bacterial gene expression (Wayne, 1994; Rachman, Strong, Schaible, *et al.*, 2006; Bold *et al.*, 2011), and resistance to the macrophage-activating effects of interferon (Rachman, Strong, Schaible, *et al.*, 2006; Zhai *et al.*, 2019), to name a few. Mtb continues to be metabolically active during this phase of the infection (Ford *et al.*, 2011), however, the largely hypoxic environment created as the infection is walled-off in the developing granuloma (Remot, Doz and Winter, 2019) and bacilli remain engulfed significantly reducing the rate of reproduction, if not stalling replication altogether (Rachman, Strong, Ulrichs, *et al.*, 2006; Davis and Ramakrishnan, 2009). *In vitro*, hypoxia-induced arrest of Mtb replication renders the bacteria phenotypically tolerant to nearly all TB drugs (Warner and Mizrahi, 2006; Rustad *et al.*, 2009). *In vivo*, it is regularly observed in non-human primate models, Mtb accumulate mutations during this period of metabolic rate reduction, or latency (Ford *et al.*, 2011). This indicates that systems regulating latency may drive mutations to allow sporadic re-entry to the life cycle, or host defences such as macrophage ROI may, over time affect the bacteria; this could, however, be advantageous to Mtb if mutations occur to permit drug tolerance or resistance, or further aid survival in the host.

Like the early stages of infection, Mtb employs several mechanisms to promote latent infection, likely in response to the environmental stresses of granuloma formation such as hypoxia or nitric oxide stress (Remot, Doz and Winter, 2019; Zhai *et al.*, 2019). The dormancy survival regulator protein (DosR) is upregulated in such environments and controls the dormancy regulon (Black *et al.*, 2009), an estimated total of 48 co-regulated genes essential for persistence (Sharma and Tyagi, 2016). For example, constituent genes permit Mtb to use alternate energy sources, notably triacylglycerol (TAG) and isocitric acid through control of dormancy regulon genes such as *tgs1* and *lipY* (TAG synthase and lipase, respectively) and *icl1* (isocitrate lyase) (Daniel *et al.*, 2011). Genes encoding T+ cell stimulating factors can be downregulated to further reduce antigen presentation and further immune system activation (Rachman, Strong, Schaible, *et al.*, 2006).

Secondary to the dormancy regulon, Mtb has a complement of no fewer than 88 toxin-antitoxin (TA) systems (Sala, Bordes and Genevaux, 2014). The balance of expression of many of these systems is predicted to influence the growth rates of Mtb; notably, many of these TA systems have been identified to be regulated throughout the infection of human macrophages (Keren *et al.*, 2011; Torrey *et al.*, 2016), and in response to a number of environmental stresses encountered during infection (Ramage, Connolly and Cox, 2009). These systems are highly varied

in the targets and effects of the toxin components (Van Melderen and De Bast, 2009; Leplae *et al.*, 2011), but there is strong evidence to suggest several systems contribute throughout the infection, notably in the latent stage as they impact whether a bacterium remains bacteriostatic or replicative (Ramage, Connolly and Cox, 2009). The exact roles of individual systems are up for debate, however the link between the number of TA systems within the mycobacteria and apparent pathogenicity (Sala, Bordes and Genevaux, 2014; Slayden, Dawson and Cummings, 2018) indicates a significant role for these systems in TB.

Additionally, Mtb encodes 5 proteins resembling well characterised resuscitation promoting factors (Rpf) (Mukamolova *et al.*, 2002; Chao and Rubin, 2010). These peptidoglycan-hydrolysing enzymes have been shown to be important for normal growth *in vivo* and implicated in the resuscitation of Mtb from the dormant state, permitting progression to disease state *in vitro* (Tufariello *et al.*, 2006; Russell-Goldman *et al.*, 2008). The mechanism of proposed resuscitation, alongside their regulation and direct influence of these genes on human TB, is not well understood and has been questioned several times (Kana *et al.*, 2008; Kana, Mizrahi and Gordhan, 2010). A major discussion point questions the stimulation of these factors during latent infection.

Together, these mechanisms indicate that Mtb has evolved to adopt this state in response to the host immune response, rather than latency being simply the suppressive effect of the immune response. Mtb possesses a highly responsive dormancy regulon and a number of environmentally stimulated TA systems to assist in controlling growth rates to counteract the host immune response for a prolonged period of time. The granuloma itself provides an environment that may keep the bacteria in check through a proposed equilibrium (Ernst, 2012; Remot, Doz and Winter, 2019), alternatively the nature of the hypoxic granuloma environment may simply drive bacterially adapted mechanisms. While this structure is initially protective to the host and limits the proliferation of bacteria, as a last resort due to the difficulties in clearing the infection, when the infection progresses the granuloma may become advantageous to the infection.

1.2.3 TB Disease: activation and transmission

The eventual progression to an active symptomatic state is characterised by shedding of Mtb in respiratory secretions, especially during coughing (facilitating transmission) (Ernst, 2012). This stage can indeed occur decades after the initial infection as the interplay between host immune system and persistent Mtb continues to progress within the granuloma structure (Ramakrishnan, 2012). Essentially, the balance between the host immune system factors controlling the Mtb infection and the bacterial factors assisting in controlling growth rate is destabilised to create a more favourable environmental for bacterial population expansion.

The role of the CD4+ T cell in controlling infection is further exemplified by the link between the qualitative and quantitative defects in this immune cell population with age (Lefebvre and Haynes, 2012), and more aggressively in HIV (Geldmacher *et al.*, 2010), and the transition to active TB. Interestingly, in HIV coinfection, HIV can specifically target and deplete the Mtb antigen-specific population at a higher rate than for other primed CD4+ T cells, resulting in the proliferation of Mtb and giving rise to the increased mortality rates in these cases (Geldmacher et al., 2010). The blocking of tumour necrosis factors (TNF), often unintentionally via therapeutics, leads to lowered macrophage anti-Mtb activity and subsequent macrophage death (Clay, Volkman and Ramakrishnan, 2008) and depletion of a subset of granulysin-containing CD8+ memory T cells that likely contribute to Mtb containment and killing (Nadkarni, Mauri and Ehrenstein, 2007). This can be exacerbated by T-cell exhaustion from antigen overexposure; cells displaying the exhaustion phenotype progress to a total loss in secretion of TNF-α, IL-2, and IFN-γ (Khan *et al.*, 2017), all of which play an important role in the immune response (Huse *et al.*, 2006). Altogether, the deterioration of the immune system provides an opportunity for the infection to develop toward the disease state.

Throughout the infection Mtb has been shown to alter its gene expression of both bacterial growth controlling systems and antigens in response to numerous stimuli (Shi, North and Gennaro, 2004; Rogerson *et al.*, 2006; Ramage, Connolly and Cox, 2009). It has been debated that Mtb may also drive the emergence from its dormant state (Ernst, 2012), however, it is likely that the changes in expression patterns are responsive to the deteriorating immunological response as a more favourable niche develops (Rustad *et al.*, 2009).

Once Mtb becomes reactivated and can dominate the deteriorating immunological response, disease state ensues and the infection progresses toward an obligate step in all infectious diseases; transmission (Ernst, 2012). Deterioration of cells and tissues containing the infection essentially decreases the hypoxic state and provides a more suitable niche for bacterial proliferation, and this is accompanied by symptoms, including coughing to expel bacteria in droplets, aiding transmission (Rodrigo *et al.*, 1997). This stage is particularly variable and the destruction of granulomas is not consistent, however, the most infectious cases arise from macroscopic destruction of tissues in the progression to cavitary TB (Rodrigo *et al.*, 1997). This form of disease permits the highest rate of bacterial proliferation and subsequent transmission as once dormant cells are exposed to larger open spaces and airways. Ultimately, deterioration and subsequent destruction of structures containing infection and the exposure to an oxygen rich niche drive TB infection toward disease. Due to the complex lifecycle of Mtb with prolonged periods of latent infection and slow growth, and the mortality rates in disease, equally complex and typically lengthy treatment regimens are required (Torres *et al.*, 2019).

1.3 Treatment of *M. tuberculosis* infection

In its simplest terms, clinicians have employed a historically binomial approach to the treatment of TB; if the active disease is confirmed microbiologically patients are treated with a 6-month multi drug regimen, while those with the latent infection and no evidence of disease are given shorter, one or two drug regimens. The success of treatment in these cases is generally high at about 87 %. Treatment is greatly complicated, and the success considerably lowered, in cases of drug-resistant TB.

1.3.1 First-line treatment of drug-susceptible TB

The treatment of pulmonary TB in the active disease is a 6-month treatment initially consisting of isoniazid, rifampicin, pyrazinamide, and ethambutol (**Table 1.1**) for 2 months, followed by 4 months of isoniazid and rifampicin, only. The goal of this treatment is to cure the disease and prevent relapse (WHO, 2022). In cases of latent TB, individuals would regularly receive either rifampicin or isoniazid, or both, for a period of 3 – 6 months with the aim of preventing the development to disease (WHO, 2022). From these two regimes, the global dependency on isoniazid and rifampicin is quite clear, however, these are antibiotics to which *M. tuberculosis* has developed wide-spread resistance.

1.3.2 Development of treatments for drug-resistant TB

The fight against drug-resistant TB is a complex and ever evolving one due to the number of medicines used in the first-line treatment and issues surrounding regimen compliance (Gebreweld *et al.*, 2018). It is estimated that the burden of wide-spread drug resistant latent tuberculosis is high in both adults and children and increasing year-on-year; this reduces the chances of effective treatment and promotes the evolution of further resistance in the population (Knight *et al.*, 2019). For these reasons, it is imperative that efficient and accurate diagnostic techniques are employed to correctly identify the resistant strain and tailor treatments accordingly, alongside the development of compliable efficacious regimens and better anti-Mtb drugs (Nguyen *et al.*, 2019).

In 2015, the WHO released rapid communications on the treatment of TB resistant to rifampicin (rifampicin resistant TB, RR-TB), and isoniazid and rifampicin (multidrug resistant TB, MDR-TB) (WHO, 2022b). At this point, it was estimated that around 21 % of previously treated, and 3.9 % of new cases, had RR-TB or MDR-TB (WHO, 2015) and required second-line treatment combinations of 4 effective drugs including pyrazinamide, an injectable agent (such as amikacin or kanamycin), and a later generation fluoroquinolone (FQ) (such as levofloxacin (LFX) or moxifloxacin (MXF)) (**Table 1.1**). This was subsequently updated again in 2016 amidst growing concerns of MDR-TB to a regimen of at least 5 effective drugs. This included pyrazinamide

alongside 4 second-line TB drugs in an appropriate combination; a fluoroquinolone, an injectable, ethionamide or prothionamide, and cycloserine or *para*-aminosalicyclic acid (**Table 1.1**) for 9 – >18 months (WHO, 2016). Even in 2016, treatment was becoming increasingly complicated relying on effective diagnostics of resistance alongside elongated courses of multiple medications in appropriate combinations.

In 2019, guidance on the treatment of MDR-TB was again updated and included 17 different medications (**Table 1.1**) to be used in specific combinations. The typical treatment included: A) LFX or MFX with bedaquiline and linezolid; and B) one or both of clofazimine or cycloserine (or teririzidone) (unless they cannot be used). In cases where medications from A or B could not be used, regimens were supplemented with: C) Ethambutol, delaminid, pyrazinamide, imipenem-cilastatin or meropenem, amikacin or streptomycin, ethionamide or prothionamide, or *para*-aminosalicyclic acid, to create 4- to 5 drug combinations. At this point, there was a high reliance of fluroquinolones as a base of the therapy, alongside various combinations of medications taken orally or injected, and in specific combination depending on resistance and patient sensitivities. Beyond the complex nature of the treatment at this point is the lengthy time-course of 9 months, and the increasing cost of such combinatorial therapies.

1.3.3 Recent guidance on the treatment of drug-resistant TB

Rapid communications disseminated by the WHO in 2022 (WHO, 2022b) outlines the development and spread of drug-resistant TB alongside further guidance on treatment. These developments are likely also a result of the COVID-19 pandemic due to the recent lack of effective diagnosis and treatment. This further emphasises the important of effective microbiological confirmation of infecting strains from patient to patient to select appropriate treatment regimens for this ever-developing, complicated infection. Alongside Mtb strains resistant to isoniazid, rifampicin (RR-TB), and isoniazid and rifampicin (MDR-TB), we now face the challenges of pre-extensively drug-resistant TB (pre-XDR-TB) and extensively drug-resistant TB (XDR-TB); these are strains resistant to rifampicin and any fluoroquinolone, and those resistant to rifampicin, any fluoroquinolone, and at least of bedaquiline or linezolid. The combinations of resistance are driven by the challenges associated with compliance to the previous regimens alongside the efficacy of the treatments (Millet *et al.*, 2013; Gebreweld *et al.*, 2018).

Currently, for MDR/RR-TB a 6-month course of bedquiline, pretomanid, linezolid, and moxifloxacin (BPaLM combination) is advised but only in cases where the individual has had no exposure (> 1 month) to the bedaquiline, pretomanid, and linezolid, or the previous 9-month regimen (WHO, 2022b). Moxifloxacin can apparently be dropped from the regimen in cases of resistance to fluroquinolones (pre-XDR-TB), however the WHO currently strongly recommends

drug-susceptibility testing. Importantly, these treatments are designed to be all-oral, and it is hoped to increase compliance rates, specifically in countries with higher TB-burden.

Globally, in 2021, 2.4 million (71 %) cases of microbiologically confirmed pulmonary TB were tested for resistance; around 5.2 % of these were found to be MDR/RR-TB, while around 1.1 % were found to be pre-XDR/XDR-TB. This was also an increase in total number of resistant cases by 6.4 % from 2020. While these figures themselves indicate the general rise of resistance, the figures become even more concerning when considering recent reports from high-burden countries, especially those contributing to over two thirds of the global cases.

In Pakistan, a study of 562 microbiologically confirmed pulmonary TB samples revealed FQ resistance in 18.5 % (Kabir *et al.*, 2020). In China, specifically Shanghai, the rate of FQ resistance was 8.4 % among all TB samples confirmed microbiologically, 6.2 % in those still susceptible to rifampicin but considerably higher at 42.9 % in samples of MDR/RR-TB (Zhang *et al.,* 2022). Finally, in India, a study revealed that 55 % of strains collected from microbiologically confirmed patients were MDR-TB, of which 69.2 % were pre-XDR-TB and 4.4% of which were XDR-TB (Dreyer *et al.,* 2022). In areas of higher TB burden, the emergence of specifically FQ resistance is concerning given our apparent reliance on these medicines in the treatment of drug resistant TB and overall, indicates a trend toward pre-XDR and XDR-TB in high-burden countries.

1.3.4 Targets of current TB drug treatments

The current multidrug treatment regimens associated with all forms of TB aim to inhibit several essential bacterial pathways and enzymes due to the slow growth and apparent resilience of Mtb, however, can largely be divided into 4 major categories: cell wall synthesis inhibitors, protein synthesis inhibitors, nucleic acid synthesis inhibitors, and those that interfere with cell integrity or division. These, alongside their targets and inhibitory effects, are summarised in **Table 1.1**. Additionally, many of the drugs used throughout are prodrugs, requiring activation by mycobacterial enzymes or pathways for a highly specific mode of action.

1.3.4.1 Mycobacterial cell wall component synthesis inhibitors

The Mycobacterial cell wall is more elaborate than those typically seen in bacterial species. Beyond the inner membrane lies a peptidoglycan layer, a layer composed of the arabinogalactan and lipoarabinomannan, followed by an inner leaflet of mycolic acids and an outer leaflet of external lipids (Alderwick *et al.*, 2015; Vincent *et al.*, 2018). Each of these components has a specific biosynthetic pathway that can be targeted by antibiotics. Due to the mycobacteria specific nature of this structure, most drugs forming TB treatments target these pathways.

The first-line therapeutic trio of isoniazid, pyrazinamide, and ethambutol each inhibit cell wall synthesis with varied mechanisms. Isoniazid, a prodrug, is activated by the mycobacterial

catalase-peroxidase KatG (mutations to which can cause isoniazid resistance) (Suarez *et al.*, 2009). The activated drug inhibits InhA, a 2-trans-enoyl-acyl carrier protein reductase, subsequently preventing fatty acid elongation preventing mycolic acid synthesis (Dessen *et al.*, 1995; Rozwarski *et al.*, 1998). Pyrazinamide is activated by pyrazinamidase at the weakly acidic pH tolerated by Mtb, the product of which binds aspartate decarboxylase, an essential enzyme in coenzyme A production (Leonardi and Jackowski, 2007), targeting it for degradation by caseinolytic protease ClpC1-ClpP (Gopal *et al.*, 2020). Subsequently, fatty acid synthesis is inhibited. Contrastingly, ethambutol inhibits arabinosyltransferase enzymes (EmbA, EmbB, and EmbC) which prevents the formation of arabinogalactan and lipoarabinomannan (Belanger *et al.*, 1996; Goude *et al.*, 2009).

Ethionamide and prothionamide are both prodrugs activated by the monooxygenase EthA (Schaaf *et al.*, 2009), and, like isoniazid, target InhA to prevent fatty acid elongation and mycolic acid synthesis (Brossier *et al.*, 2011). Delaminid, also a prodrug, undergoes biotransformation by the mycobacterial F420 coenzyme system, mutations to several genes in which may cause resistance (Lewis and Sloan, 2015). In a novel mechanism, the radical intermediates formed may have dual function; one is believed to inhibit methoxy-mycolic and keto-mycolic acid synthesis (Lewis and Sloan, 2015) whereas another derivative is thought to contribute to the generation of reactive nitrogen species and contribute to destabilising redox cycling (Xavier and Lakshmanan, 2014). Similarly, the recently approved drug pretomanid may have dual antimicrobial action. Several metabolites are produced after activation by a mycobacterial nitroreductase, some have been shown to be important in reactive nitrogen species (NO) production under anaerobic conditions (Baptista *et al.*, 2018). Other metabolites are thought to block the oxidation of hydroxymycolate to ketomycolate, inhibiting mycolic acid synthesis (Thompson *et al.*, 2017).

A number of drugs also target the synthesis of the peptidoglycan layer; cycloserine and terizidone both act as competitive inhibitors of L-alanine racemase and D-alanylalanine synthetase which are essential to the production and incorporation of D-alanine into the pentapeptide necessary for peptidoglycan formation (Lambert and Neuhaus, 1972; Prosser and De Carvalho, 2013). Terizidone, however, must initially undergo an as yet undefined biotransformation/activation pathway to form cycloserine through hydrolysation (Chirehwa *et al.*, 2020). Imipenem, meropenem, and amoxicillin are also included in specific regimens, each showing different preferences for penicillin-binding proteins (PBPs) responsible for the glycosyltransferase and transpeptidase reactions in cross-linking D-alanine and D-aspartic acid in peptidoglycan layers (Sauvage and Terrak, 2016). Resistance mechanisms to these drugs such as the production of beta lactamase and dihydropeptidase enzymes (Rullas *et al.*, 2015) are inhibited using

clavulanate (Danishuddin and Khan, 2012) and cilastatin (Balfour, Bryson and Brogden, 1996), respectively, for an effective method of resistant TB treatment (Hugonnet *et al.*, 2009).

1.3.4.2 Protein synthesis inhibitors

Fewer protein synthesis inhibiting drugs are used in the treatment of TB. Four, ribosome binding, aminoglycoside antibiotics have been used in a variety of regimens as injectables in treating drug resistant TB: amikacin, capreomycin, kanamycin, and streptomycin. Amikacin, kanamycin, and streptomycin all bind to the 30S subunit and cause misreading of transcripts (Alangaden *et al.*, 1998; Demirci *et al.*, 2013). Capreomycin differs slightly and is believed to bind across the interface of the 30S and 50S subunits involving the 23S rRNA and 16S rRNA (Akbergenov *et al.*, 2011).

Linezolid is distinct in that it binds to the 23S rRNA of the 50S subunit and blocks the interaction with the 30S subunit, inhibiting protein synthesis by preventing the formation of the 70S ribosome and ultimately the initiation complex (Stalker and Jungbluth, 2003).

1.3.4.3 Nucleic acid synthesis inhibitors

The first-line drug rifampicin, historically the most used and currently that for which the highest resistance is seen, inhibits the initiation of RNA synthesis through binding to DNA-dependent RNA polymerase (Campbell *et al.*, 2001). Mutations to the now well-characterised rifampicin-resistance determining region in the *rpoB* gene are currently wide-spread and prevent the binding of rifampicin to the enzyme.

In a contrasting and more nucleic acid precursor focused mechanism, *para*-aminosalicyclic acid (PAS) targets dihydrofolate reductase (DHFR); for activation, PAS is incorporated into the folate pathway by dihydropteroate synthase and dihydrofolate synthase, generating an antimetabolite to inhibit DHFR activity (Zheng *et al.*, 2013). Inhibition of the folate synthesis pathway in turn inhibits the production of deoxythymidine monophosphate (dTMP) from deoxyuridine monophosphate, alongside synthesis of certain amino acids such as methionine (Skipper, Mitchell and Bennett, 1950; Bermingham and Derrick, 2002).

Table 1.1 Drugs used in the treatment of TB

Drug	Prodrug activated by…	Target/effect	Reference
First line			
Isoniazid	KatG	Enoyl-acyl carrier protein reductase (InhA). Inhibits mycolic acid synthesis.	(Timmins *et al.*, 2004)
Rifampicin		DNA-dependent RNA polymerase. Inhibits initiation of RNA synthesis.	(Zaw, Emran and Lin, 2018)
Pyrazinamide	Pyrazinamidase	Aspartate decarboxylase (PanD). Inhibits CoA synthesis.	(Gopal *et al.*, 2020)
Ethambutol		Arabinosyl transferase. Inhibits mycolic acid synthesis.	(Goude *et al.*, 2009)
Fluoroquinolones			
Levofloxacin		DNA gyrase, topoisomerase IV. Inhibits replication.	(Kumar, McHugh and Lipman, 2017)
Moxifloxacin		DNA gyrase, topoisomerase IV. Inhibits replication.	(Kumar, McHugh and Lipman, 2017)
Gatifloxacin		DNA gyrase, topoisomerase IV. Inhibits replication.	(Kumar, McHugh and Lipman, 2017)
Second-line injectables			
Amikacin		30S ribosome subunit. tRNA misreading, inhibits protein synthesis.	(Alangaden *et al.*, 1998)
Capreomycin		70S ribosome interface. Inhibits protein synthesis.	(Akbergenov *et al.*, 2011)
Kanamycin		30S ribosome subunit. tRNA misreading, inhibits protein synthesis.	(Alangaden *et al.*, 1998)
Streptomycin		30S ribosome subunit. tRNA misreading, inhibits protein synthesis.	(Demirci *et al.*, 2013)
Other core second-line agents			
Ethionamide	EthA	InhA. Inhibits mycolic acid synthesis.	(Brossier *et al.*, 2011)
Prothionamide	EthA	InhA. Inhibits mycolic acid synthesis.	(Schaaf *et al.*, 2009)

Table 1.1 Drugs used in the treatment of TB

Drug	Prodrug activated by	Target/effect	Reference
Cycloserine		Alanine racemase (Alr), D-alanine:D-alanine ligase (Ddl). Inhibits peptidoglycan synthesis.	(Prosser and De Carvalho, 2013)
Terizidone	Unknown	Alanine racemase (Alr), D-alanine:D-alanine ligase (Ddl). Inhibits peptidoglycan synthesis.	(Chirehwa *et al.*, 2020)
Linezolid		50S subunit. Prevents 70S ribosome formation, inhibits protein synthesis.	(Stalker and Jungbluth, 2003)
Clofazimine		Bacterial membrane. Inhibits ATP production, impairs redox cycling.	(Cholo *et al.*, 2017)
Add-on agents			
Bedaquiline		Proton pump for ATP synthase. Inhibits ATP production.	(Sarathy, Gruber and Dick, 2019)
Delamanid	F420 coenzyme system	Inhibits mycolic acid synthesis. Generates NO.	(Xavier and Lakshmanan, 2014)
Para-aminosalicyclic acid		Dihydrofolate reductase (DHFR). Inhibits folic acid synthesis.	(Zheng *et al.*, 2013)
Imipenem (Requires cilastatin)		Penicillin-binding proteins (PBPs). Inhibits peptidoglycan synthesis.	(Balfour, Bryson and Brogden, 1996)
Cilastatin		Dihydropeptidase. Prevents beta lactam degradation.	(Balfour, Bryson and Brogden, 1996)
Meropenem (Requires clavulanate)		Penicillin-binding proteins (PBPs). Inhibits peptidoglycan synthesis.	(Hugonnet *et al.*, 2009)
Amoxicillin (Requires clavulanate)		Penicillin-binding proteins (PBPs). Inhibits peptidoglycan synthesis.	(Sauvage and Terrak, 2016)
Clavulanate		Beta lactamase. Prevents beta lactam degradation.	(Hugonnet *et al.*, 2009)
Pretomanid	Nitroreductase (Ddn)	Metabolites inhibit mycolic acid synthesis.	(Thompson *et al.*, 2017)

1.3.4.4 Energy generation inhibitors

While some of the effects of the drugs mentioned above would include perturbation of the integrity of the Mtb cell or ultimately prevent cell division, these are secondary effects and the result of inhibited cell wall component or protein synthesis. Clofazimine and bedaquiline destabilise membrane potential and therefore ATP production (Cholo *et al.*, 2017). Clofazimine may have multiple mechanisms; the drug may associate with the cell membrane and subsequently increase ROI beyond tolerable levels (Yano *et al.*, 2011), alternatively the membrane interaction may generate lysophospholipds and subsequently interfere with potassium ion uptake and reduce ATP production (Cholo *et al.*, 2012). Bedaquiline has been shown to inhibit the proton pump of mycobacterial ATP synthase, resulting in inhibited energy generation and cell death (Sarathy, Gruber and Dick, 2019). This is supported by mechanisms of resistance involving phenotypic variations in the F1/F0-ATP synthase enzyme (Chahine, Karaoui and Mansour, 2014).

1.3.4.5 Inhibition of DNA replication and maintenance of supercoiling

Fluoroquinolones are an outlier and directly inhibit DNA gyrase, the essential enzyme (Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017) responsible for resolving topological issues associated with DNA transcription and replication. For gyrase to function, a double-stranded DNA break is introduced through which another duplex is passed (strand passage) to manipulate and convert specific DNA topologies (Liu and Wang, 1987; Corbett, 2006) (the mechanisms of topoisomerase enzymes, including gyrase, are discussed in detail later in this chapter). Specifically, FQ drugs bind to the gyrase-DNA complex during the topoisomerase reaction close to the gyrase active site tyrosine that is transiently covalently linked to DNA during strand passage (Laponogov *et al.*, 2009; Bax *et al.*, 2010). Resistance has been shown to arise through mutation to GyrA Ser83 and Asp87, residues that support the position of the quinolone structure through hydrogen bonding via a bridge of water and a non-catalytic Mg^{2+} ion (Wohlkonig *et al.*, 2010; Blower *et al.*, 2016). The stabilised positioning of the fluoroquinolone antibiotic is base-stacked with the enzymatically cleaved DNA during the reaction cycle, preventing the re-ligation of the substrate and stalling the reaction in the cleavage state (Drlica *et al.*, 2008; Aldred *et al.*, 2016; Blower *et al.*, 2016).

Not only is the progression of transcription/replication machinery stalled due to overwinding ahead of the moving fork, and daughter chromosomes are not separated for replication (Ashley *et al.*, 2017; Bax *et al.*, 2019), fluoroquinolones stabilise a cleavage intermediate from the topoisomerase reaction; the accumulation of double stranded breaks from this poisoning induces the SOS response for repair and can be cytotoxic if repair is incomplete (Piddock, Walters and Diver, 1990; Drlica *et al.*, 2008).

The importance of the fluoroquinolone class of antibiotics in the treatment of TB is quite clear as a central component of the second-line therapies used to treat DR-TB (WHO, 2022) (**Table 1.1**). The unique nature of their mechanism, inhibiting gyrase to induce genotoxic DNA damage and stall multiple cellular processes(Blower *et al.*, 2016; Pranger *et al.*, 2019), highlights topoisomerases as valuable therapeutic targets in both antimicrobial and anti-cancer research (Pommier, 2013). Emerging resistance to fluoroquinolones (WHO 2022b) underlines the need for further research into this enzyme class and to identify new ways of inhibiting Mtb gyrase.

1.3 Topoisomerases

DNA gyrase, studied in this thesis, is an important drug target in the treatment of TB and belongs to a class of enzymes known as topoisomerases (topos), specifically the subclass type II A. This class of essential enzyme is responsible for the maintenance of DNA topology within the cell and generally, organisms have evolved to encode varying complements of topoisomerases, each resolving different topological issues associated with the processes of transcription and DNA replication (Corbett, 2006). For example, *E. coli* has four different topos, whereas *M. tuberculosis* encodes only two (Nagaraja *et al.*, 2017). As a generalisation, topos are classified by whether they create transient single-stranded (ss) (Type I) (Champoux, 2001), or double-stranded (ds) breaks in DNA (Type II) (Wang, 1998). The creation, and importantly re-ligation, of the ss or ds DNA break is the central feature to all topos as a covalent DNA-topo intermediate is created through an active site tyrosine residue forming a phosphotyrosyl linkage to the phosphate group in the DNA backbone through nucleophilic attack (Corbett, 2006). The existence of this potentially cytotoxic intermediate, alongside the essentiality of specific topo enzymes, has made this enzyme class incredibly attractive in both antimicrobial (Bax *et al.*, 2019) and anti-cancer (Pommier, 2013) drug research.

Topos are essential to all cells as they play a crucial role in the processes of transcription and translation, ultimately contributing to cell replication and division (Liu and Wang, 1987; Racko *et al.*, 2018). The duplex structure of DNA permits both under- and over-winding, subsequently promoting melting or stability, respectively (Schvartzman and Stasiak, 2004). In underwound DNA, negative supercoils are generated, whereas in overwound DNA, positive supercoils are generated; negatively supercoiled DNA is beneficial as it promotes access for proteins, whereas positive supercoiling has a higher stability at increased temperatures (preferred by thermophiles) (Schvartzman and Stasiak, 2004). The existence of DNA in the supercoiled state is important in compaction of the organism's genome, and topos are subsequently responsible for the interconversion of a variety of topologies (Woldringh, Jensen and Westerhoff, 1995).

During transcription and DNA replication, strands of DNA must become separated (Postow *et al.*, 2001). In replication, positive supercoils are generated ahead of the replication fork as DNA becomes overwound, while pre-catenane structures develop behind the replication fork as daughter strands become intertwined (Postow *et al.*, 2001; Corbett, 2006) (**Figure 1.2 i** – adapted from (McKie, Neuman and Maxwell, 2021). Similarly, in transcription, positive supercoils generate ahead of the translocating protein machinery, while increased negative supercoiling occurs behind (Liu and Wang, 1987) (**Figure 1.2 ii**). The overwinding of DNA ahead of the separating strands stalls the replication fork as DNA melting becomes incredibly difficult due to increased stability (López-García, 1999). Failure to separate the daughter strands in DNA replication ultimately leads to permanently interlinked genomes and prevents cell division (Racko *et al.*, 2018), while increased negative supercoiling facilitates the generation of R-loops as RNA can become hybridised with DNA (Drolet *et al.*, 2003). Altogether, failure to resolve these topologies and convert the DNA back to an optimal state of winding and supercoiling leads to inhibition of DNA and protein synthesis, ultimately preventing replication, hence the importance of topoisomerase enzymes. In recent years, there is mounting evidence that topos form part of multi-enzyme complexes responsible for a variety of cellular processes including both transcription and replication, but also potentially in DNA repair (McKie, Neuman and Maxwell, 2021).

The conversion of a specific topology is generally completed within the cell by a single topo due to structural and mechanistic preferences for DNA geometries (Crisona *et al.*, 2000; Corbett *et al.*, 2005; Ashley *et al.*, 2017), however, most topos are functional in several processes. These are summarised in **Figure 1.2 B** and the specific topoisomerase classes and their roles are discussed below. Structural biology has played a central role in understanding the different mechanisms of action of the topoisomerases, elucidating the organisation of holo and apo enzymes as well as reaction intermediates and mechanisms of inhibition (Champoux, 2001; Piton *et al.*, 2010; Karkare, Yousafzai, Lesley A Mitchenall, *et al.*, 2012; Agrawal *et al.*, 2013; Tan *et al.*, 2015; Ashley *et al.*, 2017; Vanden Broeck *et al.*, 2019).

1.3.1 Type I Topoisomerases

This class of topoisomerase creates ss breaks and operates via either strand passage (Champoux, 2001) or controlled rotation mechanisms (Stivers, Harris and Mildvan, 1997; Tan *et al.*, 2015) to resolve specific topologies. The type I enzymes are further subclassified by whether they generate covalent intermediates with the 5'-phopshate (Type IA) or 3'-phosphate (Type IB and IC) of DNA (Champoux, 2001; Corbett, 2006; Tan *et al.*, 2015).

1.3.1.1 Type IA

Within the type IA sub-class of topoisomerases is the canonical type IA enzyme, alongside topo III and reverse gyrase, each of which has a specific role within the cell and in specific organisms, creating a ss 5'-phopshate orientated DNA break in the process. Operating via a strand-passage mechanism, a single strand of duplexed DNA (G-segment) is bound at a DNA-gate, cleaved by the enzymes N-terminal domain (NTD), and the other strand is bound by the C-terminal domain (CTD) and transported (as the T-segment) through the cleaved G-segment (Champoux, 2001).

The type IA enzyme is a monomeric and responsible for the relaxation of negative supercoils *in vivo* in an Mg^{2+}, ATP-independent manner (Wang, 1971), to prevent the accumulation of overly underwound DNA behind the transcription bubble (Massé and Drolet, 1999). IA enzymes have been shown capable of catenation and decatenation of circular DNA *in vitro* in the same manner (Kirkegaard and Wang, 1985; Stivers, Harris and Mildvan, 1997).

Topo III closely resembles topo IA, is highly conserved across both prokaryotes and eukaryotes (Bizard and Hickson, 2020) and functions primarily as a decatenase *in vivo*, resolving interlinked DNA replication intermediates in an Mg^{2+}, ATP independent manner (Nurse *et al.*, 2003). The enzyme has been shown to localise to the *E. coli* replication fork *in vivo* with knockout mutants having significantly decreased efficiency in chromosome segregation (Perez-Cheeks *et al.*, 2012; Lee *et al.*, 2019). Additionally, topo III has been shown to cooperate with RecQ in resolving stalled replication forks (Suski and Marians, 2008) and can perform strand passage on RNA, implying that certain topos may have a role in RNA metabolism (DiGate and Marians, 1992).

Reverse gyrase is an outlier to the type IA subclass in that it utilises ATP alongside Mg^{2+} to perform its strand passage mechanism; the enzyme is unique in that it introduces positive supercoils *in vivo* (Kikuchi and Asai, 1984). Reverse gyrase is only found in thermophilic and hyperthermophillic archae and eubacteria (Mirambeau, Duguet and Forterre, 1984; Forterre, 2002), converting negatively supercoiled and relaxed DNA to the more thermostable, overwound form to enhance genome integrity at increased temperatures (López-García, 1999; Ogawa *et al.*, 2015).

1.3.1.2 Type IB and IC

Both type IB and IC enzymes operate via a controlled rotation mechanism, independent of both Mg^{2+} and ATP, via a 3'-phosphate covalently linked topo-DNA intermediate (Stivers, Harris and Mildvan, 1997; Tan *et al.*, 2015; Soren *et al.*, 2020). Mechanistically, a G-segment is bound and cleaved, and rotation occurs around the single intact strand with the speed of rotation proposed to be controlled by friction in the enzyme cavity before re-ligation (Tan *et al.*, 2015).

Type IB topos have been found in eukaryotes and viruses (Champoux and Dulbecco, 1972; Stivers, Harris and Mildvan, 1997) and are believed to relieve torsional stress imposed on the duplex

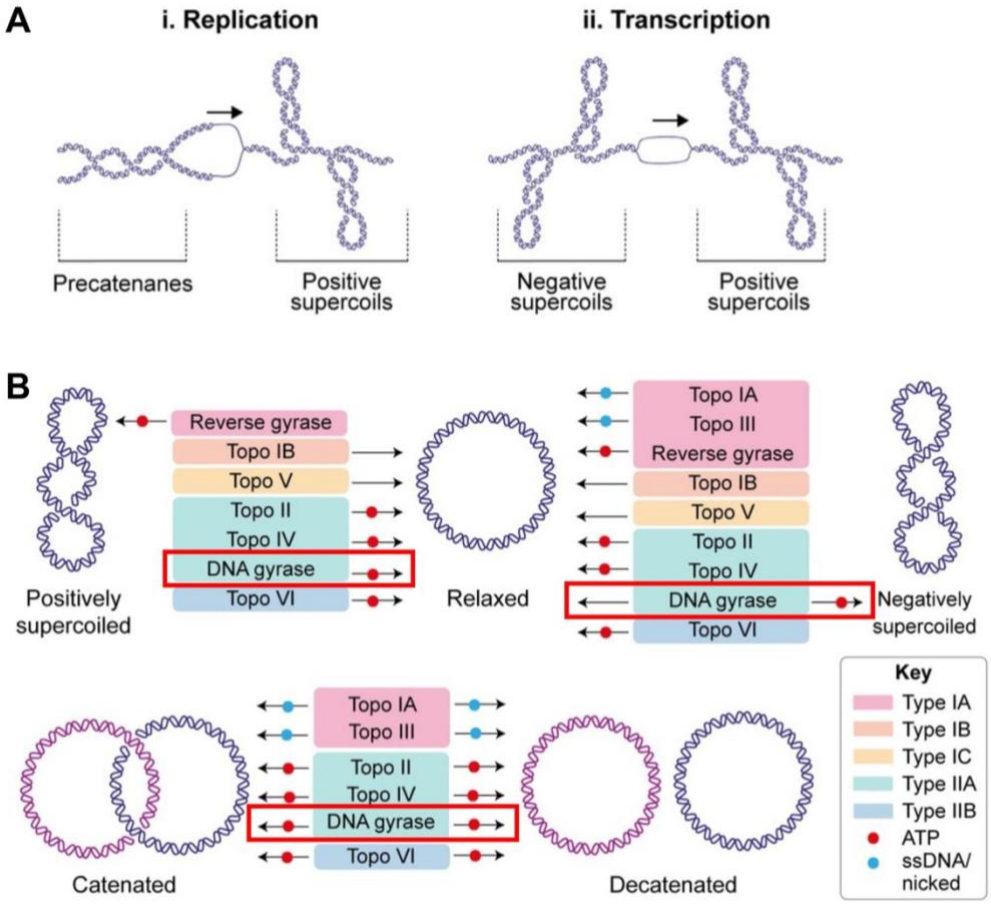
during transcription (Soren *et al.*, 2020). Physical interactions between RNA pol II and topo IB, alongside topoisomerase activation, *in vitro* (Baranello *et al.*, 2016), further indicate a strong association with transcription processes. Interestingly, this class is also associated with transcription-linked mutations causing a number of human diseases such as spinocerebellar ataxia and is a useful target enzyme in the treatment of cancer, suppressing tumorigenesis (Li and Liu, 2016).

Type IC enzymes, also known as topo V, are found in both prokaryotes and eukaryotes and are involved in the relaxation of both positive and negative DNA supercoils *in vivo* (Slesarev *et al.*, 1993) and has been shown to exhibit repair activity of abasic DNA-damage *in vitro*; a function independent from its topoisomerase activity (Rajan, Osterman and Mondragón, 2016). Operating via the same mechanism as type IB, the enzyme class is distinct through an atypical active site and unique protein folds (Taneja *et al.*, 2006).

1.3.2 Type II Topoisomerases

As previously stated, this enzyme class is defined by their mechanism relying on the generation of a ds break, creating a four-base overhang (Type IIA) (Cabral *et al.*, 1997; Blower *et al.*, 2016) or a two-base overhang (Type IIB) (Nichols *et al.*, 1999). Typically, type II topos create covalent, 5'-phopshate orientated, intermediates in an ATP/Mg^{2+} dependent manner and function via strand passage (Kampranis, Bates and Maxwell, 1999; Corbett and Berger, 2004; Bax *et al.*, 2019). Mechanistically, a duplex is bound (G-segment) at the DNA-gate and another duplex is captured (T-segment) by the ATP-operated clamp (N-gate). The transient break in the G-segment permits transport of the T-segment before re-ligation, release, and re-set of the cycle as the ATPase N-gate reopens (Kampranis, Bates and Maxwell, 1999) (**Figure 1.3**). The re-ligation of the ds break is imperative to cell-survival (and is indeed highly efficient), not only as the enzymes prevent challenges associated with specific topologies, but also in preventing extensive genotoxic damage; this feature has been exploited by numerous antimicrobial and anti-cancer drugs (Pommier *et al.*, 2010; Pommier, 2013; Bax *et al.*, 2019). While the type IIA and type IIB enzymes have a conserved strand passage mechanism (albeit via different dsDNA cleavage patterns), they are structurally distinct, and each enzyme subclass exhibits preference for a specific process (Weidlich and Klostermeier, 2020).

Figure 1.2 DNA Topology and topoisomerase activity summary



A) Cartoon representations of topological issues in DNA transcription (i) and translation (ii); **B**) Cartoon summary of role of topoisomerase enzymes. The capabilities of DNA gyrase, the focus of this study, are highlighted using red boxes. Adapted from McKie *et al.* (2021).

1.3.2.1 Type IIA

Within this subclass of enzyme is the prokaryotic DNA gyrase and topo IV, and eukaryotic topo II. DNA gyrase is heterotetrameric and formed of two GyrA and two GyrB subunits (Gellert *et al.*, 1976; Corbett and Berger, 2004), similarly topo IV is formed of equivalent subunits, two ParC and two ParE, respectively (Kato, Suzuki and Ikeda, 1992; Corbett *et al.*, 2005) (**Figure 1.3 A**). Topo II is dimeric and is formed of two Top2 subunits; a full-length fusion of the equivalent A/B or C/E subunits of gyrase and topo IV, however, lacking the A/C C-terminal domain (Goto and Wang, 1982) (**Figure 1.3 A**). Each of these share key domains present in pairs within the holoenzyme; a winged-helix domain (WHD), a topoisomerase-primase domain (TOPRIM), and a Gyrase, Hsp90, Histidine Kinase, MutL domain (GHKL), alongside motifs specific to each enzyme (**Figure 1.3 A**) (Weidlich and Klostermeier, 2020; McKie, Neuman and Maxwell, 2021).

The A/C subunits consist of an N-terminal breakage reunion domain (BRD), formed of the WHD, the tower, and the coiled-coil domains, followed by the C-terminal domain (Corbett and Berger, 2004; Petrella *et al.*, 2019) (**Figure 1.3 A**). The WHD is a common helix-turn-helix (HTH) type DNA-binding protein fold which houses the catalytic, phosphotyrosyl bond-forming, tyrosine residue central to topoisomerases alongside an isoleucine which intercalates with G-segment DNA to promote DNA bending and subsequent cleavage (Dong and Berger, 2007). The B/E subunits consist of an ATPase domain formed of the GHKL and a transducer, followed by the TOPRIM (Vanden Broeck *et al.*, 2019) (**Figure 1.3 A**). The TOPRIM domain chelates the essential Mg^{2+} ions via a DxD motif (Aravind, Leipe and Koonin, 1998) and contains a glutamate residue that serves as the general acid in the cleavage step by donating a proton to the sugar hydroxyl (Sissi and Palumbo, 2009). The same glutamate subsequently serves as a general base and accepts the proton from the 3'-hydroxyl during re-ligation (Sissi and Palumbo, 2009). Additionally, this domain contains conserved motifs which interact with the G-segment and assist in DNA-binding (Dong and Berger, 2007). The GHKL is part of the ATPase domain (Corbett and Berger, 2004), albeit, the exact role of ATP-binding is unclear and regularly debated; some reports suggest ATP-hydrolysis may be important for cleavage/re-ligation with the free-energy of ATP enabling the formation of stable protein-protein interfaces in the holoenzymes to efficiently processing DNA, preventing the persistence of genotoxic dsDNA breaks formed at the DNA-gate (Bates, Berger and Maxwell, 2011; Hobson and Berger, 2019).

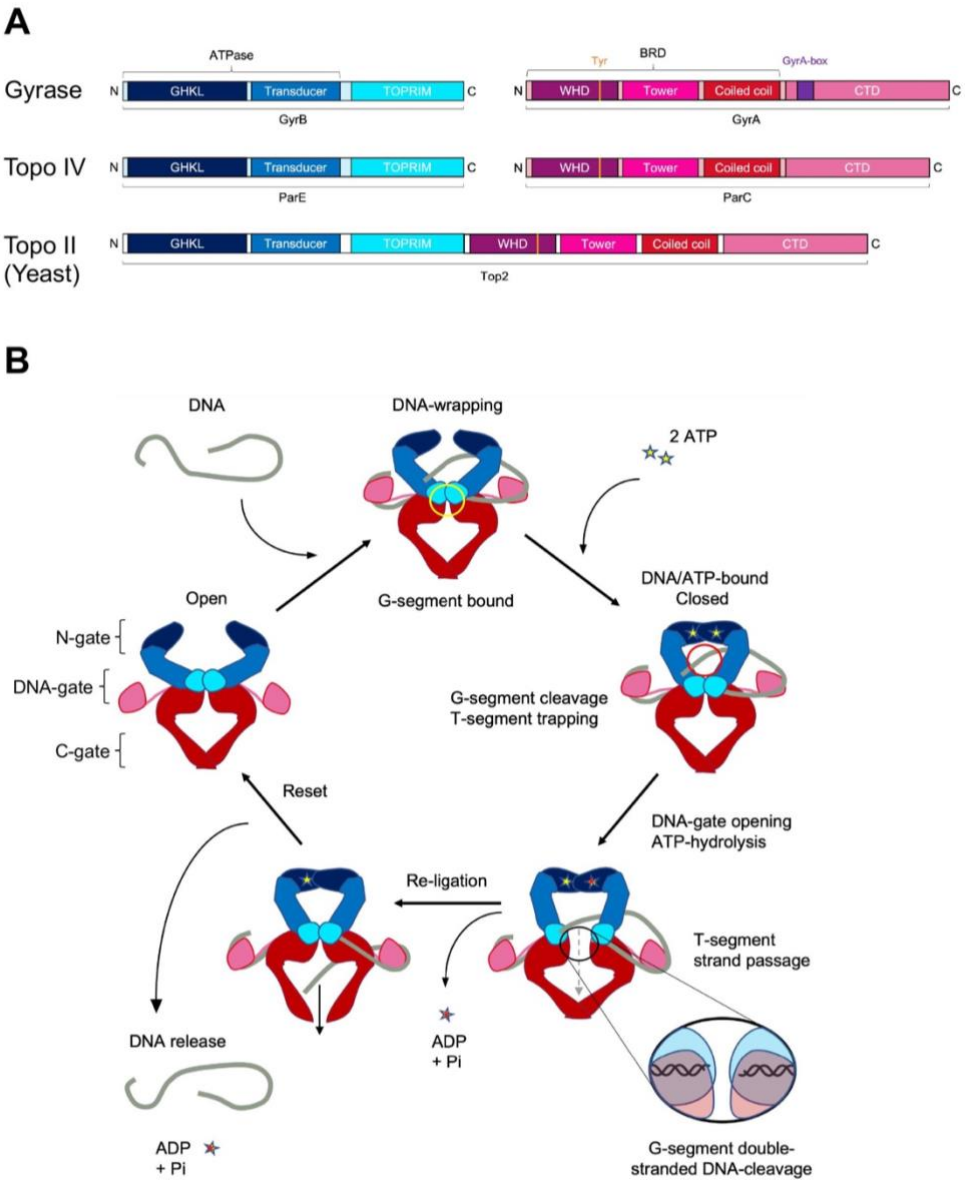
DNA gyrase, formed by two GyrA subunits and two GyrB subunits, is predominantly found in bacteria and is unique in that it is the only topoisomerase capable of converting DNA into the preferred, negatively supercoiled state, its dominant function *in vivo* (Gellert *et al.*, 1976). Gyrase can also catenate, decatenate, and relax positive supercoils in an ATP/Mg^{2+} dependent manner and relax negative supercoils in an Mg^{2+} dependent ATP-independent manner, *in vitro* (Liu and

Wang, 1987; Kampranis, Bates and Maxwell, 1999; Karkare, Yousafzai, Lesley A Mitchenall, *et al.*, 2012). Due to the preference of gyrase for the removal of positive supercoils, the enzyme is regularly seen to be highly active downstream of highly transcribed operons (Hiasa and Marians, 1996) and is considered to be essential in the elongation phase of replication and transcription, acting ahead of the translocation machinery (Kreuzer and Cozzarelli, 1979).

Mechanistically, the unique supercoiling activity arises from the enzymes ability to wrap DNA via the conserved CTD of the GyrA subunit (Corbett, Shultzaberger and Berger, 2004; Hobson, Bryant and Berger, 2020) (**Figure 1.3 B** and **Figure 1.4**). The GyrA CTD is a beta pinwheel structure formed of 6 beta strands with a largely basic outer surface promoting DNA binding (Hobson, Bryant and Berger, 2020), alongside a sequence termed the 'GyrA-box'; a conserved 7 amino acid motif essential for correct DNA substrate wrapping and coupling of ATP binding and supercoiling (Lanz and Klostermeier, 2012). Gyrase enzymes have apparent species-specific variations alongside the highly conserved features which contribute to organism specific functions; the *E. coli* gyrase enzyme contains a TOPRIM insertion essential for activity (Dar *et al.*, 2009) which is believed to contribute to configuration of the DNA-gate for substrate binding, coupling processes between the different functional domains (Schoeffler, May and Berger, 2010). This insertion is not seen in Mtb gyrase which instead has a DEEE-loop insertion in its tower domain and a *Corynbacteriales* specific insertion (C-loop) in its GHKL sequence; these interact to effectively couple ATP hydrolysis and DNA duplex transport (Agarwal and Tyagi, 2006).

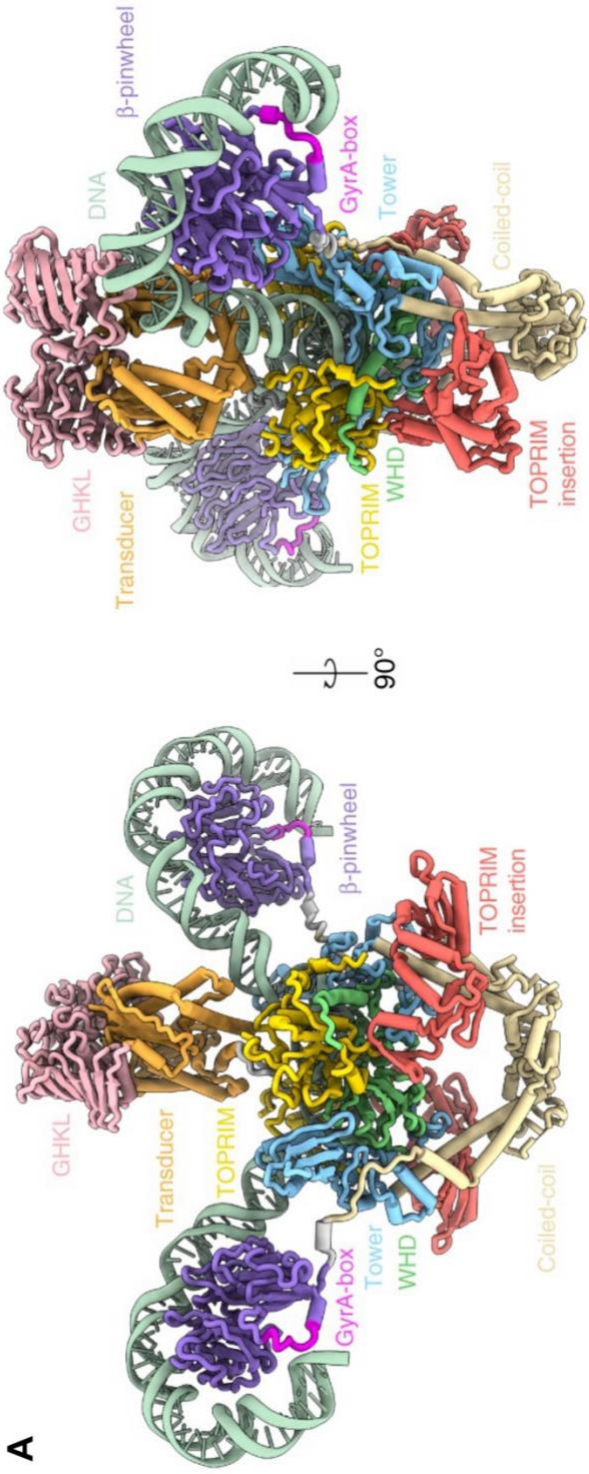
The reaction cycle (**Figure 1.3 B**) is thought to be initiated with G-segment binding at the DNA-gate with the substrate wrapping around a GyrA-CTD. Wrapping of DNA around the second GyrA-CTD correctly positions the T-segment over the G-segment in a positive supercoil for trapping by dimerisation of the GyrB GHKL ATPase domain (Kampranis, Bates and Maxwell, 1999). Approximately 130 bp of substrate DNA is wrapped by the GyrA-CTD and creates the G-segment (Orphanides and Maxwell, 1994; Vanden Broeck *et al.*, 2019) (**Figure 1.4**). Transport of the T-segment through the cleaved G-segment is likely coupled with ATP-hydrolysis (Corbett and Berger, 2004; Petrella *et al.*, 2019) and, after re-ligation, the product is released through the GyrA coiled-coil C-gate as negatively supercoiled product (**Figure 1.3 B**). The full-length, gepotidacin-inhibited, *E. coli* gyrase nucleoprotein complex structure was recently solved by Cryo-EM and provided insights into gyrase protein domain organisation and DNA binding (Vanden Broeck *et al.*, 2019) (**Figure 1.4**). Essentially, this structure highlighted the wrapping of substrate DNA and the dimerisation of the GyrB ATPase, alongside the positioning of the TOPRIM insertion (specific, but essential, to *E. coli*) (Schoeffler, May and Berger, 2010; Vanden Broeck *et al.*, 2019); this study lays the foundations for future gyrase structure studies in complex with a range of different inhibitors.

Figure 1.3 Type IIA topoisomerase domain organisation and mechanism



A) Domain maps of typical type II topoisomerases from bacteria and yeast. N- and C- termini are labelled. GHKL – Gyrase, Hsp90, Histidine Kinase, MutL; TOPRIM – topoisomerase-primase; BRD – breakage-reunion domain; WHD – winged-helix domain; CTD – C-terminal domain. Catalytic tyrosine residue is highlighted in orange, and the GyrA box highlighted in purple; **B**) Proposed catalytic cycle of type II topoisomerase enzymes (colouring as per A, with the breakage reunion domain (WHD-Tower-Coiled coil) coloured red). DNA is represented in green and ATP as star icons. For clarity, the DNA G-segment is highlighted at the "G-segment bound" step with a yellow circle, and the DNA T-segment is highlight at the "T-segment trapping" step with a red circle.

Figure 1.4 Cryo-EM structure of *E. coli* gyrase



A) Cryo-EM structure of the *E. coli* gyrase nucleoprotein complex inhibited by gepotidacin, adapted from vanden Broeck *et al*. (2019). Domains and motifs are labelled with appropriate colouring.

Topo IV is structurally similar to gyrase as the two share a similar domain organisation (**Figure 1.3 A**) (McKie, Neuman and Maxwell, 2021), however, unlike gyrase, topo IV has no GyrA-box like motif in its GyrA-equivalent subunit and thus is devoid of supercoiling capabilities (Kato, Suzuki and Ikeda, 1992; Lanz and Klostermeier, 2012). The enzyme is capable of relaxing positive and negative supercoils and can catenate and decatenate DNA in an ATP/Mg^{2+} dependent manner (Crisona *et al.*, 2000; Corbett *et al.*, 2005; Pitts *et al.*, 2011). *In vivo*, topo IV has been observed to be enriched at the *E. coli* dif site where chromosomes are unlinked (Levine, Hiasa and Marians, 1998) and has been shown to interact with XerCD recombinases and MukBEF at the origin of replication to enhance its activity, regulated by MatP (Li *et al.*, 2010; el Sayyed *et al.*, 2016). Topo IV has a preference for decatenation through its structure (containing no GyrA-box), binding preference to catenane geometries (Corbett *et al.*, 2005; Neuman *et al.*, 2009), and apparent protein-protein interactions associating the enzyme with sites requiring decatenation. The inhibition of topo IV and gyrase would clearly be detrimental to several cellular processes, ultimately preventing replication, and if dsDNA breaks are stabilised they can be cytotoxic (such as when FQ antibiotics bind, previously discussed) (Aldred *et al.*, 2013).

Yeast cells encode topo II (**Figure 1.3 A**), a homodimer which acts via a similar ATP/Mg^{2+}, 5′-phosphate orientated dsDNA break mechanism to gyrase and topo IV; the enzyme is responsible for relieving torsional strain ahead of the DNA replication for having exhibited strong preference for the relaxation of positive supercoils (Goto and Wang, 1982). Topo II can, however, also relax negative supercoils, decatenate and unknot DNA (Goto and Wang, 1982; Holm *et al.*, 1985; Baxter and Diffley, 2008). In vertebrates topo II exists as two isoforms, topo IIα which is essential for cell viability with roles throughout mitosis, regulation of gene expression (Grue *et al.*, 1998; Lee and Berger, 2019), and topo IIβ which has been implicated in neuronal development and transcription having been associated with sensory, motor, and brain defects (Linka *et al.*, 2007; Li and Liu, 2016; Lam, Yeung and Law, 2017).

1.3.2.2 Type IIB

Within the subclass of type IIB enzymes is currently topo VI (Sugimoto-Shirasu *et al.*, 2002) and, the relatively novel and emerging class, type VIII (Gadelle *et al.*, 2014). Both operate strand passage through a dsDNA break, however with a 2-base overhang (Corbett and Berger, 2003; Gadelle *et al.*, 2014).

Topo VI has been found in both prokaryotes and eukaryotes and forms a heterotetramer structure distinct from the IIA enzymes as only two interfaces form: that of the N-gate and the DNA-gate (lacking the C-gate). At these interfaces, topo VI has minimal sequence homology to the type IIA WHD, TOPRIM, and GHKL domains (Corbett and Berger, 2003). Mechanistically, the enzyme is thought to operate in a clamp-like structure passing a T-segment through a cleaved G-segment, coupled to ATPase activity (Graille *et al.*, 2008; Wendorff and Berger, 2018); the transducer domain alternates between restrained and relaxed states as mediated by the association of residues with DNA. The movement of this domain is thought to be coupled to DNA-gate opening and subsequently T-segment transport; due to the structurally conserved motifs between the topo VI transducer and the GyrB transducer, these insights are likely also applicable to type IIA enzymes (Corbett and Berger, 2003).

Relatively little is known about topo VIII, and it requires further study for proper characterisation; currently enzymes of this class have been characterised as homodimers, demonstrating relaxation capabilities on both positive and negative supercoils in an Mg^{2+} dependent manner (Gadelle *et al.*, 2014). Interestingly, one of these enzymes was capable of this function independent of ATP (Gadelle *et al.*, 2014). Due to the variable states of inactivation seen in the currently identified topo VIII enzymes, the extent of the *in vivo* roles of this sub-class remains unclear (T. S. Takahashi et al., 2020).

1.3.3 *M. tuberculosis* topoisomerases

As highlighted so far, topoisomerases are structurally and mechanistically diverse and within the cell have preference for a specific topological conversion (Crisona *et al.*, 2000; Stone *et al.*, 2003; Corbett and Berger, 2004; Ashley *et al.*, 2017; Weidlich and Klostermeier, 2020). That being said, the majority of the topoisomerases are capable of multiple functions (**Figure 1.3 B**), and these may either be compensatory within the cell, or artefactual as their roles have been determined by sequence and structural preferences for certain DNA geometries (Weidlich and Klostermeier, 2020; McKie, Neuman and Maxwell, 2021).

The *E. coli* genome encodes 4 topoisomerase enzymes, two type I and two type II; TopIA, TopoIII, gyrase (EcGyr), and Topo IV (**Table 1.3**). The roles of each of these are generally as determined by the enzyme preferences and potentially allows for compensation if one topoisomerase

becomes non-functional or is inhibited (Schmid and Sawitzke, 1993). In contrast, *M. tuberculosis* H37Rv has a decreased complement of topoisomerase enzymes, encoding only a single type I and a single type II; TopA (Rv3646c) and gyrase (GyrA: Rv0006; GyrB: Rv0005), respectively (**Table 1.3**) (Cole *et al.*, 1998). Possessing more than one type of each topo allows for a division of labour, and a certain degree of overlap in functions creates a level of redundancy, permitting inhibition or loss of function to an enzyme (Nagaraja *et al.*, 2017). *M. tuberculosis*, however, increases the burden put on TopA and gyrase, with both being essential to growth (Sassetti, Boyd and Rubin, 2003; Griffin *et al.*, 2011; Dejesus *et al.*, 2017) and adapted to expand their intracellular repertoire of functions.

1.3.3.1 *M. tuberculosis* TopA

Mtb TopA, the sole type I enzyme and a validated drug target in the treatment of TB (Nagaraja *et al.*, 2017), has novel protein folds making it distinct from other IA type topos, suggesting roles beyond just relaxation (Tan *et al.*, 2016). Conditional silencing of TopA resulted in reduced levels of genome decompaction alongside an irregular, bulbous cell morphology; together, these are believed to be due to the role of TopA in relaxing negative supercoils and assisting in decatenation (Ahmed *et al.*, 2014). Additionally, TopA has been reported to physically interact with 3-methyladenine DNA glycosylase (MAG), with TopA stimulating MAG activity on damaged DNA substrates, indicating a role in DNA repair mechanisms (Yang *et al.*, 2012).

Table 1.2 Topoisomerases of *M. tuberculosis* and *E. coli*

Enzyme	Type	Gene(s)	Function*	Reference
M. tuberculosis				
Gyrase	IIA	*gyrA, gyrB*	Introduces negative supercoils, removes positive (and negative) supercoils, (decatenates interlinked DNA)	(Corbett and Berger, 2004; Karkare et al., 2012)
Topoisomerase I (TopA)	IA	*topA*	Removes negative supercoils (decatenates interlinked nicked or gapped DNA)	(Aubry *et al.*, 2006)
E. coli				
Topoisomerase I	IA	*topA*	Removes negative supercoils, (decatenates interlinked nicked or gapped DNA)	(Corbett and Berger, 2004; Tan *et al.*, 2015)
Gyrase	IIA	*gyrA, gyrB*	Introduces negative supercoils, removes positive (and negative) supercoils, (decatenates interlinked DNA)	(Zechiedrich, Khodursky and Cozzarelli, 1997; Corbett and Berger, 2004)
Topoisomerase III	IA	*topB*	Chromosome segregation (relaxes negative supercoils)	(Perez-Cheeks *et al.*, 2012)
Topoisomerase IV	IIA	*parC, parE*	Decatenates interlinked DNA, removed positive and negative supercoils	(Kato, Suzuki and Ikeda, 1992; Stone *et al.*, 2003)

*Functions in parenthesis have been demonstrated *in vitro* using purified enzymes but have not been shown *in vivo*.

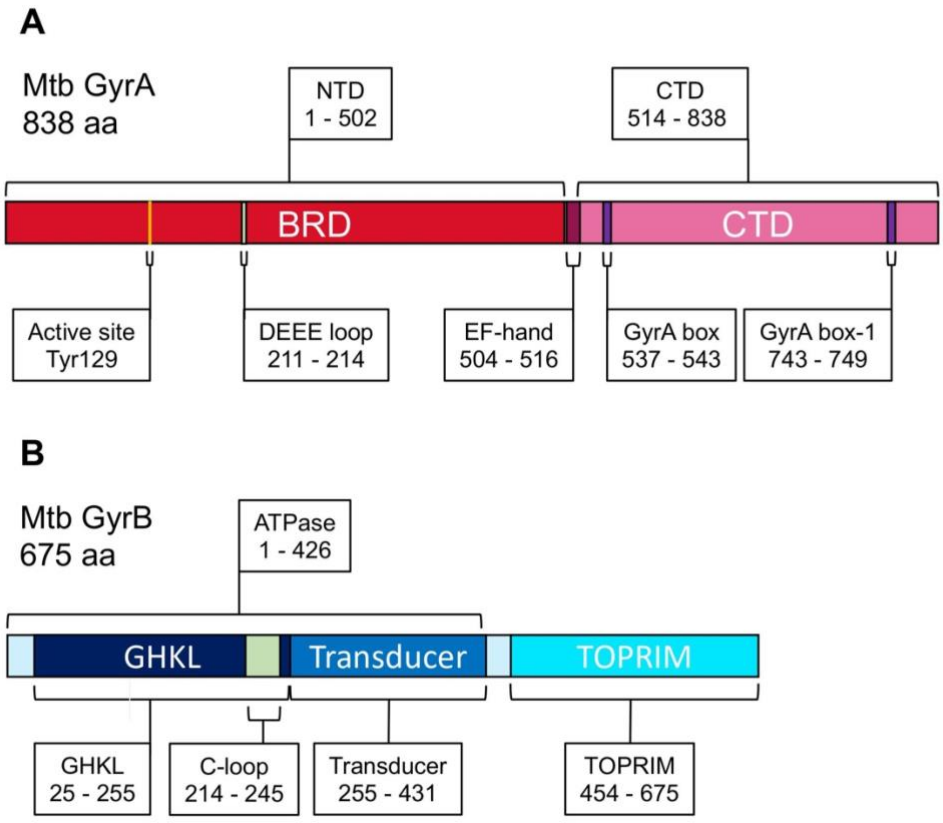
1.3.3.2 *M. tuberculosis* gyrase

The heterotetrameric $GyrB_2GyrA_2$ enzyme demonstrates an activity spectrum not seen in other gyrases; while its negative supercoiling activity is comparable to others, it demonstrates enhanced relaxation (Karkare et al., 2012), DNA cleavage, and decatenation activity (Aubry *et al.*, 2006). Interestingly, while the *E. coli* gyrase is known to prefer supercoiling over decatenation (Zechiedrich, Khodursky and Cozzarelli, 1997), the opposite has been observed for Mtb gyrase with a preference for decatenation (Aubry *et al.*, 2006), alongside a lower rate of ATP turnover (Tretter and Berger, 2012). These observations have been, in part, explained by several motifs present throughout the various domains of Mtb GyrA and GyrB (**Figure 1.5**). Many of these insights have been gained through structural studies of individual domains alongside biochemical analyses (Piton et al., 2010; Karkare et al., 2012; Agrawal et al., 2013; Bouige et al., 2013; Blower et al., 2016; Petrella et al., 2019).

It has been proposed that the increased relaxation capabilities are due to a Ca^{2+} binding site located between the GyrA BRD and CTD (EF-hand motif, **Figure 1.5 A**). Removal of the Ca^{2+} from the enzyme through EGTA dialysis resulted in a reduction in the relaxation capabilities, while mutation of the EF-hand motif led to loss of gyrase activity (Karkare et al., 2012). The presence of the GyrA DEEE-loop in the tower of the BRD (**Figure 1.5 A**) (Piton *et al.*, 2010) was shown to interact with a GyrB *Corynbacteriales* specific insert (C-loop) (Agarwal and Tyagi, 2006) within the GHKL of the ATPase (**Figure 1.5 B**) allowing the enzymes to exist in an 'extremely open' conformation (Petrella *et al.*, 2019). This association is overridden by the binding of both DNA and ATP, but not ATP alone, and is predicted to prevent the futile cycling of nucleotide, adding a mechanism to promote ATPase/strand-passage coupling (Petrella *et al.*, 2019). The presence of the GyrA-box sequence is typically believed to be essential to the supercoiling activity of gyrase enzymes (Lanz and Klostermeier, 2012). Alongside the conserved Gyr-box sequence, the Mtb gyrase CTD contains an additional GyrA-box sequence implicated in the enhanced decatenation activity (Bouige *et al.*, 2013).

Altogether, these motifs allow for Mtb gyrase to be well adapted to compensate for the lack of Topo IV. Many of these adaptations likely have important roles in the efficient coupling of processes between domains as interfering with, or removing them entirely, affects more than just their proposed or dominant function (Bouige et al., 2013; Karkare et al., 2012; Petrella et al., 2019). The apparent preferences for the decatenation reaction, alongside the reduced ATPase rate, may therefore contribute to overall storage and metabolism of DNA, and as a result impact growth and replication rates. As gyrase is therefore central to most DNA metabolism, and Mtb gyrase has an increased burden, it is an attractive and well-exploited drug target. Discovery of new inhibitors is, however, a dwindling field.

Figure 1.5 Domain and motif architecture of the *M. tuberculosis* gyrase subunits



A) 2D schematic of the *M. tuberculosis* GyrA protein domain and motif organisation; **B**) 2D schematic of the *M. tuberculosis* GyrB protein domain and motif organisation. Both schematics are to scale based on amino acid numbers.

1.3.3.3 Research into new *M. tuberculosis* gyrase inhibitors

Currently we have a high dependency on gyrase targeting FQ antibiotics in the treatment of TB. Rates of resistance to FQ second-line therapeutics predominantly used in the treatment of drug-resistant TB is unfortunately on the rise, therefore, novel ways of inhibiting the gyrase target need to be explored; especially those that remain effective even against FQ-resistant gyrase.

One approach is to continue to develop FQ antibiotics to increase their efficacy to more efficiently eradicate the Mtb infection; this method has been used in the generation of our currently used FQs from nalidixic acid (Takahashi, Hayakawa and Akimoto, 2003), such as moxifloxacin, a fourth-generation quinolone (King, Malone and Lilley, 2000). Modifications to existing quinolone antibiotics, such as generating C8-Me-moxifloxacin, have been shown to produce enhanced DNA cleavage (Blower *et al.*, 2016); these do not, however, always address the issues associated with FQ-resistant strains.

Rather than the continued development of the FQ antibiotic series, there has been a recent effort to explore novel gyrase-inhibition through drug-repurposing (Balasubramani *et al.*, 2020) or exploring novel synthetic compound libraries (Blanco *et al.*, 2015). Echinacoside (predominantly a neuroprotective drug (Liu *et al.*, 2018)) and epirubicin (a known human topo IIα inhibitor used in cancer treatment (Petit *et al.*, 2004)) have been shown to exhibit good affinity for, and inhibition of, Mtb gyrase via the ATPase domain (Balasubramani *et al.*, 2020); these may subsequently act as scaffolds for the optimisation of novel anti-tubercular drugs.

In 2015, the *Mycobacterium tuberculosis* gyrase inhibitor (MGI) compound series stemmed from the GSK novel bacterial topoisomerase inhibitor library; importantly, these compounds were devoid of FQ cross-resistance (Blanco *et al.*, 2015). Two of these compounds (GSK000 and GSK325) were subsequently shown to induce persistent ssDNA breaks with specificity for Mtb gyrase, maintained activity against Mtb gyrase with the most common FQ resistance mutations, and had no activity against the human topo IIα (Gibson *et al.*, 2018). Gepotidacin ((Bax *et al.*, 2010), a compound from the NBTI library and used as the inhibitor in the recent full-length *E. coli* gyrase cryo-EM structure (Vanden Broeck *et al.*, 2019), entered phase III clinical trials in October 2019 to explore its use for uncomplicated urinary tract infection and urogenital gonorrhoea (infections not addressed by new oral antibiotics for 20 years) (GSK, 2019). Excitingly, these trials were recently stopped due to high efficacy (GSK, 2022) indicating a positive move toward a new antibiotic stemming from the NBTI library; this should give hope for the MGI series.

Additionally, in 2022 the Specs Library was explored for Mtb gyrase targeting compounds and identified two ATPase inhibitors by virtual screening and microbiological inhibition assays (Pakamwong *et al.*, 2022).

While these are exciting avenues for drug development, options are in limited supply and we must find more novel mechanisms to inhibit gyrase enzymes, ideally with specificity for the Mtb gyrase. One method is to explore naturally occurring inhibitors.

Aminocoumarin antibiotics, such as novobiocin, and the related simocyclinones have been derived from *Streptomyces* (Smith et al., 1956; Flatman et al., 2005). Novobiocin competes with ATP at the GyrB GHKL domain, however, has an inhibitory concentration against Mtb that is too high for consideration as a treatment option (Chopra *et al.*, 2012). Simocyclinone D8, in contrast, appears to inhibit the gyrase catalytic cycle by preventing the binding of substrate DNA (Flatman *et al.*, 2005), however, little is known about its affinity for Mtb gyrase.

The currently known proteinacious inhibitors of gyrase can be divided into the small peptides microcin B17 (MccB17) (Herrero and Moreno, 1986), albicidin (Hashimi *et al.*, 2007), and evybactin (Imai *et al.*, 2022), the quinolone resistance proteins (Qnr) (Tran and Jacoby, 2002), and the CcdB, FicT, and ParE toxins of toxin-antitoxin systems (Jaffé, Ogura and Hiraga, 1985; Jiang *et al.*, 2002; Harms *et al.*, 2017). Note that the ParE toxins should not be confused with ParE subunits of type IV topos.

MccB17 has been shown to stabilise the gyrase cleavage complex at 37 °C in an ATP-dependent manner (Heddle *et al.*, 2001) and at lower temperatures simply slows both supercoiling and relaxation reactions *in vitro* (Pierrat and Maxwell, 2003). Together, these findings suggest that MccB17 interrupts strand passage, rather than completely inhibiting the enzyme (Collin and Maxwell, 2019). Interestingly, several quinolone-resistant mutants were also cross-resistant to MccB17-induced cleavage, indicating some overlap in their binding (Heddle *et al.*, 2001). The affinity of this *Enterobacteriaceae*-produced class of gyrase inhibiting peptides for the Mtb gyrase is yet to be shown. Albicidin has been shown to have comparable inhibitory concentrations to the quinolone antibiotics and similarly traps gyrase cleavage complexes in both supercoiling and relaxation reactions (Hashimi *et al.*, 2007). Like MccB17, specific quinolone-resistance mutations in *E. coli* also conferred resistance to albicidin (Hashimi *et al.*, 2007; Hashimi, 2019). Again, the affinity for this class of peptide to Mtb gyrase has yet to be shown. With regards to being effective against Mtb, both MccB17 and albicidin would also rely on being transported into the cell by the Mtb complement of OmpA-like porins (Sarathy et al., 2012; Lavina et al., 1986); given that MccB17 partial resistance did not map to the homologous *E. coli* OmpA (Lavina et al., 1986), transport may not occur *in vivo*.

The recently discovered evybactin inhibitory peptide does, however, show high affinity and selectivity for Mtb gyrase (Imai *et al.*, 2022) and can be transported into the cell by BacA, an Mtb multi-solute ABC-type transporter of hydrophilic molecules (Rempel *et al.*, 2020; Imai *et al.*,

2022). The hydrophilic peptide was shown to be a gyrase poison, generating dsDNA breaks in a strictly ATP-dependent manner (Imai *et al.*, 2022). The crystal structure of the Mtb gyrase cleavage core ($GyrB^{426\_675}A^{2\_500}$) inhibited by evybactin (PDB: 7UGW) presented binding of evybactin at the GyrA-GyrB transducer interface, away from the quinolone binding pocket (Imai *et al.*, 2022). This is a similar site to thiophene poisons, a previously identified *E. coli* gyrase inhibiting compound from the NBTI library (Bax *et al.*, 2010; Chan *et al.*, 2017). Mtb cells exposed to evybactin had an elongated cell morphology, indicating that DNA synthesis and chromosome segregation had been impaired (Kumar *et al.*, 2010; Nonejuie *et al.*, 2013; Imai *et al.*, 2022).

Bacterial species, including Mtb, have also evolved to produce larger, fully folded protein inhibitors of their own topoisomerase repertoire (Yuan *et al.*, 2010; Aghera *et al.*, 2020; Mazurek *et al.*, 2021). One such type is pentapeptide repeat proteins (PRPs), structural mimics of B-DNA (Hegde *et al.*, 2005) which have been shown to inhibit gyrase supercoiling activity and protect the enzyme from FC-mediated poisoning (Feng *et al.*, 2021), and at the expense of ATP, can rejuvenate FQ-poisoned gyrase (Mazurek *et al.*, 2021). While PRPs are typically plasmid-borne offering a communicable mechanism of FQ protection (Li *et al.*, 2019), the Mtb chromosome encodes a PRP, mycobacterial FQ resistance protein A (*mfpA, rv3361c*). PRPs appear to function differently by species with the *E. coli* quinolone resistance protein QnrB1 being capable of protecting and rejuvenating *E. coli* gyrase from FQs by binding to the B subunit operating as a T-segment mimic (Shah and Heddle, 2014). Mtb MfpA was originally proposed inhibit gyrase activity via acting as a G-segment mimic (Hegde *et al.*, 2005) and has been shown to offer no protection against FQs (Mérens *et al.*, 2009). Contrastingly, the *M. smegmatis* homologue demonstrated similar biochemistry to QnrB1, inhibiting gyrase supercoiling, protecting its cognate enzyme from FQs (CFX and MXF), stimulated ATPase activity, and structural biology (PDB: 6ZT5) allowed for the development of a model supporting T-segment mimicry (Feng *et al.*, 2021). Given the similarity in the Mtb and *M. smegmatis* MfpA sequences and structures (Hegde *et al.*, 2005; Feng *et al.*, 2021), it is probable that their functionality is shared, however Mtb MfpA has been shown to require a small GTPase, mycobacterium FQ resistance protein B (MfpB, Rv3362c) to promote its interaction with gyrase (Tao *et al.*, 2013).

In contrast to this mechanism, toxin-antitoxin (TA) systems toxin proteins CcdB and ParE have been shown to poison gyrase, inducing DNA cleavage similar to FQs (Jiang *et al.*, 2002; Dao-Thi *et al.*, 2005), rather than offering protection to them, while the FicT toxin reversibly adenylates the gyrase ATP-binding site to prevent nucleotide binding and ATP-dependent activities (Harms *et al.*, 2015). Mtb encodes two *parE* toxins (*parE1* and *parE2*) and given previous success in generating gyrase-inhibiting CcdB and ParE peptides (Trovatti *et al.*, 2008; Barbosa *et al.*, 2012),

TA systems, and particularly the Mtb ParE toxins, are candidates for further study regarding their Mtb gyrase inhibiting capabilities.

1.4 Toxin-Antitoxin systems

Toxin-Antitoxin (TA) systems are found in most bacterial genomes as small bicistronic loci and are comprised, typically, of a protein toxin and a protein or RNA antitoxin (Beck et al., 2020; Fineran et al., 2009; Gerdes et al., 1986; Jaffé et al., 1985). Toxins of these systems have highly varied targets and result in the inhibition or destabilisation of a range of cellular processes or features (Yuan *et al.*, 2010; Arcus *et al.*, 2011; Griffin, Davis and Strobel, 2013; Fernández-García *et al.*, 2016). The physiological roles of these systems are often debated, and varies organism to organism, but they have been associated with genomic stability, bacteriophage defence, programmed cell death, biofilm formation, and the hotly debated theory of bacterial persistence (Kedzierska and Hayes, 2016; Rocker and Meinhart, 2016; Song and Wood, 2020). TA systems do have an apparent association with pathogenicity, especially in the Mycobacteria as *M. tuberculosis* H37Rv encodes an estimated 88 systems (Ramage, Connolly and Cox, 2009) (~ 2% of the proteome), whereas the typically non-pathogenic *M. smegmatis* encodes an estimated 5 systems (~ 0.08% of the proteome) (Sala, Bordes and Genevaux, 2014; Slayden, Dawson and Cummings, 2018). The best-studied organism in relation to TA systems in *E. coli* contains around 35 – 40 (Yamaguchi and Inouye, 2011), indicating their varied distribution.

TA systems were originally identified as plasmid maintenance systems in the 1980s, encoded on the R1 and F plasmids; both early studies laid the foundations for the generalised theory of antitoxin instability (Gerdes et al., 1986; Ogura & Hiraga, 1983). These systems became the founding members of their respective classifications, and while they essentially performed the same role in ensuring proper genome segregation and plasmid inheritance, their mechanisms differed. The F plasmid is maintained by a protein-protein TA system (CcdBA), with the antitoxin CcdA directly interacting with the CcdB toxin to inhibit toxicity (Tam and Kline, 1989; Dao-Thi *et al.*, 2005). The CcdB toxin, when expressed alone, was shown to induce cell filamentation and the SOS response (Jaffé, Ogura and Hiraga, 1985); it has since been shown that the CcdB toxin binds to and inhibits DNA gyrase at the coiled-coil C-gate of the gyrase A subunit (Dao-Thi *et al.*, 2005). The *hok-sok* system maintained the R1 plasmid via a different mechanism, whereby the Hok toxin induced a 'ghost cell' phenotype, indicative of membrane pore forming and loss of membrane potential (Thisted and Gerdes, 1992). Toxicity in this case was prevented by an antisense RNA complementary to the *hok* toxin transcript, thereby preventing Hok synthesis (Thisted and Gerdes, 1992). Both systems become activated in toxicity if plasmids are not inherited, predicted to be due to the more stable toxin persisting in the daughter cells while the less stable antitoxins

are somehow degraded not replenished (Jaffé, Ogura and Hiraga, 1985; Thisted and Gerdes, 1992).

These early studies were merely the beginning of the discovery of a multitude of systems across prokaryotes with highly varied cellular targets, mechanisms of antitoxicity, toxicity, regulation and activation, and subsequently heavily debated roles in bacterial physiology (Guglielmini and van Melderen, 2011; Sala, Bordes and Genevaux, 2014; Yamaguchi and Inouye, 2016; Fraikin, Goormaghtigh and van Melderen, 2020). Research into TA systems has not only advanced our understanding of bacterial adaptation and pathogenicity (Wen, Behiels and Devreese, 2014), it has also been theorised they can be exploited for therapeutic purposes (Trovatti *et al.*, 2008; Chono *et al.*, 2011) and have yielded several biotechnological applications (Kristoffersen *et al.*, 2000; Stieber, Gabant and Szpirer, 2008). Due to the ever-increasing number and variety of TA systems, several methods of classification have been introduced to simplify and efficiently group them.

1.4.1 Classification of TA systems

Since the identification and study of the CcdBA and *hok-sok* systems, largely due to the advances in genome sequencing technologies (and homology searching), we have been able to identify numerous other TA systems with a variety of established and predicted functions (Anantharaman and Aravind, 2003; Pandey and Gerdes, 2005). Furthermore, the ever-expanding complement of TA system protein structures in the PDB has provided a wealth of information regarding the evolutionary structural relationships between systems, with fold sharing demonstrated between toxin types and antitoxin types (Blower, Salmond and Luisi, 2011; Chan, Espinosa and Yeo, 2016). While the toxins and antitoxins can be grouped into 'families' based on sequence and structural homology, the mechanisms of antitoxicity (toxin neutralisation) are relatively limited; this has led to our current, but constantly developing, TA system classifications (**Figure 1.6**).

1.4.1.1 Classification by mechanisms of antitoxicity

Type I systems encode a small antisense RNA antitoxin which silences their cognate toxin transcript (Gerdes et al., 1986) (**Figure 1.6 A**); type II systems are neutralised through a protein antitoxin forming a complex with the cognate toxin (Jaffé, Ogura and Hiraga, 1985; Fraikin, Goormaghtigh and van Melderen, 2020) (**Figure 1.6 B**); type III systems also encode an RNA antitoxin, however, these form a pseudoknot structure with binds to the cognate toxin to form a neutralised protein-RNA complex (Blower *et al.*, 2009; Fineran *et al.*, 2009) (**Figure 1.6 C**); type IV antitoxins are proteins which do not interact with their cognate toxin, rather, they act to protect or detoxify the target of the toxin (Masuda, Tan, Awano, Yamaguchi, *et al.*, 2012) (**Figure 1.6 D**); type V systems encode an RNase antitoxin which specifically degrades the cognate toxin

transcripts to prevent toxin accumulation (Wang *et al.*, 2012) (**Figure 1.6 E**); type VI system antitoxins are proteins which target their cognate toxin for protease degradation (Aakre *et al.*, 2013) (**Figure 1.6 F**); type VII antitoxicity occurs through protein antitoxins post-translationally modifying their cognate toxin (Cai *et al.*, 2020; Songailiene *et al.*, 2020) (**Figure 1.6 G**).

The vast majority of antitoxins are proteins, with most systems identified to date belonging to the type II family, and unsurprisingly these have become the best studied (Fraikin, Goormaghtigh and van Melderen, 2020). Type II antitoxins neutralise their cognate toxin via a toxin-binding domain, which often remains unstructured until bound (**Figure 1.6 B**) (Chan, Espinosa and Yeo, 2016). The formation of the TA complex in type II systems (and type III) tends to block the toxins' active site or essential residues for activity (Blower *et al.*, 2011; Bøggild *et al.*, 2012). For example, the binding of the RelB antitoxin to the ribosome-dependent ribonuclease RelE (**Table 1.3**) toxin displaces a C-terminal toxin helix containing an essential catalytic tyrosine residue, thereby perturbing the structure of the catalytic site (Neubauer *et al.*, 2009; Bøggild *et al.*, 2012). Furthermore, the formation of the complex prevents the RelE toxin from accessing the ribosomal A-site, sterically preventing toxicity (Neubauer *et al.*, 2009).

Other antitoxin proteins are active enzymatically; the GhoS antitoxin is catalytically active as a *ghoT* specific endoribonuclease, degrading the toxin transcript to prevent its translation (Wang *et al.*, 2012). In contrast to this, protein antitoxins of type VII systems post-translationally modify conserved active site residues in their cognate toxins; the MenA antitoxins of Mtb phosphorylate their cognate nucleotidyltransferase toxin (MenT) (Cai *et al.*, 2020; Yu *et al.*, 2020), while the MNT antitoxin itself is a nucleotidyltransferase which modifies the RNase toxin, HEPN, by di-AMPylation (Songailiene *et al.*, 2020).

Classification by antitoxicity is the canonical method, with TA systems referred to in terms of their 'type'. It is, however, possible to simplify toxicity into digestible categories depending on toxic mechanisms and allows us to better understand the variety of TA system targets.

Figure 1.6 Antitoxicity classification of toxin-antitoxin systems

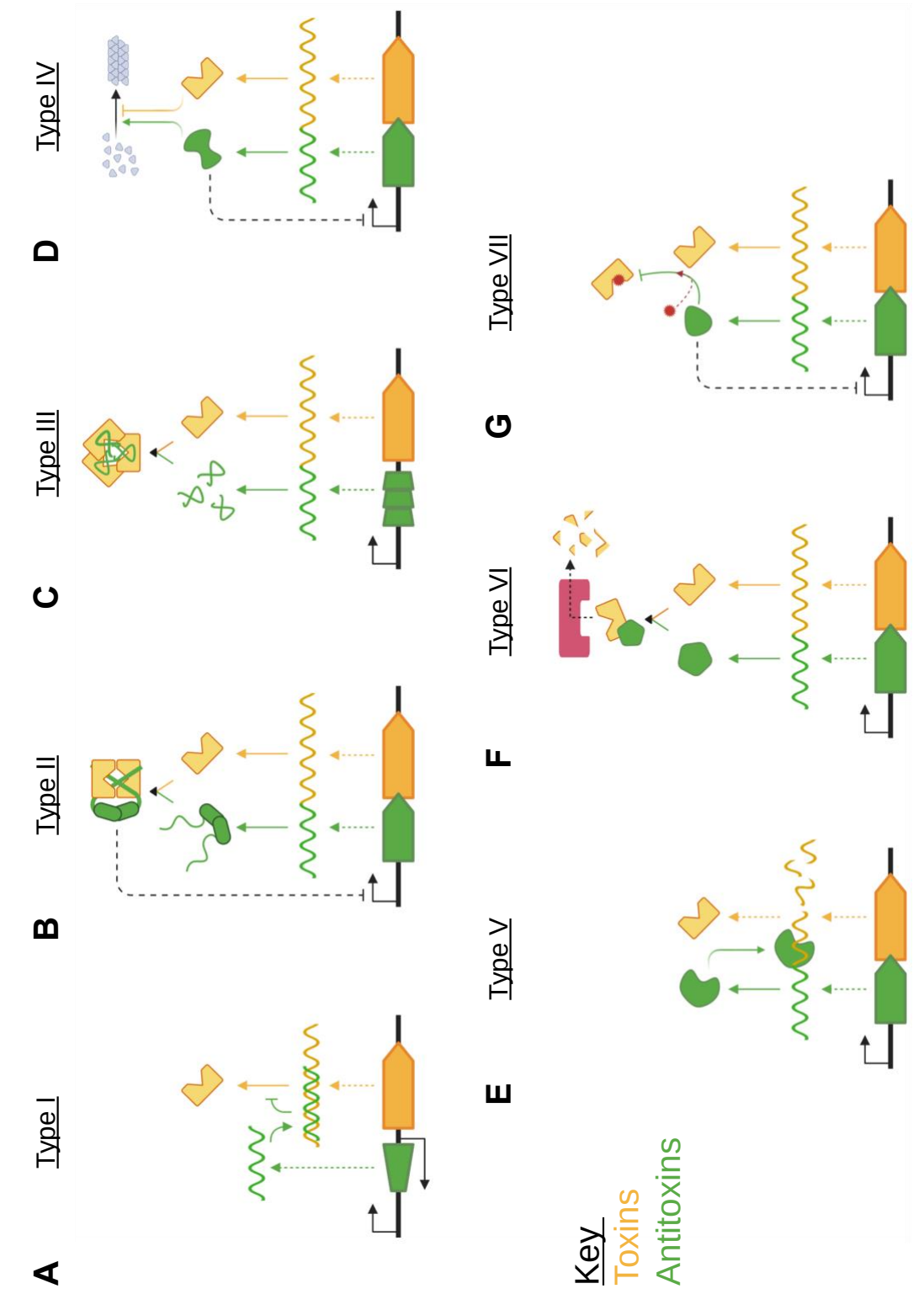


Figure legend overleaf.

Figure 1.6 Antitoxicity classification of toxin-antitoxin systems

Cartoon representation of TA system types with antitoxins coloured in green and toxins in orange. Genes encoding protein products are represented as pointed rectangles while those encoding untranslated RNAs are represented as trapezoids. **A**) Type I: An antisense antitoxin transcript silences the toxin transcript; **B**) Type II: A protein antitoxin directly binds to its cognate toxin; **C**) Type III: Pseudoknot RNAs bind directly to protein toxins; **D**) Type IV: Protein antitoxins stabilise/promote the cellular process destabilised by cognate protein toxins; **E**) Type V: Protein antitoxins act as a cognate toxin specific endoribonuclease; **F**) Type VI: Protein antitoxin acts as an adaptor targeting its cognate toxin for degradation by an ATP-dependent protease (red); **G**) Type VII: Protein antitoxin post-translationally modifies its cognate toxin, (modification represented in red). Autoregulatory function is represented by blunt-ended black dashed lines.

1.4.1.2 Classification by mechanism of toxicity

Like antimicrobial therapies, previously highlighted by the complex treatment regimens used for TB, TA systems have highly varied targets ranging from inhibiting whole cell processes by stable interaction (Thisted and Gerdes, 1992; Aakre *et al.*, 2013; Deghorain *et al.*, 2013) to modifying or degrading RNA molecules (Hazan, Sat and Engelberg-Kulka, 2004; Cai *et al.*, 2020). While TA classification occurs via the mechanism of antitoxicity, toxicity has previously been grouped into broad families, based on toxin targets, of a) Impairing DNA and replication; b) Degrading RNAs and impairing translation; c) Impairing cell envelope and cytoskeleton integrity, and; d) inducing metabolic stress (Jurėnas *et al.*, 2022). While these groupings are useful, other than the degradation of RNA, they are vague and make no comment on the actual mechanism employed by toxins.

Here, we propose a simplified classification of toxicity families based on mechanisms rather than targets; A) Degradation/cleavage; B) Modification; C) Direct interaction/inhibition (**Table 1.3**). These broad families allow for toxins to belong to multiple classes; continuing the previous example, RelE toxins (including HigB, YoeB, YafQ) which bind to the ribosome to cleave mRNA at the A-site could be considered to target both ribosome and mRNA. RelE binding directly (C) to the ribosome prevents the entry of tRNA at the A-site (Neubauer *et al.*, 2009), and this coordinates the toxin for cleavage (A) of mRNA (Griffin, Davis and Strobel, 2013). Considering these toxicity classifications alongside the antitoxicity TA families (**Table 1.3**), it is clear to see the diversity of inhibitory mechanisms employed throughout the canonical classifications shown in **Figure 1.6**, exemplified by the type II TA systems, which display toxicity mechanisms from all three proposed families.

The degradation/cleavage family of toxins (A) is by far the largest and best studied of the toxin types, with inhibition of protein synthesis being central to their toxicity and a favoured mechanism of TA system toxins (Fraikin, Goormaghtigh and van Melderen, 2020; Jurėnas and van Melderen, 2020). Overlapping with the directly interacting family, RelE type toxins (RelE, HigB, YoeB, YafQ; grouped by sequence homology (Anantharaman and Aravind, 2003)) (**Table 1.3**, type II antitoxicity) locate to the ribosome for correct positioning toward substrate mRNA (Hurley and Woychik, 2009; Neubauer *et al.*, 2009). RelE type toxins inhabit the ribosomal A-site after repositioning their catalytic tyrosine-containing C-terminal helix to reconstitute a catalytic core of atypical residues and cleave mRNA between the second and third position of targeted codons, while the HigB toxins cleave at AAA sequences (Christensen and Gerdes, 2003; Hurley and Woychik, 2009; Griffin, Davis and Strobel, 2013). Interestingly, personal correspondence with collaborators in the Genevaux group (Toulouse, France) revealed that the Mtb RelE1 toxin does not cleave mRNA in a well-established ribosome-dependent cleavage assay, despite being toxic

in *E. coli*, *M. smegmatis*, and Mtb (Sala, Bordes and Genevaux, 2014). This may indicate inhibition by direct means of blocking the ribosomal A-site to inhibit translation, or a novel mechanism of RelE activation and catalysis, or an altered target.

Contrastingly, SymE, MazF, and ToxN/AbiQ (**Table 1.3**, type I, type II, and type III antitoxicity, respectively) act to degrade free RNAs with differing levels of specificity (Kawano, Aravind and Storz, 2007; Samson *et al.*, 2013; Barth and Woychik, 2020). SymE is thought to degrade damaged mRNA (Kawano, Aravind and Storz, 2007), while some MazE toxins have been shown to have low specificity for free mRNA (Culviner and Laub, 2018) and others are tRNA specific (Barth and Woychik, 2020). The most abundant toxins identified to date, VapC (**Table 1.3**, type II antitoxicity), appear to specifically cleave the anticodon stem loop of tRNAs and in certain cases, a sarcin-rich region of 23S rRNA, likely due to structural similarities (Winther *et al.*, 2013; Cruz *et al.*, 2015). Similarly, the HEPN toxin (**Table 1.3**, type VII antitoxicity) targets specific tRNA molecules, cleaving them at the acceptor stem to prevent recruitment to the ribosome by the elongation factor Tu (Ef-Tu) (Songailiene *et al.*, 2020). Degradation/cleavage toxins are not limited to targeting RNA; other toxins have been shown to exhibit non-specific DNase activity in the case of RalR (Guo *et al.*, 2014), and the MbcT toxin hydrolyses NAD+ to impair redox reactions (Freire *et al.*, 2019).

The family of toxins capable of modifying their targets (B) is relatively smaller and exhibits higher specificity than those capable of cleavage (A) (**Table 1.3**). The zeta toxin (**Table 1.3**, type II antitoxicity) acts as a kinase, phosphorylating uridine diphosphate-N-acetylglucosamine (UNAG) and subsequently inhibiting the enzyme catalysing the initial step in peptidoglycan synthesis, MurA (Mutschler *et al.*, 2011). The Phd toxin similarly acts as a kinase, although rather unexpectedly as it contains a Fic domain typically associated with transfer of AMP moieties (AMPylation) (Worby *et al.*, 2010). The inverted substrate transfer of the phosphate, rather than AMP, to a conserved threonine residue of Ef-Tu renders the elongation factor unable to bind aminoacylated tRNA for transfer the ribosome, effectively inhibiting translation (Castro-Roa *et al.*, 2013). HipA toxins have been shown to inhibit aminoacyl-tRNA synthetases, preventing charging of tRNA molecules (Kaspy *et al.*, 2013). The MenT toxins (**Table 1.3**, type VII antitoxicity) similarly prevent charging of tRNA molecules and subsequently inhibited translation, however, they do so by ligating pyrimidine to the 3' end of uncharged tRNA acceptor stems (Cai *et al.*, 2020). DNA, rather than specific RNA molecules, is targeted by the DarT toxin (**Table 1.3**, type II/IV hybrid antitoxicity) which acts a sequence specific ADP-ribosyl transferase to impair DNA replication and induce the SOS response (Jankevicius *et al.*, 2016). In contrast to the inverted transfer of phosphate by the Doc toxin, the FicT toxin (**Table 1.3**, type II/IV hybrid antitoxicity) operates as a canonical Fic protein, inhibiting the ATPase function of gyrase and topo IV by

AMPylation of the GHKL domain (Harms *et al.*, 2015). While the ATP-dependent functions of both enzymes were inhibited *in vitro* by FicT, toxin induced elongated cell morphology suggested a potential preference for topo IV *in vivo* (Harms *et al.*, 2015).

The directly acting family of toxins (C) includes Hok, TisB, and DinQ (**Table 1.3**, Type I antitoxicity), and GhoT (**Table 1.3**, type V antitoxicity), which have been shown to resemble phage holin proteins and become membrane associated (Gurnev *et al.*, 2012; Wang *et al.*, 2012; Brielle, Pinel-Marie and Felden, 2016; Wilmaerts *et al.*, 2018). By creating membrane pores, bacterial membrane potential is lost causing the cessation of respiration and leakage of intracellular ATP (Weel-Sneve *et al.*, 2013; Wilmaerts *et al.*, 2018), ultimately leading to the 'ghost cell' phenotype and death (Thisted and Gerdes, 1992; Wang *et al.*, 2012). SocB (**Table 1.3**, type VI antitoxicity) is believed to bind directly to a hydrophobic cleft on the β sliding clamp of the DNA elongation machinery, predicted from mutants bypassing SocB mediated toxicity (Aakre *et al.*, 2013). Direct inhibition of the clamp reduces the processivity of DNA polymerase III, weakening its association with template DNA to impair DNA replication (Johnson and O'Donnell, 2005; Aakre *et al.*, 2013). The CbtA toxin (**Table 1.3**, type IV antitoxicity), also called YeeV, interacts with cytoskeletal proteins FtsZ and MreB to inhibit their GTP/ATP-dependent polymerisation, respectively, preventing cell division (Tan, Awano and Inouye, 2011). Both CcdB and ParE toxins (**Table 1.3**, type II antitoxicity) are inhibitors of gyrase (as previously mentioned), however only the mechanism of gyrase poisoning by CcdB has been elucidated; a CcdB dimer binds to the C-gate (in the space created within the GyrA dimer between the C-gate and the DNA-gate/tower, **Figure 1.4**), locking the enzyme in the cleavage state by interfering with gate dynamics and generating genotoxic DNA breaks (Tam and Kline, 1989; Dao-Thi *et al.*, 2005). While ParE toxins have been shown to inhibit gyrase activity *in vitro* (Jiang *et al.*, 2002; Yuan *et al.*, 2010), and their *in vivo* toxicity demonstrates a cell morphology consistent with gyrase poisoning (Gupta *et al.*, 2016; Ames *et al.*, 2019), the exact mechanism has not been shown. Given how widespread the ParDE systems are (Anantharaman and Aravind, 2003; Fraikin, Goormaghtigh and van Melderen, 2020), this is a perplexing inconsistency in our current knowledge of TA system toxins. The ParDE systems of *M. tuberculosis* will be the focus of this thesis in later chapters.

Table 1.3 Targets of TA system toxins

TA Type	Toxin (Family)	Target/effect	Reference
Type I			
hok-sok	HokB (C)	Membrane pore forming. Loss of membrane potential, ATP leaking.	(Gerdes *et al.*, 1986)
tisB-istR	TisB (C)	Membrane pore forming. Loss of membrane potential, ATP leaking.	(Vogel *et al.*, 2004)
dinQ-AgrAB	DinQ (C)	Membrane pore forming. Loss of membrane potential, ATP leaking.	(Weel-Sneve *et al.*, 2013)
symE-symR	SymE (A)	mRNA cleavage. Impairs translation.	(Kawano, Aravind and Storz, 2007)
Type II			
VapBC	VapC (A)	Degrades free tRNA, rRNA. Impairs translation.	(Robson *et al.*, 2009)
MazEF	MazF (A)	Degrades free RNA, rRNA, tRNA. Impairs translation.	(Engelberg-Kulka, Hazan and Amitai, 2005)
RelBE	RelE (C/A)	Ribosome-dependent mRNA cleavage. Impairs translation.	(Gotfredsen and Gerdes, 1998)
HigBA	HigB (C/A)	Ribosome-dependent mRNA cleavage. Impairs translation.	(Christensen-Dalsgaard and Gerdes, 2006)
YefM-YoeB	YoeB (C/A)	Ribosome-dependent mRNA cleavage. Impairs translation.	(Nieto *et al.*, 2007)
DinJ/YafQ	YafQ (C/A)	Ribosome-dependent mRNA cleavage. Impairs translation.	(Prysak *et al.*, 2009)
Phd-Doc	Doc (B)	Phosphorylates Tu elongation factor (EF-Tu). Inhibits tRNA presentation to ribosome and translation.	(Cruz *et al.*, 2014)
FicTA*	FicT (B)	AMPylates DNA gyrase/Topo IV. Prevents ATP binding/hydrolysis.	(Harms *et al.*, 2015)
DarTG*	DarT (B)	ADP-ribosylates DNA. Induces SOS, impairs DNA replication.	(Jankevicius *et al.*, 2016)

Table 1.3 Targets of TA system toxins

TA family	Toxin (Family)	Target/effect	Reference
Type II			
Epsilon-Zeta	Zeta (B)	Phosphorylates UDP-activated sugars. Inhibits peptidoglycan synthesis.	(Mutschler *et al.*, 2011)
CcdBA	CcdB (C)	DNA gyrase. Traps cleavage complex, inhibiting transcription/replication, ds DNA breaks induce SOS.	(Jaffé, Ogura and Hiraga, 1985)
ParDE	ParE (C)	DNA gyrase. Traps cleavage complex, inhibiting transcription/replication, ds DNA breaks induce SOS.	(Jiang *et al.*, 2002)
Type III			
ToxIN	ToxN (A)	mRNA cleavage. Impairs translation.	(Blower *et al.*, 2009)
AbiQ-antiQ	AbiQ (A)	mRNA cleavage. Impairs translation.	(Samson *et al.*, 2013)
Type IV			
CbtA-CbeA	CbtA (C)	Prevents polymerisation of MreB and FtsZ (cytoskeleton). Inhibits cell division.	(Masuda, Tan, Awano, Wu, *et al.*, 2012)
Type V			
GhoST	GhoT (C)	Membrane pore forming. Loss of membrane potential, ATP leaking.	(Wang *et al.*, 2012)
Type VI			
SocAB	SocB (C)	β sliding clamp (RNA polymerase III). Inhibits DNA replication.	(Aakre *et al.*, 2013)
Type VII			
MenTA	MenT (B)	Ligates pyrimidines to tRNA acceptor stem, preventing charging. Impairs translation.	(Cai *et al.*, 2020)
HEPN-MNT	HEPN (A)	Cleaves tRNA acceptor stem. Impairs translation.	(Songailiene *et al.*, 2020)

Toxin families: A – Direct interaction with target; B – Degradation/cleavage of target; C – Modification of target. Table highlights the different toxicity mechanisms employed within canonical TA antitoxicity type classifications. *Hybrid type II system also belonging to type IV class.

1.4.1.3 Classification by protein structure

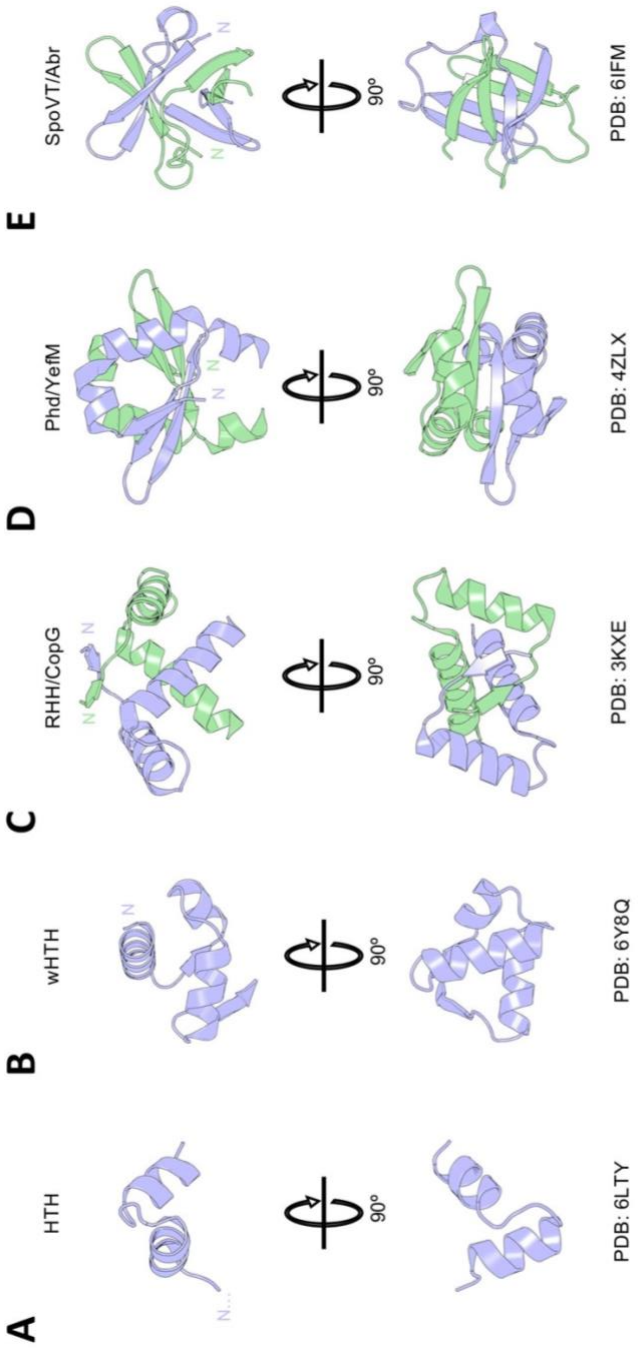
Despite the diverse mechanisms of toxicity employed by the numerous types of toxins listed in **Table 3**, only a snapshot of the TA system proteome, toxins exhibit a limited number of tertiary folds (Blower, Salmond and Luisi, 2011; Zhang *et al.*, 2020). Interestingly, a number of proteins sharing the same fold exhibit toxicity mechanisms from the different classes proposed above and low sequence similarity, notably within the type II TA systems.

The MazF fold is unsurprisingly shared by other RNAse toxins such as ToxN (type III ToxIN system) and Kid (type I *kid-kis* system), but is also shared by the gyrase inhibitor, a protein devoid of nuclease function, CcdB (Zhang *et al.*, 2020). Possibly the largest structural family is the RelE fold, shared by the each of the ribosome-dependent endoribonuclease toxins; HigB, YoeB, and YafQ, each differing slightly in their catalytic core sequences (Nieto *et al.*, 2007; Griffin, Davis and Strobel, 2013; Maehigashi *et al.*, 2015). Similar to the MazF fold, the RelE fold is also demonstrated by the ParE toxins, which like CcdB, have no nuclease activity and inhibit gyrase directly (Jiang *et al.*, 2002), albeit this has not yet been demonstrated biophysically or structurally. The RelE family has a higher degree of sequence similarity (Anantharaman and Aravind, 2003) creating a conserved mechanism of protein interaction with their cognate antitoxins via a hydrophobic patch (Dalton and Crosson, 2010).

Beyond the large MazF and RelE superfamilies, fold sharing is far lower with fewer toxins from different system types sharing a common fold. The VapC toxins create a large enough superfamily on their own due to their relative abundance (Sala, Bordes and Genevaux, 2014), but the tertiary fold is seen in the FitB toxin, also an endoribonuclease toxin (Mattison *et al.*, 2006). The Zeta and PezT toxins unsurprisingly maintain a common fold (Blower, Salmond and Luisi, 2011) with both modifying their respective targets via phosphorylation (Seok *et al.*, 2007; Mutschler *et al.*, 2011). Despite the inverted mechanisms of target modification demonstrated by the FicT and Doc toxins, they share the Fic domain and remain structurally related (Worby *et al.*, 2010). While the Zeta, DarT, MenT, and HipA toxins all appear to be bi-lobed with nucleotide binding pockets, their global tertiary folds share no similarity and they likely represent their own classes (Schumacher *et al.*, 2009; Cai *et al.*, 2020; Schuller *et al.*, 2021).

Antitoxin proteins, notably those of the type II class also, can be categorised into structural superfamilies also largely depending on their DNA-binding/dimerisation domains (Blower, Salmond and Luisi, 2011; Chan, Espinosa and Yeo, 2016). Within the type II systems autoregulation via DNA binding is a common theme, and this is routinely achieved via a small helix-turn-helix (HTH) motif, or slightly larger more complex DNA binding domains that form on antitoxin dimerisation (**Figure 1.7**).

Figure 1.7 DNA-binding domains of antitoxin proteins



A) Helix-turn-helix (HTH) of the *M. tuberculosis* HigA3 antitoxin, residues 50 – 73 (J. Y. Park *et al.*, 2020); **B**) winged HTH (wHTH) of the *S. agalactiae* AbiEi antitoxin, residues 2 – 45 (Beck *et al.*, 2020); **C**) Dimerised ribbon-helix-helix (RHH)/CopG of the *C. crescentus* ParD antitoxin, residues 2 – 45 (Dalton & Crosson, 2010); **D**) Dimerised Phd/YefM of the phage P1 Phd antitoxin, residues 2 – 45 (Garcia-Pino *et al.,* 2016); **E**) Dimerised SpoV/Abr of the S. typhimurium VapB antitoxin, residues 2 – 45 (D. W. Park *et al.*, 2020). For D – E, individual antitoxin chains are coloured light blue/pale green.

The HTH motif is the smallest found (**Figure 1.7 A**) (J. Y. Park *et al.*, 2020) and is regularly seen in HigA and MqsA type antitoxins (Chan, Espinosa and Yeo, 2016). Interestingly the HTH can be found both within antitoxin dimerisation regions (J. Y. Park *et al.*, 2020) and outside of them, closer to or within the toxin-binding domain (Yang *et al.*, 2016). A larger variation of the HTH is the winged-HTH (wHTH), (**Figure 1.7 B**) a DNA binding domain found in type VII system antitoxins (Janowski *et al.*, 2009; Beck *et al.*, 2020). Beyond the single chain HTH and wHTH domains are the DNA binding domains formed by the dimerisation between two antitoxin chains within type II systems (**Figure 1.7 C – E**). The dimeric ribbon-helix-helix (RHH) motif is formed from a topology of β- sheet – α- helix – α- helix, with dimerisation occurring through an antiparallel β-sheet (**Figure 1.7 C**). The domain was originally identified in the CopG protein (Xavier Gomis-Rüth *et al.*, 1998) and has subsequently been identified in a number of TA system antitoxin dimers including RelB (Bøggild et al., 2012), CcdA (Madl *et al.*, 2006), FitA (Mattison *et al.*, 2006), DinJ (Liang *et al.*, 2014), and ParD (Dalton and Crosson, 2010). The dimeric Phd/YefM fold constitutes two α-helices and three β-sheets through which dimerisation occurs to form a six-stranded antiparallel β-sheet (**Figure 1.7 D**) (Garcia-Pino *et al.*, 2016). These domains have only been found in the Phd antitoxin of the Phd-Doc system (Garcia-Pino *et al.*, 2016), alongside numerous YefM type antitoxins of the YefM-YoeB systems (Kumar *et al.*, 2008; Miallau *et al.*, 2013; Xue *et al.*, 2020). Finally, the SpoV/Abr dimeric domain is a layered swapped-hairpin β-barrel structure (**Figure 1.7 E**) (Bendtsen *et al.*, 2017) found in VapB and MazE antitoxins (Kamada, Hanaoka and Burley, 2003; Dienemann *et al.*, 2011). These dimerisation domains of type II antitoxins are structured and regularly form a highly stable dimer (Blower, Salmond and Luisi, 2011; Chan, Espinosa and Yeo, 2016), likely promoting the subsequent formation of the TA complex through a second, typically unstructured but toxin specific domain (Fraikin, Goormaghtigh and van Melderen, 2020). These structures contribute to the regulation of type II TA systems, often relying on toxin:antitoxin stoichiometry (Garcia-Pino *et al.*, 2010).

1.4.2 Regulation and activation of type II TA systems

Type II antitoxins are typically composed of a stable and structured dimerisation domain which, once formed, constitutes a regulatory DNA-binding domain as highlighted above (**Figure 1.7**). The very nature of antitoxicity being direct interaction in type II systems gives rise to the potential for influence of the regulation of the systems by the relative ratios, or stoichiometry, of the components. With type II system complexes regularly existing in equimolar stoichiometries (Blower, Salmond and Luisi, 2011; Chan, Espinosa and Yeo, 2016), activation of type II systems therefore occurs when the toxin exists in excess of the antitoxin, regulated and achieved either transcriptionally, post-transcriptionally, or post-translationally (Jurėnas *et al.*, 2022). Numerous TA system crystal structures have been solved in their neutralised state as equimolar stoichiometry complexes; the PDB structures of the RelE/ParE superfamily TA systems are

summarised in **Table 1.4** as an example. To complement this table, the structures of the ParDE system complexes, as these are the primary TA system focus of this study, have been summarised in supplementary **Figure S1**.

1.4.2.1 Transcriptional and post-transcriptional regulation

The general architecture of the TA operon likely plays a role in transcriptional regulation; type II systems are often encoded with the antitoxin first, the toxin following via a short intergenic region or overlapping the antitoxin transcript (Leplae *et al.*, 2011). Type III systems are organised similarly, however the antitoxin and toxin reading frames are separated by Rho-independent terminators (Blower *et al.*, 2011). The positioning of the antitoxin first may result in preferential synthesis of relatively unstable antitoxins, limiting toxin production; it has been estimated that the *E. coli* RelB antitoxin exists in a 10-fold excess of the RelE toxin in rapidly growing cells (Overgaard, Borch and Gerdes, 2009). In cases where the inverse architecture is present antitoxin-specific promoters have been described, potentially to maintain its excess (Turnbull and Gerdes, 2017). Post-transcriptionally, control of the relative levels of translation of the TA components can be regulated by the inherent stability of toxin and antitoxin transcript or by the targeted degradation of transcripts; translation has been reported to be far less efficient for toxins which results in molar excess of antitoxin (Ruiz-Echevarría, de la Cueva and Díaz-Orejas, 1995).

The resulting molar excess of antitoxin in type II TA systems ensures effective neutralisation of toxin, however, there is an intimate link between neutralisation and transcriptional repression. In antitoxins demonstrating a disorder-to-order transition upon toxin binding, negative autoregulation of their operon through the antitoxin DNA-binding domain (**Figure 1.7**) regularly occurs via the model of conditionally cooperative binding to maintain a ratio in favour of antitoxin (Garcia-Pino *et al.*, 2010). In high molar excess of antitoxin, partly unstructured antitoxin dimers demonstrate negative cooperative binding to operator sites, and when toxin levels increase and toxins are sequestered by antitoxins, the promoter affinity increases as the complex undergoes a transition to ordered and becomes structurally compatible with the operator site architecture (Garcia-Pino *et al.*, 2010), effectively repressing transcription. However, when toxin levels exceed that of the antitoxin, the resulting saturated complexes have low affinity for the promoter and either dissociate or fail to bind, de-repressing transcription to subsequently replenish antitoxin levels (Overgaard *et al.*, 2008). Conditional cooperativity in TA system autoregulation elicits tight control of toxin and antitoxin levels, preventing activation (toxin release) until somehow overridden (Garcia-Pino *et al.*, 2010; Cataudella *et al.*, 2012). This model indicates that TA complexes exist in solution in multiple stoichiometries, each with a particular role in TA regulation. However, not all antitoxins undergo the disorder-to-order transition (Kumar *et al.*,

2008; Dalton and Crosson, 2010), nor do all TA system structures exhibit DNA-binding domains (Takagi et al., 2005; Das et al., 2014).

In theory, activation of the TA system at the transcriptional level, and preferential synthesis of toxin, could be achieved through targeted degradation of antitoxin transcripts, effectively downregulating the expression of antitoxin. Given the RNA specificity of a number of endoribonuclease toxins (Fiebig *et al.*, 2010; Griffin, Davis and Strobel, 2013; Simanshu *et al.*, 2013), this is plausible and would demonstrate TA system interplay, however, it has not yet been documented. Rather, transcripts are likely degraded by cellular RNases; this is particularly important in type I TA systems whereby systems may be activated by preferential degradation of the less-stable sRNA antitoxin (Peltier *et al.*, 2020). It has been reported that the RelBE2 system of Mtb is in fact a hybrid type I/II system involving sRNA co-regulation and activation via targeted processing by RNase III (Dawson *et al.*, 2022). Contrastingly, TA systems can be regulated and activated by antitoxin degradation at the post-translational level.

1.4.2.2 Post-translational regulation

In type II systems, the act of antitoxin binding to the toxin initially regulates and inhibits the activity of its cognate toxin through appropriate neutralising mechanisms; disrupting secondary structure (Bøggild *et al.*, 2012), blocking the toxin active site (Kamada, Hanaoka and Burley, 2003), preventing toxins forming active oligomeric multimers (Jurėnas, van Melderen and Garcia-Pino, 2019), or sterically hindering the toxin from binding to its target (Aghera *et al.*, 2020).

Activation of type II systems has long been theorised to be regulated by the selective degradation of antitoxin by AAA+ proteases (Bordes and Genevaux, 2021). Interestingly, as part of the intricate co-expression network of TA systems identified in Mtb exposed to infection and treatment mimicking stresses, the Clp protease was routinely associated (Gupta *et al.*, 2017), adding weight to theorised roles of Mtb proteases in TA system activation. It is worth noting that stress-induced transcription of TA systems, likely arising from antitoxin degradation, is not consistently linked to toxin activation (LeRoux *et al.*, 2020). Indeed, however, many antitoxins contain specific protease cleavage sites (Texier *et al.*, 2021), many have partially unstructured domains which render them susceptible to targeted degradation (Lunge *et al.*, 2020), and they have been shown to interact with chaperone proteins for specific proteases (Ziemski *et al.*, 2021). Many antitoxins have been shown to be degraded *in vitro* by proteases also (Lunge *et al.*, 2020; Ziemski *et al.*, 2021), supporting this theory. However, it has also been suggested that once the TA system complex is formed, the antitoxin exists in a structured state of increased stability and are not targeted by proteases *in vivo* (Song and Wood, 2020). This would mean that proteases are likely responsible for the degradation of free excess antitoxin produced as a result of preferential transcription and translation to neutralise toxicity, otherwise, the protease must act

extremely efficiently and rapidly in tandem with chaperone proteins to sequester antitoxins before they bind with high affinity to their cognate toxin (Song and Wood, 2020; Snead, Moore and Bourne, 2022). The fate of the theoretically neutralised TA complex requires further study as regulatory mechanisms may result in their activation; it has been suggested that processes such as modification, complex remodelling, and association with protease chaperones may target antitoxins for selective degradation (Trentini *et al.*, 2016; Bordes and Genevaux, 2021; Texier *et al.*, 2021).

Type VII systems, of which both components are also proteins, are defined by their post-translational modification mechanism of toxin regulation (Cai *et al.*, 2020); in these cases, activation of the system would be achieved if the antitoxin is incapable of modifying its cognate toxin (by inhibition or degradation), or the modification is removed from the neutralised toxin. Due to this, type VII systems may be tightly controlled, balancing modification and removal of toxin inhibitor moieties for deactivation and activation, respectively. Neither of these have been demonstrated so far. Contrastingly, type II systems can be deactivated by replenishment of the antitoxin molecules or cellular quality control systems (Christensen and Gerdes, 2003; Masuda and Inouye, 2017; Fraikin, Goormaghtigh and van Melderen, 2020). Ribonuclease toxins may simply be sequestered by their cognate antitoxin, whereas directly acting toxins have been shown to be removed from their target; RelE type toxins are ejected from the ribosome by the binding of transfer messenger RNA (tmRNA) molecules (Christensen and Gerdes, 2003), while the CcdA antitoxin can rejuvenate gyrase by removing CcdB from the C-gate (Aghera *et al.*, 2020).

Due to the diversity and abundance of TA systems identified to date it is unlikely there is a single unifying model for the activation of systems. Regulation, including appropriate and controlled activation, of TA systems is imperative to their function *in vivo* and roles in bacterial physiology, notably in an organism like Mtb in which its vast complement of systems have been implicated to play a number of roles during adaptation to environmental stress during infection (Ramage, Connolly and Cox, 2009; Ernst, 2012). Mtb has been one of the most studied organisms in the TA field for this reason.

Table 1.4 RelE/ParE superfamily TA system complex structures summary

Family	Organism	PDB	A:T	Reference
ParDE				
ParDE1	*C. crescentus* NA1000	3KXE	2:2	(Dalton and Crosson, 2010)
ParDE	*P. aeruginosa*	6XRW	2:2	(Snead, Moore and Bourne, 2022)
ParDE2	*V. cholerae* O1 biovar El Tor str. N16961	7R5A	6:2	(Garcia-Rodriguez *et al.,* Unpublished*)
ParDE2	*M. opportunistum* WSM2075	5CEG	4:4	(Aakre *et al.*, 2015)
ParDE1	*M. opportunistum* WSM2075	6X0A	2:2	(Lite *et al.*, 2020)
PaaA2-ParE2	*E. coli* O157	5CZF	1:1	(Sterckx *et al.*, 2016)
ParE SO-CopA - $CopA_{SO}$	*Shewanella oneidensis* MR-1	7ETR	2:2	(Zhou *et al.*, 2021)
RelBE				
RelBE	*M. jannaschii* DSM 2661	3BPQ	2:2	(Francuski and Saenger, 2009)
RelBE	*E. coli* K-12	4FXE	6:6	(Bøggild *et al.*, 2012)
RelBE	*P. horikoshii* OT3	1WMI	2:2	(Takagi *et al.*, 2005)
RelBE2	*M. tuberculosis* H37Rv	3G5O	2:2	(Miallau *et al.*, 2013)
RelBE3	*M. tuberculosis* H37Rv	3OEI	2:2	(Miallau *et al.*, 2013)
DinJ-YafQ	*E. coli* B str. REL606	4ML0	2:2	(Liang *et al.*, 2014)
YoeB-YefM	*S. aureus*	7BWF	2:2	(Eun *et al.*, 2020)
DinJ-YafQ	*E. coli* K-12	4Q2U	2:2	(Maehigashi *et al.*, 2015)
YoeB-YefM	*S. aureus* subsp. Aureus NCTC 8325	6L8F	2:2	(Xue *et al.*, 2020)

Summary of the type II RelE/ParE superfamily TA system complex structures deposited in the PDB. A: Antitoxin; T: Toxin. *Unpublished structures have been deposited and released, awaiting publication.

1.5 Type II TA systems in *M. tuberculosis*

TA systems are evidently highly varied in their sequences, structures, mechanisms of antitoxicity, toxicity, and regulation, and are interestingly present in unusually high numbers in *M. tuberculosis* (Ramage, Connolly and Cox, 2009; Sala, Bordes and Genevaux, 2014).

Multiple genomic studies of Mtb and other species have indicated that the Mtb genome is devoid of type I, III, and V systems with these being more specific to bacterial species outside of the mycobacteria (Fozo *et al.*, 2010; Blower *et al.*, 2012; Sala, Bordes and Genevaux, 2014; Solano-Gutierrez, Pino and Robledo, 2019), albeit the RelBE2 system exists as a hybrid type I/II system coregulated by a toxin specific sRNA (Dawson *et al.,* 2022). Of the 79 TA systems identified by Sala *et al*. (2014) (**Figure 1.8**) and Slayden *et al*. (2018), 68 were type II systems, 66 of which contained cleavage/degradation family toxins; 50 VapC, 10 MazF, and 6 were part of the RelE superfamily comprised of 2 RelE, 1 YoeB (RelE3), and 3 HigB. Only 2 systems of the type II cohort were identified as direct inhibition family toxins, those being the gyrase targeting ParE toxins which are associated with the RelE superfamily (although personal correspondence with the Genevaux lab has implicated RelE1 as a potential direct inhibition family member, also).

The type II systems also demonstrate remarkable genomic stability across lineages of Mtb isolated from TB patients across multiple countries (Solano-Gutierrez, Pino and Robledo, 2019). Of the 45 TA system proteins identified to be completely conserved, 42 were from type II class of systems; many antitoxin sequences were maintained likely for toxin regulation, while endoribonuclease sequences are likely conserved to maintain their catalytic core sequence and structure for substrate specificity (Griffin, Davis and Strobel, 2013). Interestingly, 8 type II systems were completely conserved in both antitoxin and toxin sequence and while 6 of these contained cleavage/degradation family toxins, the two ParDE systems, ParDE1 and ParDE2 (highlighted in **Figure 1.8**), were also unaltered, representing the direct inhibition family (**Table 1.3**). The complete conservation of these systems may indicate important roles in Mtb physiology.

It has become apparent that TA systems, notably the extensive type II complement as the best studied subset in Mtb, link DNA replication, protein synthesis, and cell division to environmental stress encountered during infection such as hypoxia, low pH, nutrient starvation, and antibiotics (Sala, Bordes and Genevaux, 2014; Gupta *et al.*, 2017; Solano-Gutierrez, Pino and Robledo, 2019). The differential expression of TA system toxin and antitoxin proteins of Mtb appears to create a highly responsive regulatory network contributing to early, latent, and disease stages of infection and persistence in the face of antibiotic intervention (Gupta *et al.*, 2017).

Figure 1.8 Chromosomal map of *M. tuberculosis* TA systems

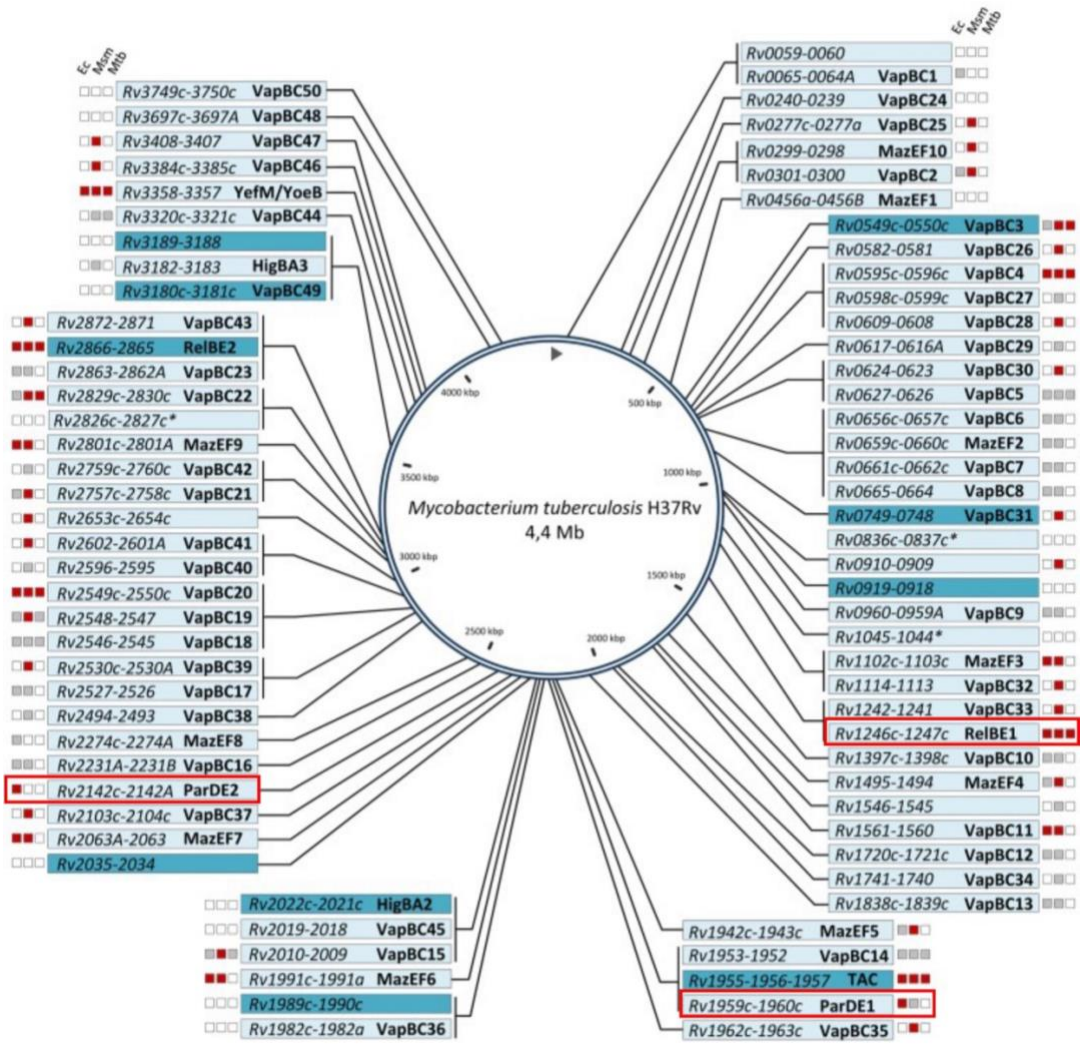


Diagram of the chromosomally encoded TA systems of *M. tuberculosis* H37Rv identified by Sala *et al.* (2014). Systems are labelled appropriately with gene name ('Rv' identifier) with type II systems also labelled by family names. The YefM/YoeB system is also known as RelBE3. For each system, the functionality in *E. coli* (Ec), *M. smegmatis* (Msm), and *M. tuberculosis* (Mtb), is depicted: red color stands for "inhibition of growth", grey for "no inhibition of growth", and white for "not tested". The 10 most induced TA systems in drug-tolerant persister cells are highlighted on dark blue background. Data are correct as of 2014, and where appropriate updated in text. Systems studied in this thesis have been highlighted in red boxes.

1.5.1 Physiological roles of type II TA systems in Mtb

Physiological roles of TA systems span from well characterised and generally accepted mechanisms of plasmid maintenance and phage defence (Gerdes, Rasmussen and Molin, 1986; Fineran *et al.*, 2009), to emerging roles in biofilm formation (Soo and Wood, 2013; Wen, Behiels and Devreese, 2014), to the more debated topics of cell death and theory of persistence (Gurnev *et al.*, 2012; Torrey *et al.*, 2016). Mtb however, encodes a complement of TA systems likely maintained to assist in adaptation to environmental stress, including antibiotic tolerance and persistence, and genomic stability; Mtb lineages have very low levels of TA gene mutation and deletion (Solano-Gutierrez, Pino and Robledo, 2019) and do not regularly acquire extrachromosomal DNA (Zainuddin and Dale, 1990). Additionally, TA systems in general have previously been described as selfish entities, making them reasonably difficult to delete from the chromosome (Magnuson, 2007; Van Melderen and De Bast, 2009). The roles of TA systems in mycobacteriophage defence are understudied in comparison to other organisms, despite the number of Mtb bacteriophages identified (Allué-Guardia *et al.*, 2021).

1.5.1.1 Chromosomal stability

Mtb TA systems can be considered to prevent chromosomal deletions, protecting regions containing genes associated with virulence, potentially by a similar mechanism as in plasmid maintenance whereby antitoxin replenishment is a constitutive requirement for cell survival (Roberts, Ström and Helinski, 1994). Analysis of the Mtb chromosome has previously highlighted a number of genomic islands rich in TA systems of various types (Stinear *et al.*, 2008). Further investigation of these islands has highlighted several genes implicated in Mtb virulence of stress adaptation; the most TA system saturated island (*rv1988 – rv1991c*) encodes *higBA*, *parDE1*, *vapBC35*, *vapBC36*, *mazEF6*, and *vapBC15* alongside the virulence associated *otsB1*, *dosT*, and *fdxA* (Stinear *et al.*, 2008; Ramage, Connolly and Cox, 2009). OtsB1 is involved in trehalose biosynthesis (Murphy *et al.*, 2005), an important sugar in mycolic acid synthesis which also acts as an alternate carbon source in stress (Takayama, Wang and Besra, 2005), protects Mtb against osmotic stress, desiccation and freezing, and has a key role in virulence (Thanna and Sucheck, 2016). DosT is a regulatory two component sensor protein inactivated by oxygen binding (Sousa *et al.*, 2007) and is a key component of the dormancy regulon implicated in hypoxia driven Mtb latency (Rustad *et al.*, 2009; Sivaramakrishnan and de Montellano, 2013). FdxA is also a gas sensor ferredoxin protein upregulated in response to carbon monoxide, nitric oxide, and hypoxia during macrophage infection as part of the DosS/DosT/DosR dormancy regulon (Sousa *et al.*, 2007; Shiloh, Manzanillo and Cox, 2008). TA systems evidently protect genomic regions contributing to its highly successful lifecycle, dominated by a period of dormancy (Wayne, 1994; Ernst, 2012).

Indeed, ParDE systems have previously been implicated in plasmid maintenance systems (Ogura and Hiraga, 1983; Roberts, Ström and Helinski, 1994) .

1.5.1.2 Intracellular survival and disease progression

As part of the infection cycle, Mtb must be capable of surviving and replicating with the cells they infect (typically macrophages) (Schnappinger *et al.*, 2003; Ernst, 2012). A number of VapC toxins (VapC4, VapC5, VapC20, VapC26, and VapC45) have been shown to be important for long-term intracellular survival (Sassetti, Boyd and Rubin, 2003; Rengarajan, Bloom and Rubin, 2005). The HigA1 and HigA2 toxins have been implicated in promoting Mtb survival in acidified phagosomes (Stewart *et al.*, 2005; Jain *et al.*, 2007) and transcripts for the *relE1*, *relE2*, and *relE3* genes can be detected in human macrophages in late stages of Mtb infection, despite their cognate antitoxin transcripts being present during *in vitro* growth (Korch, Contreras and Clark-Curtiss, 2009), indicating selected degradation of antitoxin transcripts. Additionally, the *parE1* gene was identified through mutational studies to be important for survival in the activated macrophage and in the dissemination to spleen (Sassetti, Boyd and Rubin, 2003; Rengarajan, Bloom and Rubin, 2005), potentially contributing to extrapulmonary TB. During the infection of immune cells, Mtb encounters a number of different stresses to which it must respond and adapt and TA systems may have roles in this adaptation and survival process. Intriguingly, while most observations implicate TA systems to be beneficial in the early and latent stages of infection, the MazF3, MazF6, and MazF9 systems have been associated with disease progression with knockouts leading to non-necrotic granuloma tissues hindering the progression to TB disease and transmission (Kaplan *et al.*, 2003; Churchyard *et al.*, 2017). Although, we can infer from the activation of a large number of systems during stress that they may become deactivated once a more suitable niche presents itself, contributing to replication and transmission as toxicity subsides and obstacles to growth are removed.

1.5.1.3 Response to host associated environmental stress during infection

Several TA systems across multiple studies have been implicated in the response of Mtb to environmental stresses associated with inhabiting phagocytes (Ramage, Connolly and Cox, 2009; Torrey *et al.*, 2016; Gupta *et al.*, 2017). Hypoxia is suspected to be one of the main drivers of Mtb into the dormant state throughout latent infection (Rustad *et al.*, 2009). Unsurprisingly, several VapBC and MazEF systems are transcriptionally activated by hypoxia, potentially as part of the dormancy regulon (Commandeur *et al.*, 2013; Hudock *et al.*, 2017). These systems, and other VapBC and MazEF systems, are also regulated in response to IFN-γ stimulated macrophages, oxidative stress, and low pH, all encountered during infection (Schnappinger *et al.*, 2003; Hudock *et al.*, 2017). An *in vitro* model of fatty acid and dextrose induced hypoxia, attempting to mimic the intracellular environment of the macrophage, suggested the ParDE1 system is important in

Mtb adaptation (del Portillo *et al.*, 2019); interestingly the *M. smegmatis* ParDE2 system has been implicated in a similar role (Oren and Garrity, 2020).

Transcriptome remodelling is believed to be a key feature of Mtb in existing within the host for prolonged periods of time, responding to a wide variety of ever-changing challenges. Mtb utilises its plethora of responsive endoribonuclease toxins to constantly modify its expression profile, increasing its adaptability. For example, MazF toxins have been shown to have variable preferences for the Mtb transcriptome, ranging from 20 % to 70 % allowing the bacterium to regulate a range of processes using TA systems (Zhu *et al.*, 2008; Sala, Bordes and Genevaux, 2014; Masuda and Inouye, 2017). The utilisation of alternate carbon sources, and regulated carbon transport and metabolism, is essential to Mtb survival in macrophages (Borah *et al.*, 2021; Shan Chang and Guan, 2021); several VapC toxins have been shown to be upregulated in nutrient starvation models (Albrethsen *et al.*, 2013) and to post-transcriptionally regulate the expression of genes in sugar transport and glycerol metabolism (McKenzie *et al.*, 2012).

Nutrient starvation, as a general stress encountered by Mtb infecting macrophages, has a general up-regulatory effect on numerous TA systems (Gupta *et al.*, 2017). Highest levels of induction were observed in the three HigBA systems, three of the MazEF systems, all three RelBE systems, and eighteen VapBC systems in conditions mimicking starvation (Gupta *et al.*, 2017). This likely has strong links with modifying the Mtb transcriptome for adaptation, however, the ParDE1 system was also highly induced (Gupta *et al.*, 2017); given this system targets gyrase and ParE toxins induce SOS responses (Jiang *et al.*, 2002; Ames *et al.*, 2019); it may be that gyrase inhibition slows transcription and DNA replication to assist the endoribonuclease toxins or increase mutation rates via SOS to assist in adaptability. Interestingly, a number of TA systems are responsive to SOS (Vogel *et al.*, 2004; Kawano, Aravind and Storz, 2007; Singletary *et al.*, 2009), indicating a potential role for ParE toxins within TA networks. Furthermore, in studies focussing on individual systems the MazF3 toxin is up-regulated in response to amino acid starvation, while the MazF2 toxin has been shown to be down-regulated in nutrient starvation also (Betts *et al.*, 2002; Dahl *et al.*, 2003) indicating a dynamic post-transcriptional control network involving TA systems in response to several conditions.

1.5.1.4 Response to antibiotic pressure – tolerance and persister cells

Beyond host-associated stresses as part of the immune response to Mtb infection (Ernst, 2012) antibiotic pressure has also been shown to elicit several TA system responses (Keren *et al.*, 2011; Gupta *et al.*, 2017). Mtb is remarkably tolerant to drug treatment, as highlighted previously, requiring lengthy multidrug treatment regimens inhibiting multiple processes for any chance of a cure (WHO 2022b). It is believed this is due to the dormancy of Mtb during latent infection (Wayne, 1994; Ernst, 2012), with cellular processes targeted by drugs generally downregulated,

but also due to the formation of persister cells, a sub-population of phenotypically drug-resistant cells (Keren *et al.*, 2011; Wen, Behiels and Devreese, 2014; Torrey *et al.*, 2016).

The formation of these cells in response to environmental and antibiotic pressure without undergoing genetic changes creates an insurance policy for when more favourable conditions arise (Wood, Knabel and Kwan, 2013); such as increasing oxygen concentration in later stages of TB disease progression as granulomas necrotise (Ramakrishnan, 2012). The exact mechanisms of persister cell generation are not fully understood and hotly debated in the TA field; TA systems have been heavily linked to both the entry to and exit from the persister state as responsive elements, with the deletion of one or multiple systems drastically impacting their frequency of generation and the progression of TB disease (Kim and Wood, 2016; Torrey *et al.*, 2016). It is possible that both entry and exit are stochastic processes, highlighted by the variety of TA combinations reported to influence drug-resistance (Betts *et al.*, 2002; Keren *et al.*, 2011; Page and Peti, 2016; Torrey *et al.*, 2016; Hudock *et al.*, 2017); essentially, once an optimal combination of activated and deactivated systems is achieved, cells may simply become persistent. Mtb is remarkably responsive to its environment encoding the stress-sensing dormancy regulon (Sharma and Tyagi, 2016), and a complement of toxins capable of inhibiting numerous cellular processes and remodelling transcription (Sala, Bordes and Genevaux, 2014; Page and Peti, 2016). While the exact roles of each TA system have not been elucidated, it is likely that the inhibition of multiple and overlapping cellular processes contributes to effective persister cell formation (Wood, Knabel and Kwan, 2013; Page and Peti, 2016). Ironically, this in many ways mimics the effects of a multidrug regimen, however Mtb somehow regulates this to permit long-term survival and subsequent progression back to a proliferative state (Ernst, 2012).

Drug-tolerant persister cells have been associated with the upregulation of endoribonuclease toxins VapC3, VapC31, VapC49, MazF1, MazF5, MazF6, and all three RelE toxins (Keren *et al.*, 2011). Deleting the *mazF* genes has been shown to produce a diminished population of Mtb persister cells (Singh, Barry and Boshoff, 2010). In response to antibiotic treatment, studies have revealed that all three *relE* toxins are upregulated, and *relBE2* is one of the highest induced TA systems in Mtb drug-tolerant persisters (Singh, Barry and Boshoff, 2010).

Mtb has adapted responses to exposure from first-line treatment regimen drugs; rifampicin was shown to upregulate 26 TA genes, and downregulate 22, with the highest upregulation after 24 hours of exposure seen in 8 VapBC systems, HigBA3, MazEF1, MazEF6, MazEF9, RelBE1 and RelBE2, and ParDE2 (Gupta *et al.*, 2017). Similar to the association of ParDE1 with endoribonuclease toxins in starvation response, ParDE2 may be activated to temporarily inhibit transcription to allow for a reset in the transcription profile. The ParDE2 system was also shown to be activated under isoniazid stress, with the VapBC13 system uniquely activated by the

combination of isoniazid and ethambutol (Gupta *et al.*, 2017). Interestingly, expression profiling highlighted the ParE toxins, *parE1* and *parE2*, as some of the highest differentially regulated genes (second only to *rv1045,* a type VII system MenT toxin (Cai *et al.*, 2020)) when Mtb was subjected to starvation, acidification, and first-line drug exposure in combinations for varying time-lengths (Gupta *et al.*, 2017). Taken together, if the TA system responses *in vivo* are similar, this indicates a differential expression pattern favouring toxicity and subsequent growth inhibition when Mtb is exposed to the four first-line drugs used in TB treatment.

A great deal of the systems regulated throughout the TA life cycle appear to be the endoribonuclease class (by our definition, family A toxins which cleave/degrade their targets). This is not surprising given that out of the minimum 68 type II systems, at least 66 are endoribonuclease class toxins (Sala, Bordes and Genevaux, 2014; Slayden, Dawson and Cummings, 2018). Interestingly, the two ParDE systems, ParDE1 and ParDE2, have been shown to be highly regulated in response to fatty acid hypoxia, starvation, and antibiotic pressures (Sassetti, Boyd and Rubin, 2003; Gupta *et al.*, 2017), potentially complementing the activities of simultaneously activated endoribonuclease toxins. The ParDE1 system may also maintain an important virulence region and contributing to survival in phagocytic cells (Sassetti, Boyd and Rubin, 2003; Rengarajan, Bloom and Rubin, 2005).

It is worth noting that across the studies examining the response of Mtb to environmental stresses, both *in vitro* and *in vivo*, while there are overlaps in observations there is no true consensus (Sassetti, Boyd and Rubin, 2003; Schnappinger *et al.*, 2003; Rengarajan, Bloom and Rubin, 2005; Ramage, Connolly and Cox, 2009; Keren *et al.*, 2011; Gupta *et al.*, 2017). This indicates high variability in the activation of TA systems, which may be due to plasticity or highly sensitive mechanisms of activation that are not replicated by, and indeed between, *in vitro* studies. This may also be indicative of a stochastic response by Mtb, randomly up- and downregulating genes until an ideal, 'goldilocks' combination is achieved; while this may also explain phenomena like random entry and exit from persister states, it seems an exhausting and unlikely theory.

1.6 *M. tuberculosis* ParDE systems as targets for study

Despite their widespread and highly conserved nature (Anantharaman and Aravind, 2003), ParDE systems have received little attention in the TA field, likely due to the abundance of endoribonuclease type toxins in the type II class and their associations with pathogenicity (Ramage, Connolly and Cox, 2009). The Mtb ParDE systems have clear roles in the physiology and infection cycle of Mtb as some of the highest conserved and regulated gene sequences (Gupta *et al.*, 2017; del Portillo *et al.*, 2019; Solano-Gutierrez, Pino and Robledo, 2019), likely contributing to TB disease.

Most interestingly, the target of these systems is greatly different from the RelE structural superfamily of which they are a part; ParE toxins target DNA gyrase, however, in contrast to the CcdB toxins, their exact mechanism has not yet been elucidated (Dao-Thi *et al.*, 2005). Several studies on the Mtb ParE1 and ParE2 toxins have shown their abilities to inhibit gyrase enzymes from *E. coli*, *M. smegmatis*, and Mtb (Ramage, Connolly and Cox, 2009; Gupta *et al.*, 2016; Ames *et al.*, 2019). Interestingly, The ParE1 toxin was not toxic to *M. smegmatis* and was shown to not inhibit the *M. smegmatis* gyrase (Ramage, Connolly and Cox, 2009). Due to the high degree of conservation in the mycobacterial gyrase sequence this may indicate that ParE1 has a novel target. Despite this, ParE1 was shown to be potently toxic to *E. coli*, triggering a loss of viable cells with filamentous cell morphology and marked increases in intracellular *oriC* sequences (Ames *et al.*, 2019). In contrast, these results indicate that ParE1 targets DNA replication machinery, most likely gyrase.

Due to the structural similarities in the ParE family we would expect that the ParE2 toxin would not be toxic to *M. smegmatis* through gyrase inhibition, however this was not the case and ParE2 was shown to inhibit *M. smegmatis* growth and was capable of trapping Mtb gyrase cleavage complexes (Gupta *et al.*, 2016). This indicates that the ParE toxins may be distinct in their folds or there may be specificity for amino acid motifs between ParE and gyrase species for interaction. The ability of ParE2 to trap cleavage complexes, like the toxic effects of FQ inhibitors, alongside the lack of structural relationship with CcdB toxins (Blower, Salmond and Luisi, 2011), indicates a novel mechanism of gyrase inhibition. Given the current state of the development of NBTI research, we believe that ParE toxins require further research, classifying their mechanisms of toxicity via biochemical, biophysical, and structural analyses.

Previous work in the TA and topoisomerase fields has set precedent for these studies; Gupta *et al.* (2016) previously showed that ParE2 could inhibit the supercoiling reaction of Mtb gyrase *in vitro*, and this toxicity could not be rescued by the cognate antitoxin ParD2, contrasting the rescue of CcdB poisoning gyrase by CcdA (Aghera *et al.*, 2020). The structural relationship

between the CcdB toxin and the GyrA C-gate has previously been demonstrated (Dao-Thi *et al.*, 2005), subsequently informing the design of gyrase-inhibiting CcdB peptides (Trovatti *et al.*, 2008), and CcdBA TA systems have been exploited in biotechnology (Stieber, Gabant and Szpirer, 2008). The structurally homologous RelE toxins have been crystallised in the ribosome A-site, providing structures contributing to the elucidation of their ribosome-dependent mRNA cleavage mechanism (Neubauer *et al.*, 2009). Together, these structural studies highlight that interesting insights can be gained from toxin-target structures; ParE toxins lie at the intersection between RelE and CcdB toxins and would benefit from similar structural investigation. Gyrase enzymes have proven difficult to crystallise in their fully reconstituted form, however the recent full-length Cryo-EM structure of the *E. coli* gyrase (**Figure 1.4**) (Vanden Broeck *et al.*, 2019) provides an excellent example for the possibilities of Cryo-EM toxin-gyrase complex structures. Recent gyrase structures bound to inhibitory compounds have also advanced our understanding of gate-dynamics in the gyrase reaction, with allosteric sites potentially governing efficient strand cleavage and re-ligation (Chan *et al.*, 2017).

Altogether, these studies indicate that ParE toxins may advance our understanding of gyrase gate dynamics by revealing novel mechanisms of inhibition, could contribute to biotechnology applications, and even inform downstream drug design for novel inhibitors of Mtb gyrase in the face of emerging FQ resistance.

1.7 Research aims

Mycobacterium tuberculosis remains a threat to global health, especially considering the emerging patterns of resistance to first and second-line therapeutics (Gygli *et al.*, 2017; Walker *et al.*, 2022); especially concerning is the rise in resistance to gyrase-targeting fluoroquinolone antibiotics (Maruri *et al.*, 2021), a central drug class to treatment regimens for drug-resistant TB (WHO 2022b). The identification of novel inhibitors of topoisomerases in general is not currently a fruitful field (Ganapathy et al., 2021), however there have been recent and exciting discoveries specific to Mtb gyrase (Gibson *et al.*, 2018; Imai *et al.*, 2022); naturally occurring inhibitors may not only better our understanding of this essential class of enzyme, but they may also provide a novel basis for drug discovery.

This project initially focussed on the gyrase targeting ParDE toxin-antitoxin systems of *M. tuberculosis* (**Table 1.5**); both systems are highly conserved across Mtb lineages and regulated in response to environmental stresses associated with infection (Gupta *et al.*, 2017; del Portillo *et al.*, 2019; Solano-Gutierrez, Pino and Robledo, 2019). Their apparent lack of essentiality for growth *in vitro* is perplexing (Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017) (**Table 1.5**), although this may indicate roles highly specific to the ability of Mtb to adapt *in vivo*. Historically, ParDE systems have been associated with genomic stability (Ogura and Hiraga, 1983) and this may also be the case for those encoded by Mtb, protecting regions essential for virulence from large deletions (Stinear *et al.*, 2008). We therefore highlight both systems as important in the life cycle of a major cause of morbidity and mortality worldwide.

We aim to demonstrate the molecular mechanisms underpinning the inhibition of *M. tuberculosis* gyrase by the ParE1 and ParE2 toxins using a combination of biochemical, biophysical, and structural studies (**Chapters 3 – 5**). Biochemical studies aim to demonstrate the effects of the ParE toxins on the activity of Mtb gyrase *in vitro* (**Chapter 3**); we expect that, given previous results (Gupta *et al.*, 2016), both ParE toxins will trap cleavage complexes to generate dsDNA breaks. Using a range of Mtb gyrase constructs we plan to highlight essential domains and motifs for the interaction of ParE toxins with DNA gyrase. Analytical size exclusion chromatography (SEC) will be used to estimate the state of gyrase and TA system proteins in solution and will assist in the confirmation of structural models for proteins of interest (**Chapter 3**). Crystallographic studies of the ParDE1 and ParDE2 system complexes (**Chapter 4**) will be important to develop our understanding of mechanisms of toxin regulation in type II TA systems and build on the repertoire of RelE/ParE structures in the PDB (**Table 1.4**). These studies will provide the basis for interaction studies between gyrase and ParE toxins via both analytical SEC and likely future structural work through crystallography and cryoEM. We aim to potential to

identify novel mechanisms of gyrase inhibition to inform future drug-design, and potentially further our exploitation of TA systems in biotechnology and as potential therapeutic targets.

Phylogenetic analyses of the ParE toxin family (**Chapter 5**) will assist in understanding potential differences between the Mtb ParE1 and ParE2 toxins and these studies will be expanded to include the RelE toxin family as part of an international collaboration with the Genevaux group (Toulouse, France). The structurally related RelE1 toxin of Mtb is apparently not active in ribosome-dependent mRNA cleavage (personal correspondence) but is regulated during infection and toxic in Mtb, *M. smegmatis*, and *E. coli* (Sala, Bordes and Genevaux, 2014) indicating a conserved target. Developing these sequence-based analyses into structural studies of the RelBE1 TA system complex will allow us to highlight possible differences in the RelE1 catalytic residues or tertiary structure (**Chapter 5**). This will inform future studies of the RelE1 toxin to explore possible altered targets or toxicity mechanism from cleavage to direct, or even requirement for activation via modification. Altogether, solving the ParDE1 ParDE2, and RelBE1 crystal structures will complete the structural characterisation of Mtb RelE/ParE superfamily TA systems (**Chapters 4 and 5**) (**Table 1.4**) and hopefully provide insights into toxin functions and mechanisms of regulation.

Table 1.5 *M. tuberculosis* Toxin-Antitoxin systems studied in this thesis

Gene Summary			Protein information				
Name	Locus	Length (bp)	Length (aa)	M_r (kDa)	pI	Essentiality *in vitro*	Reference
parE1	Rv1959c	297	98	11.27	6.94	Non-essential/ growth advantage	(Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017)
parD1	Rv1960c	252	83	9.21	7.54	Non-essential	(Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017)
parE2	Rv2142c	318	105	12.31	6.52	Growth advantage	(Dejesus *et al.*, 2017)
parD2	Rv2142A	216	71	7.88	4.41	Non-essential/ Growth advantage	(Dejesus *et al.*, 2017)
relE1	Rv1246c	294	97	11.03	9.66	Non-essential	(Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017)
relB1	Rv1247c	270	89	9.77	4.86	Non-essential	(Sassetti, Boyd and Rubin, 2003; Dejesus *et al.*, 2017)

Information is relevant to the *M. tuberculosis* H37Rv chromosome. Gene and amino acid lengths include start codons. Essentiality is based on *in vitro* transposon mutagenesis studies, disrupting the genes presented. bp – base pair; kDa – kilodalton; M_r – molecular weight; pI – isoelectric point.

Chapter 2. Materials & Methods

2.1 Media, reagents, and solutions

All media, antibiotics and other supplements, and solutions used in this study are detailed in **Tables 2.1, 2.2** and **2.3**, respectively. Where appropriate these were sterilised via autoclaving at 121 °C for 20 minutes or filtration via 0.22 µm filter.

2.2 Bacterial strains and culture

Escherichia coli strains used in this study are listed in **Table 2.4**. *E. coli* were grown at 37 °C in liquid culture shaken at 180 rpm, or on agar plates. Culture temperature and shaking varied only in protein expression protocols as stated later. Growth was monitored using a cell density meter (WPA Biowave C08000) to give OD_{600}.

2.2.1 Bacterial transformation

Appropriate chemically competent *E. coli* strains were transformed by heat-shock. Bacterial cells were incubated on ice with 50 – 100 ng plasmid DNA for 30 minutes. Cells were incubated at 42 °C for 45 seconds then returned to ice for 2 minutes prior to the addition of 1 mL LB broth (**Table 2.1**). Cells were incubated at 37 °C for 1 hour before pelleting by centrifugation, resuspension in 100 µL LB broth, and plating on LBA (**Table 2.1**) containing the appropriate antibiotic selection. Bacteria were grown overnight at 37 °C.

2.3 Recombinant DNA techniques

Molecular biology techniques involving DNA were performed by standard methods (Berger and Kimmel, 1987). Plasmids used in this study can be found in **Table 2.5**. All oligonucleotide primers were obtained from Integrated DNA Technologies (IDT) and can be found in Table 2.6.

2.3.1 Small-scale plasmid DNA purification - MiniPrep

Plasmid DNA was purified from transformed DH5α cells using a NEB Monarch® Plasmid MiniPrep kit following the manufacturer's instructions. Plasmids were eluted in dH_2O for storage at -20 °C.

2.3.2 Medium-scale plasmid DNA purification – MaxiPrep

Negatively supercoiled plasmid (pSG483) DNA was purified from transformed DH5α cells using as a Machery-Nagel NucleoBond Xtra Midi Plus EF kit following the manufacturer's instructions. Plasmids were eluted in dH_2O for storage at -20 °C.

Table 2.1 Media used in this study

Medium	Ingredients per litre
Luria-Broth	10 g Casein Digest Peptone 5 g Yeast Extract 10 g NaCl
Luria-Broth Agar	10 g Casein Digest Peptone 5 g Yeast Extract 10 g NaCl 12 g Agar
2 x YT Broth (Nutrient rich broth)	16 g Casein Digest Peptone 10 g Yeast Extract 5 g NaCl

Table 2.2 Antibiotics and supplements used in this study

Chemical (abbreviation)	Stock solution	Working concentration
Antibiotic		
Ampicillin (Ap)	1000 x stock, 100 mg/mL in dH_2O, stored at -20 °C	100 µg/mL
Chloramphenicol (Cm)	1000 x stock, 25 mg/mL in EtOH, stored at -20 °C	25 µg/mL
Kanamycin (Km)	1000 x stock, 50 mg/mL in dH_2O, stored at -20 °C	50 µg/mL
Moxifloxacin (MXF)	5 mM in DMSO, stored at -20 °C	Variable
Supplement		
Isopropyl β – D – thiogalactopyranoside (IPTG)	1000 x stock, 1 M in dH_2O, stored at -20 °C	1 mM (unless otherwise stated)

Table 2.3 Solutions used in this study

Solution	Components
DNA work	
50 x TAE Buffer (per L)	242 g Tris base 57.1 mL Glacial acetic acid (17.4 M) 100 mL EDTA pH 8.0 (0.5 M)
Agarose gel mix	1 – 2 % Agarose in 1 x TAE 0.5 µg/mL gel ethidium bromide (EtBr) (if appropriate)
1D SDS-PAGE	
10 x Stock electrode buffer (per L)	30.2 g Tris base 141 g glycine 10 g SDS pH 8.3
Resolving acrylamide gel (variable %) (enough for four 10 cm BioRad gels)	X mL 40 % Acrylamide/Bis-Acrylamide (w/v) Ratio 19:1 12 mL 0.75 M Tris base pH 8.8 240 µL 10 % SDS 240 µL 10 % APS 24 µL TEMED 11.5 – X mL dH_2O
6% (stacking) acrylamide gel (enough for four 10 cm BioRad gels)	1.5 mL 40 % Acrylamide/Bis-Acrylamide (w/v) Ratio 19:1 1.5 mL 1.25 M Tris base pH 6.8 150 µL 10 % SDS 150 µL 10 % APS 15 µL TEMED 11.7 mL dH_2O
4 x SDS-PAGE loading dye (10 mL stock)	1 mL 1 M Tris base pH 6.8 0.8 g SDS 4 mL 100 % glycerol 0.4 mL 14.7 M β-mercaptoethanol 1 mL 0.5 M EDTA pH 8.0 8 mg bromophenol blue To 10 mL with dH_2O

Table 2.3 Solutions used in this study

Solution	Components
Protein purification and crystallisation	
A500 (lysis buffer)	20 mM Tris base pH 8.0 500 mM NaCl 30 mM Imidazole 10 % (vol/vol) glycerol
A800 (lysis buffer)	20 mM Tris base pH 8.0 800 mM NaCl 30 mM Imidazole 10 % (vol/vol) glycerol
B500 (Ni-NTA elution buffer)	20 mM Tris base pH 8.0 500 mM NaCl 250 mM Imidazole 10 % (vol/vol) glycerol
B100 (Low-salt Ni-NTA elution buffer)	20 mM Tris base pH 8.0 100 mM NaCl 250 mM Imidazole 10 % (vol/vol) glycerol
A100 (FPLC low salt buffer)	20 mM Tris base pH 8.0 100 mM NaCl 10 % (vol/vol) glycerol
C1000 (FPLC high salt buffer)	20 mM Tris base pH 8.0 1000 mM NaCl 10 % (vol/vol) glycerol
S500 (FPLC SEC buffer)	50 mM Tris base pH 8.0 500 mM KCl 10 % (vol/vol) glycerol
S300-A (FPLC analytical SEC buffer)	20 mM Tris base pH 8.0 300 mM NaCl
Storage buffer (Protein sample storage)	50 mM Tris base pH 8.0 500 mM KCl 70 % (vol/vol) glycerol
X (Crystallisation buffer)	20 mM Tris base pH 8.0 150 mM NaCl 2.5 mM DTT
X-5 (RelBE crystallisation buffer)	20 mM Tris base pH 8.0 150 mM NaCl 2.5 mM DTT 5 % (vol/vol) glycerol

Table 2.3 Solutions used in this study

Solution	Components
Topoisomerase assay buffers	
Gyrase dilution buffer	50 mM Tris base pH 8.0 500 mM KOAc 2 mM MgOAc 50 µg/mL BSA 1 mM DTT 10 % (vol/vol) glycerol
4 x gyrase reaction buffer	40 mM Tris base pH 8.0 38.4 mM MgOAc 100 µg/mL BSA 4 mM DTT 32 % (vol/vol) glycerol
Stopping buffer	5 % (wt/vol) SDS 125 mM EDTA pH 8.0

2.3.3 Preparation of nicked and linear form pSG483

For nicking, 10 µg pSG483 (purified as per 2.3.2) was incubated with 10 units of Nb.Bpu10I (ThermoFisher) in 1 x Buffer R (ThermoFisher) for 1 hr at 37 °C. The enzyme was deactivated by a further incubation step at 80 °C for 20 minutes For linearisation 10 µg pSG483 (purified as per 2.3.2) was incubated with 10 units of BamHI-HF® (NEB) in 1 x rCutSmart (NEB) for 1 hour at 37 °C. The enzyme was deactivated after incubation by a further incubation step at 65 °C for 10 minutes. Completion of both reactions, that is, conversion of all supercoiled pSG483 into the appropriate products, was assessed by agarose gel electrophoresis. Both nicked and linear form pSG483 were subsequently store at -20 °C.

2.3.4 Preparation of relaxed form pSG483

Initially, 50 µg pSG483 was nicked by incubation with 10 units Nb.Bpu10I (ThermoFisher) in 1 x Buffer R (ThermoFisher) for 4 hours at 37 °C. The enzyme was deactivated by a further incubation step at 80 °C for 20 minutes. The reaction was allowed to cool to room temperature before being supplemented with ATP to a final concentration of 1 mM. 10 µL T4 DNA ligase was added and the reaction was left at room temperature for 16 hours.

After ligation, ethanol precipitation was performed to remove proteins. An equal volume of UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, vol/vol) (ThermoFisher) was added to the reaction mixture before vortexing briefly. The sample was centrifuged at 16,000 x g for 2 minutes and the resulting aqueous layer was removed and carried forward. An equal volume of chloroform (ThermoFisher) was added to the aqueous layer before centrifugation at 16,000 x g for 2 minutes. The resulting aqueous layer was carried forward and 1/10 volume 3 M sodium acetate pH 5.2 was added. Then, 2 volumes of 100 % ethanol was added, briefly mixed by pipetting, and stored at -80 °C for 30 minutes. The sample was centrifuged at 16,000 x g and 4 °C for 20 minutes. The aqueous layer of ethanol was removed, and the DNA pellet dried at room temperature. The DNA pellet was resuspended in room temperature dH_2O to approximately 300 ng/µL.

2.3.5 Agarose gel electrophoresis and DNA extraction

Agarose gel preparation is described in **Table 2.3**. A 6x DNA loading dye (New England Biolabs) was added to the DNA sample in the appropriate volume prior to loading into the agarose gel. The DNA molecules were separated by electrophoresis at 120 V until necessary resolution was achieved. Molecular weights were compared to a 1 kb ruler (ThermoFisher) and bands at the appropriate molecular weight cut out for DNA extraction using a NEB Monarch® Gel Extraction kit, following the manufacturer's instructions. DNA was eluted in dH_2O for storage at -20 °C.

Table 2.4 Bacterial strains used in this study

Strain	Genotype	Source
Escherichia coli		
DH5α	*F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1*	Invitrogen
ER2566 (Expression strain)	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
Rosetta™ 2 (DE3) pLysS (Expression strain)	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysSRARE2 (CamR)*	Novagen

2.3.6 DNA visualisation

Agarose gels containing ethidium bromide (EtBr) required no further processing until imaging. Gels without EtBr (see later, Topoisomerase assays), were post-stained in TAE + 0.5 µg/mL EtBr for 30 minutes, before rinsing and de-staining in TAE for 1 hour.

Gels were imaged on a BioRad ChemiDoc™ XRS+ with ImageLab™ software on the ethidium bromide setting (BioRad).

2.3.7 Polymerase chain reaction (PCR)

Q5 DNA polymerase was used for DNA amplification via PCR according to the components and steps outlined below. Each component can be scaled to alter the total reaction volume and the annealing temperature was dependent on specific primer pairs.

Q5 DNA polymerase PCR

Component	Volume (µL)	Step	Temp (^{o}C)	Time
10 x Q5 Buffer	5	1) Initial denaturation	95	30 sec
10 mM dNTP's	1	2) Denaturation	95	10 sec
10 µM Primer 1	2.5	3) Annealing	Varies*	30 sec
10 µM Primer 2	2.5	4) Extension	72	30 sec
DNA sample (template)	1	Repeat 2 – 4 x 40 times		
Q5 polymerase	0.5	5) Final extension	72	2 minutes
dH_2O	37.5	6) Hold	10	∞

2.3.8 Cloning

Ligation independent cloning (LIC) (Aslanidis and de Jong, 1990) was performed in order to generate expression constructs from which the genes of interest can be selectively expressed. *M. tuberculosis* H37Rv antitoxin genes *parD1* (*rv1960c*) and *parD2* (*rv2142A*) were cloned into pSAT1-LIC, fused to a 6His-hSUMO-2 gene under control of the T7 promoter.

2.3.8.1 Vector digest

1 µg of pSAT1-LIC was digested with StuI producing a linearised plasmid with exposed blunt ended LIC sites. The linearised DNA was run on agarose gel and extracted as detailed in 2.3.5. Below is a summary of the digestion reaction.

pSAT1-LIC digestion with *Stu*I

Component	Volume (µL)	Temperature	Time
pSAT1-LIC	Variable (1 µg)	37 ^{o}C	2 hr
10 x Buffer 2.1 (NEB)	2.5		
Stu1 (NEB)	1		
dH_2O	to 25 µL		

2.3.8.2 Insert amplification

A LIC site flanked-*gene-of-interest* was amplified via the Q5 polymerase chain reaction (PCR, 2.3.5) using the method and components described previously with template *M. tuberculosis* H37Rv genome. Primers were designed using the respective antitoxin gene sequences with the addition of sequences complementary to the LIC sites of pSAT1-LIC and can be found in **Table 2.6**. Amplified inserts were then run on an agarose gel and purified as detailed in 2.3.5.

2.3.8.3 LIC reaction

Purified linearised vector and amplified insert, both with LIC sites, were carried forward to the LIC reaction to expose complementary 'sticky ends' on both vector and insert. The reaction for each differs slightly and is detailed below. The products of the LIC reaction were combines in a ratio of 1:1 as 40 % of the reaction volume and annealed overnight at room temperature. *E. coli* DH5α cells were transformed with the annealed product as per 2.2.1.

LIC reaction

Vector		Insert		Thermocycle	
Component	Volume (µL)	Component	Volume (µL)	Temperature	Time
Vector	25	Insert	10	1) 22 ^{o}C	30 minutes
dTTP (25 mM)	5	dATP (25 mM)		2) 75 ^{o}C	20 minutes
10 x 2.1 Buffer (NEB)	5	10 x 2.1 Buffer (NEB)	2		
DTT (100 mM)	2.5	DTT (100 mM)	1		
T4 DNA Polymerase	1	T4 DNA Polymerase	0.4		
(NEB)		(NEB)			
dH_2O	11.5	dH_2O	4.6		
	50		20		

2.3.9 Sequencing and sequence analysis

Extracted plasmids, and those outsourced, containing genes of interest were sequenced in-house by DBS Genomics, Durham University Biosciences Department, using the ABI 3730 DNA sequencer via primer walking (primers listed in **Table 2.6**). Confirmation of gene sequences was performed using 4Peaks and BLASTn sequence alignment (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.4 Large scale protein expression

Proteins were expressed and purified following published protocols (Blower *et al.*, 2016), with small variations as appropriate. Transformed *E. coli* strains were grown using pre-defined conditions in nutrient rich 2 x YT media (**Table 2.1**) supplemented with appropriate antibiotic selection to an optical density (OD_{600}) of 0.6 – 0.8. The cultures were cooled to the appropriate temperature for expression and IPTG (**Table 2.2**) was added to a final, appropriate concentration to induce overexpression of the target protein. Cultures were shaken at 160 rpm for between 4 - 16 hours to maximise expression.

2.4.1 Expression of gyrase protein subunits and fusions

For the expression of the gyrase subunit protein, GyrA and GyrB, Rosetta™ 2 pLysS cells were transformed with pTRB696 and pTRB316, respectively. Both gyrase subunits were expressed with a TEV protease cleavable N-terminal hexahistidine (6His) tag for purification. For the expression of the full-length gyrase fusion, GyrBA, Rosetta™ 2 pLysS were transformed with pTRB697 (Kindly supplied by the Maxwell group). GyrBA is expressed with a TEV-protease cleavable N-terminal 6His tag (6His-TEV site) for purification. For the expression of the truncated gyrase fusion proteins, $GyrBA^{56}$ and $GyrB^{28}A^{56}$, ER2566 cells were transformed with pTRB642 and pTRB643, respectively. Both truncated fusion proteins were expressed with a human Sentrin Protease (hSENP-2) cleavable N-terminal 6His-SUMO tag for purification. All strains and plasmids can be found in **Table 2.4** and **Table 2.5**, respectively. Cells were grown at 37 °C with shaking at 180 rpm to an optical density (OD_{600}) of 0.6 at which point the incubation temperature was reduced to 30 °C and IPTG was added to a final concentration of 0.8 mM to induce overexpression. Cells were subsequently grown for a further 4 hours at 30 °C with shaking at 160 rpm.

2.4.2 Expression of toxin-antitoxin system complexes

Each of the toxin-antitoxin systems in this study were expressed from Duet vectors which can be found in **Table 2.5**. Rosetta™ 2 pLysS cells were transformed with the appropriate plasmids for the expression of the ParDE1 complex (pTRB569), ParDE2 complex (pTRB570), and the RelBE1 complex (pTRB638). Toxins, ParE1, ParE2, and RelE1, were expressed with a hSENP-2 cleavable N-terminal 6His-SUMO tag for purification of the complexes. Cells were grown at 37 °C with

shaking at 180 rpm to an optical density (OD_{600}) of 0.6 at which point the incubation temperature was reduced to 18 °C and IPTG was added to a final concentration of 1 mM to induce overexpression. Cells were subsequently grown for a further 16 hours (overnight) at 18 °C with shaking at 160 rpm.

2.4.2.1 Expression of heterotetrameric ParDE1

To express the ParDE1 complex in the heterotetramer stoichiometry, the protocol as per 2.4.2 is followed with the following adjustments: IPTG is added to a final concentration of 0.5 mM and expression temperature is lowered to 16 °C.

Table 2.5 Plasmids used in this study

Plasmid	Backbone plasmid	Resistance	Inserted gene / region	Origin organism / notes	Source
pTRB312	1B	Km	*rv0005 (gyrB)*	*M. tuberculosis,* gyrase subunit B	*Blower Lab*
pTRB696	1B	Km	*rv0006 (gyrA)*	*M. tuberculosis,* gyrase subunit A	*Blower Lab*
pTRB697	pET28-MHL	Km	*$gyrB^{2-675}A^{2-838}$ full-length fusion*	*M. tuberculosis,* gyrase fusion	*Maxwell Lab*
pTRB642	pET-Duet1	Ap	*$gyrB^{2-675}A^{2-500}$ fusion*	*M. tuberculosis,* gyrase fusion	Genscript *(This study)*
pTRB643	pET-Duet1	Ap	*$gyrB^{2-426}A^{2-500}$ fusion*	*M. tuberculosis,* gyrase fusion	Genscript *(This study)*
pTRB567	pET-Duet1	Ap	*rv1960c (parD1)*	*M. tuberculosis,* antitoxin	Genscript *(This study)*
pTRB568	pET-Duet1	Ap	*rv2142A (parD2)*	*M. tuberculosis,* antitoxin	Genscript *(This study)*
pTRB569	pTRB567	Ap	*6His-SUMO-rv1959c (parE1)*	*M. tuberculosis, tagged* toxin	Genscript *(This study)*
pTRB570	pTRB568	Ap	*6His-SUMO-rv2142c (parE2)*	*M. tuberculosis, tagged* toxin	Genscript *(This study)*
pTRB637	pET-Duet1	Ap	*rv1247c (relB1)*	*M. tuberculosis,* antitoxin	Genscript *(This study)*
pTRB638	pTRB637	Ap	*6His-SUMO-rv1246c (relE1)*	*M. tuberculosis,* tagged toxin	Genscript *(This study)*

Sequence files with corresponding open reading frames (ORFs) for constructs used in this study are available as supplementary files S1 – S8.

2.5 Protein purification

2.5.1 Isolation of the soluble fraction and Nickel-affinity chromatography

Bacterial cells were pelleted from liquid culture by centrifugation at 4200 x g for 30 minutes at 4 ^{o}C. Cell pellets were resuspended in lysis buffer A500 [20 mM Tris base pH 8.0, 500 mM NaCl, 30 mM imidazole pH 8.0, 10 % (vol/vol) glycerol] (**Table 2.3**), except for cultures expressing GyrA, which were resuspended in A800 [20 mM Tris base pH 8.0, 800 mM NaCl, 30 mM imidazole pH 8.0, 10 % (vol/vol) glycerol] (**Table 2.3**) and sonicated using a Vibracell™ VCX500 ultrasonicator with medium tip (Sonics) for a total of 2 minutes (10 seconds on/10 seconds off). The sonicated sample was centrifuged at 20,000 x g for 1 hour at 4 °C to isolate the soluble fraction from cell debris.

The protein rich isolated soluble fraction was passed through a Ni-NTA His-Trap^{TM} HP 5 mL column (Cytiva) at slow speed to maximise recombinant protein binding via N-terminal hexahistidine (6His) tags. A 10 cv wash step was performed using lysis buffer. From this stage onward, purifications were optimised for each protein, detailed in 2.5.3 – 2.5.10. Fast protein liquid chromatography (FPLC) steps were carried out using an Åkta™ Pure protein chromatography system (Cytiva) at 4 °C. This allowed for programmable, precise and semi-quantifiable fractionation with multiple buffers when required.

2.5.2 Fast Protein Liquid Chromatography (FPLC)

2.5.2.1 Anion exchange chromatography

Protein samples were loaded on to a pre-equilibrated HiTrap Q HP anion exchange 5 mL column (Cytiva) in low salt buffer A100 [20 mM Tris base pH 8.0, 100 mM NaCl, 10 % (vol/vol) glycerol] (**Table 2.3**). This column was then subjected to an increasing salt gradient using the Åkta™ system, titrating in high salt buffer C1000 [20 mM Tris base pH 8.0, 1000 mM NaCl, 10 % (vol/vol) glycerol] (**Table 2.3**) until a final salt concentration of 600 mM NaCl was achieved. 2 mL fractions were collected and analysed by SDS-PAGE. Fractions containing the protein of interest were carried forward for further purification or dialysed into an appropriate buffer for storage. An example anion exchange fractionation chromatogram can be seen in **Figure 3.1**.

2.5.2.2 Heparin affinity chromatography

Protein samples were loaded on to a pre-equilibrated HiTrap Heparin HP 5 mL column (Cytiva) in low salt buffer A100 (**Table 2.3**). Following this, purification from this column was performed in the same manner as for anion exchange chromatography (2.5.2).

2.5.2.3 Size-exclusion chromatography (SEC)

HiPrep 16/60 Sephacryl S-200, S-300, and S-500 HR SEC columns (Cytiva) were used to partially separate proteins by size. Column selection depended on the column fractionation range and size of the target protein. The column was pre-equilibrated in sizing column buffer, S500 [50 mM Tris base pH 8.0, 500 mM KCl, 10 % (vol/vol) glycerol] (**Table 2.3**), prior to a concentrated protein sample being applied via capillary loops at a rate of 0.5 mL/min. Fractionation occurred at 0.5 ml/min and the resulting chromatographic peaks were sampled and analysed by SDS-PAGE. Fractions containing the protein of interest were carried forward for further purification if needed, dialysed into an appropriate buffer, or stored. An example SEC chromatogram can be seen in **Figure 3.1**.

2.5.3 Purification method and storage conditions for *M. tuberculosis* GyrA

Once bound to the initial Ni-NTA and washed with 10 cv A800, the column was washed with a further 5 cv A100. The sample was eluted directly on to a pre-equilibrated anion exchange column with 10 cv B100 [20 mM Tris base pH 8.0, 100 mM NaCl, 250 mM imidazole pH 8.0, 10 % (vol/vol) glycerol] (**Table 2.3**) before washing again in A100 to remove the high imidazole. The anion exchange column was run as in 2.5.2.1. Fractions were analysed for protein purity by SDS-PAGE, and appropriate fractions were pooled before the addition of 0.4 mg 6His-TEV protease to cleave the 6His-TEV site tag. The sample was rolled overnight at 4 °C then passed down a second Ni-NTA column (ortho Ni-NTA) to remove the 6His-TEV protease and 6His-TEV site tag. The flow through was collected and concentration in a 10 kDa cut-off centrifugal concentrator (Sartorius) to 2 mL. The 2 mL sample was injected into a 2 mL capillary loop on the Åkta™ Pure system before fractionation by SEC using the Sephacryl S-300 column, as per 2.5.2.3. Fractions were analysed for purity by SDS-PAGE, appropriate fractions were pooled and concentrated to >300 µM before diluting by one third volume with storage buffer [50 mM Tris base pH 8.0, 500 mM KCl, 70 % (vol/vol) glycerol] (**Table 2.3**) for a final glycerol (cryoprotectant) concentration of 30 %, and final protein concentration of >200 µM. Appropriate volume aliquots were made and snap frozen in liquid nitrogen before storage at -80 °C.

2.5.4 Purification method and storage conditions for *M. tuberculosis* GyrB

Once bound to the initial Ni-NTA and washed with 10 cv A500, the column was washed with a further 5 cv A100. The sample was eluted directly on to a pre-equilibrated anion exchange column with 10 cv B100 before washing again in A100 to remove the high imidazole. The anion exchange column was run as in 2.5.2.1. Fractions were analysed for protein purity by SDS-PAGE, and appropriate fractions were pooled before the addition of 0.4 mg 6His-TEV protease to cleave the 6His-TEV site tag. The sample was rolled overnight at 4 °C then passed down a second Ni-NTA column (ortho Ni-NTA) to remove the 6His-TEV protease and 6His-TEV site tag. The flow through

was collected and concentration in a 10 kDa cut-off centrifugal concentrator (Sartorius) to 2 mL. The 2 mL sample was injected into a 2 mL capillary loop on the Åkta™ Pure system before fractionation by SEC using the S-300 column, as per 2.5.2.3. Fractions were analysed for purity by SDS-PAGE, appropriate fractions were pooled and concentrated to >300 µM before diluting by one third volume with storage buffer for a final glycerol (cryoprotectant) concentration of 30 %, and final protein concentration of >200 µM. Appropriate volume aliquots were made and snap frozen in liquid nitrogen before storage at -80 °C.

2.5.5 Purification method and storage conditions for *M. tuberculosis* GyrBA

Once bound to the initial Ni-NTA and washed with 10 cv A500, the column was washed with a further 5 cv A100. The sample was eluted directly on to a pre-equilibrated anion exchange column with 10 cv B100 before washing again in A100 to remove the high imidazole. The anion exchange column was run as in 2.5.2.1. Fractions were analysed for protein purity by SDS-PAGE, and appropriate fractions were pooled before the addition of 0.4 mg 6His-TEV protease to cleave the 6His-TEV-site tag. The sample was dialysed overnight at 4 °C into A100 then passed down a second Ni-NTA column (ortho Ni-NTA) to remove the 6His-TEV protease and 6His-TEV-site tag. The flow through was passed directly on to a Heparin affinity column for fractionation as per 2.5.2.2. Fractions were analysed for purity before appropriate fractions were pooled and concentrated in a 10 kDa cut-off centrifugal concentrator (Sartorius) to 2 mL. The 2 mL sample was injected into a 2 mL capillary loop on the Åkta™ Pure system before fractionation by SEC using the Sephacryl S-500 column, as per 2.5.2.3. Fractions were analysed for purity by SDS-PAGE, appropriate fractions were pooled and concentrated to >75 µM before diluting by one third volume with storage buffer for a final glycerol (cryoprotectant) concentration of 30 %, and final protein concentration of >50 µM. Appropriate volume aliquots were made and snap frozen in liquid nitrogen before storage at -80 °C.

2.5.6 Purification method and storage conditions for *M. tuberculosis* GyrBA^{56}

This process was identical to 2.5.5 with a single exception; tag cleavage occurred using the hSENP-2 enzyme to remove the 6His-SUMO tag.

2.5.7 Purification method and storage conditions for *M. tuberculosis* GyrB^{28}A^{56}

This process was identical to 2.5.5 with the following exceptions: tag cleavage occurred using the hSENP-2 enzyme to remove the 6His-SUMO tag and SEC was performed using the Sephacryl S-300 column, as per 2.5.2.3.

2.5.8 Purification method and storage conditions for *M. tuberculosis* ParDE1

This process was identical to 2.5.4 with the following exceptions: tag cleavage occurred using the hSENP-2 enzyme to remove the 6His-SUMO tag and SEC was performed using the Sephacryl S-

200 column, as per 2.5.2.3. Protein was stored in SEC buffer only (10 % glycerol) at a concentration of >100 µM.

2.5.8.1 Purification method for the heterotetrameric ParDE1 complex

This process was identical to 2.5.4 with the following exceptions: tag cleavage occurred using the hSENP-2 enzyme to remove the 6His-SUMO tag concentration and SEC via the HiPrep 16/60 Sephacryl columns was not performed. Protein was of sufficient purity and concentration after anion exchange for biophysical studies.

2.5.9 Purification method and storage conditions for *M. tuberculosis* ParDE2

This process was identical to 2.5.4 with the following exceptions: tag cleavage occurred using the hSENP-2 enzyme to remove the 6His-SUMO tag and SEC was performed using the S-200 column, as per 2.5.2.3. Protein was stored in SEC buffer only (10 % glycerol) at a concentration of >100 µM.

2.5.9.1 Purification method and storage conditions for *M. tuberculosis* ParE2

Once bound to the initial Ni-NTA and washed with 10 cv A500, the sample was eluted in 5 cv B500 [20 mM Tris base pH 8.0, 500 mM NaCl, 250 mM imidazole pH8.0, 10 % (vol/vol) glycerol] (**Table 2.3**) and 0.4 mg 6His-hSENP-2 was added to cleave the 6His-SUMO tag. The sample was dialysed into A100 overnight at 4 °C before being passed down a second Ni-NTA column (ortho Ni-NTA) to remove the 6His-hSENP-2 and 6His-SUMO. The flow through was passed directly on to an anion exchange column for fractionation as per 2.5.2.1. Fractions were analysed for protein purity by SDS PAGE; routinely ParE2 eluted in an early 'shoulder' peak before the full ParDE2 complex. Appropriately pure ParE2 fractions were not subjected to SEC due to low yields, rather, the sample was pooled and concentrated in a 10 kDa cut-off centrifugal concentrator (Sartorius) to >100 µM before snap freezing in aliquots for storage at -80 °C.

2.5.10 Purification method and storage conditions for *M. tuberculosis* RelBE1

This process was identical to 2.5.4 with the following exceptions: after overnight incubation a high amount of precipitate must be removed by centrifugation, the supernatant is carried forward and SEC was performed using the Sephacryl S-200 column, as per 2.5.2.3. Protein was stored in SEC buffer only (10 % glycerol) at a concentration of >100 µM.

2.6 Analysis of protein purity and solution state

2.6.1 SDS-PAGE

2.6.1.1 Gel casting

Depending on the sample being analysed, the appropriate percentage resolving gel solution (**Table 2.3**) was made and 4.5 mL poured into a 1 mm casting plate (BioRad) leaving room for 1.5 mL 4 % stacking gel (**Table 2.3**). The resolving gel was levelled off with a layer of isopropanol and allowed to set prior to removing the isopropanol layer, pouring the stacking gel and adding the appropriate gel comb.

2.6.1.2 Sample preparation

Samples were prepared in 20 µL volumes including 5 µL of 4 x SDS-PAGE loading dye (**Table 2.3**). Samples containing resuspended cells were initially incubated at 95 °C for 10 minutes. Sample amount added to the loading dye varied by sample and purity. Once ready, the 20 µL samples are incubated at 95 °C for 5 minutes before gel loading.

2.6.1.3 Sample loading and gel running

Generally, 10 µL of the prepared sample was loaded into the gel. A PageRuler™ Plus Prestained protein ladder was always run alongside samples for estimation of molecular weights and identification of proteins of interest. Gels were initially run at 80 V through the stacking gel until the resolving gel was reached, at which point the voltage was increased to 160 V until the dye front reached the end of the resolving gel. Gels were run in 1 x stock electrode buffer (1/10 dilution of 10 x electrode buffer with dH_2O) (Table 2.3) in BioRad cassettes and tanks.

2.6.1.4 Gel staining and imaging

Once fully resolved, gels were stained for 1 hour in Quick Coomassie stain (Protein Ark) before de-staining for 1 hour in dH_2O. Gels were imaged on a BioRad ChemiDoc™ XRS+ with ImageLab™ software on the Coomassie setting, using a white background (BioRad).

2.6.2 Mass spectrometry

Matrix-Assisted Laser Desorption/Ionisation Time of Flight (MALDI-TOF) mass spectrometry of protein samples was kindly performed in-house at Durham University Chemistry Department by Mr Peter Stokes. 100 µL protein samples were supplied at 1 mg/mL in 10 mM ammonium bicarbonate.

2.6.3 Circular dichroism spectroscopy and thermal denaturation

Both CD and thermal denaturation were performed in-house in Durham University Physics department very kindly by Professor Elizabeth Bromley. CD was performed at 20 °C pre and post

melting to analyse secondary structure of TA complexes. Thermal denaturation was performed between 20 °C and 80 °C and normalised to 222 nm for melt curves. Proteins were analysed in A500 buffer. Both CD and thermal denaturation curves are plotted in GraphPad Prism (Version 9.4.1) as an XY table, with X as 'Numbers' and Y as a 'single Y value for each point'. Graphs are presented with the connecting line only.

2.6.4 Analytical size-exclusion chromatography

The Superose 6 10/300 GL SEC column (Cytiva, discontinued) was selected for its broad fractionation range and short run time, allowing for analysis and purification on the Åkta™ pure system (Cytiva). The column was equilibrated in buffer S300-A [20 mM Tris base pH 8.0, 300 mM NaCl) (**Table 2.3**). For analysis, protein samples were manually loaded into a 100 µL capillary loop in their respective storage buffers at appropriate concentrations to generate UV signal, generally 1 mg/mL was sufficient. Samples were injected onto the column using S300-A buffer at a flowrate of 0.5 mL/min for fractionation across 1.2 column volumes (column volume (V_c) of 24 mL). Where appropriate, samples were collected for further analysis in 250 µL fractions. Elution volumes (V_e) were calculated using the Peaks function in Unicorn™ 7 (Cytiva).

2.6.4.1 Generation of calibration curves

Calibration curves were generated for the Superose 10/300 GL SEC column using appropriate combinations of commercially available low and high molecular weight kit proteins (Cytiva) for best resolution. Two calibration protein mixes were created and are detailed below, with molecular weight (M_r, kDa) and Stokes radius (R_{st}, Å). Calibration mixes were fractionated as per 2.6.2 without collection. Column void volume (V_o) was obtainable from Mix 2.

SEC calibration protein mixes

Calibration Mix 1	Calibration Mix 2
Thyroglobulin (669 kDa, 85 Å)	Ferritin (440 kDa, 61 Å)
Aldolase (158 kDa, 48.1 Å)	Conalbumin (75 kDa, 36.4 Å)
Ovalbumin (44 kDa, 30.5 Å)	Carbonic anhydrase (29 kDa, 23 Å)
Cytochrome C (12.9, 17.8 Å)	Aprotinin (6.5 kDa, 13.5 Å)

Elution volumes (V_e) were converted into the partitioning coefficient (K_{av}) for each sample using the following equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

The molecular weight calibration curve is subsequently plotted as K_{av} vs $Log_{10}(M_r$, kDa). The Stokes radius calibration curve is subsequently plotted as $Log_{10}(R_{st}$, Å) vs K_{av}.

2.6.4.2 Sample preparation and application

Protein samples for analytical sizing were thawed on ice from stocks stored at -80 °C. Generally, samples were diluted in S300-A buffer to > 100 µL at 1 mg/mL for manual loading into a 100 µL capillary loop on the Åkta™ pure system. If the resulting UV signal was too low, the sample concentration was increased. Fractionation occurred as per 2.6.2.

2.6.4.3 Molecular weight and Stokes radius estimation

For estimates of M_r and R_{st}, linear regression was performed on the respective plots. The resulting line equations (y = mx + c) were used to calculate the observed M_r and R_{st} through the following rearrangements:

$$M_r = 10\text{^}\left(\frac{K_{av} - c}{m}\right)$$

$$R_{st} = 10\text{^}((m(K_{av}) + C)$$

Observed values were then compared to calculated values of M_r and R_{st} and presented as a ratio of calculated:observed. M_r values were calculated using the online ProtParam tool (Expasy) (Gasteiger *et al.*, 1999). R_{st} values were calculated using crystal structures of AlphaFold generated models using the HullRad calculator (Fluidic Analytics) (Fleming and Fleming, 2018).

2.6.4.4 ParDE1 complex remodelling

ParDE1 expressed and purified as detailed in 2.4.2.1 and 2.5.8.1 respectively provides the starting material (theoretical heterotetramer) for remodelling experiments. Once samples were ready for analysis they were subjected to analytical SEC as described in 2.6.2.

For initial incubation and buffer alteration experiments, ParDE1 concentration remained at 2.5 mg/mL (~62.5 µM). Incubation at 4 °C was performed in the fridge, while 37 °C and 45 °C incubation was performed in a thermocycler. For concentration dependent studies, ParDE1 was concentrated in a 5 kDa cut-off centrifugal concentrator column (Sartorius) from 2.5 mg/mL and 100 µL sampled at the appropriate concentrations.

For the 37 °C time-course, ParDE1 concentration begun at 4 mg/mL (100 µM) to allow for coupling with cleavage assays. 100 µL was sampled at each time-point and subjected to analytical SEC. Incubation was controlled in a thermocycler.

2.6.4.5 Data normalisation and presentation

Chromatograms from SEC experiments were exported and observed mAU (mAU_{obs}) values normalised between 0 and 1 for presentation and comparison. Datasets were initially cropped from 1 – 24 mL to remove an artefactual chromatographic peak consistent in all experiments (attributed to valve repositioning). Corresponding mAU values were normalised (mAU_{norm}) in Microsoft Excel by identifying the minimum (mAU_{min}) and maximum (mAU_{max}) mAU values and assigning these the values of 0 and 1 respectively, with the remaining values being normalised between them. The following equation is used to normalise the entire dataset:

$$mAU_{norm} = \frac{mAU_{obs} - mAU_{min}}{mAU_{max} - mAU_{min}}$$

The resulting normalised dataset was imported into GraphPad Prism (Version 9.4.1) as an XY table, with X as 'Numbers' and Y as a 'single Y value for each point'. Graphs are presented with the connecting line only and cropped on the X axis to the appropriate mL for each experiment.

2.7 Topoisomerase assays

Topoisomerase assays were performed using published protocols (Aldred *et al.*, 2016; Blower *et al.*, 2016), adapted where appropriate.

2.7.1 Reconstitution and dilution of gyrase enzymes

The DNA gyrase holoenzyme was reconstituted by incubating equimolar amounts of GyrB and GyrA to a final heterotetramer ($GyrB_2A_2$) concentration of 10 µM on ice for 5 minutes. Gyrase fusion proteins were incubated at a final dimer concentration of 10 µM on ice for 5 minutes. Gyrase enzymes were then serially diluted in twofold steps using gyrase dilution buffer [50 mM Tris base pH 8.0, 2 mM MgOAc, 1 mM DTT, 500 mM KOAc, 50 µg/mL BSA, 10 % (vol/vol) glycerol] (**Table 2.3**), down to the appropriate concentration for assays (stated below).

2.7.2 ATP-independent removal of supercoils

Each reaction contained 5 µL of 4 x gyrase reaction buffer [40 mM Tris base pH 8.0, 38.4 mM MgOAc, 4 mM DTT, 100 µg/mL BSA, 32 % (vol/vol) glycerol)] (**Table 2.3**) and 1 µL of a 250 ng/µL solution of negatively supercoiled pSG483 (see 2.3.2). 4 µL of the appropriate gyrase enzyme dilution was added before incubation on ice for 5 minutes. Reactions were then diluted to 20 µL with dH_2O and incubated at 37 °C for 30 minutes. Final reaction conditions were 12.5 nM plasmid, variable gyrase enzyme, 20 mM Tris base pH 8.0, 10 mM MgOAc, 1.2 mM DTT, 100 mM KOAc, 35 µg/mL BSA and 10% (vol/vol) glycerol.

2.7.3 ATP-dependent generation of negative supercoils

Each reaction contained 5 µL of 4 x gyrase reaction buffer and 1 µL of a 250 ng/µL solution of relaxed form pSG483 (see 2.3.4). 4 µL of the appropriate gyrase enzyme dilution was added before incubation on ice for 5 minutes. Reactions were then diluted to 20 µL with 1 µL 20 mM ATP and 9 µL dH_2O and incubated at 37 °C for 30 minutes.

2.7.4 Cleavage assay

Each reaction contained 5 µL of 4 x gyrase reaction buffer and 1 µL of a 250 ng/µL solution of negatively supercoiled pSG483 (see 2.3.2). 4 µL of 0.15625 µM gyrase enzyme is added before incubation on ice for 5 minutes. 2 µL of drug or protein dilution was added, or solvent/buffer where appropriate, before incubation on ice for a further 5 minutes. Reactions were diluted to 20 µL with 8 µL dH_2O and incubated at 37 °C for 30 minutes.

Drug (Moxifloxacin) (**Table 2.2**) dilutions were prepared by twofold dilution in solvent (DMSO) to the appropriate assay concentrations. Protein additive (TA system components and complexes) dilutions were prepared by twofold dilution in respective storage buffers to appropriate assay concentrations.

2.7.5 ParE2 nuclease analysis

2.7.5.3 ParE2 titration against supercoiled pSG483

Each reaction contained 5 µL of 4 x gyrase reaction buffer, 1 µL of a 250 ng/µL solution of negatively supercoiled pSG483 (see 2.3.2), and 4 µL gyrase dilution buffer. ParE2 was diluted in twofold steps before adding 2 µL of the appropriate dilution to the independent reactions and incubating on ice for 5 minutes. Reactions were diluted to 20 µL with 8 µL dH_2O and incubated at 37 °C for 30 minutes.

2.7.5.4 ParE2 endonuclease/exonuclease analysis

For endonuclease analysis, each reaction contained 5 µL of 4 x gyrase reaction buffer, 1 µL of a 250 ng/µL solution of negatively supercoiled pSG483 (see 2.3.2). For exonuclease analysis each reaction contained 5 µL of 4 x gyrase reaction buffer, 1 µL of a 250 ng/µL solution of negatively supercoiled pSG483 (see 2.3.3).

Gyrase dimer subunits and reconstituted heterotetramer were added where appropriate at 31.25 nM final concentration. ParE2 was added where appropriate at 2.5 µM final concentration. ParD2 was added where appropriate at 10 µM final concentration. EDTA was added where appropriate at 1 mM final concentration. Reactions were incubated on ice for 5 minutes before being diluted to 20 µL with dH_2O and incubated at 37 °C.

2.7.6 Reaction stopping and gel loading

Following incubation, reactions were first quenched with 2 µL of stopping buffer [5 % (wt/vol) SDS, 125 mM EDTA] (**Table 2.3**), followed by adding 1 µL of 12 mg/mL proteinase K and further incubation at 37 °C for 1 hour. Reactions were stored at 4 °C until immediately before gel loading, whereupon a 6 x agarose gel loading dye was added to the samples and the samples were warmed to 37 °C for 5 minutes. Samples were separated by electrophoresis in 1.4 % (wt/vol) TAE agarose gels (**Table 2.3**) (containing 0.5 µg/mL EtBr when appropriate) for 16 hours at 45 v. Agarose gels were post-stained in TAE containing 0.5 µg/mL EtBr (when appropriate) and visualised by UV illumination as per 2.3.6.

2.7.7 Quantification and graphical analysis

Gel images were analysed using ImageJ2 (Rueden *et al.*, 2017) with background subtracted. For ATP-independent removal of supercoils, supercoiled band intensity was measured throughout the titration and converted to percentage of the '0' gyrase lane supercoiled band. For ATP-dependent generation of supercoils, supercoiled band intensity was measured and converted to percentage of the control lane supercoiled band as this contained the same amount of plasmid (250 ng).

Mean values and standard deviation were calculated from triplicate data (unless otherwise stated in figure legends) for the band of interest. Data were plotted in GraphPad Prism (Version 9.4.1) and presented with connecting line and error bars.

2.7.7.1 Quantification of linear products in cleavage assays

Cleavage assay measurements were taken from gels containing EtBr (when possible); supercoiled, linear, and nicked band intensities were calculated per lane. Linear band percentage was subsequently calculated per lane and normalised to the '0' lane linear percentage, per assay.

Measurements for the gyrase fusion proteins ($GyrBA^{56}$ and $GyrB_{28}A_{56}$) were taken from gels not containing EtBr as only supercoiled, linear, and nicked bands were present. Calculations occurred as per the cleavage assay.

2.7.7.2 Quantification of linear products and DNA loss in ParDE1 assays

Measurements for the DNA damage induced by thermal remodelling of ParDE1 were performed on gels containing EtBr. Linear and nicked product estimates were calculated as per cleavage assays. To estimate the amount of DNA loss per lane the total band intensity of supercoiled + linear + nicked per lane was compared as a percentage to the band intensity of the control supercoiled (S) lane. The difference in percentage between the experimental lane and control lane is presented as DNA loss.

2.8 Protein crystallisation

2.8.1 Crystallisation screening

Once optimum purity had been reached, proteins for crystallography were dialysed into buffer X [20 mM Tris base pH 8.0, 150 mM NaCl, 2.5 mM DTT] (**Table 2.3**) and concentrated to 12 mg/mL for initial trials. Commercially available 96-well crystallisation screens (Molecular Dimensions) were used to assay preferred crystallisation conditions. Sitting drop crystallisation trials were set-up using an SPT Labtech Mosquito® crystal robot. Each condition was tested with a 2:1 and 1:1 ratio of protein to mother-liquor on the crystallisation stage. Crystal screens were left at 18 °C, unless otherwise stated during optimisation.

2.8.2 Crystal optimisation

ParDE1 required no optimisation, datasets of sufficient quality were collected from needle shaped crystals grown in 0.1 M Bis Tris Propane pH 7.5, 20 % PEG 3350, 0.2 M $NaNO_3$ and harvested directly from the crystal screen.

2.8.2.1 Optimisation of ParDE2 crystals

For ParDE2, initial crystal screens set at 12 mg/mL yielded no results with low numbers of precipitated conditions. To optimise the screening process, ParDE2 starting concentration was increased to 16 mg/mL and 20 mg/mL. Best crystals (large hexagonal planar) grew after 3 months in 0.1 M MES pH 6.2, 15 % wt/vol PEG 3350 at 20 mg/mL starting concentration and were harvested directly from the crystal screen.

2.8.2.2 Optimisation of RelBE crystals

Early crystal screens resulted in many conditions with high levels of precipitation, or what appeared to be micro-crystals, even when incubated at 4 °C. To slow the rate of crystallisation the crystal buffer was supplemented with 5 % (vol/vol) glycerol to give buffer X-5 [20 mM Tris base pH 8.0, 150 mM NaCl, 2.5 mM DTT, 5 % (vol/vol) glycerol] (**Table 2.3**), and the starting RelBE1 concentration was reduced to 8 mg/mL.

4 M ammonium acetate, 0.1 M Bis-Tris Propane pH 7.0 was subsequently identified as the best 'hit', yielding small diamond-like crystals. For optimisation, the drop size was increased to 2 µL in a 1:1 ratio of protein to mother liquor. Ammonium acetate concentrations above 2.5 M produced incredibly durable hexagonal bifrustum shaped RelBE1 crystals, while reducing the ammonium acetate below 2 M resulted in needle shaped RelBE1 crystals.

2.8.3 Crystal harvesting

Positive individual crystallisation conditions were identified by microscopy. Mother liquor from the appropriate condition and 100% glycerol were mixed in a ratio of 1:1 and an equal volume of this mixture was added to the sitting drop. The crystal was visualised under a microscope to check for durability or deformity at this point. Surviving crystals were harvested using cryo-loops of appropriate sizes and snap-frozen in liquid nitrogen before adding to a puck submerged in liquid nitrogen for storage and transport.

2.9 X-ray crystallography

2.9.1 Data collection

Data collection was performed at Diamond Light Source, Oxford, UK, via remote access on i04. Data sets were collected from a number of native protein crystals. These datasets were merged in downstream processing, if necessary.

2.9.2 Data processing

Initial data processing was automated by Diamond Light Source iSpyB using the X-ray image integration programs Xia2 and Xia2-DIALS (Winter and IUCr, 2010). Image integration and space group selection were carried out manually using the same programs as well as Mosflm (Leslie and Powell, 2007) to confirm maximum resolution and space group. Data reduction and generation of R sets was carried out using AIMLESS (Evans, Murshudov and IUCr, 2013).

2.9.3 ParDE1 structure determination

Diffraction data were collected at Diamond Light Source on microfocus beamline i24 (**Table 4.1**). Single, 360°, datasets were collected from three native ParDE1 crystals and merged using iSpyB (Diamond Light Source). Diffraction data were processed with XDS (Kabsch, 2010b, 2010a), and then AIMLESS from CCP4 (Winn *et al.*, 2011) was used to corroborate the spacegroups (**Table 4.1**). The crystal structure of ParDE1 was solved by molecular replacement using PHASER (McCoy and IUCr, 2007) using the starting model 3KXE (heterotetrameric $ParD1_2ParE1_2$) (Dalton and Crosson, 2010) split into individual protomers ParD1 and ParE1 and input as individual assemblies. The solved starting model for ParDE1 was built in REFMAC (Vagin *et al.*, 2004) and BUCCANEER (Cowtan and IUCr, 2006). The model was then iteratively refined and built using PHENIX (Adams *et al.*, 2010) and Coot (Emsley and Cowtan, 2004a), respectively. The quality of the final model was assessed using Coot and the wwPDB validation server (Gore, Velankar and Kleywegt, 2012). Structural figures were generated using PyMol (Schrödinger).

2.9.4 ParDE2 structure determination

Diffraction data were collected at Diamond Light Source on microfocus beamline i24 (**Table 4.3**). Six, 360°, datasets were collected from two native ParDE2 crystals and merged using iSpyB (Diamond Light Source). Diffraction data were processed with XDS (Kabsch, 2010b, 2010a), and then AIMLESS from CCP4 (Winn *et al.*, 2011) was used to corroborate the spacegroups (**Table 4.3**). The crystal structure of ParDE2 was solved by molecular replacement using PHASER (McCoy and IUCr, 2007) using the *M. tuberculosis* H37Rv ParE2 AlphaFold structure prediction (Jumper *et al.*, 2021; Varadi *et al.*, 2022) as the search model. The solved starting model for ParDE2 was built in REFMAC (Vagin *et al.*, 2004) and BUCCANEER (Cowtan and IUCr, 2006). Initially, ParD2 could not be placed by PHASER. The ParE2 AlphaFold search model was edited to remove the C-terminal 12 amino acids to allow for subsequent manual building of the ParD2 chain in Coot (Emsley and Cowtan, 2004b). The model was then iteratively refined and built using PHENIX (Adams *et al.*, 2010) and Coot (Emsley and Cowtan, 2004a), respectively. The quality of the final model was assessed using Coot and the wwPDB validation server (Gore, Velankar and Kleywegt, 2012). Structural figures were generated using PyMol (Schrödinger).

2.9.5 RelBE1 structure determination

Diffraction data were collected at Diamond Light Source on microfocus beamline i04 (**Table 5.1**). Twelve, 360°, datasets were collected from a single RelBE1 crystal and merged using iSpyB (Diamond Light Source). Diffraction data were processed with XDS (Kabsch, 2010b, 2010a), and then AIMLESS from CCP4 (Winn *et al.*, 2011) was used to corroborate the spacegroups (**Table 5.1**). The crystal structure of RelBE1 was solved by molecular replacement using PHASER (McCoy and IUCr, 2007) using the starting model 3G5O split into individual protomers RelB1 and RelE1 and input as individual assemblies. The solved starting model for RelBE1 was built in REFMAC (Vagin *et al.*, 2004) and BUCCANEER (Cowtan and IUCr, 2006). The model was then iteratively refined and built using PHENIX (Adams *et al.*, 2010) and COOT (Emsley and Cowtan, 2004a), respectively. The quality of the final model was assessed using COOT and the wwPDB validation server (Gore, Velankar and Kleywegt, 2012). Structural figures were generated using PyMol (Schrödinger).

2.10 Generation of AlphaFold multimer models

Protein structure predictions for the monomers of the TA system proteins are readily available in the published AlphaFold database, accessible online (Varadi *et al.*, 2022). For multimer models, the Google Colaboratory (ColabFold) (Jumper *et al.*, 2021; Mirdita *et al.*, 2022) was used.

This allowed for the input of multiple protein sequences and the subsequent automated generation of multimer models. Protein sequences for *M. tuberculosis* TA system components

were sourced from Mycobrowser (Kapopoulou, Lew and Cole, 2011). The highest scoring models from these structure predictions are presented in this thesis. Structural figures were generated using PyMol (Schrödinger).

2.11 Phylogenetic analyses of protein sequences

2.11.1 Protein sequence identification

The toxin antitoxin system database (TADB 2.0) (Xie *et al.*, 2018) was used to identify both RelE and ParE family protein sequences. Sequences were visually inspected, alongside inspection of subsequent alignments, to identify potentially mis-annotated sequences (these were removed if necessary). Sequences are presented alongside their gene name for identification.

The full sequence datasets are available as Supplementary files S1 (RelE), S2 (ParE), S3 (RelE/ParE), and S4 (RNase T1/RelE/ParE).

2.11.2 Multiple sequence alignment

Multiple sequence alignment (MSA) was performed on RelE, ParE, RelE/ParE, and RNase T1/RelE/ParE datasets using ClutalOmega (Sievers and Higgins, 2014). Output format was selected as ClustalW with character counts.

2.11.3 Generation of phylogenetic trees

Phylogenetic trees were generated from ClustalOmega generated MSA. Tree data was input to FigTree (Cummings, 2004) and presented in a circular format with colouring and highlighting where appropriate. For RelE, ParE, and RelE/ParE alignments, resulting phylogenetic trees were midpoint rooted. For the RNase T1/RelE/ParE alignment, the resulting phylogenetics tree was re-rooted in FigTree to the RNase T1 sequence.

Chapter 3. Biochemical and biophysical investigation of the effects of ParE toxins on DNA gyrase

Chapter 3. Introduction

ParDE toxin-antitoxin systems have previously been shown to target the essential enzyme, DNA gyrase (Jiang *et al.*, 2002; Gupta *et al.*, 2016; Muthuramalingam *et al.*, 2018; Ames *et al.*, 2019; Kamruzzaman and Iredell, 2019). The biological importance of the ParDE systems is as yet unknown, although they have been implicated in stress responses and promote survival when cells are challenged with increased temperature and antibiotics (Ames *et al.*, 2019; Kamruzzaman and Iredell, 2019). Inhibition of the *M. tuberculosis* gyrase is a potent mechanism in arresting cell growth as it is the sole type II topoisomerase and contributes to a number of cellular processes; notably transcription and DNA replication, and cell division (Karkare et al., 2012; Ashley et al., 2017; Hobson, Bryant and Berger, 2020).

Biochemically, the ParE toxins have been shown to be capable of stopping the enzyme in the cleavage state, trapping the cleavage complex and creating linearised DNA *in vitro* (Jiang *et al.*, 2002; Gupta *et al.*, 2016). *Mycobacterium tuberculosi*s harbours two such ParDE systems (ParDE1 and ParDE2), one of which (ParDE2) has been shown to be active against gyrase (Gupta *et al.*, 2016). Refolded ParE2 was shown to be capable of trapping cleavage complexes in a negative supercoiling reaction (Gupta *et al.*, 2016).

To begin untangling the molecular mechanisms by which ParE toxins inhibit DNA gyrase, we must first demonstrate their activity against the enzyme. As previous studies have indicated that ParE toxins trap cleavage complexes, we can use a cleavage assay as demonstrated for a number of antibiotics (Aldred *et al.*, 2016; Blower *et al.*, 2016). Prior to this, it will be important to express, purify, and test *M. tuberculosis* gyrase; activity will be shown for the ability of the enzyme to both remove and introduce supercoils in ATP-independent and ATP-dependent manners, respectively. Further to this, the purified holoenzyme enzyme will be tested against Moxifloxacin to demonstrate the efficacy of the cleavage assay. Expression and purification methods will be developed for the ParDE system components to allow for thorough testing of the effects of their toxins on gyrase *in vitro*. Further to this, we can test the effects of the ParE toxins on a number of recombinant gyrase fusion proteins that are readily used in structural biology (Blower *et al.*, 2016; Petrella *et al.*, 2019; Vanden Broeck *et al.*, 2019).

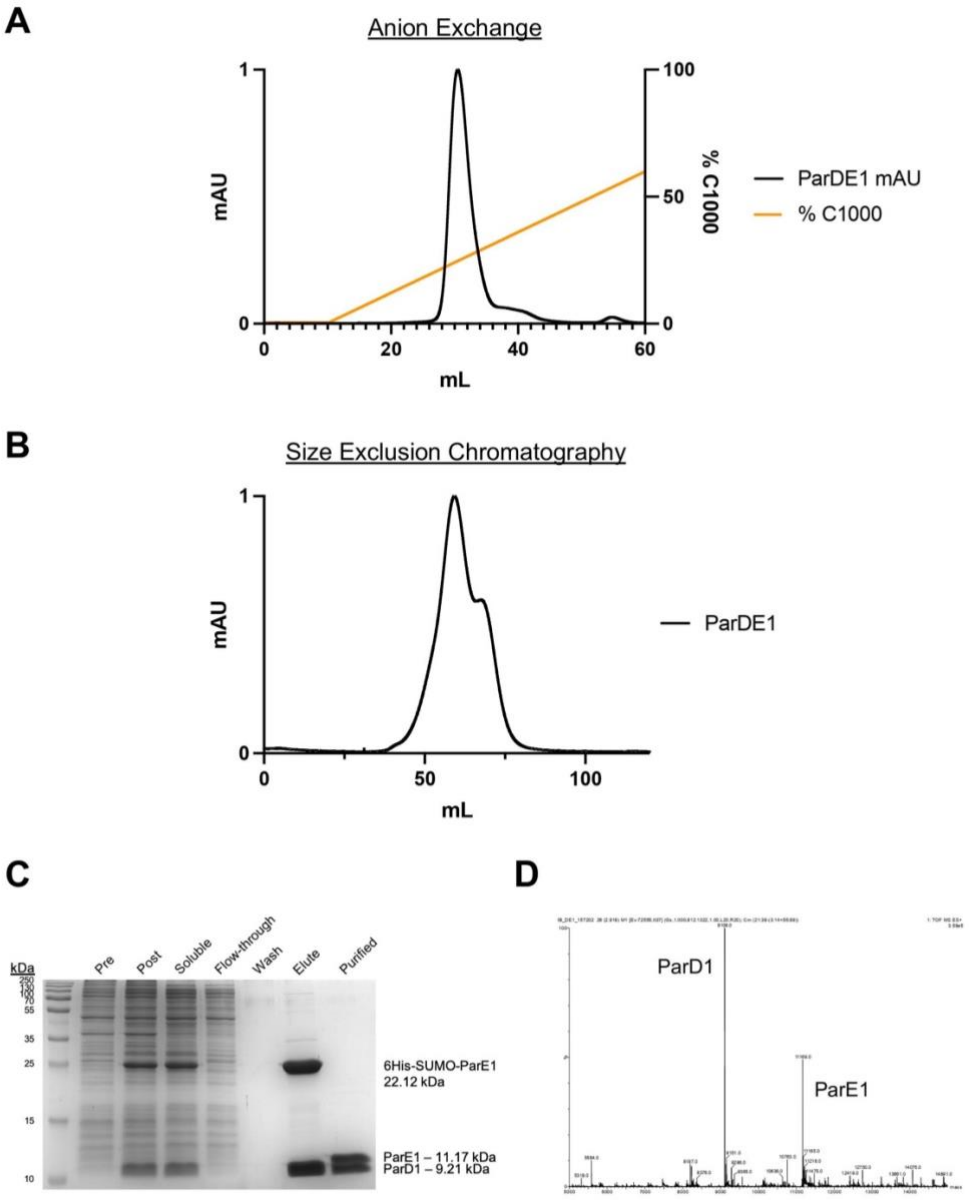
Once biochemically tested, biophysical analyses will allow us to develop our understanding of the stability of the ParDE systems in solution, and build toward testing for efficient methods of purifying a gyrase-ParE complex. Elucidating the molecular mechanism of ParE-mediated gyrase inhibition will start with protein purification, biochemistry to confirm activity, and a range of biophysical studies to analyse solution states, building toward structural studies of the cleavage complexes formed.

3.1 Protein expression and purification

Characterisation of ParDE proteins and DNA gyrase targets was started by preparing sufficient purified protein for study. Expression constructs were cloned, and the specifics for expression and purification of each recombinant protein used in biochemical, biophysical, and later crystallographic studies are detailed in **Chapter 2, Materials & Methods**. Generally, the appropriate bacterial strain (**Table 2.4**) was grown in liquid culture with shaking until mid-log phase before supplementation with IPTG to a final concentration of 1 mM to induce protein expression via the pET system (Mierendorf *et al.*, 2003). Cells were grown overnight at 18 °C with shaking before cells were isolated through centrifugation. Pelleted cells were resuspended in A500 buffer (**Table 2.3**) and lysed by sonication before centrifugation to isolate the soluble protein fraction. Proteins of interest were pulled down via a cleavable His tag by passing the soluble fraction through a Ni-NTA column.

As an example, **Figure 3.1** demonstrates the chromatographic stages of protein purification for the ParDE1 protein complex. Each of the purified proteins presented in this thesis was subjected to these methods of purification with changes to the order or additional steps as detailed in **Materials & Methods 2.5**. Ni-NTA bound ParDE1 protein was washed with 10 cv A500 before equilibrating into low salt buffer, A100, and elution on to an anion exchange column with B100 (Table 2.3). Once re-equilibrated into A100 the anion exchange column was plumbed into an Åkta™ pure system and subjected to a gradient of buffer C1000 to elute the ParDE1 complex (**Figure 3.1 A**). SDS-PAGE analysis was performed to check the purity of fractions from anion exchange and for ParDE1, 24 – 27 mL was pooled (**Figure 3.1 A**). The pooled sample was incubated overnight with the appropriate protease to remove the cleavable tag, and passed through a second Ni-NTA column to purify the ParDE1 complex (now in the unbound or flow-through fraction) from the cleavage protease and cleaved His tag. The flow-through fraction was concentrated to 2 mL and subjected to size exclusion chromatography (SEC) with an Åkta™ Pure system, and in this case for ParDE1, using a HiPrep™ Sephacryl® S-200 HR column (Cytiva) (**Figure 3.1 B**). Fractions were analysed via SDS-PAGE and those of appropriate purity were pooled (**Figure 3.1 B**, 52 – 74 mL) and stored as per **Materials & Methods 2.5**. Both ParDE1 peaks contained high-purity protein complex (discussed in section 3.4.3), however, for other proteins a single peak was obtained in SEC. An SDS-PAGE summary of the protein purification procedure is presented in **Figure 3.1 C**, demonstrating the purity in each fraction and final, purified ParDE1 sample. This purified sample was analysed via MALDI-ToF mass spectrometry (Gundry *et al.*, 2009) (**Figure 3.1 D**) to demonstrate that the dominant species present in the purified product are of the appropriate and expected molecular weights.

Figure 3.1 Protein Purification



(**A**) Example Anion Exchange chromatogram from the purification of the ParDE1 complex; (**B**) Example Size Exclusion chromatogram from the purification of the ParDE1 complex; (**C**) SDS-PAGE summary of the protein expression and purification procedure of the ParDE1 complex. Pre – whole cell lysate prior to protein expression induction; Post – whole cell lysate after protein expression induction showing the appearance of expression recombinant protein; Soluble – soluble protein fraction; Flow-through – protein fraction not bound to the initial Ni-NTA step; Wash – sample collected mid-way through the wash step from the initial Ni-NTA step; Elute – un-cleaved recombinant protein purity after initial Ni-NTA step; Purified – cleaved final product after Anion Exchange, tag cleavage and removal, and Size Exclusion; (**D**) MALDI-ToF Mass Spectrometry of the purified ParDE complex sample.

3.2 Biochemical analyses of the inhibitory effects of ParE toxins on *M. tuberculosis* gyrase

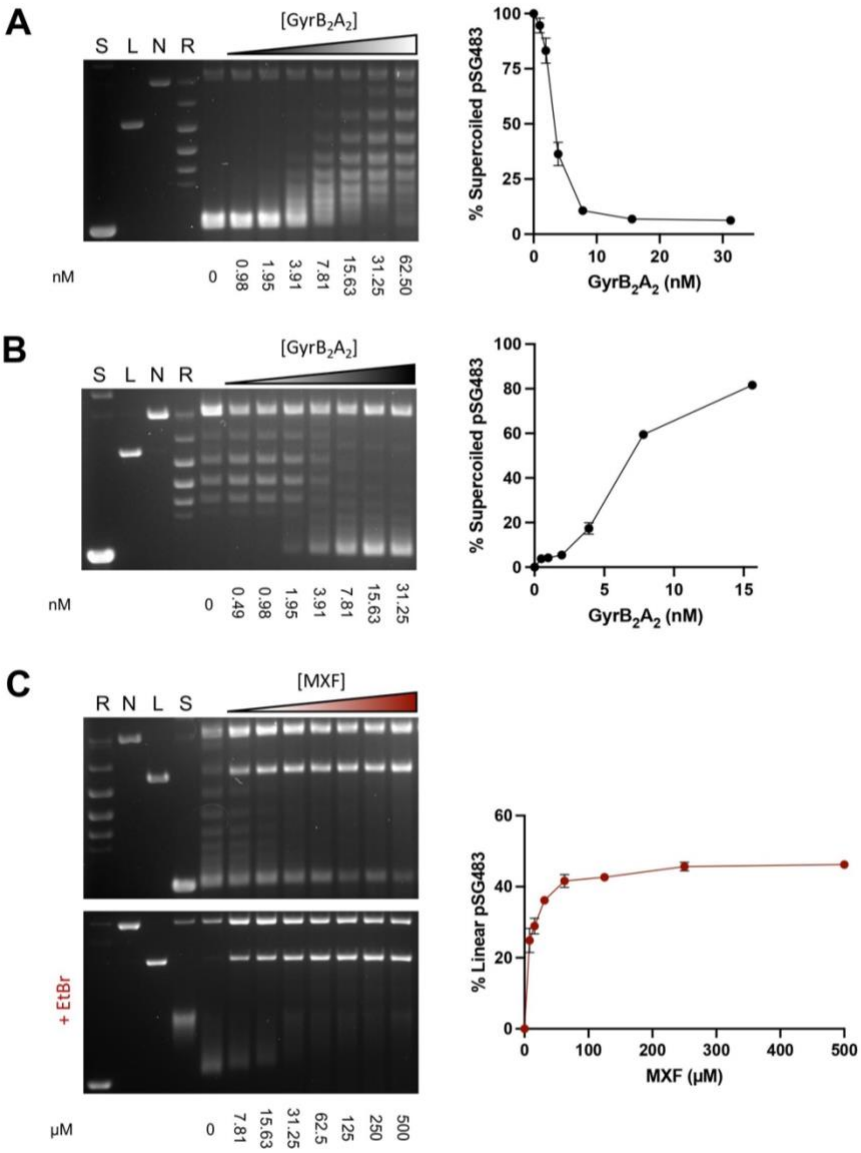
3.2.1 *M. tuberculosis* gyrase is functional for DNA relaxation and supercoiling, and is inhibited by fluoroquinolones *in vitro*

Having purified independent GyrB and GyrA subunits of *M. tuberculosis* DNA gyrase, their activity was confirmed by reconstituting the $GyrB_2A_2$ holoenzyme *in vitro* and testing for activity in DNA relaxation, DNA supercoiling, and fluoroquinolone-based DNA cleavage assays, as per **Materials & Methods 2.7**. The results confirm that the reconstituted holoenzyme functions as expected (**Figure 3.2**).

Briefly, purified GyrA and GyrB were mixed in equimolar amounts prior to 2-fold serial dilution to a desired concentration range prior to addition to the appropriate substrate; either supercoiled pSG483 for the DNA relaxation reactions, or relaxed pSG483 in the DNA supercoiling reactions performed with the supplementation of 1 mM ATP (**Figure 3.2 A and B**). *M. tuberculosis* gyrase successfully converted >90 % of the supercoiled pSG483 substrate into multiple topoisomers at a concentration of 31.25 nM (**Figure 3.2 A**). The reaction appears to reverse at 62.5 nM, potentially due to gyrase-saturated substrate leading to difficulties in strand-passage (**Figure 3.2 A**). The holoenzyme can also convert relaxed substrate into supercoiled product *in vitro* in an ATP-dependent manner; > 80% of the plasmid substrate was converted into supercoiled product as demonstrated by the appearance of the supercoiled band over the course of the titration (**Figure 3.2 B**). Both reactions are comparable to previously demonstrated ATP-independent relaxation and ATP-dependent supercoiling capabilities of *M. tuberculosis* gyrase (Karkare et al., 2012).

The relaxation reaction can be used to demonstrate the effects of inhibitors, as is demonstrated in **Figure 3.2 C** wherein the addition of Moxifloxacin successfully traps cleavage complexes, causing the production of a high percentage of linearised plasmid DNA. A constant concentration of 31.25 nM DNA gyrase was used alongside supercoiled plasmid (relaxation/cleavage reaction) against a titration of Moxifloxacin (**Figure 3.2 C**). The high-affinity fluoroquinolone antibiotic sterically hinders the resealing of DNA at the G-gate by pi-stacking between the exposed bases during DNA cleavage (Aldred *et al.*, 2016; Blower *et al.*, 2016). This results in release of cleaved linear form DNA upon subsequent proteinase K digestion of the poisoned fluoroquinolone-gyrase complexes.

Figure 3.2 Reconstituted *M. tuberculosis* gyrase is functional *in vitro*



(**A**) $GyrB_2A_2$ ATP-independent DNA relaxation reaction titration (removal of plasmid DNA negative supercoils); (**B**) $GyrB_2A_2$ ATP-dependent DNA supercoiling reaction titration (generation of plasmid DNA negative supercoils); (**C**) Moxifloxacin induced DNA cleavage assay with drug titrated against constant $GyrB_2A_2$ (31.25 nM) and Supercoiled (S) plasmid DNA (12.5 nM). Control lanes represent Supercoiled (S), Linear (L), Nicked (S), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage supercoiled DNA (**A** and **B**) or percentage linear DNA (**C**, +EtBr) per lane (obtained by densitometry). Concentration gradients are quantified below each gel. Assays are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

3.2.2 The ParDE1 complex has minimal effect on gyrase activity

Due to the difficulty of cloning *parE1* (*rv1959c*) alone, *parE1* and the cognate antitoxin *parD1* (*rv1960c*) were cloned into a Duet vector (Novagen) (Tolia and Joshua-Tor, 2006) (**Table 2.5**) to circumvent toxicity issues. This allowed for the expression and purification of the ParDE1 protein complex, as shown in **Figure 3.1**. The ParE1 toxin could not be subsequently separated from the complex during purification, therefore, the ParDE1 complex was tested against DNA gyrase for any inhibitory effects.

At the highest concentration tested (10 µM), ParDE1 caused a small (~ 4%) amount of linear product (**Figure 3.3 A**). This potentially indicates that over the course of the assay a small amount of the toxin component, ParE1, may have been released (**Figure 3.3 A**). When taking the error into account, however, this amount of linear product is only just slightly greater than the basal amount and could simply be attributed to a high amount of protein in the reaction. As was expected, the effects of the ParDE1 protein complex are minimal to negligible as the ParE1 toxin is likely effectively sequestered by the ParD1 antitoxin.

3.2.3 Effects of the ParDE2 system on gyrase activity

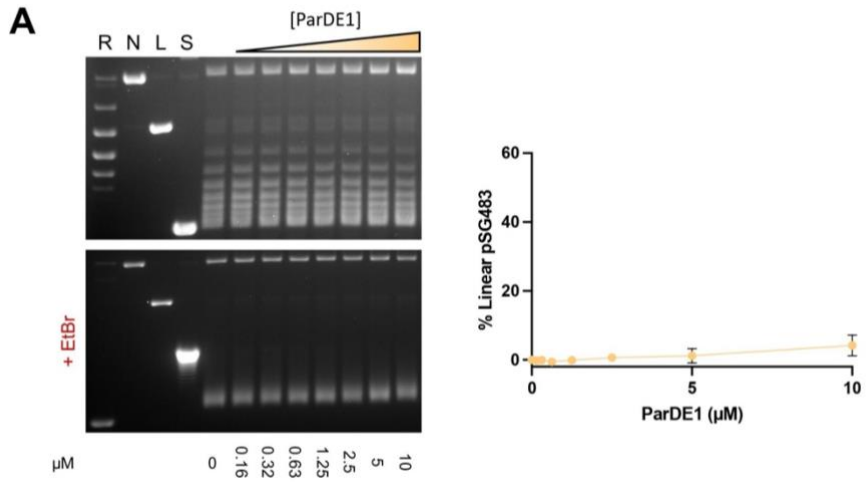
3.2.3.1 Purification of ParDE2 system components for biochemical analysis

Like the *parDE1* system, *parE2* (*rv2142c*) and *parD2* (*rv2142A*) were cloned into a Duet vector (Novagen) (Tolia and Joshua-Tor, 2006) (**Table 2.5**) for co-expression and purification of the ParDE2 protein complex. Similarly, the protein complex could be expressed and purified, however, during purification it was possible to separate and purify the ParE2 toxin from the complex (**Materials & Methods 2.5.9.1**) (**Figure 3.4 A and B**). During anion exchange of the cleaved ParDE2 complex, ParE2 was isolated in a distinct chromatographic peak. SDS-PAGE analysis (**Figure 3.4 A**) shows the purified 12.18 kDa toxin alongside the ParDE2 complex; the sample was determined to be purified from the 7.78 kDa ParD2 antitoxin by MALDI-ToF mass spectrometry due to the abundance of the ParE2 peak and lack of an appropriate molecular weight ion for the antitoxin (**Figure 3.4 B**). The ParD2 antitoxin could be independently expressed and purified as a recombinant protein, as detailed in **Materials & Methods 2.5.9**.

Prior to testing the components of the ParDE2 system against DNA gyrase, ParE2 was analysed for any basal levels of nuclease function, or for potential nuclease function due to a contaminant protein in the ParE2 sample (**Figure 3.4 B**). The basal level of linearisation caused by the purified ParE2 sample was calculated at each titration concentration, thereby allowing data from future ParE2 experiments to be effectively normalised. Due to the activity seen in **Figure 3.4 C**, ParE2 was further analysed for nuclease activity as seen in **Figure 3.4 D**. Both supercoiled and linear DNA substrates were used to determine endonuclease or exonuclease activity respectively. Due

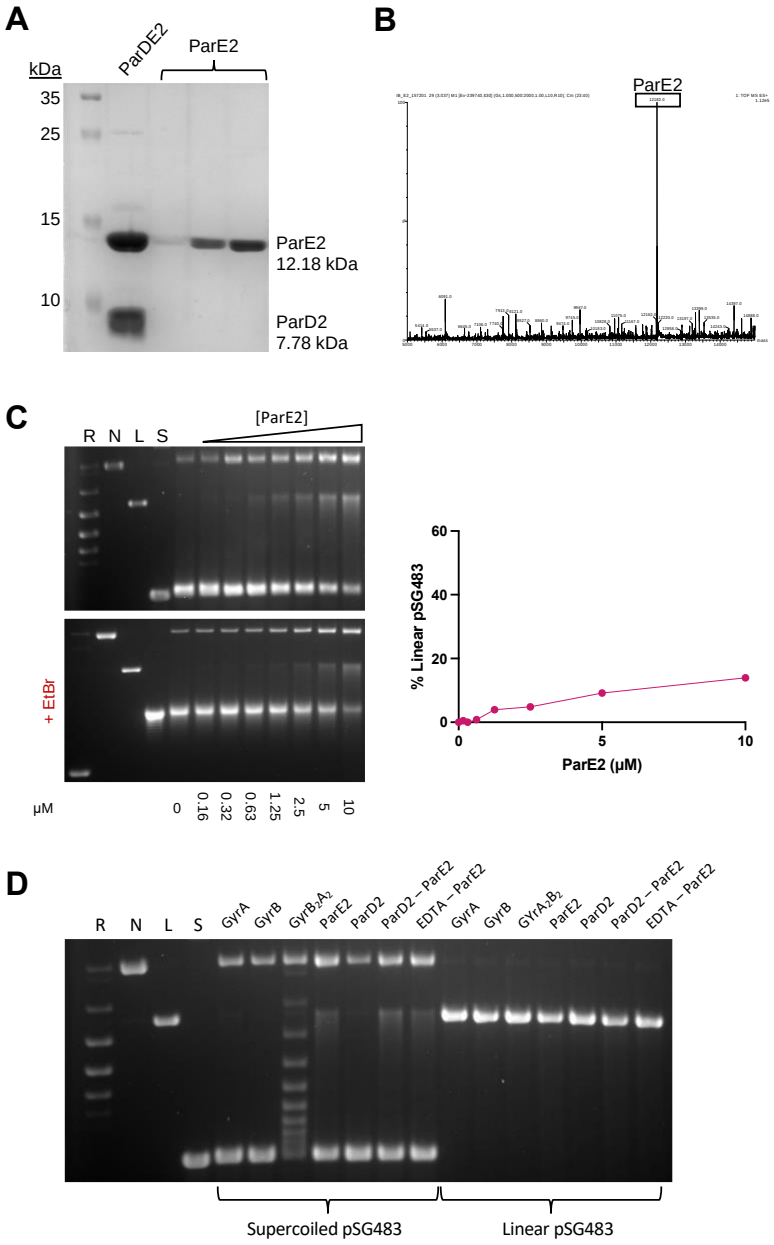
to the persistence of linearisation from the ParE2 sample in the presence of 10-fold excess ParD2, it is possible a small amount of metal-dependent endonuclease is present as a contaminant (highlighted by the reduced activity in the presence of EDTA) (**Figure 3.4 C**).

Figure 3.3 ParDE1 induced cleavage assay



(**A**) ParDE1 induced DNA cleavage assay with ParDE1 titrated against constant $GyrB_2A_2$ (31.25 nM) and Supercoiled (S) plasmid DNA (12.5 nM). Control lanes represent Supercoiled (S), Linear (L), Nicked (N), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage linear DNA per lane (**A**, +EtBr, obtained by densitometry). Concentration gradients are quantified below the gels. Assays are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

Figure 3.4 ParE2 isolation and nuclease analysis



(**A**) SDS-PAGE analysis of purified ParDE2 and isolated ParE2 fractions from anion exchange that were subsequently pooled; (**B**) MALDI-ToF Mass Spectrometry analysis of the isolated ParE2 toxin; (**C**) Initial nuclease analysis of ParE2 by titration against supercoiled plasmid; **D**) Continued nuclease analysis of ParE2 (1.25 μM). Endonuclease analysis is performed against supercoiled plasmid (left) and exonuclease analysis is performed against linear plasmid (right). Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage linear DNA (+EtBr) per lane. Concentration gradients are quantified below each gel, where appropriate. Assays presented are a single repeat. Each gel is presented with Supercoiled (S), Linear (L), Nicked (N) and Relaxed (multiple topoisomers) (R) plasmid DNA. ParE2 protein concentration per lane is presented below the agarose gels.

3.2.3.2 ParE2 induced linearisation is gyrase dependent

Similar to ParDE1, the purified ParDE2 complex (**Figure 3.4 A**) was also tested for inhibition of *M. tuberculosis* gyrase activity in a DNA relaxation reaction (**Figure 3.5 A**). The ParDE2 protein caused minimal, but above the basal rate, linearisation of pSG483 (~ 5 – 7 % at the highest concentrations of 5 and 10 µM respectively) (**Figure 3.5 A**). This may be due to the dissociation of the complex over the course of the assay, and at such a high concentration only a small amount of ParE2 release could be required to obtain this result. Alternatively, the sample may have a small fraction of free toxin as a result of the purification process and/or minor contamination from the isolated ParE2 fraction.

Next, ParE2 was tested for impact within an *M. tuberculosis* gyrase DNA relaxation reaction (**Figure 3.5 B**). At the highest concentration tested, ParE2 caused the persistence of ~50% linearised pSG483 indicating a good affinity for *M. tuberculosis* gyrase and ability to stall the reaction in the cleavage state. This was substantially above the level of linearisation seen for the ParDE2 complex (**Figure 3.5 A and B**). This result also corroborates, and indeed builds upon, what has been previously shown for ParE2; Gupta *et al.* (2016) demonstrated the ability of isolated *M. tuberculosis* ParE2 toxin to trap cleavage complexes in a DNA supercoiling reaction. Taken together, these results indicate that ParE2 recognises the structure of the gyrase complex in the cleavage state (consistent between both supercoiling and relaxation reactions) and has the ability to trap the reaction in this state. The exact mechanism of ParE2-dependent gyrase poisoning requires further study. This also indicates that ParE2 may have affinity for the cleavage complex formed during a DNA decatenation reaction, but it is unclear yet which of the reactions are preferred, if any.

Finally, purified ParD2 showed no significant increase in linear pSG483 in the DNA relaxation reaction (**Figure 3.5 C**). Collectively, these results demonstrate the ability of the ParE2 toxin to inhibit DNA gyrase by trapping the cleavage complex, causing the persistence of linearised DNA. The results obtained in the ParDE2 assay (**Figure 3.5 A**) are likely caused by release or contamination of the ParE2 toxin as ParD2 alone had no effect on the relaxation reaction (**Figure 3.5 C**).

Figure 3.5 ParDE2 system induced cleavage assays

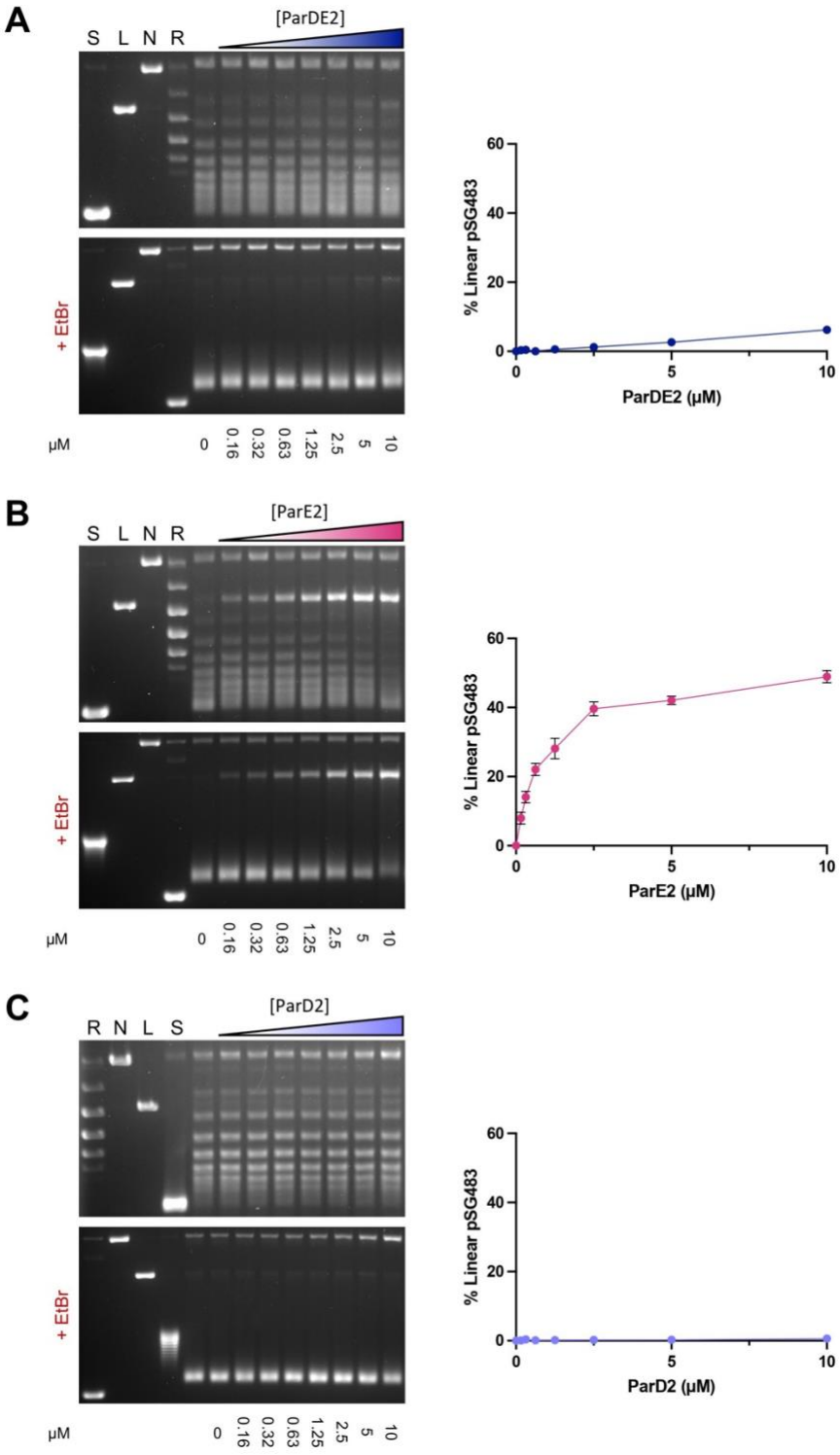


Figure legend overleaf

Figure 3.5 ParDE2 system induced cleavage assays

(**A**) ParDE2 induced cleavage assay; (**B**) ParE2 induced cleavage assay; (**C**) ParD2 induced cleavage assay. ParDE system components were titrated against constant $GyrB_2A_2$ (31.25 nM) and Supercoiled (S) plasmid DNA (12.5 nM). ParDE protein concentration per lane is presented below the agarose gels. Control lanes represent Supercoiled (S), Linear (L), Nicked (S), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage linear DNA per lane (+EtBr, obtained by densitometry). Concentration gradients are quantified below each gel. Assays are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

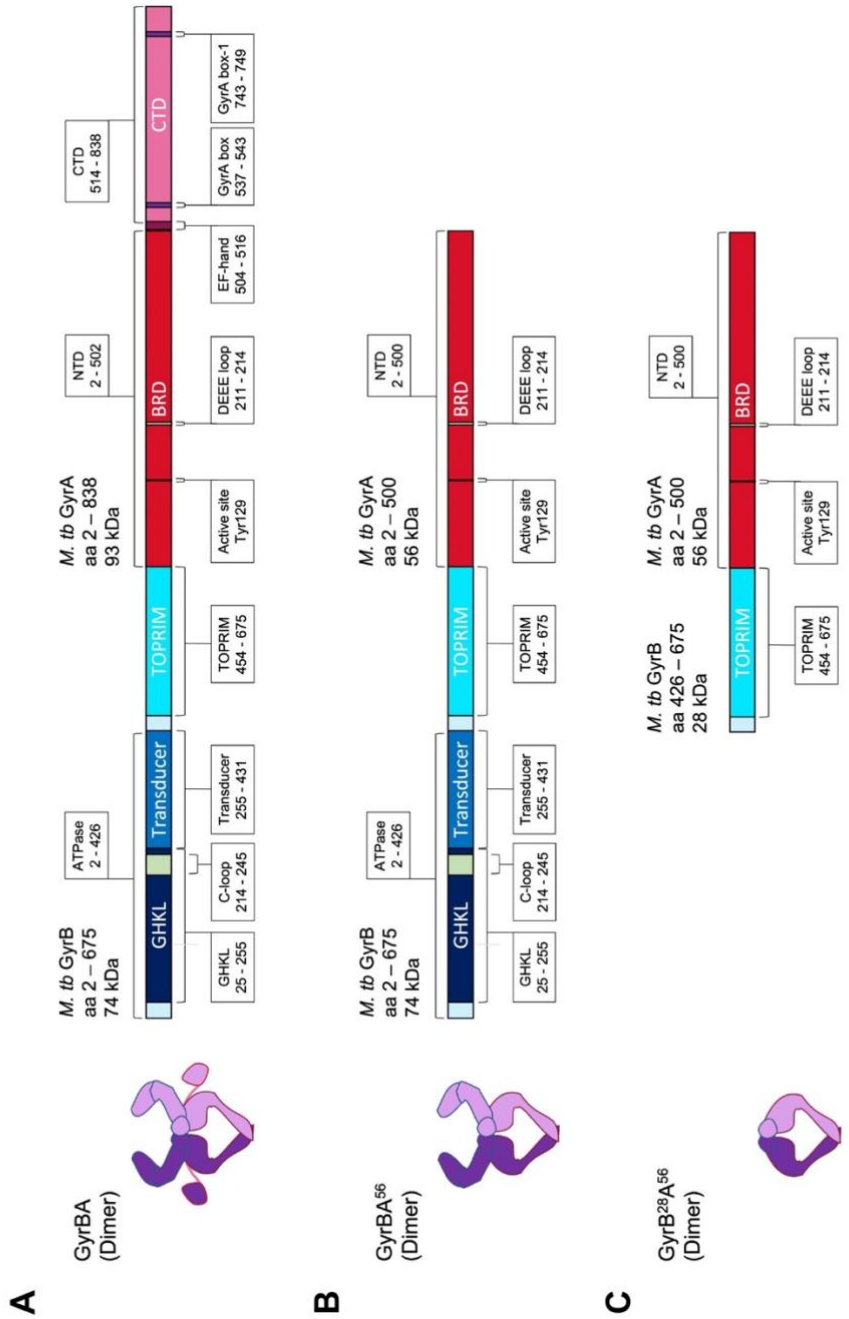
3.3 Analysis of the effects of ParE2 on *M. tuberculosis* gyrase fusion proteins – guidance for structural studies

Structural studies will be an important avenue in advancing our understanding of how exactly the ParE toxins trap gyrase cleavage complexes. Not only would this be the first example of a ParE:gyrase complex structure, but it may also provide new and exciting insights into the mechanisms of DNA gyrase and a novel basis for downstream drug design. In order to progress this project toward this goal, three *M. tuberculosis* gyrase fusion proteins were selected due to either their previous use in structural work, or having matching construct design to solved gyrase fusions from other species (Cabral *et al.*, 1997; Blower *et al.*, 2016; Bax *et al.*, 2019; Hobson and Berger, 2019; Petrella *et al.*, 2019).

The full-length DNA gyrase fusion (GyrBA) is comprised of the 74 kDa GyrB subunit (amino acids 2 – 675) fused at its C-terminus to the N-terminus of the 93 kDa GyrA subunit (amino acids 2 – 838) (**Figure 3.6 A**). In theory, this protein is capable of completing all of the functions of the gyrase holoenzyme and is analogous to the *E. coli* gyrase fusion protein used in the highly successful and recent Cryo-EM studies (Petrella *et al.*, 2019). This *M. tuberculosis* full-length fusion has the potential to be an excellent option for similar studies with the addition of the ParE toxins.

The two smaller fusions are truncations of the full-length fusion. The GyrBA^{56} fusion protein is truncated to remove the C-terminal domain of GyrA, leaving the 56 kDa N-terminal BRD region (amino acids 2 – 500) intact, fused C-terminally to the full-length GyrB subunit (**Figure 3.6 B**). The GyrB^{28}A^{56} is a truncated version of this, removing the N-terminal GHKL and Transducer domains of GyrB to leave the 28 kDa TOPRIM fused to the BRD of GyrA (**Figure 3.6 C**). The functionality of these two truncated fusion proteins in a DNA relaxation assay (our model for studying the effects of ParE toxins) was, as yet, unknown. The GyrBA^{56} fusion may prove useful in Cryo-EM studies or in crystallography; notably Petrella *et al.* (Petrella *et al.*, 2019) used an almost identical, single kDa extended, *M. tuberculosis* GyrBA^{57} fusion in crystallographic studies, publishing two high resolution structures (6GAV; 6GAU). This single kDa addition to the GyrA subunit should not impact structure or function of this fusion as it is part of a linker region to the GyrA CTD. The GyrB^{28}A^{56} will be an excellent crystallographic start-point due to the number of crystal structures deposited in the PDB and published of this fusion (Blower *et al.*, 2016; Imai *et al.*, 2022).

Figure 3.6 Schematic of *M. tuberculosis* gyrase fusion proteins



(**A**) Cartoon representation and schematic of the full-length GyrBA fusion protein. Residues 2 – 675 of GyrB are fused at the N-terminus to residues 2 – 838 of GyrA at the C-terminus; (**B**) Cartoon representation and schematic of the GyrBA^{56} fusion protein. Residues 2 – 675 of GyrB are fused at the N-terminus to residues 2 – 500 of GyrA (N-terminal 56 kDa) at the C-terminus; (**C**) Cartoon representation and schematic of the GyrB^{28}A^{56} fusion protein. Residues 426 – 675 of GyrB (N-terminal 28 kDa) are fused at the N-terminus to residues 2 – 500 of GyrA (N-terminal 56 kDa) at the C-terminus.

3.3.1 Expression, purification, and activity of three gyrase fusion proteins in the relaxation reaction

All three fusion proteins were successfully expressed and purified as per **Materials & Methods 2.4.1 and 2.5**. The general protein purification procedure is detailed in **Figure 3.1**; a significant difference to assist in the purification of the gyrase fusion proteins was the addition of a HiTrap Heparin HP column after tag cleavage and 'Ortho-Ni' step. This allowed for higher specificity purification of DNA-binding proteins (Bolten, Rinas and Scheper, 2018).

The purified fusion proteins were tested for activity in DNA relaxation reactions, following **Materials & Methods 2.7**, as per the holoenzyme. The GyrBA full-length fusion showed comparable relaxation activity to the $GyrB_2A_2$ holoenzyme across the same titration (**Figure 3.7 A**; **Figure 3.2 A**). There was no substantial difference in the percentage of removed supercoiled DNA at 31.25 nM of holoenzyme or fusion, indicating highly similar activity.

The truncated fusions, $GyrBA^{56}$ and $GyrB^{28}A^{56}$, showed no ability to relax supercoiled DNA (**Figure 3.7 B and C**). Rather, both fusion proteins caused marked increases in linear and nicked form DNA, albeit at different rates. The $GyrBA^{56}$ fusion converted ~8 % of the supercoiled substrate into linear form, while nicked form increased to ~50 % of the product DNA at the top concentration of 250 nM (**Figure 3.7 B**). In contrast, the $GyrB^{28}A^{56}$ converted only ~4 % of the supercoiled substrate into linear form, while nicked form increased to ~65 % of the product DNA at the top concentration of 250 nM (and is not significantly different from the nicking seen at 31.25 nM) (**Figure 3.7 C**).

Taken together, these results informed us as to the appropriate concentration to use when testing inhibitors against the fusion proteins. In the case of the full-length GyrBA fusion protein, the same concentration as used in the holoenzyme assays at 31.25 nM is appropriate, as maximal removal of supercoils was seen at this concentration. This also allows for direct comparison between the two full-length gyrase forms. To maintain this option for direct comparison, 31.25 nM was also selected for both truncated fusion proteins. As previously mentioned, the $GyrBA^{56}$ maintains reasonable low levels of DNA cleavage at this concentration in the relaxation reaction, allowing us to normalise against the low background and visualise the effects of the ParE toxin. The $GyrB^{28}A^{56}$, however, demonstrated large amounts of cleavage at 31.25 nM with most of the product becoming nicked DNA. It will therefore be interesting to determine whether the ParE2 toxin, being capable of producing linear species DNA in a gyrase-dependent manner, will cause the $GyrB^{28}A^{56}$ to preferentially produce linear species also, stabilising the symmetric cleavage or not.

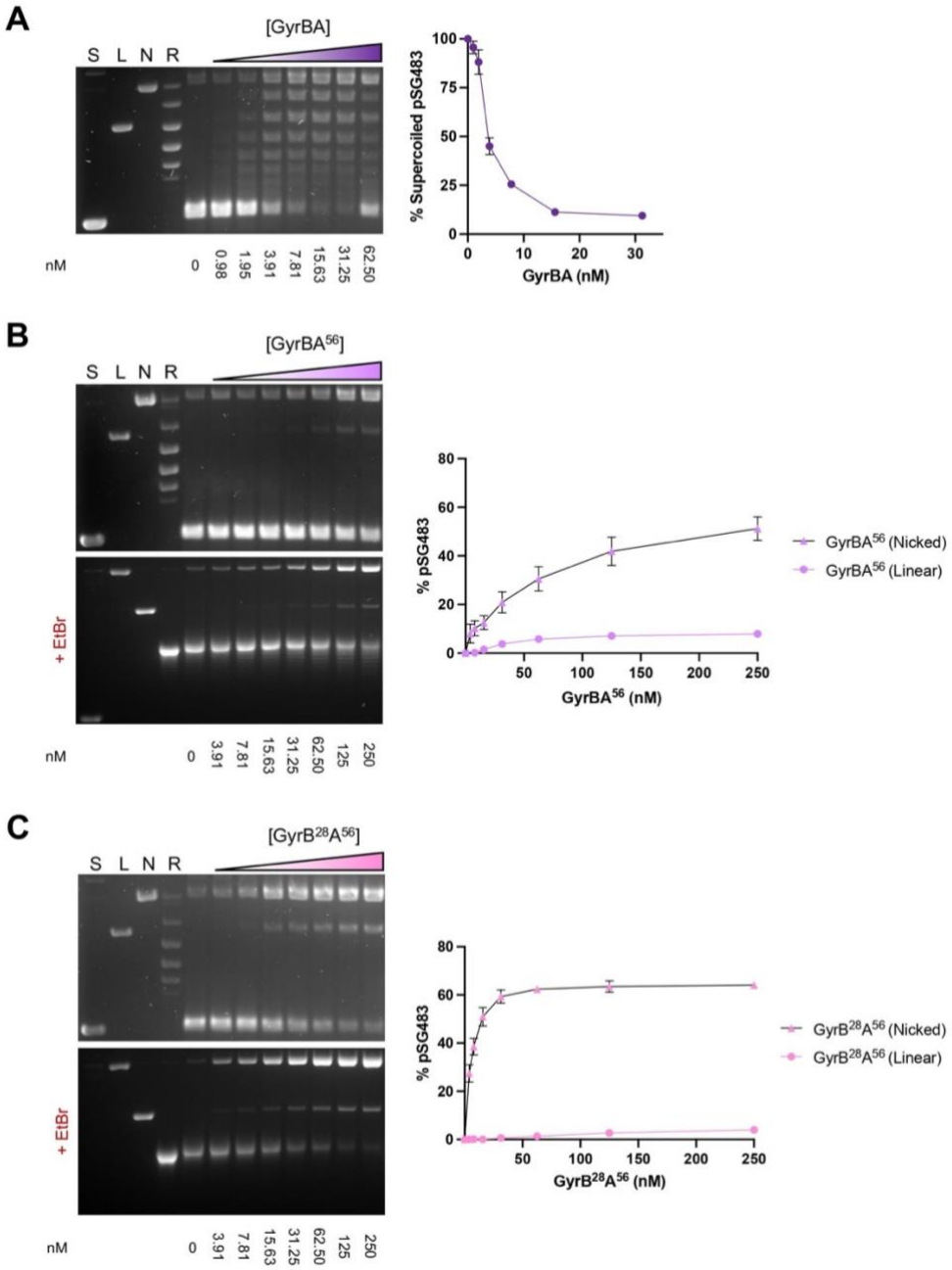
Alternatively, experiments with the truncated fusions will also offer potential insights into the interactions between the gyrase subunits and domains, and the ParE2 toxin. We may indeed see large differences when domains are removed, indicating a potential role in stabilising the suspected toxin:gyrase interaction that causes the linearisation shown previously for the holoenzyme (**Figure 3.5 B**).

3.3.2 ParE2 traps gyrase fusion cleavage complexes

Having characterised the gyrase fusions, it was then possible to test the fusions against a titration of the purified ParE2 toxin, as shown in **Figure 3.4 A** and tested against the gyrase holoenzyme in **Figure 3.5 B**. Unsurprisingly, ParE2 caused a significant increase in linear product when tested against a constant concentration of GyrBA in the DNA relaxation reaction (**Figure 3.8 A**). This was expected due to comparable activity between the full-length fusion and holoenzyme (**Figure 3.2 A**, **Figure 3.7 A**). Interestingly, ParE2 caused a markedly higher percentage of linear product in the assay with the full-length GyrBA fusion at lower concentrations, when compared with the holoenzyme; ~50 % of the product became linear species DNA at 0.3125 µM (**Figure 3.8 A**) whereas 50 % linearisation was not achieved in the $GyrB_2A_2$ DNA relaxation assay (**Figure 3.5 B**). A peak linear species percentage was achieved in the GyrBA assay at a ParE2 concentration of 2.5 µM, causing ~64 % linearised pSG483 (**Figure 3.8 A**). This does however tail off slightly in the assay at the higher concentrations but remains stable around 60 %. Unfortunately, repeats were not available in this assay due to shortages of ParE2 and difficulties in purification, therefore only one replicate was completed. Given the previous repeatability of the effects of ParE2 on DNA gyrase (**Figure 3.7 B**), it is expected that this result would also be replicated.

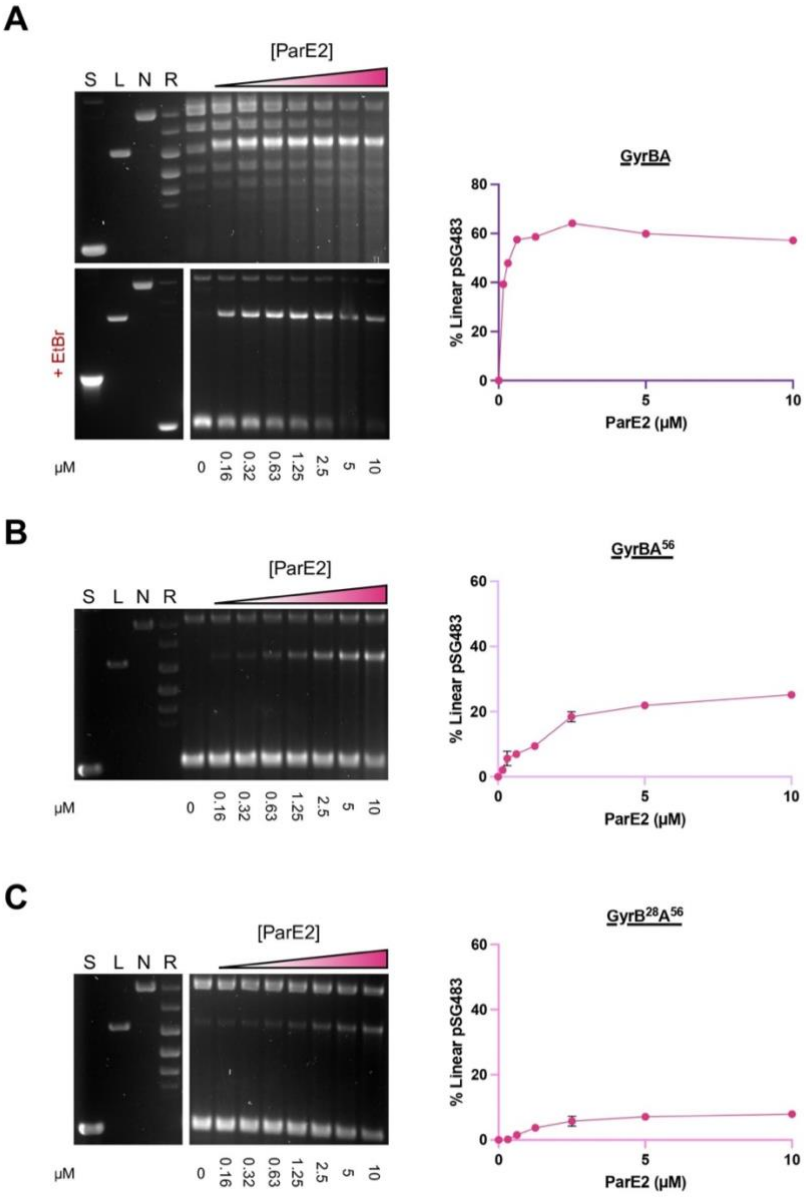
We initially expected that all three fusion proteins would be behave similarly due to the conservation of the breakage-reunion domain and TOPRIM (**Figure 3.6**). Interestingly, when testing ParE2 against the truncated gyrase fusions, a far lower percentage linear product was achieved (**Figure 3.8 B and C**). In the case of $GyrBA^{56}$, when only the GyrA CTD is removed, linearised DNA reached only ~21 % at 10 µM ParE2 (**Figure 3.8 B**). Further to this, for $GyrB^{28}A^{56}$, when both the GyrA CTD and GyrB ATPase are removed, linear product dropped to ~8% at 10 µM ParE2 (**Figure 3.8 C**). These observations are potentially due to progressive loss of essential interfaces for ParE2-gyrase interaction, or because of the lack of functionality of the enzymes in these experiments; neither truncated fusion has been shown to be capable of relaxation under our conditions (**Figure 3.7**).

Figure 3.7 Activity of the *M. tuberculosis* gyrase fusion proteins



(**A**) GyrBA ATP-independent DNA relaxation reaction titration (removal of negative pSG483 supercoils); (**B**) GyrBA^{56} ATP-independent DNA relaxation reaction titration; (**C**) GyrB^{28}A^{56} ATP-independent DNA relaxation reaction titration. Control lanes represent Supercoiled (S), Linear (L), Nicked (S), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage Supercoiled DNA (**A**) or percentage Nicked and Linear DNA (**B** and **C**, +EtBr) per lane (obtained by densitometry). Concentration gradients are quantified below each gel. Assays are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

Figure 3.8 ParE2 induced cleavage assays with gyrase fusion proteins



(**A**) ParE2 induced DNA cleavage assay with GyrBA (single repeat); (**B**) ParE2 induced DNA cleavage assay with GyrBA^{56}; (**C**) ParE2 induced DNA cleavage assay with GyrB^{28}A^{56}. ParE2 was titrated against constant gyrase protein (31.25 nM) and Supercoiled (S) plasmid DNA (12.5 nM). ParE2 protein concentration per lane is presented below the agarose gels. Control lanes represent Supercoiled (S), Linear (L), Nicked (S), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage linear DNA per lane (+EtBr, obtained by densitometry). Concentration gradients are quantified below each gel. Assays in **B** and **C** are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

3.4 Estimation of protein molecular weights and Stokes radii via analytical size exclusion chromatography

3.4.1 Superose 6 10/300 GL calibration for analytical purposes

As a precursor to effectively analysing the interactions between gyrase subunits, and the gyrase proteins and ParE toxins, an appropriate gel filtration column was selected based on a combination of purification and analytical capabilities, alongside molecular weight (M_r, kDa) range. The Superose 6 10/300 GL column was equilibrated and calibrated as per **Materials & Methods 2.6.4**. Two separate calibration runs were performed as per manufacturer's instructions to maximise resolution; the constitution of each calibration mix is detailed in **Figure 3.9 A** alongside the sample M_r and R_{st} values and resultant elution volumes (V_e).

Briefly, to estimate the M_r (kDa) and Stokes Radii (R_{st}, Å) of samples, elution volumes (V_e) are used to calculate the partitioning coefficient (K_{av}) using the following equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

Where V_c is the geometric column volume and V_o is the column void volume. The K_{av} values obtained from the calibration samples are then plotted against $Log_{10}(M_r, kDa)$ or $Log_{10}(Rst, Å)$ of the respective sample. Calibration run chromatograms and subsequent plots for M_r and R_{st} are presented in **Figure 3.9 B – D** respectively. The equation of the line for both plots is subsequently used to calculate K_{av} from the V_e of a sample, which can in turn be used to estimate either M_r or R_{st}.

3.4.2 Analytical sizing analysis of gyrase subunits and fusion proteins

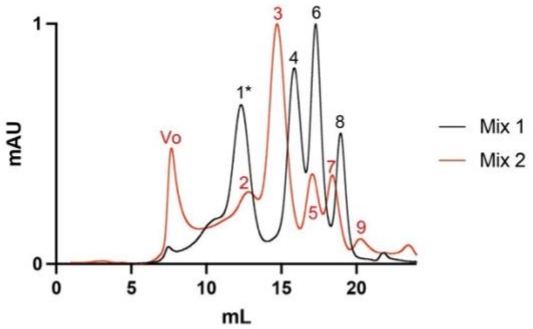
Analytical sizing analysis of protein samples was performed as per **Materials & Methods 2.6.4.2**. Briefly, a 150 µL sample of each protein, diluted from stocks using buffer S500 to approximately 1 mg/mL, was manually injected into a 100 µL capillary loop on an Åkta Pure system (Cytiva). This sample was subsequently injected onto the pre-equilibrated and calibrated Superose 6 10/300 GL column and run at 0.5 mL/min. Samples were run in triplicate and a representative UV trace is presented in each respective figure. Peak elution volumes (V_e) were deduced using the peaks function in the evaluation software for the Åkta system (Unicorn™, Cytiva). V_e was subsequently used to calculate K_{av} to estimate M_r and R_{st}.

Figure 3.9 Calibration of the Superose 6 10/300 GL Gel Filtration Column

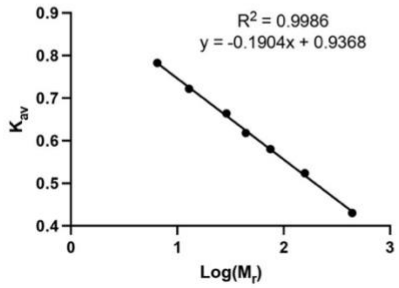
A

Protein (Calibration Mix)	M_r (kDa)	R_{st} (Å)	V_e (mL)
Thyroglobulin (1)*	669	85	12.46
Ferritin (2)	440	61	14.4
Aldolase (1)	158	48.1	15.9
Conalbumin (2)	75	36.4	16.81
Ovalbumin (1)	44	30.5	17.42
Carbonic anhydrase (2)	29	23	18.16
Cytochrome C (1)	12.9	17.8	19.09
Aprotinin (2)	6.5	13.5	20.07

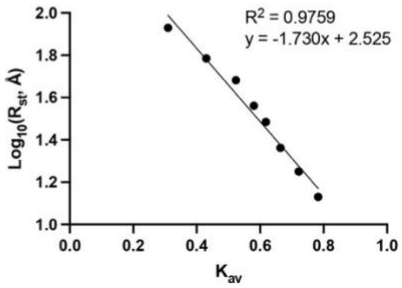
B



C



D



(**A**) Table of combined low and high molecular weight calibration kit proteins (Cytiva) with pre-calculated respective molecular weights (M_r) and Stokes Radii (R_{st}). Elution volume (V_e, mL) results from gel filtration of Mix 1 and Mix 2 (indicated by brackets); (**B**) Overlaid chromatograms of Calibration Mix 1 (black) and Mix 2 (red). Peaks are observed for Thyroglobulin (1), Ferritin (2/3), Aldolase (4), Conalbumin (5), Ovalbumin (6) Carbonic anhydrase (7), Cytochrome C (8), Aprotinin (9). V_e for Ferritin is calculated from peak 3. V_o – Column void volume results from Mix 2; (**C**) Linear plot of K_{av} against Log(M_r) presented with equation of line and R^2 value; (**D**) Linear plot of $Log_{10}(R_{st})$ against K_{av} presented with equation of line and R^2 value. * Thyroglobulin exhibits a non-linear relationship with the selected column and is omitted from M_r calculations, however, can be included in R_{st} calculations.

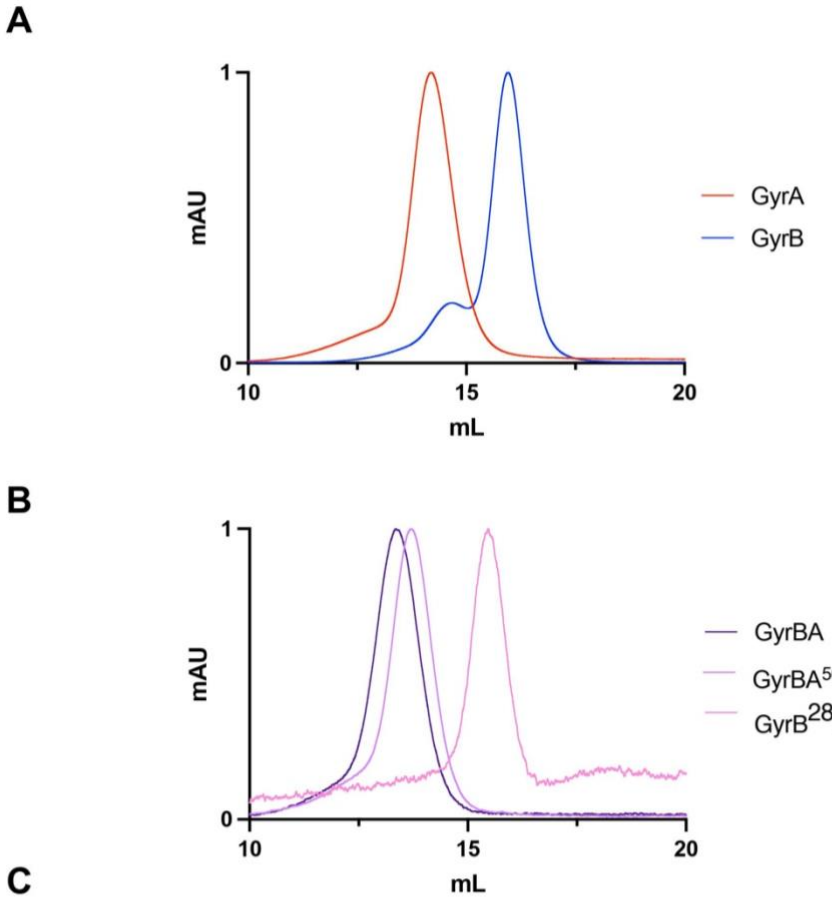
3.4.2.1 Multimeric state of *M. tuberculosis* gyrase subunits

Having calibrated the column, target proteins were characterised by analytical SEC (**Figure 3.10**). The *M. tuberculosis* gyrase subunit A (GyrA) protein eluted in a single peak, while GyrB produced a major peak alongside a minor peak at a slightly lower V_e, **Figure 3.10 A**. The amino acid sequences of GyrA (2 – 838) and GyrB (2 – 675) can be used to calculate the respective M_r values of 92.14 kDa and 73.96 kDa respectively (**Figure 3.10 C**). Initial M_r calculations and predictions were performed using the monomer M_r; this subsequently allows prediction of the multimeric state once the ratio of observed/calculated is obtained (**Figure 3.10 C**).

The observed M_r for GyrA was 531.53 kDa, significantly greater than the monomeric M_r for this protein; the multimeric state was subsequently calculated to be 5.77 (**Figure 3.10 A and C**), thus indicating that the multimeric state is closer to 6, rather than the predicted dimer (ratio of 2). In contrast, the observed M_r for GyrB was 142.56 kDa, which returned a multimeric ratio of 1.93; this is highly suggestive that GyrB, or the major peak, is a dimer (**Figure 3.10 A and C**). Size exclusion chromatography resins rely more on the shape of proteins rather than the M_r (O'Fágáin, Cummins and O'Connor, 2011), and therefore Stokes radius (R_{st}) (hydrodynamic radius) can be a much better estimate for the size of a given sample (**Figure 3.9 D**). One difficulty in calculating the Stokes radius is the need for a protein structure, however, gyrase protein structure has been well studied and values can subsequently be calculated (Piton *et al.*, 2010; Agrawal *et al.*, 2013; Blower *et al.*, 2016; Petrella *et al.*, 2019; Vanden Broeck *et al.*, 2019).

Stokes radii for both GyrA and GyrB can be estimated using crystal structures deposited in the PDB; GyrA was calculated, using an *E. coli* structure, to possess a radius of 66.5 Å, (PDB: 6RKW, edited - PyMol), whereas GyrB could only be partly calculated due to a truncated *M. tuberculosis* crystal structure (PDB: 3ZKD) (Agrawal *et al.*, 2013) which provided a value of 36.6 Å. Both PDB crystal structures are dimeric and will therefore indicate dimeric state by a ratio of closer to 1 (**Figure 3.10 C**). Indeed, considering the Stokes radius, and therefore the shape and size of the GyrA protein, the prediction of solution state is resolved; the observed R_{st} was determined to be 63.32 Å and provided a ratio of 0.95. This strongly supports that GyrA exists as a dimer in solution and highlights the potential inaccuracy of using M_r as an indicator of multimeric state in SEC. Contrastingly the observed R_{st} for GyrB was 1.12 times larger than the calculated value at 41.05 Å, however, we must consider that the model used for calculating the predicted R_{st} of 36.6 Å is GyrB 1 – 426, missing the N-terminal 249 residues (TOPRIM domain) (Agrawal *et al.*, 2013). Taken together with the Mr prediction of a dimer, (**Figure 3.10 A and C**), we can conclude that GyrB also exists as a dimer in solution, likely with the ATPase dimerised as in PDB: 3ZKD.

Figure 3.10 Analytical SEC of *M. tuberculosis* gyrase proteins



C

Sample	Calculated M_r (kDa)	Observed M_r (kDa)	Observed M_r/ Calculated M_r	Calculated R_{st} (Å)	Observed R_{st} (Å)	Observed R_{st}/ Calculated R_{st}
GyrA	92.14	531.53	5.77	66.5[1]	63.32	0.95
GyrB	73.96	142.56	1.93	36.6[2]	41.05	1.12
GyrBA	166.10	984.75	5.93	76.8[3]	77.58	1.01
GyrBA^{56}	129.89	785.87	6.05	66[4]	72.03	1.09
GyrB^{28}A^{56}	83.88	207.63	2.48	49.2[5]	46.46	0.94

(**A**) Analytical SEC traces of purified *M. tuberculosis* GyrA (red) and GyrB (blue); (**B**) Analytical SEC traces of purified *M. tuberculosis* fusion proteins GyrBA (dark purple), GyrBA^{56} (lilac), and GyrB^{28}A^{56} (pink); (**C**) Table of gyrase protein molecular weight (M_r) and Stokes Radius (R_{st}) calculations, observations, and comparisons. Comparison of observed/calculated is coloured green if within 10 % of the predicted ratio, yellow if > 10 ≥ 25 %, and red if > 25 %. Calculated R_{st} values were generated using HullRad (Fleming and Fleming, 2018) and structures from the PDB: [1, 3, 6] 6RKW; [2] 3ZKD; [5] 5BS8. Chromatograms are representative of duplicate data and are normalised between 0 – 1 for presentation and comparison. Graphs are cropped to the appropriate scale (10 – 20 mL).

3.4.2.2 Multimeric state of *M. tuberculosis* gyrase fusion proteins

Gyrase fusion proteins, GyrBA, GyrBA^{56}, and GyrB^{28}A^{56} were also subjected to analytical sizing analysis as previously described. The combination of M_r and R_{st} calculation and prediction was used to determine the multimeric solution states for these proteins.

The GyrBA fusion protein (**Figure 3.6 A**) is calculated to possess an M_r of 166.10 kDa as a monomer, and an estimated R_{st} of 76.8 Å using the dimeric PDB structure from *E. coli* (PDB: 6RKW) (Vanden Broeck *et al.*, 2019) with the DNA and drug molecule removed in manual editing (PyMol). GyrBA eluted as a single peak (**Figure 3.10 B**) with an observed M_r of 984.75 kDa, 5.93 times that of the monomer (**Figure 3.10 C**). Prior studies indicate that this protein would not exist as a hexamer, and this is supported by the observed R_{st} value of 77.58 Å (**Figure 3.10 C**), that is, 1.01 times that calculated R_{st} from the full-length gyrase fusion PDB structure. This strongly indicates that the *M. tuberculosis* GyrBA fusion exists as a dimer in solution. Structural alignments of PDB 6GAU and PDB 6RKW to generate a homodimer GyrBA fusion with an 'extremely open' GyrB ATPase (Petrella *et al.*, 2019) was calculated to possess and R_{st} of 73.3 Å. The ratio of this to the observed is 1.06, indicating close similarity to the observed R_{st}. Together, it is not possible to deduce with certainty whether or not the GyrBA fusion exists in the 'extremely open' state, whereby the C loop of GyrB and DEEE loop of GyrA interact (Petrella *et al.*, 2019), or whether the GyrB ATPase is dimerised as seen for the GyrB subunit (Agrawal *et al.*, 2013), and in the full-length *E. coli* Cryo-EM structure (Vanden Broeck *et al.*, 2019).

The GyrBA^{56} fusion protein (**Figure 3.6 B**) is calculated to possess an M_r of 129.89 kDa as a monomer, and an R_{st} of 66 Å calculated from a manually modified (PyMol) structure of the *E. coli* full-length gyrase Cryo-EM model (PDB: 6RKW) (Vanden Broeck *et al.*, 2019), edited to remove the homologous C-terminal domain of the GyrA protein. Again, the observed M_r for GyrBA^{56} is far larger than the monomer (and indeed the dimer), at 785.87 kDa, which predicts a hexameric state (**Figure 3.10 C**). Comparison of the observed R_{st}, 72.03 Å (**Figure 3.10 C**), to the edited *E. coli* structure returned an R_{st} ratio of 1.09, strongly supporting the existence of this fusion as a dimer as seen in structural work. Interestingly, this fusion has been studied structurally (Petrella *et al.*, 2019) (PDB: 6GAU) and was shown to exist in the open ATPase state as previously mentioned. The 6GAU structure was calculated to have an R_{st} value of 58.5 Å, which is further from the observed value of 72.03 Å than the edited *E. coli* structure. This may indicate that this sample exists in the closed ATPase state, rather than the extremely open state as has been previously shown for this fusion in structural studies (Petrella *et al.*, 2019).

The GyrB^{28}A^{56} fusion protein (**Figure 3.6 C**) is calculated to possess an M_r of 83.88 kDa as a monomer, and an R_{st} of 49.2 Å calculated using the dimeric crystal structure of the same sequence, PDB: 5BS8 (Blower *et al.*, 2016) (DNA and drug removed – PyMol). Analytical SEC

returned an observed Mr of 207.63 kDa, indicating a multimeric state of 2.48 times. While this is closer to the expected dimer (2 times), it is necessary to confirm the solution state using the R_{st} calculations also, especially given the accessibility of the crystal structure. The observed R_{st} for $GyrB^{28}A^{56}$ returned a ratio of 0.94 times the calculated value at 46.46 Å. The slight discrepancy may be explained by the presence of DNA in the crystal structure, while removed for calculations, this may have shifted the fusion monomers slightly at the G-gate resulting in a more open structure. Taken together, these data strongly support that the $GyrB^{28}A^{56}$ fusion protein also exists as a dimer in solution.

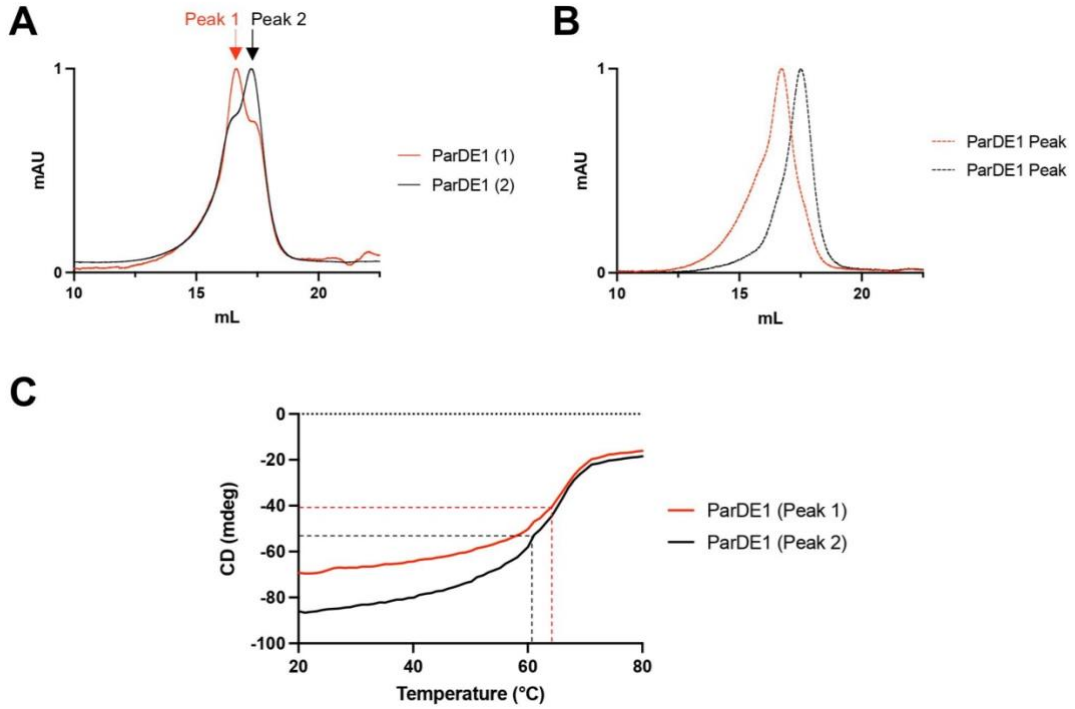
3.4.3 Biophysical analyses of the *M. tuberculosis* ParDE1 TA system

To date, ParDE structures have typically been solved as equimolar heterotetramers (**Table 1.4**) (3KXE; 6X0A; 6XRW). Biophysical analyses of ParDE systems also indicate that the affinity between the antitoxin, ParD, and cognate toxin, ParE, is in the picomolar range; that is, extremely high affinity (Muthuramalingam *et al.*, 2018). It is therefore expected that the *M. tuberculosis* ParDE systems will exhibit similar properties in that they are incredibly stable and well folded in solution when analysed via analytical SEC (Irvine, 2000), thermal denaturation (melting) (Niklasson *et al.*, 2015) and circular dichroism (CD) spectroscopy (Greenfield, 2007).

Analytical SEC of two different purifications of ParDE1 produced distinctly different, but overlapping, UV traces (**Figure 3.11 A**). It is very clear that there are two peaks present in both samples of ParDE1 in **Figure 3.11 A**; this is consistent with large scale purification of the complex whereby two peaks are usually present, both containing proteins of the appropriate size and of high purity via SDS-PAGE. This indicates the presence of ParDE1 complexes in multiple states or sizes; conversion of the respective V_e to M_r indicates a complex of 84.22 kDa to be dominant in the red curve (**Figure 3.11 A**) and a complex of 43.77 kDa to be dominant in the black curve (**Figure 3.11 A**) with the M_r of the heterotetramer ($ParD1_2ParE1_2$) calculated to be 40.76 kDa.

The two peaks present can be largely separated through SEC using a HiPrep™ Sephacryl® S-200 HR (Cytiva), creating largely purified ParDE1 samples that are theoretically of different molecular weights, or significantly different Stokes radii (**Figure 3.11 B**). Separating these ParDE1 peaks allowed for thermal stability analysis and circular dichroism to check the fold of the respective protein complexes (**Figure 3.11 C – E**). Thermal denaturation analysis (melting) (Niklasson *et al.*, 2015) provided insights into the thermal stability of the ParDE1 species isolated in peaks 1 and 2 (**Figure 3.11 C**). Estimating T_m (melting temperature) from the inflection point of the respective melt curves indicates that the species present in peak 1 is approximately 3 - 5 °C more thermostable. For peak 2 (black curve), T_m is estimated to be 60 - 61 °C, whereas for peak 1 (red curve), T_m is estimated to be 64 - 65 °C (**Figure 3.11 C**). The separate nature of the curves indicates two different ParDE1 species present with differing levels of thermal stability.

Figure 3.11 Biophysical analyses of the ParDE1 protein complex



F

Sample	Calculated M_r (kDa)	Observed M_r (kDa)	Observed M_r/ Calculated M_r	Calculated R_{st} (Å)	Observed R_{st} (Å)	Observed R_{st}/ Calculated R_{st}
Peak 1	-	84.22	-	-	34.51	-
Peak 2*	40.76[1]	43.77	1.07	29.50[2]	27.82	0.94

Figure legend overleaf

Figure 3.11 Biophysical analyses of the ParDE1 protein complex

(**A**) Analytical SEC traces of two independent purifications of the ParDE1 complex; (**B**) Analytical SEC traces of separated ParDE1 peaks from a mixed sample as presented in A; (**C**) Protein thermal denaturation curves for the separated ParDE1 samples normalised to 200 nm and 20 µM protein with background subtraction. Appropriately coloured dashed lines mark the inflection points for each curve and are used to estimate the T_m for each sample; (**D**) Circular dichroism spectroscopy scans for ParDE1 peak 1, before (solid red) and after (dashed red) melting; (E) Circular dichroism spectroscopy scans for ParDE1 peak 2, before (solid black) and after (dashed black) melting; (**F**) Table of ParDE1 protein molecular weight (M_r) and Stokes Radius (R_{st}) calculations, observations, and comparisons. Comparison of observed/calculated is coloured green if within 10 % of the predicted ratio, yellow if > 10 ≥ 25 %, and red if > 25 %. [1] M_r for $ParD1_4ParE1_2$ (heterohexamer); [2] M_r for $ParD1_2ParE1_2$ (heterotetramer); [3] R_{st} for the AlphaFold solution of $ParD1_4ParE1_2$; [4] R_{st} for the AlphaFold solution of $ParD1_2ParE1_2$. R_{st} values were generated using HullRad (Fleming and Fleming, 2018). Chromatograms are representative of duplicate data and are normalised between 0 – 1 for presentation and comparison. Graphs are cropped to the appropriate scale (10 – 22.5 mL).

Analysis of the fold of the complexes within each of the peaks via circular dichroism (CD) spectroscopy before and after melting returned highly similar data (**Figure 3.11 D and E**). Before melting, both protein complexes isolated from their respective peaks are largely helical in structure (**Figure 3.11 D and E**, 'pre' curves), which aligns well to the deposited structures for ParDE systems in the PDB (3KXE; 6X0A; 6XRW). The post-melt CD analysis indicates that both complexes regained minimal structure (**Figure 3.11 D and E**, 'post' curves) once the samples were cooled back to 20 °C from the peak melting temperature of 80 °C (where both samples were fully melted). This is indicated by the shallowness of the post curve; had more helical structure been regained the curves would have been more similar. This result can potentially be explained due to the highly helical nature of the ParD antitoxins which have likely become partially refolded. It is unlikely that the toxins regained much structure considering their more compact and globular fold.

Similar to previous analyses performed to estimate the solution states for the gyrase proteins, V_e from **Figure 3.11 A** (both peaks) was used to estimate the M_r and R_{st} for the dominant ParDE1 species in each peak. The observed M_r and R_{st} for peak 1 were 84.22 kDa and 34.51 Å respectively, and the observed M_r and R_{st} for peak 2 were 43.77 kDa and 27.82 Å respectively (**Figure 3.11 F**). The data obtained for peak 2 can be relatively easily explained as the calculated M_r for the ParDE1 complex in a heterotetrameric stoichiometry ($ParD1_2ParE1_2$), as would be consistent for ParDE systems, is 40.76 kDa. This provides an Observed/Calculated ratio of 1.07 and indicates that the dominant species within peak 2 (**Figure 3.11 A**) is indeed a heterotetrameric complex. The existence of the ParDE1 heterotetramer can be further supported using protein complex modelling via AlphaFold (Jumper *et al.*, 2021; Evans *et al.*, 2022); a $ParD1_2ParE1_2$ protein complex structure can be solved by AlphaFold (presented and discussed in **Chapter 5**), providing us with a good model to calculate a theoretical R_{st} value of 29.50 Å. This aligns well with the observed R_{st} value of 27.82 Å and provides a ratio of 0.94 (**Figure 3.11 F**), complementing the M_r prediction that peak 2 is likely the expected ParDE1 heterotetramer.

Investigation into the ParDE complex species present in 'peak 1' (**Figure 3.11 A and B**, red curve) indicates the existence of a higher order complex, rather than a heterotetrameric complex of significantly different shape. Again, using AlphaFold (Jumper *et al.*, 2021; Evans *et al.*, 2022), the heterotetramer solutions (of which 5 were generated) do not present anything of larger R_{st} to match the observed R_{st} for this chromatographic peak. The observed M_r of 84.22 kDa (**Figure 3.11 F**) indicates that the complex present may be hetero-octameric, potentially generated by the dimerization of two ParDE1 heterotetramers or by the generation of a whole new complex structure. Despite this, an octameric complex could not be generated using AlphaFold, therefore alternative stoichiometries were input to generate models of the appropriate R_{st} rather than

focus on the M_r prediction; we have seen previously that SEC relies more on the R_{st} than the M_r (**Figure 3.10 C**).

Interestingly, a heterohexameric complex of $ParD1_4ParE1_2$ could successfully be generated (presented and discussed in **Chapter 5**), providing an M_r of only 59.18 kDa. While this resulted in the observed M_r being 1.4 times larger than the calculated, the R_{st} value obtained for the $ParD1_4ParE1_2$ was 32.6 Å and returned an observed/calculated ratio of 1.06 (**Figure 3.11 F**). This is suggestive that the species present in 'peak 1' is a heterohexameric ParDE1 complex, whereby two ParD1 antitoxins have somehow been donated to the heterotetramer.

Taken together, these data indicate that the ParDE1 complex may exist in multiple stoichiometries in solution. Interestingly, the different purifications of ParDE1 presented (**Figure 3.11 A**) resulted in peaks of different intensities for the two ParDE1 complexes. This indicates that there may be an equilibrium between the two species and that one may become the other, rather than the selective complexing of the distinct ParDE1 stoichiometries. It is therefore important that this equilibrium is explored as this may elucidate a driving factor of post-translational ParDE complex reorganisation.

3.4.3.1 Conditional ParDE1 complex remodelling provides insights into a novel mechanism of post-translational ParE toxin release

To explore the theory that ParDE1 complexes exist in multiple stoichiometries and that these may be part of an equilibrium, a method of ParDE1 complex purification was developed (**Material & Methods 2.5.8.1**) whereby the purified fraction existed predominantly in the theoretical $ParD1_2ParE1_2$ (heterotetramer) without need for separating the species via SEC. This provided us with the heterotetramer as a starting material, importantly containing ParD1 antitoxin and ParE1 toxin in theoretically equimolar amounts. The purified complex was subjected to numerous conditions (**Figure 3.12**) in order to explore whether $ParD1_2ParE1_2$ is remodelled into the predicted $ParD1_4ParE1_2$, signified by the emergence of 'peak 1' from 'peak 2' during analytical SEC. The hypothesis was that should this occur, and the stoichiometry of the complex is therefore capable of being altered, there should be a point at which ParE1 toxin becomes free in solution. In theory, as $ParD1_2ParE1_2$ donates two ParD1 molecules to another heterotetramer, forming $ParD1_4ParE1_2$, two ParE1 toxin molecules should be released. If this occurs, a chromatographic peak may be evident for the free toxin and this would appear as 'peak 1' grows from 'peak 2'.

High-yield purification of ParDE1 (> 50 mL at 2.5 mg/mL) in the heterotetramer stoichiometry (**Figure 3.12 A**, black curve/0 hr) allowed for the exploration of multiple factors that may influence protein complex states, notably concentration, temperature, pH, and salt concentration. Following incubation at 4 °C for 48 hr, the chromatographic trace does indeed change and the theoretical heterohexameric ParDE1 complex 'peak 1' does grow from the original heterotetrameric 'peak 2'observed at 0 hr (**Figure 3.12 A**). It is worth noting that the traces are all normalised between 0 and 1 mAU to better align the respective peaks, whereas in reality the intensity of the peaks does indeed change.

Initial hypotheses were that protein concentration would have the greatest influence on the predicted equilibrium between $ParD1_2ParE1_2$ and $ParD1_4ParE1_2$ due to the likely existence of a dissociation constant (K_D) when considering protein-protein interaction. The starting material, at 2.5 mg/mL (or 0 hr) was concentrated and sequentially analysed via SEC (**Figure 3.12 B**). The 48 hr 'mixed' ParDE1 sample is presented alongside the increasing concentrations to identify the respective peaks. It is clear to see that concentrating ParDE1, even as high as 45 mg/mL, does not have a large effect on the predicted stoichiometries as at each concentration the dominant species remains aligned with the starting material at 2.5 mg/mL. There is a minor shift at the higher concentrations with the appearance of a 'shoulder' peak that aligns with the heterohexamer peak. Considering the evolution of the second peak over time (**Figure 3.12 A**), it is possible that this is an artefact of the experiment as concentration and the sequential analytical SEC analysis took over 6 hours.

Given that protein concentration seemingly had minimal effect on the emergence of the heterohexamer peak, it was hypothesised that the complex remodelling process may be thermodynamically driven. Given that we have seen the isolated heterohexameric peak (peak 1) appears to be around 5 °C more thermostable via melt analysis (**Figure 3.11 C**), it would make sense that temperature has some effect on this equilibrium. Incubating the ParDE1 heterotetramer at the constant starting concentration of 2.5 mg/mL at 37 °C results in a more rapid evolution of the heterohexameric peak (peak 1) (**Figure 3.12 C**, 16 hr). Not only is the emergence of the peak apparently more rapid than the 4 °C incubation over 48-hour, it is also more dominant in terms of the respective intensities of the two peaks with the heterohexamer peak being the dominant species in 37 °C incubation. At the 48-hour time-point the entire peak has shifted from the heterotetramer to the heterohexamer, and in fact a shoulder appears on the left of the heterohexamer, indicating that the entire fraction has been remodelled into a higher order complex from the starting $ParD1_2ParE1_2$ heterotetramer (**Figure 3.12 C**, 48 hr). Crucially, an additional small peak was observed just below 20 mL, most evident on the 16-hour

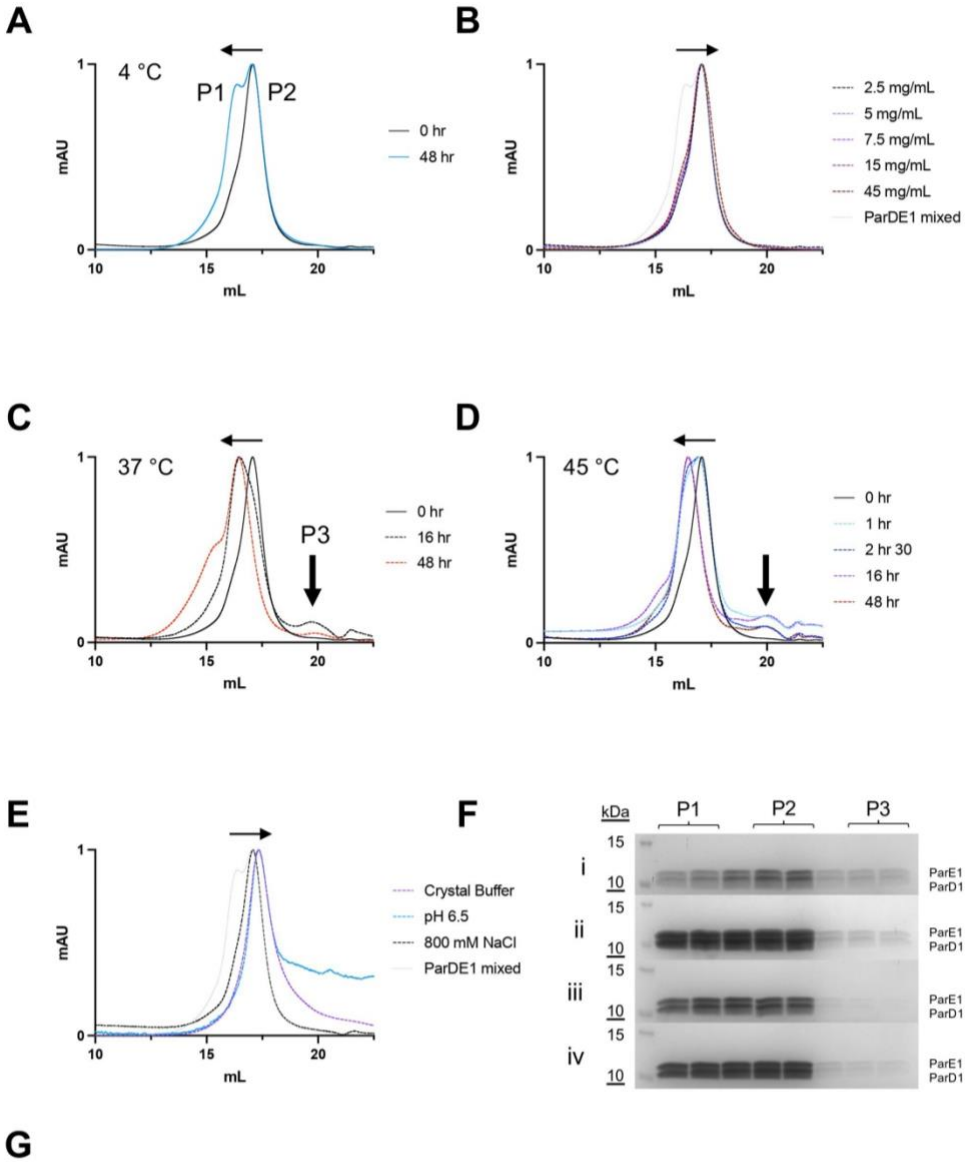
curve, which could be formed by free ParE1 toxin (thick black arrow) (**Figure 3.12 C**), as previously hypothesised.

Considering the melt curve data (**Figure 3.11 C**), temperature could be increased to 45 °C in further experiments as both complexes appeared to be stable at this temperature (**Figure 3.12 D**). This increased temperature was used in an attempt to speed up the process of complex remodelling and optimise a purification process for potentially free ParE1 toxin. Indeed, the increase in temperature appeared to cause a more rapid complete shift to the heterohexameric peak, with the process being complete by 16-hours and clear shifting occurring as early as one hour into the time-course (**Figure 3.12 D**). Again, the potentially ParE1 peak is highlighted by the thick black arrow. Taken together with the data presented in **Figure 3.12 C**, it is likely that this process is thermally driven. Interestingly, the entire fraction of $ParD1_2ParE1_2$ (starting material peak, 0 hr) was remodelled into a higher-order complex, likely the heterohexameric $ParD1_4ParE1_2$. Whether this process is reversed by temperature decreases is yet to be explored.

In showing that the process can be driven by temperature increase we explored, briefly, the effects of reducing agent, acidic pH, and high salt (**Figure 3.12 E**), as these conditions can all effect protein complex states (Deng, Toledo and Lillard, 1976; Souza and Garcia-Rojas, 2015; Speer *et al.*, 2021) and are environmental conditions that TA systems may be responsive to (Kedzierska and Hayes, 2016). None of these conditions caused any noticeable shift in the positioning of the starting peak toward the theoretical heterohexameric peak (left). Both peaks are represented in **Figure 3.12 E** by the ParDE1 mixed sample (48 hr curve, **Figure 3.12 A**).

Given that the position of the three chromatographic peaks of interest ('starting' heterotetramer, theoretically larger heterohexamer, and potentially released ParE1 toxin) have now been established, fractionation and SDS-PAGE analysis was used to demonstrate the purification of the respective species (**Figure 3.12 F**). Resolution of two ParDE1 complex species was highly unlikely as the two peaks are never truly separate, rather they form distinct shoulders and merge during the hypothesised remodelling process (**Figure 3.12 A** 48 hr; **C** 16 hr; **D** 1 hr). This is also very clear in SDS-PAGE analysis of fractionation; 250 µL fractions were collected and two fractions aligned directly under the V_e of each peak were used to represent the constituent protein of that peak (Peak 1/P1 – Heterohexamer; Peak 2/P2 – Heterotetramer; Peak 3/P3 – ParE1) (**Figure 3.12 F**). A single fraction was run between the peaks to demonstrate levels of separation, despite this, both proteins, ParE1 (11.17 kDa), and ParD1 (9.21 kDa), were present in visually equal intensities in each fraction (**Figure 3.12 F, i – iv**).

Figure 3.12 Conditional ParDE1 complex remodelling analysis via SEC



G

Peak	Calculated M_r (kDa)	Observed M_r (kDa)	Observed M_r/ Calculated M_r	Calculated R_{st} (Å)	Observed R_{st} (Å)	Observed R_{st}/ Calculated R_{st}
ParE1	11.17	7.88	0.70	16.40	15.82	0.96

Figure legend overleaf

Figure 3.12 Conditional ParDE1 complex remodelling analysis via SEC

Analytical SEC traces for ParDE1 (**A**) Purified in the heterotetrameric fraction (0 hr, black, labelled P2 for 'Peak 2'). Incubated at 4 °C this single peak evolves into the mixed peak (48 hr, blue, labelled P1 for 'Peak 1'); (**B**) Increasing concentration presented alongside the mixed sample from A (light grey); (**C**) 37 °C incubation alongside the starting, 0 hr trace (solid black); (**D**) 45 °C incubation alongside the starting, 0 hr trace (solid black); (**E**) Buffer changes, presented alongside the mixed sample from A (light grey). Horizontal black arrows indicate general direction of peak shifting, vertical arrows indicate the generation of the third peak (labelled P3), predicted to be liberated ParE1 toxin; (**F**) SDS-PAGE analysis of fractions aligning to observed peaks 1 (P1, left), 2 (P2, central/starting), and 3 (P3, right/toxin): i – 0 hr from A; ii – 15 mg/mL from B; iii – 37 °C 16 hr from C; iv – 45 °C 16 hr from D; (**G**) Table for ParE1 protein molecular weight (M_r) and Stokes Radius (R_{st}) calculations, observations, and comparisons. Calculated R_{st} for ParE1 was generated using HullRad (Fleming and Fleming, 2018) and the structure solution available in the AlphaFold database (Varadi *et al.*, 2022). Comparison of observed/calculated is coloured green if within 10 % of the predicted ratio, yellow if > 10 ≥ 25 %, and red if > 25 %. Chromatograms are representative of single repeats and are normalised between 0 – 1 for presentation and comparison. Graphs are cropped to the appropriate scale (10 – 22.5 mL).

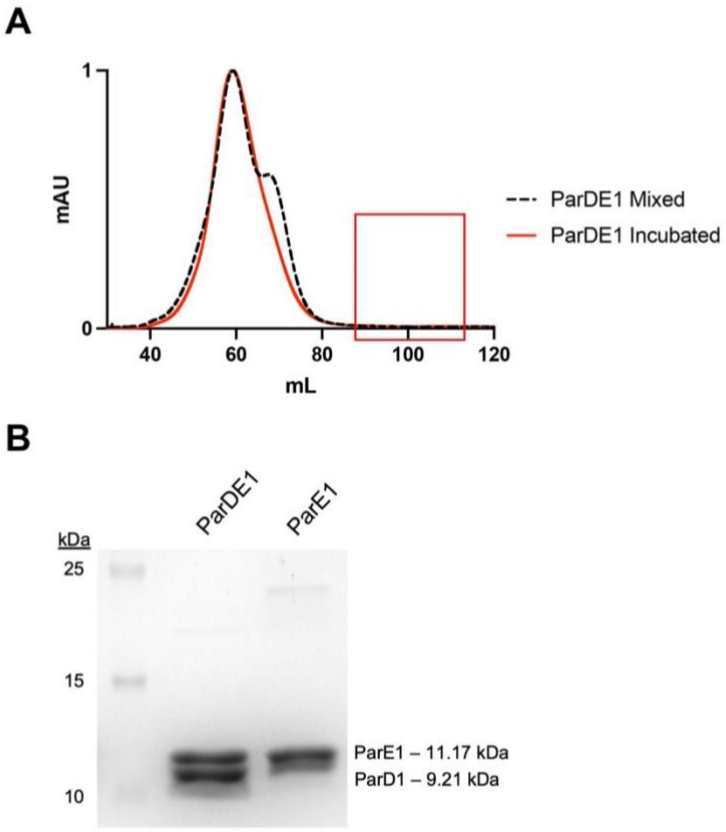
Due to the column efficiency and resolution, separation of the desired species was not possible. **Figure 3.12 F i** shows the fractions from the starting material (**Figure 3.12 A**, 0 hr), and while the intensity of the bands for P2 (heterotetramer) are greater than those in P1 and P3, there remains complex present throughout. No detectable increase in the intensity of the ParE1 toxin peak (P3) was seen in fractions taken from complex remodelling experiments (**Figure 3.12 F ii – iv**) indicating this is not a suitable method for purification of the desired ParE1 toxin.

Considering these findings, it is likely that P3 does contain the released ParE1 toxin, albeit not resolved from the tail of the central peak (P2) (**Figure 3.12 F**). Converting the obtained V_e to M_r for the theoretical toxin peak from **Figure 3.12 C and D** returned a value of 7.88 kDa; this is far less than the expected Mr of 11.17 kDa for ParE1 and provides a ratio of 0.70 (**Figure 3.12 G**). As seen previously, R_{st} may be a better estimate of the constituents of chromatographic peaks, and this combined with the availability of the predicted *M. tuberculosis* ParE1 structure from the AlphaFold database (Jumper *et al.*, 2021; Varadi *et al.*, 2022) returns a more reassuring result. The calculated R_{st} for the ParE1 model structure is 16.40 Å and the Observed R_{st} was 15.82 Å; the Observed/Calculated ratio of 0.96 (**Figure 3.12 G**) indicates that the peak produced during the hypothesised ParDE1 remodelling process could very well be free monomeric ParE1. In order to confirm this, a higher resolution FPLC column could be used in a high-scale protein purification, using the data gathered here as a guide to releasing the ParE1 toxin via incubation at higher temperature prior to SEC.

3.4.3.2 A novel ParE1 toxin purification method through thermally driven ParDE1 complex remodelling

Purification of the released ParE1 toxin would be the ideal next step in exploring the biophysical and biochemical activities of the ParDE1 complex. Using the data gathered above, alongside previously used protein expression and purification methods for the ParDE1 heterotetrametric complex (**Materials & Methods 2.5.8.1**; **Figure 3.1**; **Figure 3.12**), attempts were made to isolate the free toxin (**Figure 3.13**). The recombinant protein complex was initially purified via Ni-NTA followed directly by anion exchange, whereby the protein elutes in a single major peak (**Figure 3.1 A**). After overnight incubation at 4 °C with hSENP2 and a further Ni-NTA stage to purify the complex from the removed tag and cleavage enzyme, the sample was concentrated using a 5 kDa cut-off centrifugal concentrator column (Sartorius) to ensure no free toxin was lost. The sample was concentrated to 10 mL in volume, which was around 10 mg/mL in concentration prior to incubation at 37 °C for 16 hr (**Figure 3.12 C**).

Figure 3.13 High-scale ParE1 purification attempt



(**A**) Sephacryl S-200 HR SEC trace of the soluble fraction from the incubation of 10 mg/mL ParDE1 complex at 37 °C for 16 hr (solid red) alongside the mixed peak purification example (dashed black) (presented earlier in Figure 3.1). Red box indicates the expected range for the ParE1 toxin peak; (**B**) SDS-PAGE of the purified ParDE1 complex from SEC presented in A (ParDE1 lane) alongside the precipitated fraction from overnight incubation of ParDE1 at 37 °C (ParE1 lane).

Post-incubation, the sample was inspected prior to SEC. Unfortunately, a high amount of precipitate was present; this was removed by centrifugation and 5 mL of the supernatant was subjected to SEC on a HiPrep™ Sephacryl® S-200 HR (Cytiva) as per **Materials & Methods 2.5.2.3**. The resulting chromatographic trace was aligned to a previous ParDE1 purification (dashed black trace) from the same column showing that the entire incubated fraction (red trace) aligns well with the left-hand peak (theoretical heterohexamer) (**Figure 3.13 A**). Previous data (**Figure 3.12**) suggests that once the peak is shifted to this position, another peak (ParE1) would be obtained (theoretical position indicated by the red box). No ParE1 peak was obtained during SEC, therefore the major peak fraction was analysed alongside the precipitated fraction from overnight incubation via SDS-PAGE (**Figure 3.13 B**). This did indeed demonstrate that SEC produced purified ParDE1 complex with the presence of both ParD and ParE proteins (ParDE1 lane), whereas the liberated ParE1 toxin after 37 °C incubation appears to have precipitated with the single ParE1 band being present (ParE1 lane). The purity of the toxin with respect to potential contaminating ParD1 protein is not determined, however, ParE1 is clearly the dominant species in the precipitated sample (**Figure 3.13 B**).

This indicates that the theorised, thermally driven, complex remodelling process to liberate the ParE1 toxin is in effect. The process of ParE1 purification may require substantial optimisation, considering temperature, incubation time, protein concentration, and buffer components. It was previously shown that the ParDE1 complex had a minor effect in the gyrase relaxation assay, causing linearisation at the top concentration of 10 µM (**Figure 3.3**). As was discussed, the 37 °C assay incubation temperature may have played a role in the dissociation of the ParE1 toxin resulting in the increased linearisation; considering these biophysical studies (**Figures 3.11 – 3.13**), it is quite likely we had previously observed ParE1 toxin liberation through ParDE1 complex remodelling. The observation of low level linearisation in the assay is likely due to the purification of ParDE1 being pooled from both heterotetramer and heterohexamer peaks after SEC (**Figure 3.1 B**) where most of the peak has shifted, leaving a minority of the complex free for remodelling and toxin release. It will therefore be interesting to repeat the biochemical analysis of the effects of ParDE1 on the gyrase relaxation assay starting with the heterotetramer species to observe the effects of incubation over time.

3.4.3.3 Thermally driven ParE1 release induced cleavage

Given the difficulties in purifying the ParE1 toxin, but having obtained data supporting the release of the protein from the complex during a theorised remodelling process, we returned to biochemical analysis via the gyrase DNA relaxation assay. As previously stated, now that the protein purification for the tetrameric ParDE1 complex has been optimised, the toxic effects of any ParE1 release may be more obvious compared to that seen in **Figure 3.3**.

Coupling the relaxation assay with a ParDE1 37 °C incubation time-course, prior to addition, could allow for greater linearisation to be seen. The previously presented gyrase DNA cleavage assay was used (**Materials & Methods 2.7.4**), however, this time assays were performed following pre-incubation of ParDE1 at 37 °C to promote toxin release (**Materials & Methods 2.6.4.4**). Briefly, a stock of $GyrA_2B_2$ holoenzyme was created and kept on ice at the appropriate concentration throughout the assay. Heterotetrameric ParDE1 was diluted to 100 µM in preparation for addition to the gyrase relaxation assay to a final concentration of 10 µM. The ParDE1 stock was then incubated at 37 °C in a thermocycler across the full time-course and sampled for biochemistry and analytical sizing when appropriate (**Figure 3.14**).

ParDE1 analytical SEC was performed on the hour at the presented time points in **Figure 3.14 A**. The starting, 0 hr, sample was visually homogenous and positioned at the appropriate V_e (~17.3 mL) for the heterotetrameric ParDE1 complex as expected. Incubation at 37 °C clearly causes a gradual shift in the positioning of this peak to the left as the complex is remodelled over time until the entire peak elutes at the V_e of the heterohexamer (~16.1 mL) (**Figure 3.14 A**). It is worth noting that the 3 hr, and final 20 hr time-points were not tested in the corresponding biochemical assay (**Figure 3.14 B**), however, the bulk of the fraction is remodelled by 12 hr (**Figure 3.14 A**), alongside clear shifting throughout.

Interestingly, at the 1 hr time-point an intense and appropriate V_e peak for the ParE1 toxin evolved (**Figure 3.14 A**, brow dash). This may have spontaneously been liberated and caught at a fortunate time-point in the process as the peak disappears at the next time point and until 12-hours (where the peak is also far less obvious). This does indicate the process is dynamic and that the components may be released and re-absorbed throughout.

Figure 3.14 ParE1 induced cleavage through thermally driven toxin liberation

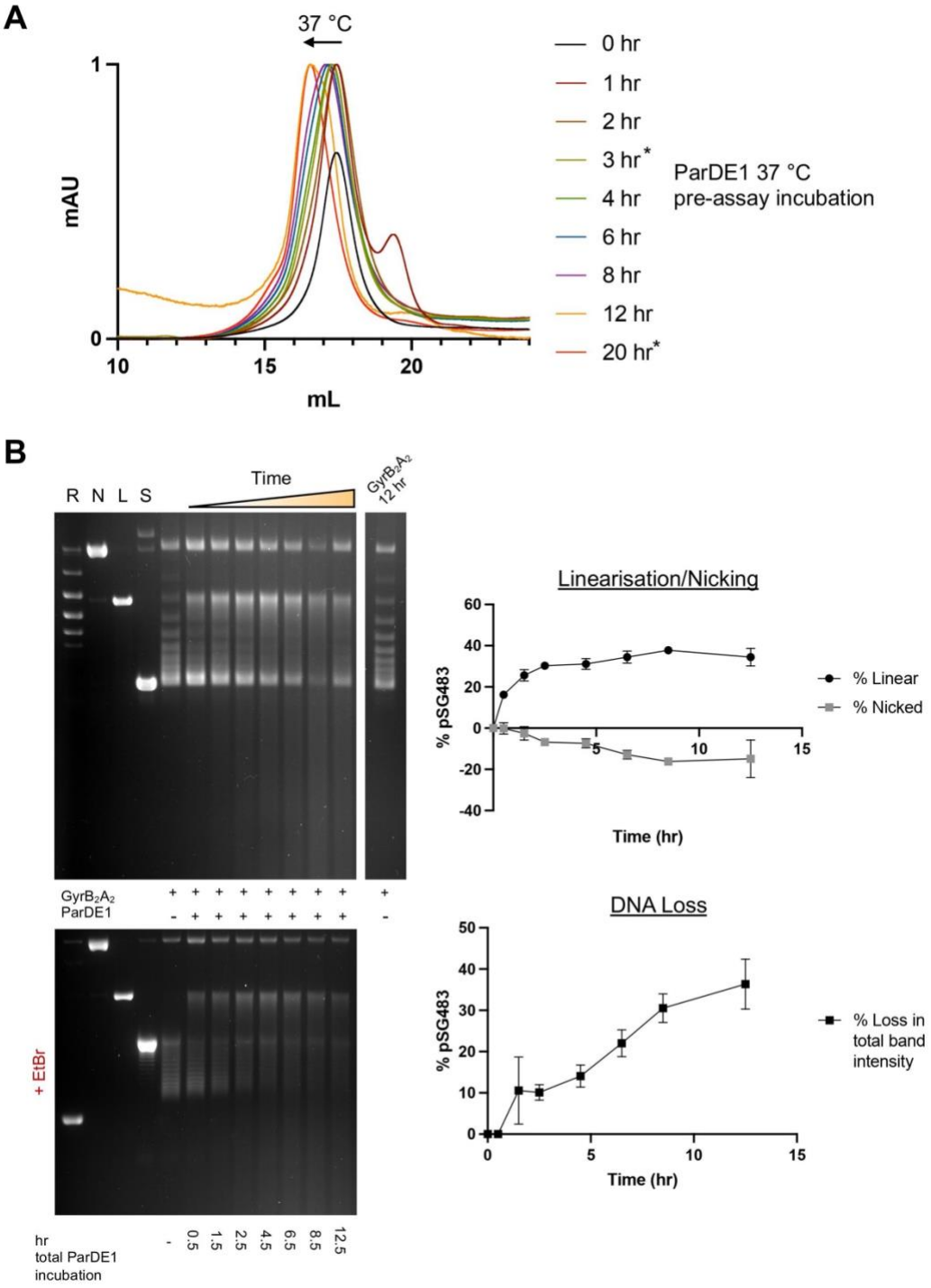


Figure legend overleaf

Figure 3.14 ParE1 induced cleavage through thermally driven toxin liberation

(**A**) Analytical SEC traces for ParDE1 37 °C incubation time course, starting with heterotetramer form complex (0 hr, solid black). A 100 µL sample of incubated ParDE1 (100 µM) was injected onto a Superose 6 10/300 GL column using an Åkta Pure system at each listed time point. Horizontal arrow indicates the general direction of peak shifting; (**B**) ParDE1 induced gyrase DNA cleavage assay. $GyrB_2A_2$ was reconstituted, diluted, and stored on ice. ParDE1 (10 µM final concentration) from each time point was added to constant $GyrB_2A_2$ (31.25 nM) and Supercoiled (S) plasmid DNA (12.5 nM). Presence (+) or omission (-) of $GyrB_2A_2$ and ParDE1 is detailed between the agarose gels for each lane. A 12 hr $GyrB_2A_2$ only assay is included to ensure enzyme stability throughout the time-course. ParDE1 total incubation time at 37 °C (pre-assay incubation time point + assay incubation) is shown below the gels (hr). Control lanes represent Supercoiled (S), Linear (L), Nicked (S), and Relaxed (multiple topoisomers) (R) plasmid DNA. Assays are presented on 1.4 % Agarose 1x TAE gels (run with ethidium bromide (+EtBr) as stated, or post-stained) alongside graphical analysis of percentage Linear/Nicked DNA and total percentage loss in band intensity per lane (+EtBr, obtained by densitometry) against time (hr). Incubation time for ParDE1 is quantified below the gels. Assays are representative of triplicate and data points and error bars represent the mean and SD of triplicate data, respectively.

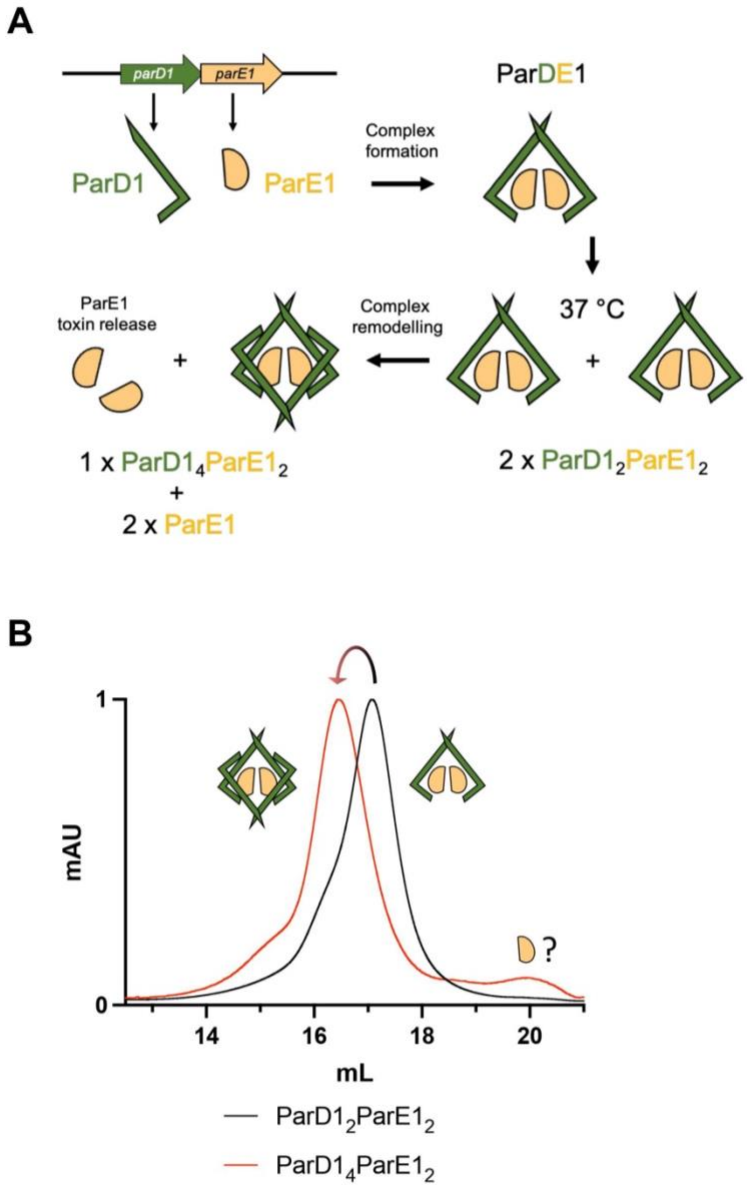
At the appropriate time-point, incubated ParDE1 was added to the relaxation assay to a final concentration of 10 µM as per the standard protocol used throughout the biochemistry presented prior (**Materials & Methods 2.7.4**). A new relaxation reaction was set-up for each time-point using the pre-diluted stock of $GyrB_2A_2$. Importantly, the reaction provided an extra 30 minutes of incubation at 37 °C, alongside a different buffer environment. The relaxation reaction is clearly inhibited at the 0-hr pre-incubation point with the notable presence of linear species DNA (0.5 hr biochemistry incubation, **Figure 3.14 B**). The reaction becomes almost fully inhibited over the pre-incubation time-course as linear species DNA appears to increase; the area below the supercoiled band is presented to demonstrate the increasing levels of non-specific DNA cleavage (evidenced by the smearing pattern within the lane) (**Figure 3.14 B**, +/- EtBr).

The linear species is apparently stabilised at around 30 % from the 2.5 hr (total incubation time) onward (**Figure 3.14 B**, Linear plot), alongside a general decline in the nicked species **Figure 3.14 B**, Nicked plot). It should be noted that there is a clear loss in the total band intensity (**Figure 3.14 B**) that likely biases the calculations for linearisation and nicking. Altogether, we can consider the observations of stabilised linearisation and increasing total DNA loss as a general increase in DNA damage over the time-course. This nicely correlates with the movement in the chromatographic peak (**Figure 3.14 A**) from the right-hand side starting point ParDE1 heterotetramer (black trace), toward the final form, left-hand side (red trace), remodelled heterohexamer. The theoretical result of the remodelling process is the gradual increase in liberated ParE1 toxin which in turn is likely responsible for the total increase in DNA damage. Interestingly, this happens as early as 30 minutes, and as more ParE1 is released we see an increase in non-specific DNA cleavage and increasing losses in total DNA band intensity in the agarose gels (**Figure 3.14 B**). It is also worth noting that these effects are independent of the increasing age of the initial gyrase stock; this was used throughout the assay at each time point and was tested, alone, for function at the 12-hr time-point (**Figure 3.14 B**). Gyrase remained stable and active throughout the experiment, thus, these results are solely due to the addition of ParDE complex.

Here we show that ParDE1 can be expressed and purified in a stoichiometry that is predicted to be a $ParD1_2ParE1_2$ heterotetramer by analytical SEC (**Figure 3.11; Figure 3.12**). Having tested several conditions, we have identified that the $ParD1_2ParE1_2$ heterotetramer re-distributes into a theoretical $ParD1_4ParE1_2$ heterohexamer when incubated at higher temperatures, such as 37 °C (**Figure 3.11; Figure 3.12**). In theory, this process liberates ParE1 toxin molecules, however, high-scale purification was unsuccessful; SDS-PAGE analysis confirmed that the ParE1 was isolated from the ParDE1 complex, albeit as precipitate (**Figure 3.13 C**). It can, however, be demonstrated that ParE1 liberation occurs during thermally driven complex remodelling, resulting in higher levels of DNA damage in the gyrase DNA relaxation assay over time (**Figure**

3.14). The remodelling process appears to be gradual as shown by incubation at 37 °C over several hours (Figure 3.14 A). The rate of the remodelling does appear to be dependent on increasing temperatures as shown by incubation at 4 °C, 37 °C, and 45 °C (Figure 3.12 A – D). The extent to which other factors such as buffer conditions and concentration play a role in complementing the process is, as yet, unknown. A schematic for the remodelling process is presented in Figure 3.15 A, demonstrating that it may be possible to convert two heterotetramer complexes into a single heterohexamer, and for every heterohexamer produced two ParE1 toxin molecules are liberated.

Figure 3.15 Schematic for ParE1 liberation through thermally driven ParDE1 complex remodelling



(**A**) Cartoon representation for the expression, initial complexing, and complex remodelling process hypothesised for ParDE1. ParD1 (green) and ParE1 (yellow) and expressed from a bicistronic operon and form a predicted ParDE1 heterotetrameric complex. Through incubation at 37 °C, the entire fraction of ParD1$_2$ParE1$_2$ heterotetramer complexes can remodel to form ParD1$_4$ParE1$_2$ heterohexamer complexes. Theoretically, two heterotetramers are required to generate a single heterohexamer, releasing two ParE1 toxin molecules; (**B**) Analytical SEC traces for the separated heterotetramer and heterohexamer fractions alongside their respective ParDE1 schematic. The heterohexamer trace is observed with a peak aligning to the size of the ParE1 toxin.

3.4.4 Biophysical analyses of the *M. tuberculosis* ParDE2 TA system

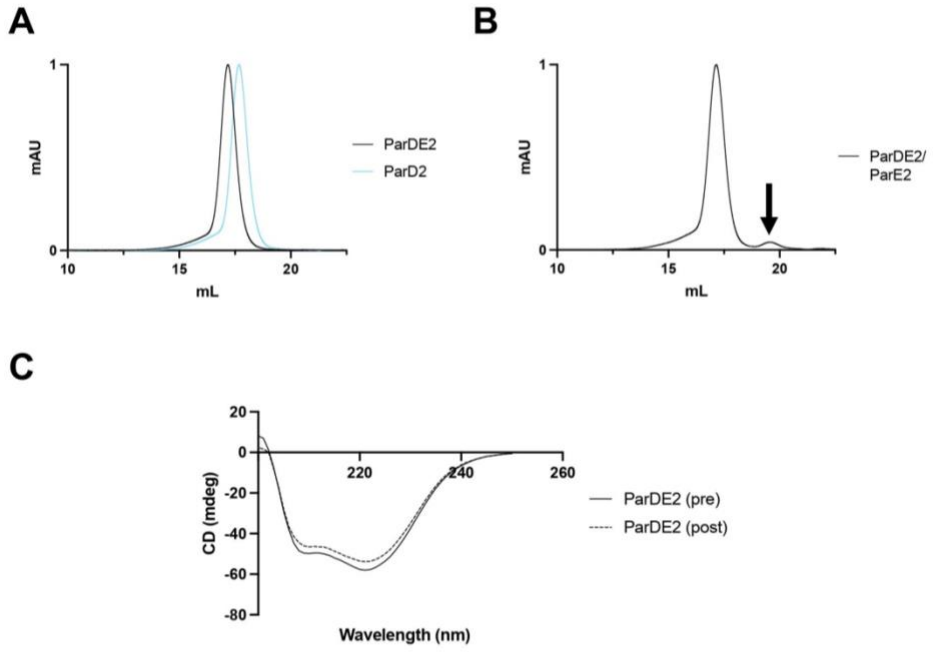
In contrast to the ParDE1 complex, analytical SEC of the purified ParDE2 complex produces a single peak (**Figure 3.16 A**, back). The recombinantly expressed and purified ParD2 antitoxin also elutes in a single peak (**Figure 3.16 A**, blue). Whilst free ParE2 toxin isolated during the purification process (**Figure 3.4**) was not analysed due to limitations with yield and prioritisation of biochemical analyses, it was possible to obtain an elution volume (V_e) for free ParE2 (**Figure 3.16 B**, black arrow) using a mixed ParDE2/ParE2 sample gathered by pooling the ParDE2 and ParE2 peaks resulting from anion exchange chromatography during purification (**Materials & Methods 2.5.9.1**).

Circular dichroism spectroscopy confirmed that the ParDE2 complex exists folded in solution with a largely alpha helical secondary structure (**Figure 3.16 C**, solid black, pre-melting). The post-melting analysis of ParDE2 via CD indicates a large degree of refolding as a highly similar curve is obtained (**Figure 3.16 C**, dashed black, post-melting). This is heavily biased by severe aggregation in the sample during thermal denaturation, hence why the melt curve is not presented for this complex (aggregation interfered heavily with readings).

Using elution volumes obtained for ParDE2, ParD2, and ParE2 (**Figure 3.16 A and B**) we can generate observed molecular weights (M_r) and Stokes radii (R_{st}) values in kilodaltons (kDa) and angstroms (Å), respectively (**Figure 3.16 D**). As seen previously for gyrase and ParDE1, when comparing the observed values to the calculated values from sequences and structures, R_{st} predictions can often be slightly more accurate.

For the ParDE2 complex, the molecular weight was calculated to be 39.92 kDa as a predicted heterotetramer (in line with most ParDE systems in the PDB). The observed M_r for the complex was 56.54 kDa, 1.42 times larger than the calculated M_r, however, when comparing the R_{st} of two AlphaFold generated $ParD2_2ParE2_2$ complexes (presented and discussed in **Chapter 4**) to the observed R_{st} value, we were able to identify the best fitting model (Model 3) (**Figure 3.16 D**). This indicates that, while AlphaFold does generate solutions of high quality and accuracy, a complement of techniques should be employed to confirm these data (discussed in **Chapter 4**).

Figure 3.16 Biophysical analyses of the ParDE2 system



D

Sample	Calculated M_r (kDa)	Observed M_r (kDa)	Observed M_r/ Calculated M_r	Calculated R_{st} (Å)	Observed R_{st} (Å)	Observed R_{st}/ Calculated R_{st}
ParDE2	39.92[1]	56.54	1.42	26.10[2]	30.27	1.16
				30.50[3]	30.27	1.05
ParD2	7.78	38.82	4.99	16.10[4]	26.74	1.66
ParE2	12.18	7.88	0.64	16.70[5]	15.82	0.95

(**A**) Analytical SEC traces for ParDE2 (black) and ParD2 (blue); (**B**) Analytical SEC trace of a ParDE2 purification containing free ParE2 toxin indicated by the vertical arrow; (**C**) Circular dichroism spectroscopy scans for ParDE2 before (solid black, pre) and after (dashed black, post) melting; (**D**) Table of ParDE2 protein molecular weight (M_r) and Stokes Radius (R_{st}) calculations, observations, and comparisons. [1] ParDE2 M_r is calculated as a $ParD2_2ParE2_2$ heterotetramer, [2] R_{st} for AlphaFold Model 2 of $ParD2_2ParE2_2$ (heterotetramer); [3] R_{st} for AlphaFold Model 3 of $ParD2_2ParE2_2$ (heterotetramer); [4/5] R_{st} for ParD2 and ParE2, respectively, are calculated from structure models available in the AlphaFold database (Varadi *et al.*, 2022). R_{st} values were generated using HullRad (Fleming and Fleming, 2018). Comparison of observed/calculated is coloured green if within 10 % of the predicted ratio, yellow if > 10 ≥ 25 %, and red if > 25 %. Chromatograms are representative of duplicate data and are normalised between 0 – 1 for presentation and comparison. Graphs are cropped to the appropriate scale (10 – 22.5 mL).

For the ParD2 antitoxin, the molecular weight was observed as 4.99 times greater (38.82 kDa) than the calculated value for the monomer (7.78 kDa) (**Figure 3.16 D**). This is not surprising given that ParD antitoxins tend to be fibrous in structure. The AlphaFold solution for the ParD2 antitoxin (presented and discussed in **Chapter 4**) is of extremely low confidence and does not resemble other ParD antitoxins. The Stokes radius was observed to be 1.66 times greater (26.74 Å) than the calculated value for the AlphaFold generated monomer (16.10 Å) (**Figure 3.16 D**).

While the ParDE2 and ParD2 observations and comparisons result in large discrepancies in both M_r and R_{st}, the ParE2 toxin V_e results in a much more reassuring result. Similar to the predicted ParE1 analytical SEC V_e (**Figure 3.12 G**), the free toxin peak appears to be far smaller in observed molecular weight compared to the calculated value; ParE2 eluted at a molecular weight of 7.88 kDa, 0.64 times the calculated Mr of 12.18 kDa (**Figure 3.16 D**). When comparing the observed R_{st} value for the free ParE2 in **Figure 3.16 B** to the R_{st} value obtained from the high-confidence AlphaFold structure of ParE2 (presented and discussed in **Chapter 4**), we achieve a ratio of 0.95 (Figure 3.16 D). This indicates that the chromatographic peak highlighted in **Figure 3.16 B** likely contains ParE2 as a monomer in solution and that, similar to ParE1, the toxins are small and compact in nature given the discrepancy in M_r prediction.

3.5 Discussion

Initial experiments confirmed the model system used in these studies acted as previously reported. *M. tuberculosis* gyrase was successfully expressed and purified as its constituent subunits, GyrA and GyrB, both of which exist as dimers in solution as analysed by size exclusion chromatography (**Figure 3.10**). Mixing these subunits in equimolar amounts reconstituted the gyrase holoenzyme, which was active *in vitro* in relaxing DNA supercoils in an ATP-independent manner (**Figure 3.2 A**), and generating DNA supercoils in an ATP-dependent manner (**Figure 3.2 B**). The activity of the enzyme in these reactions is comparable to what has previously been shown (Karkare et al., 2012). Additionally, the enzyme responded to the addition of moxifloxacin as expected; the reaction was inhibited through the trapping of cleavage complexes and increase in linear species DNA (**Figure 3.2 C**). These control experiments confirmed that the purified enzyme is functional and that the cleavage assay could be used as a proven method for exploring the increase in linear DNA as a result of cleavage complex trapping.

In addition to gyrase subunits (GyrA and GyrB), three gyrase fusion proteins were successfully expressed and purified (**Figure 3.6**), and have been routinely used in structural studies due to increased complex stability and likelihood of crystallisation (Blower *et al.*, 2016; Petrella *et al.*, 2019; vanden Broeck *et al.*, 2019). The gyrase fusion proteins display significantly different activity in ATP-independent reaction when compared to the holoenzyme (**Figure 3.2**), and to each other. The full-length *M. tuberculosis* gyrase fusion, that is the 74 kDa GyrB subunit fused at its N-terminus to the C-terminus of the 93 kDa GyrA protein (**Figure 3.2 A**), is unsurprisingly capable of removing DNA supercoils in a similar manner to the holoenzyme. There was, however, a significant difference in the topoisomer banding pattern between the agarose gels for the relaxation reactions; the appearance of higher order topoisomers (fewer supercoils) at a low concentration of GyrBA fusion indicates that the enzyme is more processive, resolving the supercoils on a single substrate sequentially, in contrast to the $GyrB_2A_2$ which looked to have more distributive activity (Morrison, Higgins and Cozzarelli, 1980).

The gyrase holoenzyme has previously been shown to be processive in its ATP-dependent supercoiling reactions, while it is, as shown (**Figure 3.2**), distributive in ATP-independent reactions (Levine, Hiasa and Marians, 1998). The observation that the full-length gyrase fusion is processive in its ATP-independent removal of supercoils (relaxation) is likely due to the dimeric nature of this protein; the fusion was shown to exist as a dimer in solution (**Figure 3.10 B**). The stable dimeric nature of the fusion may enhance the enzyme's ability to maintain association in some way with the DNA substrate, rather than dissociating and reconstituting when appropriate DNA geometry is recognised (Ashley *et al.*, 2017). This allows it to repeat through reaction cycles along the same substrate DNA repetitively.

The processive nature of ATP-dependent holoenzyme reactions could therefore be explained by presence of ATP promoting the association of the GyrB and GyrA subunits. The exact order of DNA binding and ATP binding is not fully defined. It has been speculated that the GyrB C-loop – GyrA DEEE loop interaction that causes the 'bent back' gyrase state seen in Petrella *et al.,* (2019), is a mechanism to prevent futile ATP cycling. If this were true it would seem that ATP may occupy its binding site within the ATPase, promoting association to GyrA. Once the correct DNA topology is recognised and bound, the reaction cycle is permitted and T-segment trapped for transport. In this model, DNA binding overrides the C-loop – DEEE loop interaction. In ATP-dependent reactions, ATP is in abundance and therefore gyrase could exist as the reconstituted holoenzyme (like the full-length fusion), generating supercoils on the same substrate in a processive manner.

The gyrase fusions behaved considerably differently to this. GyrBA^{56}, missing only the GyrA C-terminal domain (Pinwheel domain) (**Figure 3.6 B**), and GyrB^{28}A^{56}, missing both the GyrB ATPase and GyrA C-terminal domain (**Figure 3.6 C**), could only generate linear and nicked products in an ATP-independent reaction (**Figure 3.7 B and C**). While testing the GyrB^{28}A^{56} fusion in an ATP-dependent reaction is futile (lack of ATPase domain), the GyrBA^{56} is capable of removing supercoils in an ATP-dependent manner (not presented in this body of work) as was seen for the *E. coli* equivalent (Kampranis and Maxwell, 1996). Essentially, the fusion enzyme is converted into a conventional type II topoisomerase.

The gyrase fusions presented significantly different levels of both nicked and linear species DNA with the GyrB^{28}A^{56} converting a high percentage of the substrate into nicked form at concentrations below 50 nM (**Figure 3.7 C**). In contrast, the GyrBA^{56} fusion failed to reach the same percentage of nicked DNA even at 250 nM, but did generate a slightly higher amount of linear DNA (**Figure 3.7 B**). These significantly different patterns in DNA cleavage may indicate general stability of their respective interactions with substrate DNA; both fusion proteins convert similar amounts of supercoiled plasmid DNA into linear and nicked form, indicating they both bind DNA well. When considering the combined linear and nicked product percentages, the GyrB^{28}A^{56} did have a significantly higher total at the top concentration, potentially indicating a stronger DNA-binding capability; this may be due to the lack of the GyrB transducer and ATPase hindering access at the G-gate. This is further supported when examining the combined total at the lower enzyme concentrations; for example, at 31.25 nM the cleaved total product for GyrBA^{56} was only about 25 %, whereas the cleaved product total for GyrB^{28}A^{56} remained significantly higher at about 60 %. Alternatively, these results may indicate that the GyrBA^{56} simply has better control mechanisms regulating the cleavage step, thereby preventing the significantly higher increase in total nicked and linear species seen for the GyrB^{28}A^{56} relaxation assay (**Figure 3.7 B and C**). The large and significant increase in asymmetric cleavage, producing nicked DNA when

comparing GyrB^{28}A^{56} to GyrBA^{56}, supports that the DNA-binding and cleavage step is not as controlled, as symmetrical cleavage producing linear form is preferred to allow strand-passage in all gyrase reactions. GyrBA^{56} may indeed have a similar affinity for DNA to GyrB^{28}A^{56}, but the control over cleavage (or indeed resealing) may be greatly enhanced by the GyrB transducer and ATPase domains.

ParDE systems were expressed and purified as complexes to avoid issues with ParE toxicity. Both ParDE1 and ParDE2 were shown to have very different biophysical properties (**Figure 3.11** and **Figure 3.16**, respectively). Notably, ParDE1 is predicted to exist in multiple stoichiometries, the two dominant species studied in this chapter are predicted to be heterotetrameric ($ParD1_2ParE1_2$) and heterohexameric ($ParD1_4ParE1_2$) as demonstrated using a combination of size exclusion chromatography (**Figure 3.11 A, B, and F**) and AlphaFold generated models of the ParDE complexes (presented in **Chapter 5**). ParDE1 is expressed and initially complexes as the predicted heterotetramer (**Figure 3.11 F**) which can be purified and used as starting material in studying complex remodelling (**Figure 3.15 B**, black). It can be driven to form a higher-order protein complex, predicted to be the heterohexamer (**Figure 3.11 F**), existing in a distinctly different chromatographic peak (**Figure 3.15**, red). The growth of the heterohexameric peak coincides with the growth of a chromatographic peak predicted to be liberated ParE1 toxin (**Figure 3.12 G**), and this process has been shown to increase overall DNA damage when coupled with the gyrase relaxation/cleavage assay (**Figure 3.14**). It does appear that an even higher-order protein complex forms over time, evidenced by the left-hand shoulder consistently forming on the heterohexameric peak (**Figure 3.15 B**), the exact species of which is not yet modelled or solved. The evolution of this larger species of ParDE1 correlates with the reduction in intensity of the predicted ParE1 peak, most obvious in **Figure 3.12 C** (48 hr). This indicates that ParE1 liberation may be the by-product and mid-point of a more complicated process that is not yet understood. The process does, however, provide a window whereby ParE1 toxin is theoretically free and we have shown this to negatively affect activities of DNA gyrase (**Figure 3.14**). Interestingly, the structure of the *V. cholerae* ParDE2 system complex has recently been solved as a hetero-octamer in a $ParD2_6ParE2_2$ hetero-octameric stoichiometry (**Table 1.4**) (Garcia-Rodriguez *et al.*, 2021). This higher-order structure was not shown to evolve from a lower-order species, however, and was suggested to have a role in autoregulation. Additionally, in the hetero-octamer, we do not see equimolar ParE molecules, which we suggest may occur in the advanced stages of ParDE1 remodelling.

We are currently unable to quantify the amount of liberated toxin in this remodelling process, but given amount of visual and quantified DNA damage (**Figure 3.14**), we predict ParE1 to have high affinity for gyrase. ParE1 is capable of not only trapping cleavage complexes, but apparently

drives gyrase to cleave DNA non-specifically, indicated by the amount of 'DNA loss' or smearing on gels (**Figure 3.14**). This remodelling process warrants further study as it is likely a complex thermodynamic process. This process requires validation *in vivo* through a series of growth curves assessing the effects of temperature and growth time. Theoretically, if the heterotetrameric ParDE1 complex is present in the *M. tuberculosis* cell when the bacterium become aerosolised in droplets (lower environmental temperature), upon infection and subsequent temperature elevation to 37 °C in the body, the system may become activated through remodelling. This in turn would inhibit gyrase to an, as yet, unknown extent and potentially contribute to decreasing growth rates and increasing resistance to environmental stress. This would indeed be the first confirmed case for a post-translational mechanism of toxin release; those hypothesised to involve degradation of the antitoxin when complexed has been recently disputed (Song and Wood, 2020).

At this moment, we do not know whether the ParDE2 complex would behave in a similar manner, remodelling in response to temperature. This is unlikely as purification and analysis via size exclusion chromatography routinely results in a single peak (**Figure 3.16**, black curve). It is highly likely that ParDE2 exists in the typical 2:2 stoichiometry seen for ParDE and RelBE superfamily TA systems (**Table 1.4**) (Francuski and Saenger, 2009; Overgaard, Borch and Gerdes, 2009; Dalton and Crosson, 2010; Snead, Moore and Bourne, 2022); our AlphaFold generated multimeric models (presented in **Chapter 4**) provided two reasonable solutions, with 'model 2' being best supported by our SEC data (**Figure 3.16 D**).

What we do observe to be consistent between the ParDE1 and ParDE2 systems is a potentially lowered ParD – ParE affinity. While the ParDE1 system is responsive to environmental stimuli and the complex remodelling process theoretically causes a number of $ParD1_1ParE1_1$ dimers to dissociate, we find that ParDE2 dissociates during purification (**Figure 3.4**). Specifically, the tag cleavage step in low salt must come after the initial Ni-NTA purification step, followed by ortho-Ni-NTA directly on to anion exchange. Separation from both the ParDE2 complex and ParD2 antitoxin is possible by their distinct isoelectric points. Only a small fraction of the complex appears to dissociate and the process is inconsistent in yield. We hypothesised that tag removal somehow caused the complex to be forced apart and the low salt environment being unfavourable for ParDE2 complex reconstitution. This would indicate that concentration may impact the final yield as the ParD2 – ParE2 K_d will impact whether the complex reconstitutes. This process requires optimisation for consistency in future. Once purified, the sample is generally of high purity (**Figure 3.4 A and B**). Nuclease analysis initially indicated that the ParE2 toxin is capable of ribonuclease activity on its own (**Figure 3.4 C**), causing increased linear species DNA.

Further investigation revealed this was likely a contaminant from the purification, that is, unless antitoxin ParD2 did not neutralise the toxin in this assay (**Figure 3.4 D**).

Importantly, we show that ParE2 is capable of generating linear species in a gyrase-dependent manner (**Figure 3.5**), beyond any background level of nuclease activity; this likelihood is that this functions by trapping the cleavage complex. Previous reports suggest that *M. tuberculosis* ParE2 interacts with the gyrase holoenzyme via the GyrB subunit, while not impairing ATPase activity (Gupta *et al.*, 2016). Contrastingly, the *V. cholerae* ParE toxin has been shown to interact with GyrA (Yuan *et al.*, 2010), indicating that ParE toxins may have evolved to bind diverse regions of the gyrase enzyme. Coincidentally, this ParE toxin is part of the hetero-octameric ParDE complex referred to earlier (Garcia-Rodriguez *et al.*, 2021). Our results here indicate that toxicity is dependent on the dimerisation of $GyrB_2$ and $GyrA_2$, indicated by the significantly increased amount of linear DNA seen in the GyrBA fusion at low ParE2 concentrations (**Figure 3.8 A**) when compared to low ParE2 concentrations tested against the holoenzyme (**Figure 3.5 B**). Specifically, at a concentration of 0.625 µM ParE2, only around 20 % of the substrate is converted into linear species with the holoenzyme (**Figure 3.5 B**), whereas around 60 % is converted to linear species with the GyrBA fusion (**Figure 3.8 A**). The likelihood is that this is due to a more readily available GyrB – GyrA interface to which ParE2 may bind in the GyrBA fusion due to its dimeric nature. This also supports future study of the ParE2 – GyrBA complex in structural work due to an apparent increased affinity. This specific comparison also supports that the observed linearisation is induced by ParE2 rather than a contaminant as it has become gyrase protein dependent.

Testing ParE2 against the smaller gyrase fusion proteins (**Figure 3.8 B and C**) informed future structural work, but additionally provided mutation analysis of potential ParE2 binding sites as significant portions of the GyrBA fusion are deleted (**Figure 3.6**). We hypothesised, given the supposed affinity of ParE2 for the GyrB subunit (Gupta *et al.*, 2016), that linearisation would be equal to the full-fusion when testing the $GyrBA^{56}$ fusion as the B subunit remains intact. In fact, we see significantly decreased levels of linearisation (**Figure 3.8 B**), which would immediately indicate some level of reliance on the existence of the GyrA CTD (**Figure 3.6**). Further to this, when removing the ATPase domain in the testing of the $GyrB^{28}A^{56}$ fusion we see significantly decreased levels of linearisation again, when compared to both the full fusion (**Figure 3.8 A**) and $GyrBA^{56}$ fusion (**Figure 3.8 B**). Taken together, this indicates some level of reliance of the ParE2 toxin on both the GyrA CTD and GyrB ATPase for binding, which is unlikely. Rather, it is more likely the ParE2 toxin relies on gyrase catalytic mechanisms that are assisted by the presence of both of these domains. We hypothesise that the ParE2 toxin likely binds in the region of the GyrB transducer and is supported by the proximity of this domain with GyrA, supporting the results from the GyrBA and holoenzyme assays (**Figure 3.8 A and Figure 3.5 B**, respectively), and also

findings from Gupta *et al.,* (2016). This would be a similar mechanism to that seen for the inhibitory effects of thiophene antibacterial compounds, which allosterically stabilise the dsDNA breaks (Chan *et al.,* 2017). The effects of the removal of the GyrA CTD and ATPase in reducing the levels of ParE2 induced linearisation (**Figure 3.8 B and C**) indicate that the CTD and ATPase may both influence G-gate dynamics in cleavage and resealing, and hence alter the potency of trapping the cleavage complex on addition of ParE2.

Analytical SEC studies of ParE2-gyrase complexes using the gyrase fusion proteins will complement the biochemical data gathered in this chapter (**Figure 3.5 and Figure 3.8**). Fusion proteins may be preferred in these experiments due to the apparently higher affinity of ParE2 for the full-length GyrBA fusion (**Figure 3.8 A**). While visualising the potential Mr shift of +12.18 kDa (+ParE2 monomer) for the gyrase proteins may be difficult, we may simply be able to observe the decrease in intensity, or total loss of, the ParE2 chromatographic peak when added to gyrase proteins prior to analysis. Beyond these biochemical and biophysical studies, the generation of ParE2-resistant gyrase fusion mutants through error prone PCR (Abou-Nader and Benedik, 2010) of the expression vectors could be a useful tool in furthering our proposals regarding the ParE2 binding site. Resistance to ParE2 would be tested via initially confirming *in vitro* activity in typical gyrase functions (**Figure 3.2**), then subsequently subjecting the mutant fusion to both the biochemical cleavage studies previously presented (**Figure 3.5**), and the biophysical studies proposed. Sequencing of the mutated expression vectors would allow us to map the mutations to Mtb gyrase protein structures, and to the homologous full-length *E. coli* gyrase structure, to highlight a putative site of ParE2 interaction. This may be a 'needle in a hay stack' approach, therefore a reasonable starting point may be to generate the mutants resistant to thiophene poisoning (Chan *et al.*, 2017) as we propose ParE2 may operate via a similar binding site.

The gold standard in elucidating the binding site of the ParE toxins to gyrase is structural biology via either crystallographic or cryo-EM studies. An important first step in this process is to better understand the structures of the ParE toxins themselves and this will be possible through crystallisation of the individual ParDE complexes. This would add to the small complement of ParDE system structures in the PDB (**Table 1.4**) and further our understanding of both toxin and antitoxin structure conservation within this family. Not only will this assist in future structural studies, but it may also shed light on the proposed ParDE1 remodelling and toxin release process (**Figure 3.15**).

Chapter 4. Structural characterisation of the *Mycobacterium tuberculosis* ParDE toxin – antitoxin systems

Chapter 4. Introduction

Structural biology will be central to understanding the molecular basis of ParE toxin-induced inhibition of DNA gyrase demonstrated in **Chapter 3**. Given the previously described toxicity of ParE1 and ParE2, we proceeded with studying these proteins. Elucidating the structures of the *M. tuberculosis* ParDE complexes will add to the complement of ParDE structures available in the PDB (**Table 1.4**) (PDB: 3KXE (Dalton and Crosson, 2010); 5CEG (Aakre *et al.*, 2013); 60XA (Lite *et al.*, 2020); 7R5A (Garcia-Rodriguez et al., 2021); 6XRW (Snead, Moore and Bourne, 2022)), allow us to study the key residues involved in complex formation, and potentially highlight those essential to toxicity.

Importantly, given that we have previously demonstrated that the *M. tuberculosis* ParE toxins can be separated from their respective complexes, we may be able to highlight key toxin – antitoxin interactions that can be destabilised. Not only might this aid us in future studies to more efficiently isolate the toxins, it will also provide information regarding the dynamics of the systems in solution, thus developing beyond the biophysical studies presented in **Chapter 3**. Studies of the ParDE1 complex structure will further our developing model for ParE1 toxin release through complex remodelling. It is currently predicted that ParD1 antitoxin chains are somehow donated to another complex, with their bound ParE1 toxins being released in the process; highlighting the interactions between chains will be an important step toward understanding the molecular mechanism behind thermally driven toxin release. This may also allow us to develop our current understanding of post-translational toxin release throughout TA systems by identifying key structural motifs and domains involved in the process that are conserved between different systems.

Further to this, it will be important to utilise the data, both released in the AlphaFold database (Jumper *et al.*, 2021; Varadi *et al.*, 2022) and produced through the AlphaFold Colaboratory (Mirdita *et al.*, 2022) (**Materials & Methods 2.10**), in potentially structure solving and importantly in subsequent structural comparisons. This will allow us to continue to validate the quality of the monomer and multimer models generated by artificial intelligence. From this, we may subsequently be able to develop models for hypothesised ParDE system complexes, and in future to develop models for the ParE – gyrase interaction.

4.1 ParDE1 complex crystal structure

ParD1 and ParE1 were co-expressed and the resulting protein complex was purified as per **Materials and Methods 2.4.2 and 2.5.8**. The purified sample, containing the ParD1-ParE1 protein complex (purification demonstrated in **Chapter 3**, **Figure 3.1**) was used to set crystal screens and optimise crystal conditions, as detailed in **Materials & Methods 2.8.1 and 2.8.2.** Crystals grew as needles, visually suitable forms were harvested, cryo-protected and X-ray crystallographic data were collected as described in **Materials and Methods 2.9**. Single datasets were collected from 3 crystals and subsequently merged to a final resolution of 2.10 Å and the ParDE1 complex structure was determined by molecular replacement in space group P1 2_1 1 using PDB: 3KXE (Dalton and Crosson, 2010) as a search model (**Materials and Methods 2.9.3**, and **Table 4.1**). Interestingly, whilst phasing was successful, on viewing the unrefined density it was clear that a significant portion remained unmodeled.

The asymmetric unit contained a heterohexameric $ParD1_4ParE1_2$ stoichiometry complex, with two full-length ParD1 antitoxins resolved alongside two shorter, partially resolved ParD1 antitoxins (**Figure 4.1 A**). The structure somewhat resembles the search model (PDB: 3KXE) (Dalton and Crosson, 2010), with the ParD1 antitoxins interacting via an anti-parallel N-terminal beta sheet (only one pair of the antiparallel beta sheets was resolved in the structure) and a pair of ParE1 toxins positioned inside a cage-like structure. In the *C. crescentus* ParDE structure, the ParD antitoxins interact through an additional coiled-coiled between the corresponding second alpha helix; this structural motif is not seen between the full-length ParD1 chains in the *M. tuberculosis* ParDE1 structure, rather, the chains appear to abut one another. A coiled-coil is present, however, between the full-length – partially resolved ParD1 chains alongside the anti-parallel beta sheet (**Figure 4.1 A**). 18 non-hydrogen ligand atoms (**Table 4.1**) are built in as 3 molecules of glycerol and are not believed to contribute to complex stability or function.

The ParE1 toxin is comprised of α1 (P9 – W26), α2 (V28 – A47), α3 (P49 – I51), β1 (R61 – A67), β2 (H70 – T77), β3 (G80 – H89), and α4 (Q90 – M92) (**Figure 4.1 B**). α1 and α2 form a hairpin-like structure as a base. The three beta strands form an antiparallel sheet that sits above the hairpin, followed by a short helix (α4), which then becomes an unstructured C-terminal 6 residues (with a further 3 remaining unmodeled due to lack of density, likely due to high flexibility in this region). Using the Dali server (Holm and Sander, 1993; Holm, 2020) allows us to search for structural homologues in the PDB, thereby identifying conserved protein folds. The highest scoring Dali search results for the ParE1 monomer all belong to the RelE/ParE superfamily, with the top 10 results ranging from an RMSD 0.9 Å (PDB: 3KXE – *C. crescentus* ParE1; search model for MR), to 2.3 Å (PDB: 3OEI – *M. tuberculosis* RelE3). In contrast, the ParD1 antitoxin has a simpler secondary structure; for the purposes of simplicity, I will refer to the fully resolved ParD antitoxin chains as

'primary' and the partially resolved chains as 'auxiliary'. The primary chains are comprised of β1 (T5 – V8), α1 (E13 – A23), α2 (A29 – S60), α3 (F68 – S80). The auxiliary chains are identical to the primary chains from residues G2 – E55, with the C-terminal A56 – R83 being unresolved (**Figure 4.1 B**).

Dali searches for the ParD1 antitoxin were performed on both the primary and auxiliary chain. The top 10 results when analysing the primary (full-length) ParD1 chain ranged from ParD antitoxin chains RMSD values of 5 Å (PDB: 6X0A, *M. opportunistum* ParDE (Lite *et al.*, 2020)) to 3.0 Å (PDB: 3KXE, search model (Dalton and Crosson, 2010)) and 2.4 Å (PDB: 7R5A, *V. cholera* ParDE (Garcia-Rodriguez et al., 2021)). The top 10 results for the auxiliary ParD1 chain contained only chains from PDB: 7B22 (Garcia-Rodriguez et al., 2021) and PDB: 7R5A (Garcia-Rodriguez et al., 2021) and range from 0.9 Å to 1.4 Å RMSD; these low values are unsurprising given that this is likely the more stable structured domain. Additionally, 7B22 and 7R5A are structures of the *V. cholera* ParD antitoxin and ParDE complex, respectively. Dali searches when entering the ParD1 dimer and tetramer indicated structural similarity to CopG transcriptional regulator proteins with RMSD values of < 2.0 Å localised to the dimerised N-terminus (PDB: 6IYA (Costa *et al.*, 2001); 1EA4 (Zhao *et al.*, 2019)) due to the ribbon-helix-helix (RHH) motif created through the antiparallel beta sheets (**Figure 4.1 A; Figure 4.2**). This indicates that ParD1 belongs to the CopG/Arc/MetJ (RHH/CopG – **Figure 1.7**) family of transcriptional regulators (Xavier Gomis-Rüth *et al.*, 1998).

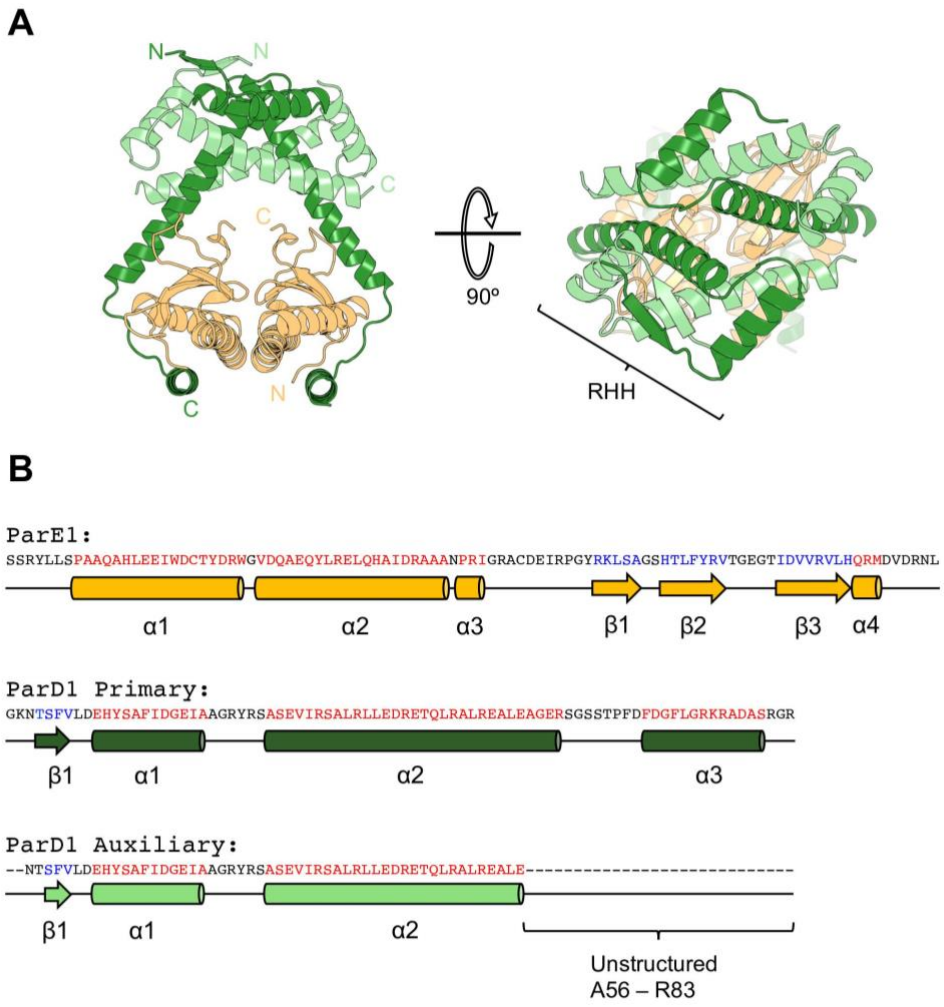
Manual searches of the ParDE structures in the PDB indicates that ParD antitoxins have a consistent fold with the CopG N-terminal domain followed by two toxin-interfacing helices. It is also evident that this domain is shared with other TA systems such as MazEF (Ahn *et al.*, 2017). Building on these results, it was apparent that ParDE complexes have increased plasticity in their stoichiometries; while 3KXE, 6X0A, and 6XRW all exist as 2:2 heterotetramers, 5CEG ($ParD_4$:$ParE_4$ stoichiometry) and 7R5A ($ParD_6$:$ParE_2$ stoichiometry) exist as heterooctamers. Even more noteworthy is that the 7R5A structure resembles the ParDE1 complex with both fully and partially resolved ParD chains forming RHH CopG N-terminal dimers, however, an additional partially resolved ParD1 dimer has places itself in-between the full-length ParD chains. Further to this, the 7B22 structure of only the *V. cholera* ParD chains forms a hetero-16mer (8 dimers) in a ring-like structure. Altogether, these results indicate that the ParD N-terminal domain permits higher-order stoichiometries to form and supports the biophysical data and model currently developing for *M. tuberculosis* ParDE1, albeit in other models generally with conditional cooperativity and autoregulatory capacities (Garcia-Rodriguez et al., 2021).

Table 4.1 Data collection and refinement statistics for ParDE1

	ParDE1
PDB ID Code	8C24
Number of crystals	3
Beamline	Diamond I04
Wavelength, Å	0.9795
Resolution range, Å	45.92 – 2.10 (2.175 – 2.10)
Space group	P 1 21 1
Unit cell, *a b c* (Å), α β γ (°)	44.56 125.45 52.26, 90 90 90
Total reflections	207238
Unique reflections	31537 (3147)
Multiplicity	1.9
Completeness (%)	99.58 (99.05)
Mean I/sigma(I)	9.13
R_{merge}	0.036 (0.478)
R_{meas}	0.050 (0.676)
$CC_{1/2}$	0.999 (0.669)
R_{work}	0.198 (0.297)
R_{free}	0.234 (0.333)
No. of non-hydrogen atoms	3634
Macromolecules	3519
Ligands	18
Solvent	97
Protein Residues	440
RMSD (bonds, Å)	0.007
RMSD (angles, °)	0.78
Ramachandran favoured (%)	98.36
Ramachandran allowed (%)	1.64
Ramachandran outliers (%)	0.00
Average B-factor	47.62
Macromolecules	47.48
Ligands	58.82
Solvent	50.65

Values in parenthesis are for the highest resolution shell. R_{free} was calculated with 5% of the reflections selected.

Figure 4.1 ParDE1 crystal structure and secondary structures of the component ParE1 and ParD1 proteins

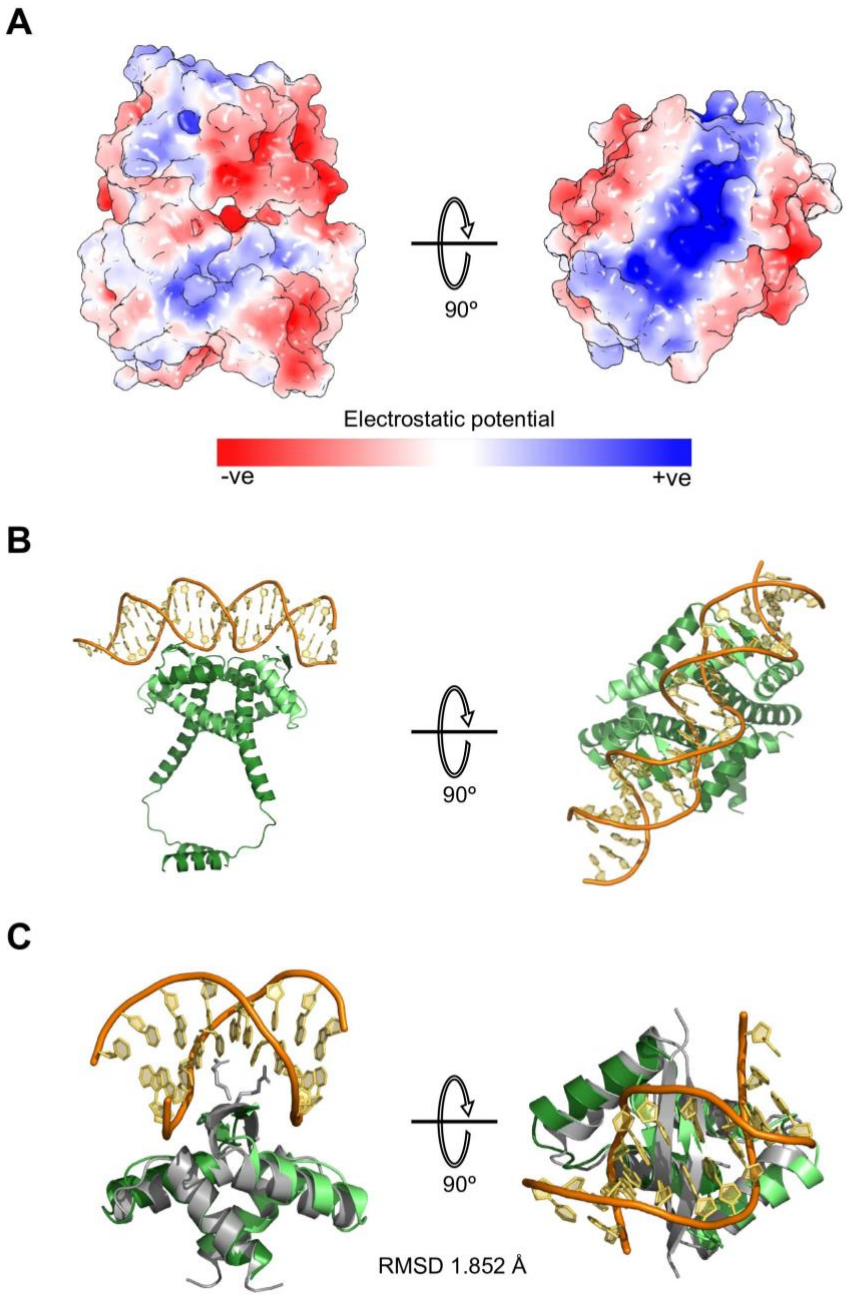


(**A**) Cartoon representation of the heterohexameric $ParD1_4ParE1_2$ complex crystal structure. ParD1 antitoxins exist in two forms within the structure; two full-length (primary) chains are coloured dark green with N and C termini labelled, and two short-length (auxiliary), partially resolved chains are coloured pale green. The two ParE1 toxins are coloured light orange with N and C termini labelled. The ParD1 tetramer is formed through a beta sheet between the full- and short-length antitoxins and creates a cage-like structure around the toxins. Rotated view (right) shows the N-terminal dimerization domain of the ParD1 antitoxins forming a ribbon-helix-helix structure (RHH); (**B**) Secondary structure schematics of ParE1, ParD1 primary, and ParD1 auxiliary proteins. Secondary structure is displayed underneath the protein sequence, whereby amino acids constituting α helices are coloured red, and those constituting β strands are coloured blue. Secondary structure elements are coloured as per their crystal structure counterpart and represented by tubes for α helices and arrows for β strands and are labelled accordingly. Dashes represent unmodelled regions.

4.1.1 The ParDE1 complex contains a putative DNA-binding domain

Surface electrostatics reveal a largely electronegative complex, with an electropositive patch created at the antitoxin complexing region (**Figure 4.2 A**). This electropositive patch, alongside the formation of a RHH motif (**Figure 1.7**) (CopG DNA-binding domain) between the antitoxins at the antiparallel beta sheet, indicates a potential for DNA-binding and autoregulation (Xavier Gomis-Rüth *et al.*, 1998). This is consistent with the literature for type II TA systems in general (Wen *et al.*, 2018; Qian *et al.*, 2019; Beck *et al.*, 2020), and autoregulation has been demonstrated in ParDE systems via both ParD alone, and the ParDE complex (Xu *et al.*, 2018). Crystal structures of the *S. agalactiae* CopG RHH have been solved as a tetramer (PDB: 2CPG), a tetramer bound to a single operator site (PDB: 1B01) (Xavier Gomis-Rüth *et al.*, 1998) and a tetramer bound to two operator sites, inducing a DNA-bend of 60° (PDB: 1EA4 – identified prior in the Dali search of ParD1) (Holm and Sander, 1993; Holm, 2020). Using these structures, we can model the ParD1 tetramer interaction with DNA whereby the antiparallel beta sheets insert into the major groove of the bent DNA from 1EA4 (**Figure 4.2 B**). While this structural superposition allows for correct positioning of the DNA at the putative DNA-binding domain of the ParDE1 complex, further analysis demonstrates a potential loss of DNA binding capabilities. Using a structural superposition of the ParD primary – ParD auxiliary N-terminal 42 amino acid dimer with 1B01 (RMSD 1.852 Å), it is evident that an arginine residue (R4) of CopG inserts into the major groove of the operator site DNA; this is not present in the ParD1 structure (**Figure 4.2 C**). The same analysis can be performed using the RHH motif containing FitAB TA system structure bound to DNA (PDB: 2BSQ) (Mattison *et al.*, 2006). The ParD1 dimer structurally superposes to a respectable RMSD (2.887 Å) positioning the antiparallel beta sheet within the major groove of the 2BSQ operator DNA, however, the FitA antitoxin also positions an arginine residue (R7) into the major groove highlighting a potentially essential interaction for DNA-binding that the ParDE1 complex apparently lacks. Furthermore, a putative operator site (usually an inverted repeat sequence, as is seen for the FitAB system) (Mattison *et al.*, 2006) could not be identified up to 2000 bp upstream of *parD1* (*rv1960c*), nor could the operator sequence from 1EA4 be aligned with the same upstream region. These findings indicate a potential loss of autoregulatory capacity in this system, despite the presence of the structurally conserved RHH DNA-binding motif and positive charge at the antitoxin dimerisation interface.

Figure 4.2 ParDE1 complex electrostatics and DNA-binding models



(**A**) Surface rendering of the ParDE1 complex crystal structure coloured by electrostatic potential using the APBS plugin (PyMol); (**B**) DNA-bound ParD1 tetramer model (coloured as per Figure 1 A) generated using PDB 1EA4. Both RHH domains are present, generated by superimposing the fully built ParD1 chains onto those missing the N-terminal residues, to demonstrate how these align with the operator sites. Left and right panels of A and B are equivalent views, respectively; (**C**) DNA-bound model for a single RHH domain from ParDE1 generated using, and aligned to, PDB 1B01 (coloured grey).

4.1.2 ParD1 – ParE1 protomer comparisons

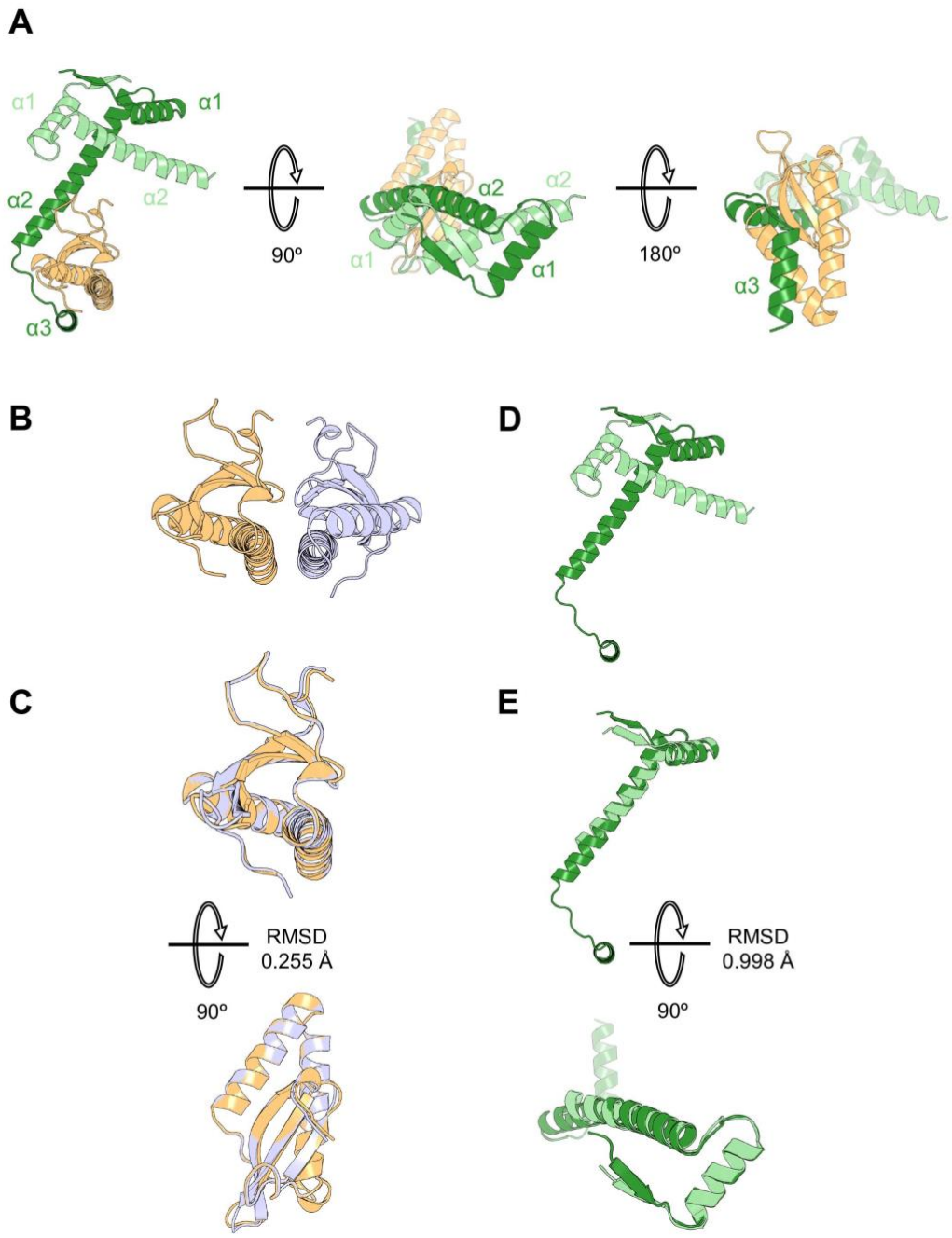
The crystal structure can be broken down into a trimer of interacting components; ParD1 primary – ParD1 auxiliary – ParE1. The major interactions appear to occur between the primary ParD1 antitoxin α2 – loop – α3 and the ParE1 toxin, alongside an antiparallel beta sheet formed between the primary and auxiliary ParD antitoxins (**Figure 4.3 A**). Examination of the toxin – antitoxin complex constituents indicates that ParE1 and ParD1 may form respective homodimers (**Figure 4.3 B and D**). Further analysis of the individual protomers within the complex demonstrates little flexibility in the toxin structure; aligning the two monomers returned an RMSD of 0.255 Å (**Figure 4.3 C**). Contrastingly, the ParD1 antitoxins have a higher degree of flexibility through their largely helical structure, with the primary – primary alignment returning an RMSD of 1.667 Å, and the primary – auxiliary alignment returning an RMSD of 0.998 Å. This indicates a high degree of structural alignment, especially in the N-terminal region. However, there is clearly flexibility throughout the chain in the with the highest degree seen in the C-terminal toxin-interacting region of ParD1 which would on its own, likely remain unstructured (**Figure 4.3 E**).

4.2 ParDE1 complex interfaces and protein recognition

4.2.1 ParDE1 complex assembly

PISA analysis (Krissinel and Henrick, 2007) of the crystal structure supports the above observations somewhat, by calculating the surface area of the interfaces (Å^2) and estimating the free-energy change (Δ^iG kcal/mol) on interfacing. Interfaces are then scored by "complex significance" from 0 – 1, with 1 being an essential interaction for complex assembly. The primary – auxiliary ParD interaction via the antiparallel beta sheet scored 1 (CSS) with an interface area of 1560.4 Å^2 and a free-energy change of -22.4 kcal/mol (**Figure 4.4 A i and B**). The primary ParD1 – ParE1 interaction through the C-terminal portion of ParD1 α2 – loop – α3 also scored 1 (CSS). The interface area was calculated as 1458 Å^2 with a free-energy change of -12.5 kcal/mol (**Figure 4.4 A ii and B**). Electron density at this essential interface is demonstrated in **Figure S2 A**, showing the 2Fo-Fc map for the relevant residues of the ParD1 antitoxin. The only other two interfaces of any significance within the structure, ParE1 – ParE1 (919.3 Å^2) and primary ParD1 – primary ParD1 (491.3 Å^2) (**Figure 4.4 A iii and iv**, respectively), are considered to be of low importance in complex formation (**Figure 4 B**). These analyses indicate not only that the essential interactions for ParDE1 complex formation are at the antitoxin antiparallel beta sheet and the primary ParD1 – ParE1 interfaces, but also that this is the respective order in which these interactions form, and how the complex assembles (indicated by free-energy change predication). Therefore, it is also likely that ParD1 exists as a dimer prior to full ParDE1 complex formation.

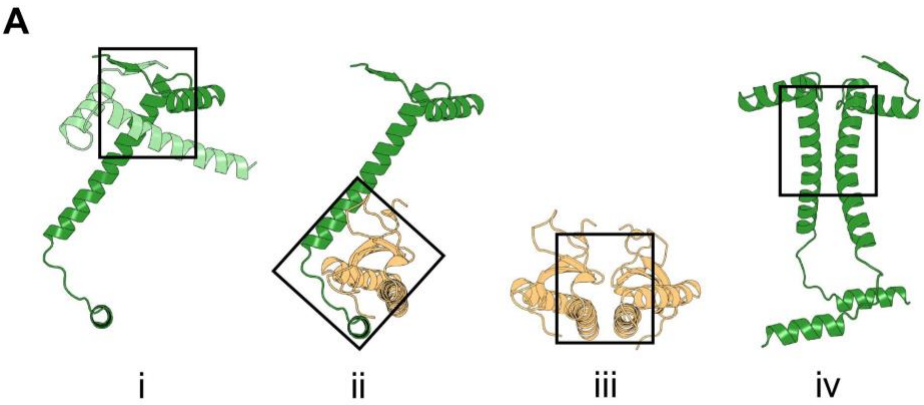
Figure 4.3 Protomer analysis of the ParDE1 complex components



(**A**) Trimer of interacting components from the ParDE1 complex. Interactions occur between the primary (dark green) and auxiliary (pale green) ParD1 proteins via an antiparallel beta sheet, and between primary ParD1 (dark green) and ParE1 (light orange); (**B**) ParE1 monomers, right side recoloured pale blue; (**C**) Alignment of the ParE1 protomers from B; (**D**) ParD1 (primary – auxiliary) dimer; (**E**) Alignment of the two ParD1 protomers from D.

Figure 4.4 Essential interfaces for assembly of the ParDE1 complex

A



B

Interface	Interface area (Å^2)	Δ^iG kcal/mol	Δ^iG P-value	N^{HB}	N^{SB}	N^{DS}	CSS
i	1560.4	-22.4	0.140	20	13	0	1.000
ii	1458.0	-12.5	0.243	27	19	0	1.000
iii	919.3	-1.4	0.481	12	6	0	0.079
iv	491.3	5.9	0.941	19	17	0	0.053

(**A**) Interfaces of the ParDE1 complex analysed by PISA. i: Primary ParD1 – Auxiliary ParD1 interface formed by an antiparallel beta sheet; ii: Primary ParD1 – ParE1 interface; iii: ParE1 – ParE1 interface; iv: Primary ParD1 – Primary ParD1 interface; (**B**) PISA analysis of interfaces i – iv. Interface area is calculated in square angstroms (Å^2); Δ^iG – solvation energy on folding (free energy gain); N^{HB} – Number of hydrogen bonds within the respective interface; N^{SB} – Number of salt bridges within the respective interface; N^{DS} – Number of disulphide bonds formed within the respective interface; CSS – Complex Significance Score ranges between 0 – 1 as interface relevance to complex formation increases.

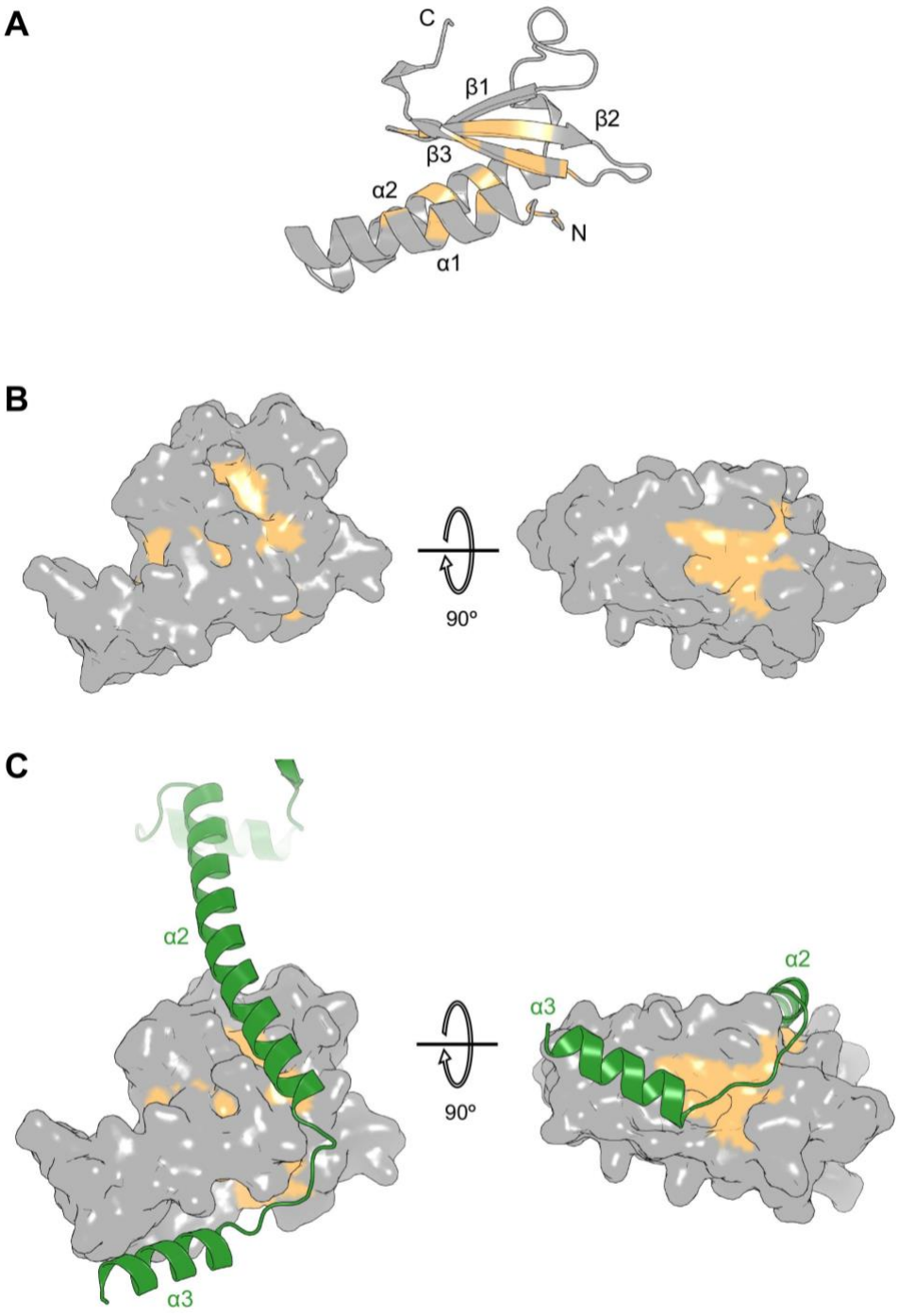
4.2.2 The ParDE1 system demonstrates a conserved mechanism of protein recognition

Analysis of the *C. crescentus* ParDE (PDB: 3KXE) crystal structure performed by Dalton & Crosson (Dalton and Crosson, 2010) allowed for the elucidation of a conserved mode of protein recognition. The conservation of potentially important residues was initially highlighted via multiple sequence alignment across TA system RelE/ParE superfamily of toxins (Anantharaman and Aravind, 2003). Highlighting the most conserved (>80%) residues across the toxin sequences and representing these on the surface of the toxin structure within the respective ParDE complex allows us to demonstrate this conserved mode of recognition (**Figure 4.5**). It is clear to see that the highest conserved residues across RelE/ParE superfamily toxins are concentrated on the internal facing portions of α1 and α2, and along β2 and β3 (**Figure 4.5 A**). Mapping these to the surface of the ParE1 toxin shows two highly conserved hydrophobic patches; one along the groove created by the antiparallel β strands (1 – 3), and the other on the underside of the toxin structure between the hairpin of α helices 1 and 2 (**Figure 4.5 B**). Representing this highlighted surface alongside the primary ParD1 antitoxin further highlights the close spatial relationship between the antitoxin and the highly conserved residues at the toxin surface (**Figure 4.5 C**).

4.2.3 Polar contacts stabilise the ParDE1 quaternary structure

The PISA analysis from **Figure 4.4** indicates that interfaces i (Primary – Auxiliary ParD1) and ii (Primary ParD1 – ParE1) form a $ParD1_2ParE1$ trimer (as shown in **Figure 4.3 A**). The ParD1 – ParE1 interaction has been shown to occur through a conserved interface, further highlighting the importance of this in complex formation (Dalton and Crosson, 2010) (**Figure 4.5**). Interfaces iii and iv appear to stabilise the heterohexameric structure (**Figure 4.4 A**). Interestingly, the ParE1 – ParE1 interface is consistent with the *C. crescentus* structure (PDB: 3KXE), however, the interface between the abutting, full-length ParD1 chains is not; this has closer resemblance to the *V. cholera* $ParD_6ParE_2$ structure, whereby three ParD dimers appear to multimerise through their N-terminal CopG domains (Garcia-Rodriguez et al., 2021). Several polar contacts are evident at interfaces iii and iv and as summarised in **Table 4.2**. The ParDE1 complex is largely stabilised via a network of ionic bonds at interface iv (between Primary ParD1 chains), partly explaining the conservation of charged residues in this area and the presentation of the electropositive surface (**Figure 4.2 A**) as arginine residues are concentrated toward the N-terminus of ParD1. Several ionic bonds and polar contacts also form between the ParE1 monomers. Both interfaces iii and iv may therefore play a key role in the hypothesised remodelling process as theoretical ParDE1 heterotetramers could interact at this region, subsequently releasing a ParE1 toxin each, while two $ParD1_2E1$ trimers (**Figure 4.3 A**) remain stabilised by the polar contacts of interface iv (**Figure 4.4 A**).

Figure 4.5 Conserved mechanism of protein recognition in the ParDE1 system



(**A**) Cartoon representation of the ParE1 toxin, coloured grey, with conserved residues from the RelE/ParE superfamily highlighted in light orange; (**B**) Surface rendered ParE1 toxin with conserved residues highlighted in light orange from A. Left side view is equivalent to the positioning in A; (**C**) Surface rendered ParE1 with conserved residues in light orange (equivalent views to B) complexed with Primary ParD1 (coloured forest green). Secondary structure elements are labelled where appropriate throughout.

Table 4.2 Polar contacts stabilising the heterohexameric ParDE1 complex

Chain 1	Chain 2	Å	Polar/Ionic
ParD1 (C)	**ParD1 (D)**		
R25	D42	3.2	Ionic
R25	E41	2.9	Ionic
R25	R34	3.5	Polar*
R25	R38	3.4	Polar*
S30	R27	2.7	Polar
E31	R34	2.8	Ionic
E31	R38	3.2	Ionic
R34	E31	2.9	Ionic
R38	E31	3.0	Ionic
E41	R25	2.4	Ionic
D42	R25	2.9	Ionic
ParE1 (A)	**ParE1 (B)**		
Q30	N48	2.7	Polar
Q30	R34	3.1	Polar
E37	R44	2.6	Ionic
H40	Q33	3.4	Polar
R44	E37	2.7	Ionic
R44	Q30	3.1	Polar*
S66	Q90	3.0	Polar
Q90	S66	3.0	Polar

*Interaction is between an amino acid side chain and main chain. List of polar interactions between designated chains within the ParDE1 crystal structure identified in PyMol. Amino acid single letter code is followed by residue number alongside the distance between the polar interacting region in Ångstroms (Å). Interactions are categorised as Polar where interaction is not charge-based, and Ionic where the interaction forms an ionic bond due to charge.

4.3 Amino acid conservation analysis of ParD1 – ParE1

To better understand the importance of the position and accompanying characteristics of specific amino acids within the ParDE1 structure a bioinformatic approach was used to analyse evolutionary conservation based on homologous sequences. The ConSurf server (Ashkenazy *et al.*, 2016) generates a multiple sequence alignment of homologous sequences using the input PDB structure, ranks individual amino acids from lowest conserved to highest, scales the scores and allows for the plotting of these scored amino acids onto a crystal structure model by colour (Landau *et al.*, 2005; Ashkenazy *et al.*, 2016). This allows us to visually demonstrate the scaled conservation scores across the individual protein structures, highlighting areas of highest and lowest conservation.

4.3.1 Multiple sequence alignment highlights conserved secondary structure components

Using the output multiple sequence alignment (MSA) from ConSurf (Ashkenazy *et al.*, 2016), we can visually represent the highest conserved amino acids across the entire alignment. Using JalView (Waterhouse *et al.*, 2009) we can colour amino acids by the Clustal colouring scheme (assigning specific colours to residues by type) and increase the conservation threshold to colour only those conserved across the MSA to a user-defined percentage.

Independent MSA outputs were generated by inputting the ParE1 structure PDB and the ParD1 (primary) structure PDB to ConSurf, followed by visualisation with a conservation threshold set at 80% (**Figure 4.6 A and B**). In both **4 A** and **4 B** the input sequence is displayed at the top of the alignment and the first 25 aligned sequences (out of the 250-sequence alignment) are shown with the consensus sequence represented below. By aligning the secondary structure of ParE1 and ParD1 to their respective MSA we can show where the most highly conserved amino acids are concentrated in each structure, alongside identifying their characteristics.

From the aligned ParE1 secondary structure we can see that most of the conserved amino acids are hydrophobic, and these are concentrated at α1, β2 and β3 (**Figure 4.6 A**). A highly conserved positively charged arginine residue is also present in β3, and interestingly, we can see a conserved negatively charged aspartic acid in α1, however, ParE1 contains a histidine at this position. There is little to no sequence conservation observed from α3 through β1, and at the C-terminal flexible region including α4.

From the aligned ParD1 secondary structure we can see that most of the conserved amino acids are, again, hydrophobic and appear highly concentrated on α2. This is unsurprising given the identification of the conserved hydrophobic protein interaction (Dalton and Crosson, 2010). We also see a highly conserved negatively charged aspartic acid/glutamic acid, and two positively

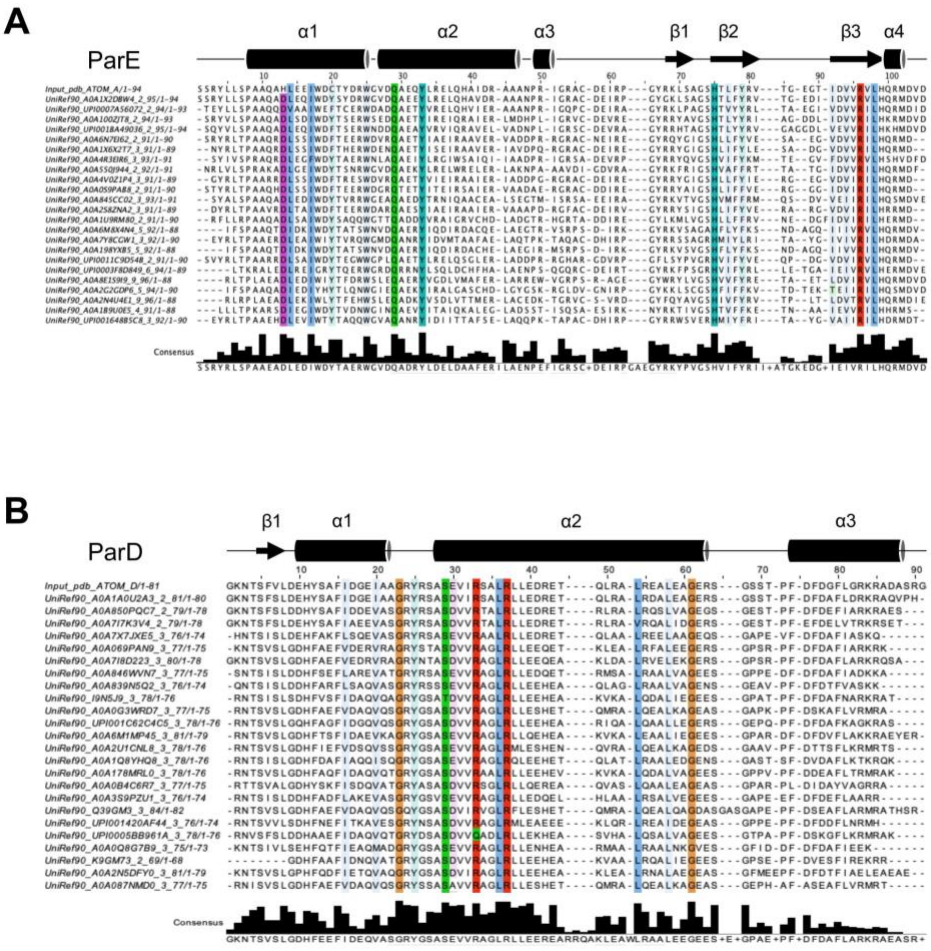
charged arginine residues at the N-terminal end of α2 (**Figure 4.6 B**). The MSA and secondary structure alignment for ParD1 also demonstrates a significant reduction in amino acid conservation towards to the C-terminus, indicated by the lack of colouring and consensus sequence bar chart.

Taken together, this indicates that the highest conserved amino acids across both ParE1 and ParD1 structures are hydrophobic, with some charged positions, and supports the model proposed by Dalton and Crosson (2010). To highlight the positioning and interactions of these highly conserved amino acids the ConSurf plot must be used to render conservation scores on to the crystal structures of ParE1 and ParD1.

4.3.2 Amino acid conservation within the ParE1 structure

Rendering the ParE1 crystal structure by amino acid conservation score allows us to show where the highest degree of sequence conservation maps to within the structurally conserved toxin protein. The ParE1 toxin shows high sequence conservation through β1 and β2 indicated by pink to purple colouring concentrated in this area, part of the interface for primary antitoxin binding (**Figure 4.7 A; Figure 4.4 A i**). Lower conservation scores are seen along the underside of the alpha helix hairpin, and the majority of α2, indicated by dark green to white colouring (**Figure 4.7 A**). Interestingly, the highest degree of conservation is seen on the top side of the hairpin, primarily α1 (notably I18) and at the N-terminal end of α2 (notably Y34) (**Figure 4.7 A and B**) and these residues appear to interact with the highly conserved H70, contributed by β2 (**Figure 4.7 B**). It is possible that the interactions between these amino acids contributes to the maintenance and stability of the ParE1 tertiary structure, whereby the three anti-parallel β strands sit atop the α helical hairpin. Rendering the surface of ParE1 demonstrates patches of high conservation and these map to the binding groove of the ParD1 antitoxin, as indicated by the boxes and pink colouring in **Figure 4 C**. We can also see that the top portion of the ParE1 toxin containing the β sheet has a far greater degree of conservation (pink – purple colouring), while the lower α helical portion is far less conserved (green to white colouring) (**Figure 4 C**).

Figure 4.6 Multiple sequence and secondary structure alignment of ParE1 and ParD1

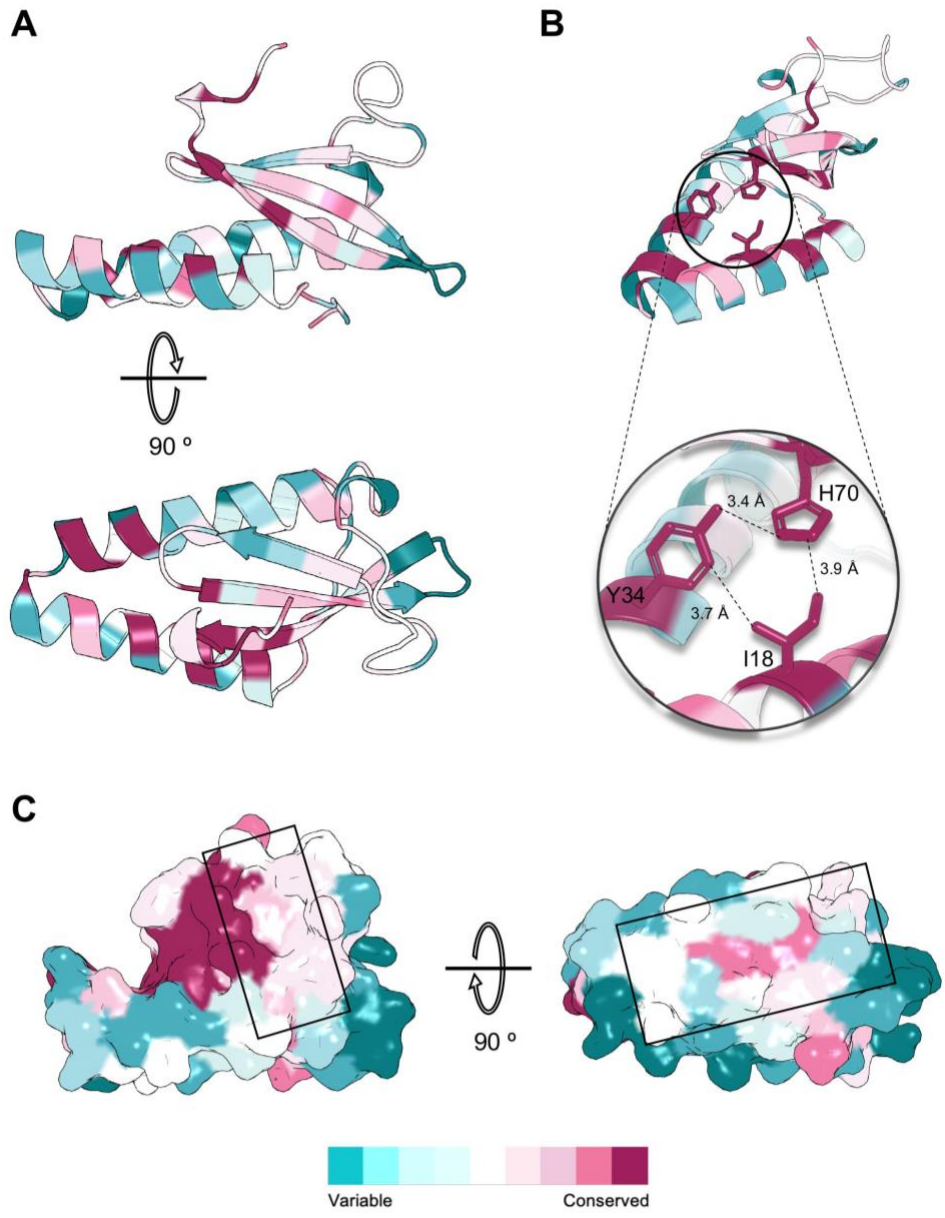


(**A**) ParE1; (**B**) ParD1. Both alignments are visualised in JalView, with Clustal colouring shown at 80% conservation. Matching secondary structure diagrams are presented above each alignment (cylinders represent α helices and arrows represent β strands). Consensus sequences and corresponding graphs are shown below each alignment. Multiple sequence alignment generated by ConSurf for both ParE1 and ParD1 using Clustal; top 25 sequences out of 125 are shown.

4.3.3 Amino acid conservation within the ParD1 structure

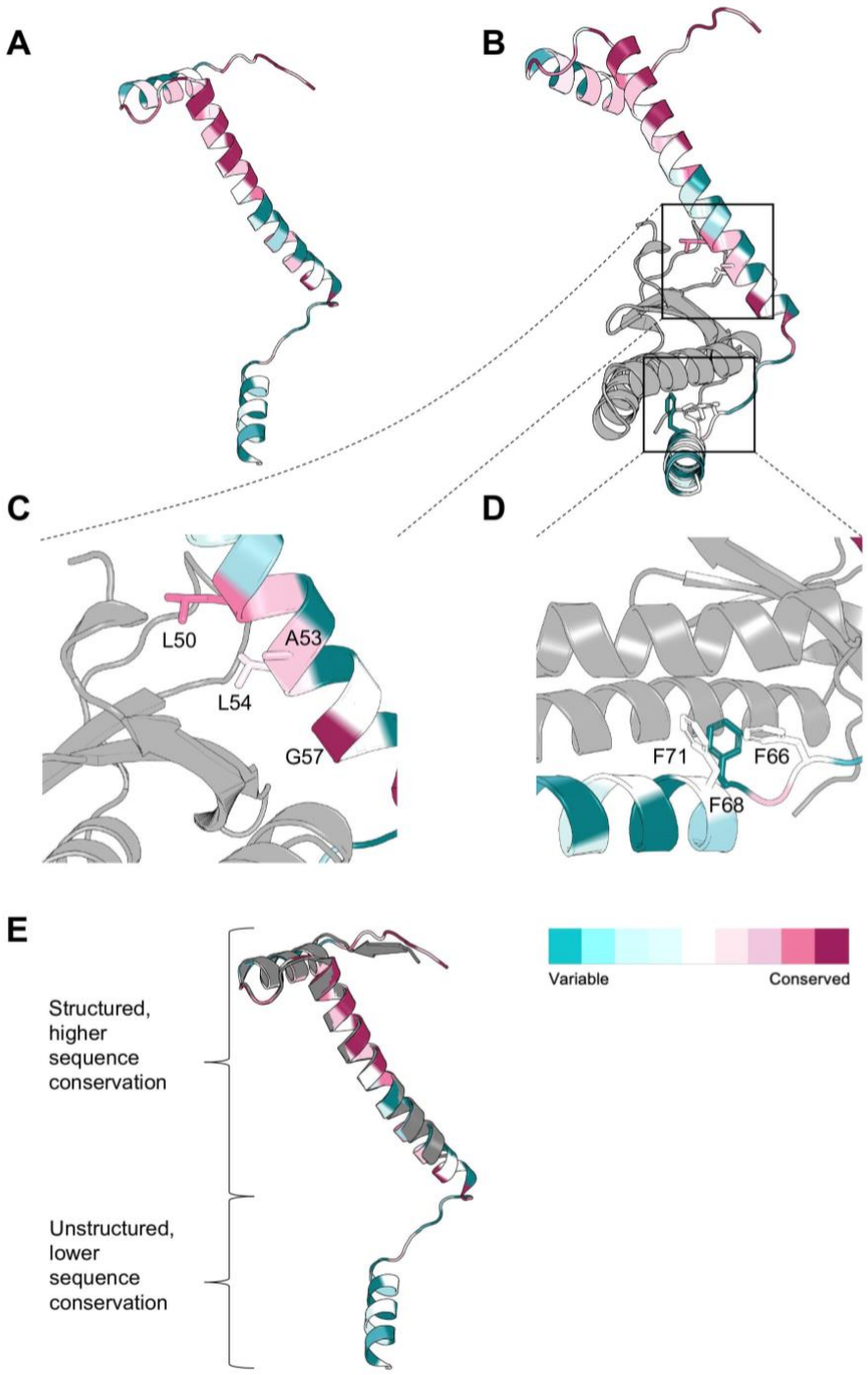
Rendering the ParD1 crystal structure by amino acid conservation indicates that the N-terminal domain is highly conserved at a region above the toxin binding region (**Figure 4.8 A and B; Figure 4.6 B**). The highly conserved amino acids at the N-terminal end of α2 likely conserve the structure of the antitoxin dimerisation domain, contributing to the most structured part of the ParD1 protein. At the C-terminal end of α2, which interacts with the β sheet interface of ParE1, we can see several conserved hydrophobic residues (L50, A53, L54, G57) (**Figure 4.8 B and C**). At the C-terminal end of the primary ParD1 structure we see a lower degree of conservation indicated by the general green – white colouring, including three phenylalanine residues (F66, F68, F71) which appear to insert into the underside space created by the α helical hairpin of ParE1 (**Figure 4.8 B and D**). This suggests that the N-terminal, structured, domain remains highly conserved in sequence and structure, while the C-terminal domain of ParD antitoxins has become variable, and potentially tuneable to its cognate toxin. This is reinforced when we align the auxiliary ParD1 structure to the ConSurf coloured primary ParD1 structure (**Figure 4.8 D**), showing that sequence conservation drops off (indicated by colouring) when the structural alignment stops at the unstructured region of the auxiliary ParD1 (A56 – R83).

Figure 4.7 ConSurf analysis of ParE1



(**A**) Rotated views of the ParE1 monomer cartoon; (**B**) Identification of closely positioned side chains, measurements in Ångstroms (Å), from the highest conserved residues within the ParE1 structure. Interactions between I18, Y34, and H70 may stabilise the overall tertiary structure; (**C**) Surface rendered ParE1 maintaining ConSurf colouring highlights highly conserved patches, displayed in boxes. Cartoon and surface representations of the isolated ParE1 monomer from the ParDE1 crystal structure are rendered in ConSurf colouring scheme (shown as a gradient below C) to represent amino acid conservation.

Figure 4.8 ConSurf analysis of ParD1



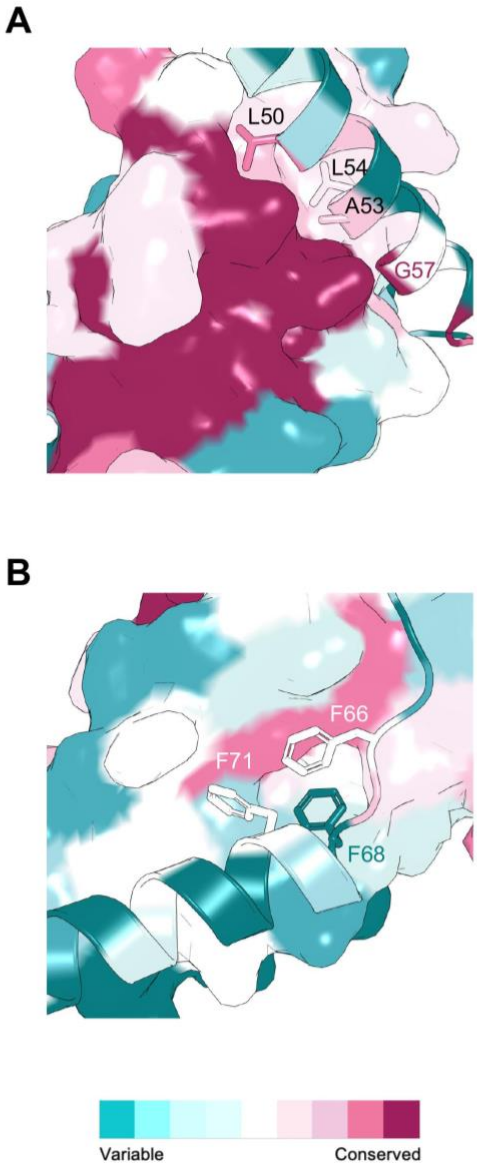
(**A**) Primary ParD1 monomer cartoon with ConSurf colouring for residue conservation; (**B**) ParD1-ParE1 (light grey) dimer highlighting potentially important ParD1 sidechains with ranging conservation scores; (**C**) Expanded view of the top black box in B showing highly conserved residues at this ParDE interface; (**D**) Expanded view of the bottom black box in B showing higher variability in ParD1 residues at this ParDE interface; (**E**) Auxiliary ParD1 monomer (dark grey) aligned to Primary ParD1 (ConSurf). ConSurf colouring scheme (shown as a gradient below D) represents amino acid conservation scores.

4.3.4 Amino acid conservation is both complementary and tuned within the ParDE1 structure

We have shown that the conservation of amino acid residues throughout both ParE1 and ParD1 is concentrated at specific areas, leaving other areas to be more variable in sequence; ParE1 appears to have a conserved β sheet core while ParD1 has a conserved N-terminal region. Most of the conserved residues in both proteins appear to be hydrophobic, and a number contribute to the interactions that form key interfaces for complex formation. If we examine 'interface ii' of **Figure 4.4 A**, that is the primary ParD1 – ParE1 interface, using the ConSurf analysis we can show complementary localisation of highly conserved residues between ParD1 and ParE1 (**Figure 4.9 A**). The highly conserved β sheet core of ParE1 complements that of the N-terminal, structured portion of ParD1 α2 with several hydrophobic residues (L50, A53, L54, G57) inserting into the conserved surface of ParE1 (reinforcing the analysis from Dalton & Crosson – (2010)). Interestingly, this is not mimicked on the underside of the ParE1 toxin, where sequence is less conserved. While there is a conserved hydrophobic patch (pink) present (created through the interaction of the β sheet core and the α helical hairpin), the surrounding residues colour green to white indicating higher variability (**Figure 4.9 B**). This is complemented by the variability of the antitoxin residues at this surface, involving those in the 'unstructured' C-terminal domain of ParD1. The three phenylalanine residues highlighted previously as conserved hydrophobic residues (Dalton and Crosson, 2010) appear to be part of this more variable, tuned interaction between the ParD1 C-terminal loop and helix, and the ParE1 α helical hairpin.

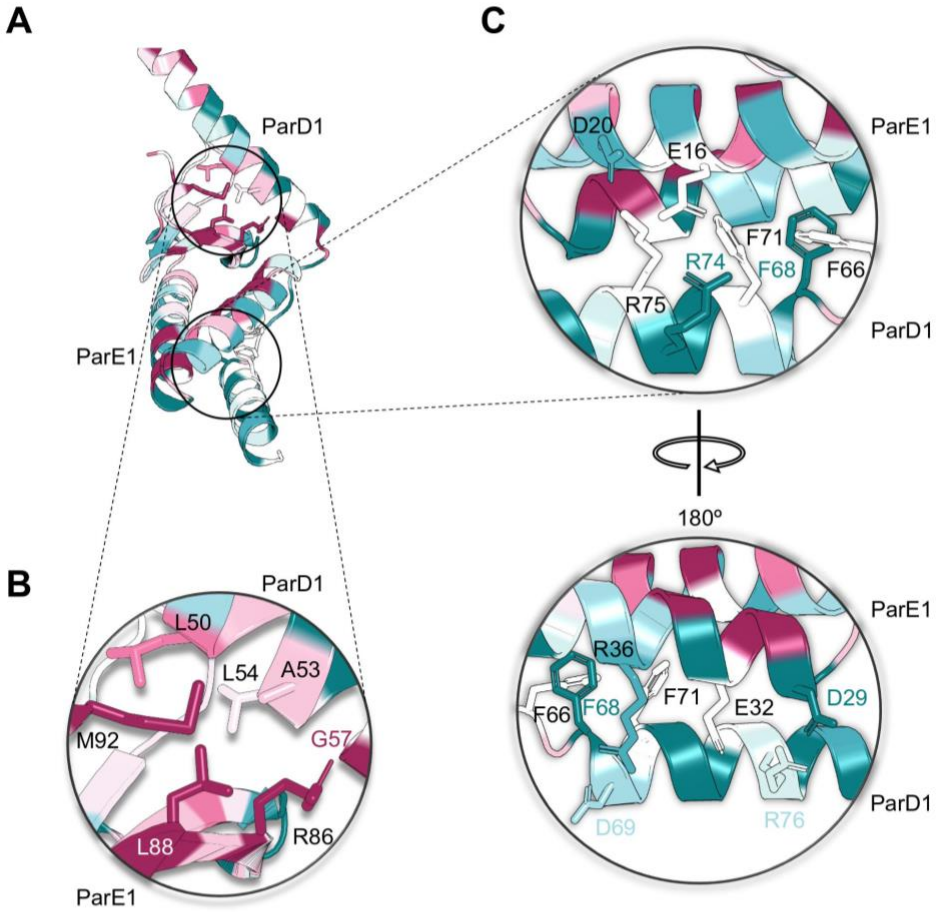
Further analysis of the interactions between these two specific surfaces indicates that the highly conserved complementary hydrophobic interaction has limited polar contacts and is largely stabilised by the hydrophobic residues (**Figure 4.10 A, B, D**). Contrastingly, the more variable complementary interaction, while containing several hydrophobic interactions, is also stabilised by a network of polar contacts (ionic bonding) (**Figure 4.10 A, C, E**). This indicates that the structured region of ParD1 has a stable sequence and complements the hydrophobic β sheet core of the ParE1 toxin, while the unstructured region of ParD1 complements the variability of the α helical hairpin region of ParE1 as they interact via an ionic network (**Figure 4.10 C and E**).

Figure 4.9 ParD1 – ParE1 amino acid surface complementation



(**A**) ParD1 (cartoon) – ParE1 (surface rendered) dimer with ConSurf colouring; (**B**) Rotated view of A to show underside of ParE1 α helical hairpin surface and C-terminal helix of ParD1, both with ConSurf colouring. Amino acid side chains are represented on the ParD1 cartoon structure with single letter codes and numbers. ConSurf colouring scheme (shown as a gradient below B) represents amino acid conservation scores.

Figure 4.10 ParD1 – ParE1 amino acid tuning for a highly specific interaction



ParD1	ParE1	Å
A53	L88	3.6
A53	M92	4.0
L50	L88	4.1
L50	M92	4.5
G57	R86	4.5
L54	L88	4.8

ParD1	ParE1	Å
D69	R36	3.2
R74	E16	3.5
K75	E16	2.7
K75	D20	2.4
R76	E32	3.9
R76	D29	3.0

(**A**) Cartoon ParDE1 dimer with ConSurf colouring, highlighting interactions between ParD1 and ParE1 chains by conservation; (**B**) Expanded view of the top black circle from A highlighting conserved amino acid interactions; (**C**) Expanded view of the lower black circle in A with additional 180 ° rotated view highlighting tuned, lower conservation, amino acid side chain interactions; (**D**) List of side chain interactions and distances from B; (**E**) List of side chain interactions and distances from C. ConSurf colouring scheme (shown as a gradient below D/E) represents amino acid conservation scores.

4.4 AlphaFold modelling of the ParDE1 complex

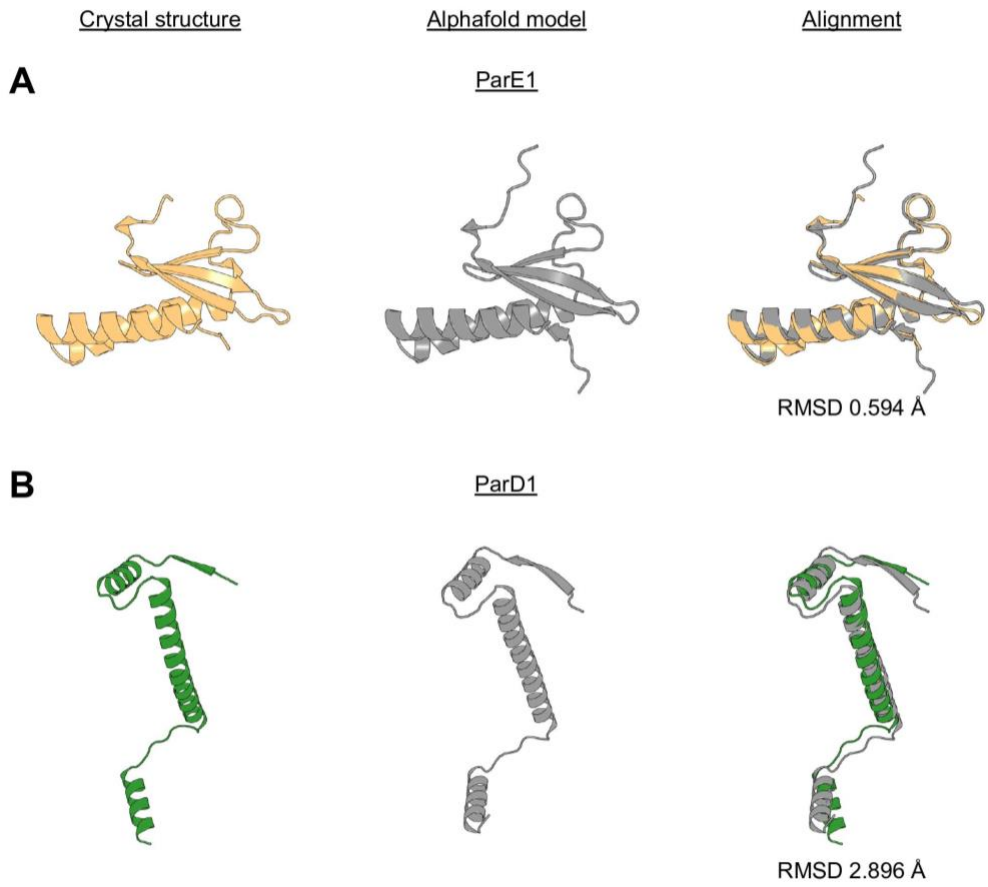
A large selection of the *Mycobacterium tuberculosis* (H37Rv) proteome was included as part of the initially released databases from AlphaFold (Jumper *et al.*, 2021; Varadi *et al.*, 2022). Amidst these selected proteins were solutions for both ParE1 and ParD1. For the ParE1 structure solution only beta 2 and beta 3 scored 'low confidence' with the E57 – R59 and T77 – G80 loops and Q90 – L98 C-terminus scoring as 'confident'. The remaining amino acids scored as 'high confidence'. For the ParD1 structure solution, only the C-terminal R83 was reported as 'low confidence' with G2, S80, and R81 scoring 'confident' and the remaining amino acids modelled with 'very high' confidence. These confidence scores provided by AlphaFold indicate that both models are predicted to be highly accurate in predicting the tertiary structure for both ParD1 and ParE1 monomers.

4.4.1 AlphaFold models of the ParE1 and ParD1 monomers

In order to validate these models, structural alignments were created for the ParE1 and ParD1 monomers (**Figure 4.11**). The ParE1 AlphaFold model aligned to the ParE1 crystal structure with an RMSD of 0.594 Å (**Figure 4.11 A**), indicating a very high degree of structural similarity. The AlphaFold model suggests the presence of an N-terminal β strand spanning Y5 – L7 and an elongated β strand spanning G80 – H89 rather than I82 – H89. Both expansions simply add to the β sheet core of the protein, but do not provide any additional information of antitoxin binding as the hydrophobic residues in this region are positioned correctly in the crystal. AlphaFold also modelled the C-terminal R96 – L98 which were unresolved in the crystal, however, this remains in a more unstructured and flexible region of the toxin and provides little additional information. The ParD1 AlphaFold model aligned to the ParD1 (Primary chain) crystal structure with an RMSD of 2.896 Å (**Figure 4.11 B**) indicating a lower degree of structural similarity which is unsurprising given the helical nature of the structure and general flexibility this allows for. Secondary structure matches extremely well with each helix and loop being a perfect match in sequence; only the N-terminal β strand is elongated in the AlphaFold model to include K3 – N4. Interestingly, AlphaFold models the C-terminal A56 – R83 as a structured region with high confidence, which somewhat contrasts our data when considering the partially unstructured auxiliary ParD1 chain.

Overall, these models are highly accurate, matching the crystal structure data with high secondary and tertiary structure similarity. Now that we have confidence in the abilities of AlphaFold to predict the monomeric structures of ParE1 and ParD1 we can investigate its abilities in predicting quaternary structure, potentially permitting a better understanding of the novel 4:2 stoichiometry seen in the crystal structure of ParDE1.

Figure 4.11 Structural comparison of AlphaFold models and crystal structures for ParDE1 monomers



(**A**) ParE1 crystal structure and AlphaFold model separately, alongside sequence-based structural alignment returning an RMSD of 0.594 Å; (**B**) ParD1 crystal structure and AlphaFold model, alongside sequence-based structural alignment returning an RMSD of 2.896 Å. Manual editing of the ParDE1 crystal structure to isolate the ParE1 and ParD1 crystal structures took place in PyMol, as did alignments to AlphaFold respective models.

4.4.2 AlphaFold models of the ParDE1 complex

Rather than using the existing database of solved structures, AlphaFold allows for the input of sequences to be run in multimer mode, solving structures in multimeric states (Bryant *et al.*, 2022; Evans *et al.*, 2022). We decided to build up from 1:1 ParDE stoichiometry to assess how stoichiometry influenced the overall quaternary structure. AlphaFold in the multimer mode allows for the fold of the input sequences to influence each other and therefore may provide slightly different outputs to the monomeric solutions (Evans *et al.*, 2022). Additionally, data collected during purification and biophysical analysis strongly indicates the existence of a 2:2 stoichiometry complex (**Chapter 3**, **Figure 3.11 – Figure 3.13**), which matches the canonical ParDE (and RelBE) system complex structure (**Table 1.4**).

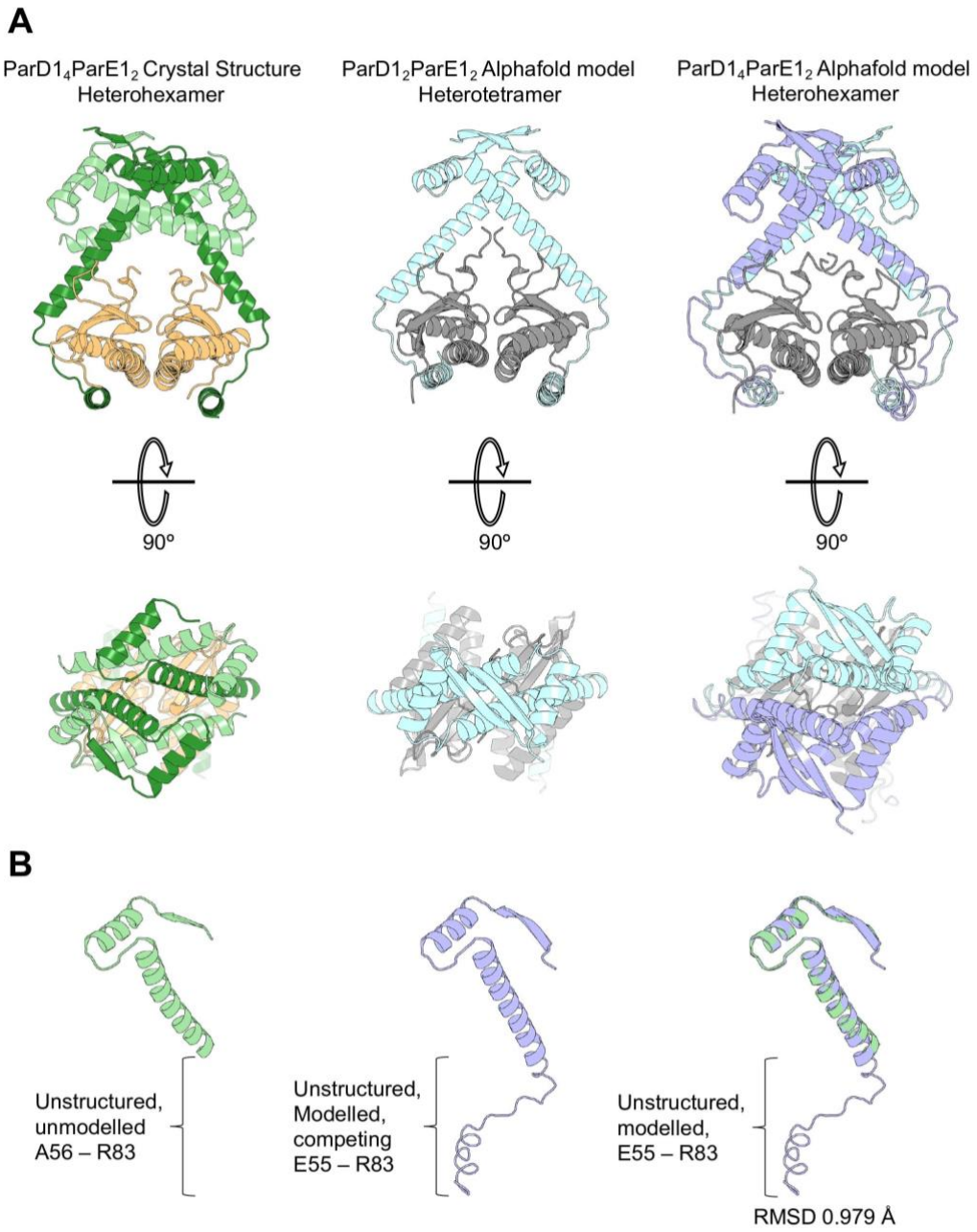
The 1:1 complex solved as expected, with the full length ParD1 antitoxin and ParE1 toxin present as seen in **4.11 A and B**, complexed as expected creating the interface shown in **4.4 A ii**. Interestingly, as per our prediction, the complex could also be solved in the heterotetrameric stoichiometry creating a highly similar quaternary structure to the search model used in MR (PDB: 3KXE) (**Figure 4.12 A**). Both full-length toxin and antitoxin are present in the structure; the ParE1 toxins interface in a highly similar manner to that seen in the crystal structure, and the ParD1 monomers dimerise via the CopG DNA-binding domain (**Figure 4.12 A and B**), with toxin binding as seen for the Primary ParD1 – ParE1 interface. This structure provided the Stokes radius value for the heterotetramer predicted to be the initial complex quaternary structure for ParDE throughout **Chapter 3.4.3**.

By increasing the number of antitoxin molecules in the input so that the stoichiometry is that seen in the crystal structure ($ParD1_4$-$ParE1_2$), initially performed to aid in predicting the second species of ParDE1 identified to evolve from the initially purified complex (**Figure 3.11**), AlphaFold successfully generates a solution highly similar to the crystal structure (**Figure 4.12 A**) with the complexed regions of the antitoxins displaced off-centre creating the tetrameric CopG ribbon-helix-helix as discussed in **4.1.1** (**Figure 4.2**). Interestingly, the AlphaFold solution presents the C-terminal ParD1 E55 – R83 region as vaguely unstructured for every ParD1 monomer, as the Primary and Auxiliary antitoxins directly compete for the ParE1 surfaces. Each of the C-terminal regions track along the ParE1 interface and form a weak helical structure resembling α3, however due to this direct competition neither is presented as fully folded. This indicates error in this solution as the crystal structure shows only the ParD1 monomers that form the interface in **Figure 4.4 A iv** (the innermost Primary ParD1 molecules that cross over at α2, forming ionic bonds) are capable of complexing with the toxin. The unstructured and competing region from the AlphaFold model is a near exact match for the unresolved region of the Auxiliary ParD1 crystal structure (**Figure 4.12 B**), with a single extra A56 amino acid present in the AlphaFold model of

each monomer. This does indicate that this region is highly flexible and likely unstructured, contrasting the confidence of the full-length solution shown in **Figure 4.11 B**.

The AlphaFold models presented here assisted in predicting the identity of the ParDE1 complex species observed in analytical sizing (**Chapter 3.4.3**). These ParDE1 complex models (**Figure 4.12**) alongside our heterohexameric ParDE1 crystal structure (**Figure 4.1**) strongly suggest that the theorised ParDE1 complex remodelling process is occurring in solution, releasing ParE1 toxin molecules as modelled in **Figure 3.15** (**Chapter 3**). Interestingly, we identified temperature as a driving force for the reorganisation of this complex, essentially converting the entire fraction of theoretical heterotetramer into heterohexamer (predicted by Stokes radius). Now we have confirmed the existence of the heterohexamer, it is more likely that our predictions are correct; ParDE1 most likely forms the heterotetramer (consistent with other ParDE and RelBE systems, **Table 1.4**) as modelled in **Figure 4.12** and is remodelled into the heterohexamer, both modelled (**Figure 4.12**) and observed in our structure (**Figure 4.1**). Confirmation of the existence of the ParDE1 heterotetramer may be a necessary next step through crystallographic studies, however, AlphaFold has proven to be highly accurate thus far. Further study will be required to understand the thermodynamics of this process and why the heterohexamer is more thermostable.

Figure 4.12 AlphaFold modelling of the ParDE1 complex



(**A**) Comparison of the ParDE1 complex crystal structure and AlphaFold models of varied stoichiometry. Rotated view shows the ParD1 complexing region. Crystal structure is coloured as previously. AlphaFold models show ParE1 in dark grey and ParD1 in light blue and lilac; (**B**) Comparison of Auxiliary ParD1 to AlphaFold ParD1 from the heterohexamer model. Alignment returns an RMSD of 0.979 Å and highlights the model accuracy at the unstructured region, which aligns well to the unmodelled region of the crystal.

4.5 ParDE2 complex crystal structure

Having solved ParDE1, efforts switched to solving ParDE2. ParD2 and ParE2 were co-expressed and the resulting protein complex was purified as per **Materials & Methods 2.4.2 and 2.5.9**. The purified ParDE2 protein complex was used to set crystal screens and optimise crystal conditions, as detailed in **Materials & Methods 2.8.1 and 2.8.2.1**. Crystals grew as planar hexagons and took 3 months to appear, even at the high concentration of 20 mg/mL. Suitable crystals were harvested, cryo-protected and X-ray crystallographic data were collected as described in **Materials & Methods 2.9**. Three datasets were collected from two crystals, the 6 datasets were merged to a final resolution of 2.35 Å.

The ParDE2 complex structure was initially determined by molecular replacement using a truncated ParE2 (T2 – G86) AlphaFold model as a search model in spacegroup R 32:H (**Materials & Methods 2.9.4**, and **Table 4.3**). Phasing including the ParD2 AlphaFold model was unsuccessful and only the toxin molecule could be positioned. On inspection, the asymmetric unit contained a single ParE2 (2 – 86) toxin molecule alongside a significant portion of unmodelled density which was subsequently built in Coot (Emsley and Cowtan, 2004) as residues I36 – G71 of ParD2, creating three α helices of varying lengths (**Figure 4.13**). Two non-hydrogen ligand atoms are built in (**Table 4.3**) as individual chloride ions and set at 1/3 occupancy due to their position at the crystallographic axis. These are not believed to contribute to the stability of function of the protein complex.

The ParDE2 crystal structure exists as a heterodimer with a largely unresolved ParD2 antitoxin (**Figure 4.13 A**). The ParE2 toxin is comprised of β1 (R4 – H8), α1 (N9 – Y22), α2 (P27 – Q47), β2 (R60 – Y63), β3 (Y69 – T75), β4 (A79 – H87), with the C-terminal M88 – E105 being unresolved (**Figure 4.13 B and D**). β1/α1 and α2 form a hairpin structure which links to the anti-parallel β sheet core of β2 – β4. The β sheet core sits on top of the hairpin with β4 and β1 interacting in parallel (**Figure 4.13 B**).

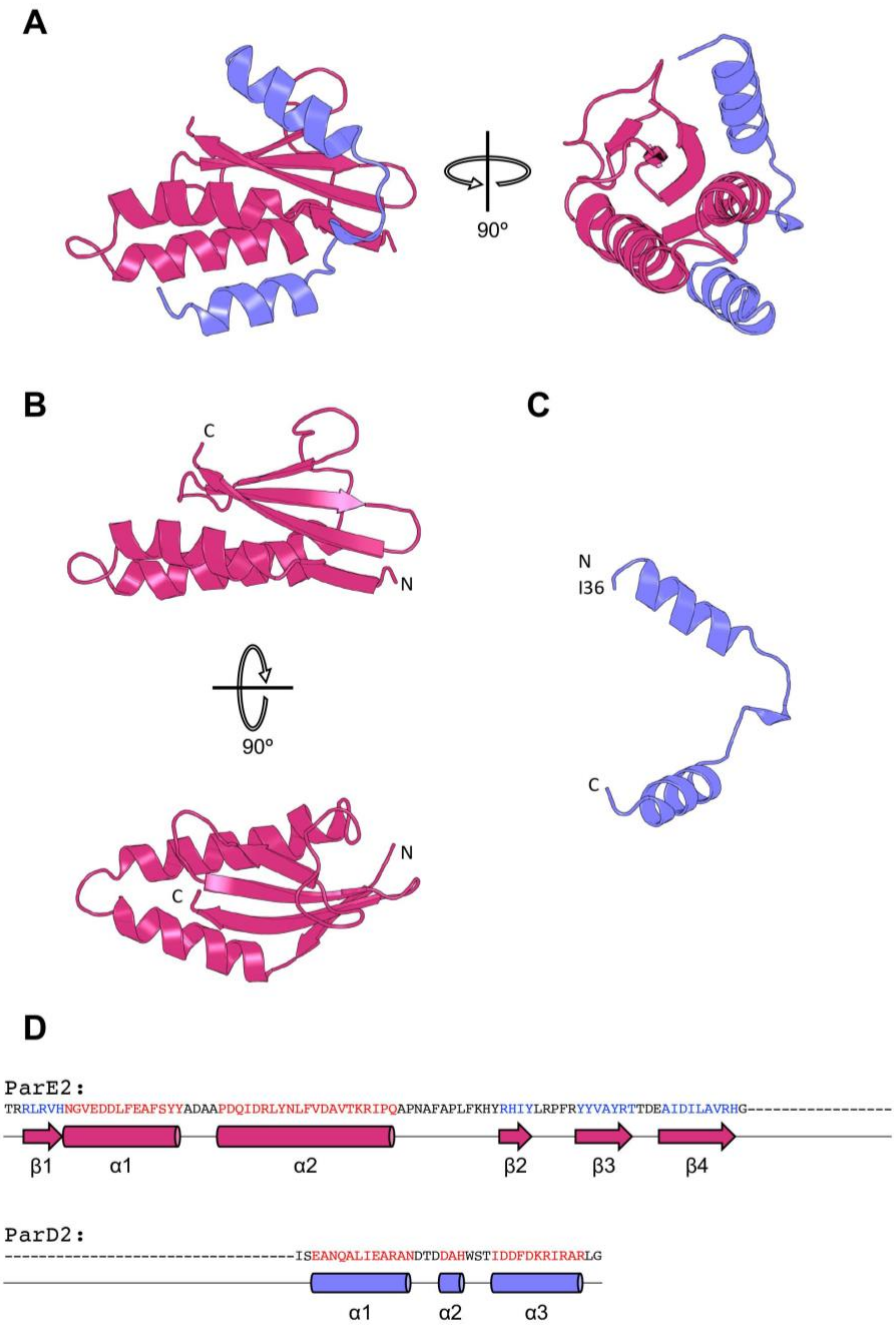
The ParD2 antitoxin is comprised of α1 (E38 – N49), α2 (D53 – H55), α3 (I59 – R69), with the N-terminal V2 – H35 being unresolved in the crystal structure (**Figure 4.13 C and D**). Given that the N-terminal region of type II antitoxins is usually structured, and the time taken to crystallise, it is unlikely that this is unresolved due to flexibility, rather, this region may have been truncated through proteolysis during the crystallisation process.

Table 4.3 Data collection and refinement statistics for ParDE2

	ParDE2
PDB ID Code	8C26
Number of crystals	2
Beamline	Diamond I04
Wavelength, Å	0.9793
Resolution range, Å	50.59 – 2.345 (2.429 – 2.345)
Space group	R 3 2 :H
Unit cell, *a b c* (Å), α β γ (°)	68.074 68.074 197.049, 90 90 120
Total reflections	160967
Unique reflections	7244 (349)
Multiplicity	19.8
Completeness (%)	100 (100)
Mean I/sigma(I)	7.5 (0.81)
R_{merge}	NA
R_{meas}	NA
$CC_{1/2}$	1.0
R_{work}	0.2539 (0.5053)
R_{free}	0.2883 (0.5020)
No. of non-hydrogen atoms	1023
Macromolecules	1021
Ligands	2
Solvent	0
Protein Residues	123
RMSD (bonds, Å)	0.008
RMSD (angles, °)	0.99
Ramachandran favoured (%)	85.71
Ramachandran allowed (%)	14.29
Ramachandran outliers (%)	0.00
Average B-factor	102.12
Macromolecules	102.09
Ligands	115.84

Values in parenthesis are for the highest resolution shell. R_{free} was calculated with 5% of the reflections selected.

Figure 4.13 Crystal and secondary structure of the ParDE2 complex



(**A**) Cartoon representation of the dimeric ParDE2 complex crystal structure with 90° rotation. ParE2 is coloured magenta, ParD2 is coloured slate blue; (**B**) Rotated views of the isolated ParE2 structure with N and C -terminal labels; (C) Isolated view of the ParD2 structure with N and C -terminal labels. The N-terminal 35 amino acids were note resolved in the structure; (**D**) Secondary structure schematics of ParE2 and ParD2. Secondary structure is displayed underneath the protein sequence, whereby amino acids constituting α helices are in red, and those constituting β strands are in blue. Secondary structure elements are coloured as per their crystal structure counterpart and represented by tubes for α helices and arrows for β strands and are labelled accordingly. Dashes represent unmodelled regions.

4.5.1 Conserved mechanism of protein recognition in the ParDE2 system

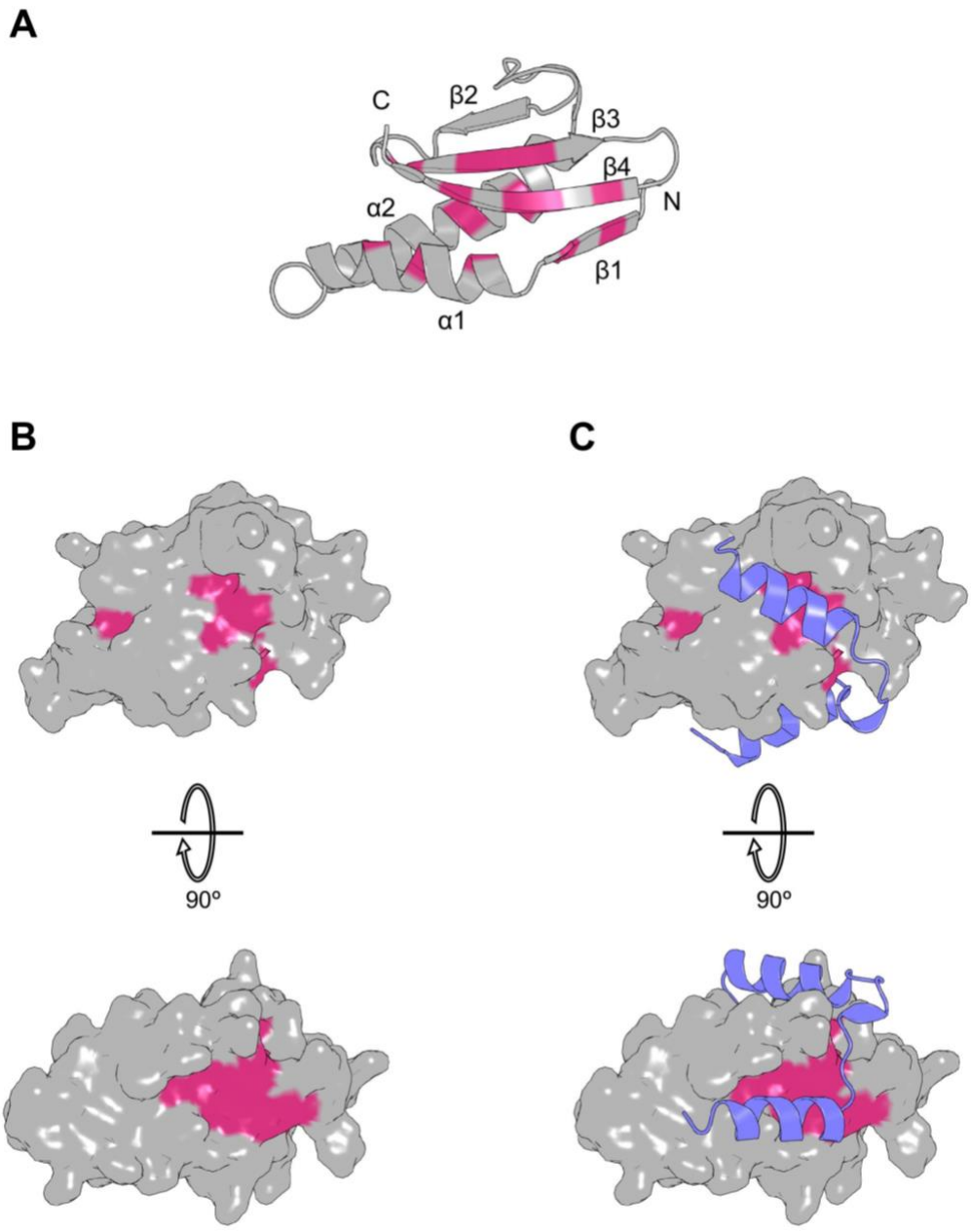
Using the same analysis completed for the ParDE1 system (**Figure 4.5**), we can demonstrate that the conserved mechanism of protein recognition is maintained within the ParDE2 system (**Figure 4.14**). The *M. tuberculosis* ParE2 sequence appeared in the RelE/ParE superfamily alignment previously discussed (Anantharaman and Aravind, 2003). This allows us to plot and visualise the conserved hydrophobic residues within the ParE2 toxin structure which are clearly concentrated on the internally facing residues of the α helical hairpin, and the throughout the β sheet core of the toxin (**Figure 4.14 A**). This positioning of residues creates two major hydrophobic grooves on the surface of the ParE2 toxin which run along the β sheet core and through to the underside of the toxin between the hairpin helices (**Figure 4.14 B**). This is highly similar to the ParDE1 structure and mimics the conserved mechanism of protein recognition identified in the *C. crescentus* ParDE structure (Dalton and Crosson, 2010). Representing the ParD2 antitoxin alongside the surface rendered ParE2 toxin clearly demonstrates the specific interaction at the conserved hydrophobic patches, as they align closely to the truncated region of ParD2 (**Figure 4.14 C**).

4.5.2 Interfaces and assembly of the ParDE2 complex

PISA analysis (Krissinel and Henrick, 2007) of the ParDE2 complex, unsurprisingly, returns only a single interface; that formed between the truncated ParD2 and ParE2 (matching the conserved hydrophobic patches) (**Figure 4.15 A**). Calculations from PISA indicate that the area of this interface is 1235 $Å^2$ and its formation leads to an energy change of -7.6 kcal/mol; together, this is shown to be an essential interface for complex formation by CSS (scoring the maximum of 1) (**Figure 4.15 B**). Electron density at this essential interface is demonstrated in **Figure S2 B**, showing the 2Fo-Fc map for the relevant residues of the ParD2 antitoxin.

Further investigation of the essential contacts made between the chains highlights several polar and ionic interactions that stabilise the largely hydrophobic interfacing. Notable contacts are the ionic bonds formed between ParD2 E45 (on α1) and ParE2 K57 (found on the loop region between α2 and β2) (**Figure 4.15 C**) and between ParD2 R47 (on α1) and ParE2 (D14) (on α1, also) (**Figure 4.15 D and E**). These are highly specific side chain interactions that demonstrate the mechanism of ParDE2 interaction extends beyond a conserved hydrophobic groove.

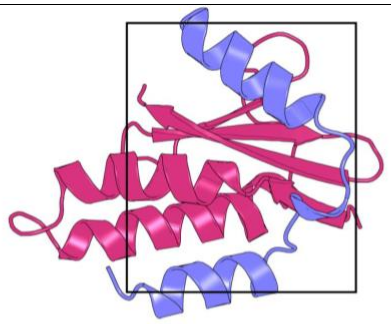
Figure 4.14 Conserved mechanism of protein recognition in the ParDE2 system



(**A**) Cartoon representation of the ParE2 toxin, coloured grey, with conserved residues from the RelE/ParE superfamily highlighted in magenta; (**B**) Surface rendered ParE2 toxin with conserved residues highlighted in magenta from A; (**C**) Surface rendered ParE2 with conserved residues in magenta complexed with ParD2 (coloured slate blue).

Figure 4.15 Interface and assembly analysis of the ParDE2 complex

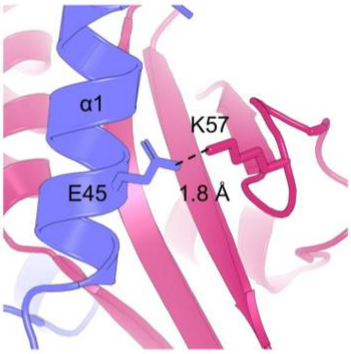
A



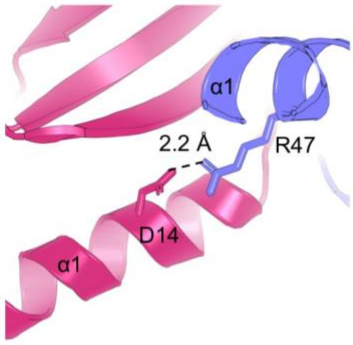
B

Interface	Interface area (Å²)	Δ^iG kcal/mol	Δ^iG P-value	N^{HB}	N^{SB}	N^{DS}	CSS
i	1235.0	-7.6	0.360	12	11	0	1.000

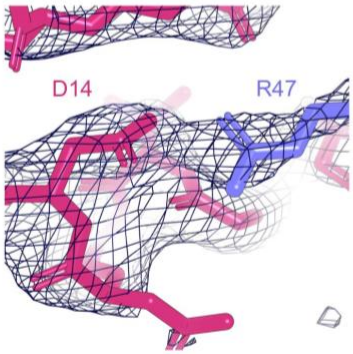
C



D



E



(**A**) Interfaces of the ParDE2 complex analysed by PISA. i: Single interface between the ParE2 toxin and truncated ParD2 antitoxin (**B**) PISA analysis of interface i. Interface area is calculated in square angstroms (Å²); Δ^iG – solvation energy on folding (free energy gain); N^{HB} – Number of hydrogen bonds within the respective interface; N^{SB} – Number of salt bridges within the respective interface; N^{DS} – Number of disulphide bonds formed within the respective interface; CSS – Complex Significance Score ranges between 0 – 1 as interface relevance to complex formation increases; (**C** and **D**) Ionic bonding stabilising the complex; (**E**) ParE1 D14 – ParD2 R47 interaction with F_o-F_c electron density map contouring at 2σ.

4.6 AlphaFold modelling of the ParDE2 complex

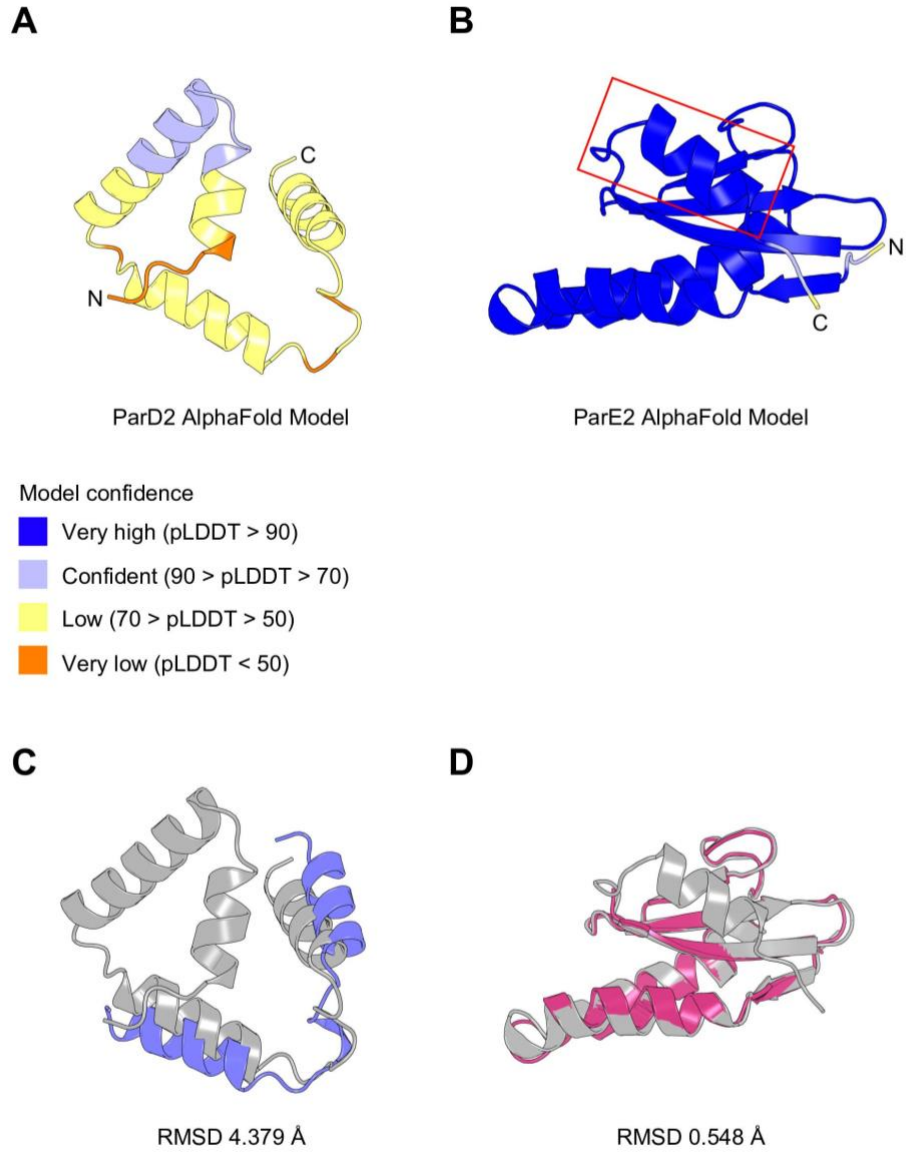
4.6.1 AlphaFold models of the ParD2 and ParE2 proteins

Both ParD2 and ParE2 can be found in the database of 'solved' protein structures released by AlphaFold (Jumper *et al.*, 2021; Varadi *et al.*, 2022). These models have been generated to contrasting levels of confidence (**Figure 4.16 A and B**), scored by pLDDT, demonstrating the potential stability of the respective monomers.

The ParD2 monomer AlphaFold model is extremely weak in confidence scores, generally being ranked as 'low' in the per residue confidence scoring system (pLDDT) (**Figure 4.16 A**). This low scoring may indicate intrinsic disorder across the majority of the ParD2 antitoxin (Jumper *et al.*, 2021), which would contrast what is generally seen with only the toxin interfacing region being disordered. The ParD2 AlphaFold model is comprised of α1 (R5 – A13), α2 (R16 – E32), α3 (I36 – N49), and α4 (I59 – L70), with only D12 – E24 being scored as 'confident' in their modelling (**Figure 4.16 A**). Interestingly, however, the secondary structure is similar to that seen for the resolved region of the ParD2 crystal structure from the ParDE2 complex, with the AlphaFold ParD2 model α3 and α4 being close matches with the ParD2 crystal structure α1 and α2 (**Figure 4.13 D**).

In contrast, the ParE2 AlphaFold model scores very highly in confidence per residue (pLDDT) across the entire structure (**Figure 4.16 B**). Even prior to alignments, this structure has largely been validated due to its success as the search model in MR to phase the ParDE2 crystallography dataset. The interesting addition we see in the ParE2 AlphaFold model is that the C-terminal P92 – G101 form α3, indicated by the red box in **Figure 4.16 B**, and occupies the interface for ParD2 binding that is bound by ParD2 α1 in the crystal structure. Indeed, phasing of the ParDE2 dataset is successful when the ParE2 C-terminal helix is included, however the linker region between β4 and α4 is not resolved. Contrastingly, when the C-terminal helix of ParE2 is removed from the search model, empty density is present in its place, and can successfully be built as residues I36 – N49 of ParD2, followed by the resolved linker region through to the N-terminus (**Figure 4.17**). This gave us confidence that the predicted ParE2 α4 is displaced in the ParDE2 complex, likely becoming disordered as is seen for other ParDE system complexes, and that the density is occupied by ParD2 across the conserved hydrophobic groove at the β sheet core.

Figure 4.16 AlphaFold models and comparisons of the ParD2 and ParE2 proteins



(**A**) AlphaFold model of ParD2, N and C -terminal labels included; (B) AlphaFold model of ParE2 with N and C -terminal labels included. Models are coloured by pLDDT confidence score. Score colouring is shown below A and is as found in the AlphaFold database (Varadi *et al.*, 2022); (**C**) Sequence-based alignment of AlphaFold ParD2 (now grey) to the crystal structure of ParD2 (slate blue) returned an RMSD of 4.379 Å; (D) Sequence-based alignment of AlphaFold ParE2 (now grey) to the crystal structure of ParE2 (magenta) returned an RMSD of 0.594 Å.

Alignments of the crystal structures of ParD2 and ParE2 with their respective AlphaFold models returns contrasting RMSD values, indicating different levels of model accuracy to complement the AlphaFold confidence scores. While the ParD2 model does somewhat match the resolved portion of the ParD2 crystal structure in secondary structure, the positioning of the respective helices is only vaguely correct, and the linker region does not present the joining 3 amino acid helix (**Figure 4.16 C**). This results in a weak RMSD value of 4.379 Å, meaning poor alignment between the two structures highlighting the inaccuracy of this AlphaFold model. The alignment between the ParE2 crystal structure and model, however, demonstrates high similarity. The RMSD value between these two structures is 0.548 Å indicating a high level of secondary and tertiary structure similarity (**Figure 4.16 D**), with only the AlphaFold model C-terminal α4 not being aligned. This reflects the confidence scores reported by AlphaFold and largely validates the accuracy of the model.

4.6.2 Validation of the ParDE2 crystal structure through electron density maps

To demonstrate that the ParE2 C-terminal helix, present in the full-length ParE2 AlphaFold solution, is not present in the crystal structure we compared the fit of the structures to the refined electron density (**Figure 4.17**). In our model, ParE2 residues T2 – G88 are built alongside ParD2 residues I36 – G71 (Figure 4.17 A); the solution required the removal of the C-terminal alpha helix from the AlphaFold solution. We have confidence in this model due to the lack of density beyond ParE2 G88 for a linker region to the ParE2 C-terminus (**Figure 4.17 B**). The density across the ParE2 beta sheet core is modelled as ParD2 I36 – N49 as previously stated, followed by strong density for the linker region between this alpha helix and the ParD2 C-terminal helix (**Figure 4.17 C**). When the AlphaFold ParE2 model (T2 – G88 coloured grey, M89 – E105 coloured purple) is aligned to the crystal structure ParE2, the linker region of M89 – N93 is clearly not supported by electron density (**Figure 4.17 D**) and the C-terminal helix, while unrefined, is not held as well as the ParD2 residues in this region, which remains unmodelled beyond the C-terminal ParE2 E105 residue (**Figure 4.17 E**). Altogether, these analyses support that the ParD2 residues are built correctly, and that the ParE2 C-terminal helix is displaced in the crystal structure. This allows us to develop models for ParDE2 using AlphaFold with confidence that ParD2 occupies the conserved protein recognition interface, as expected for RelE/ParE superfamily members (Anantharaman and Aravind, 2003; Dalton and Crosson, 2010).

Figure 4.17 Electron density of the ParDE2 crystal structure

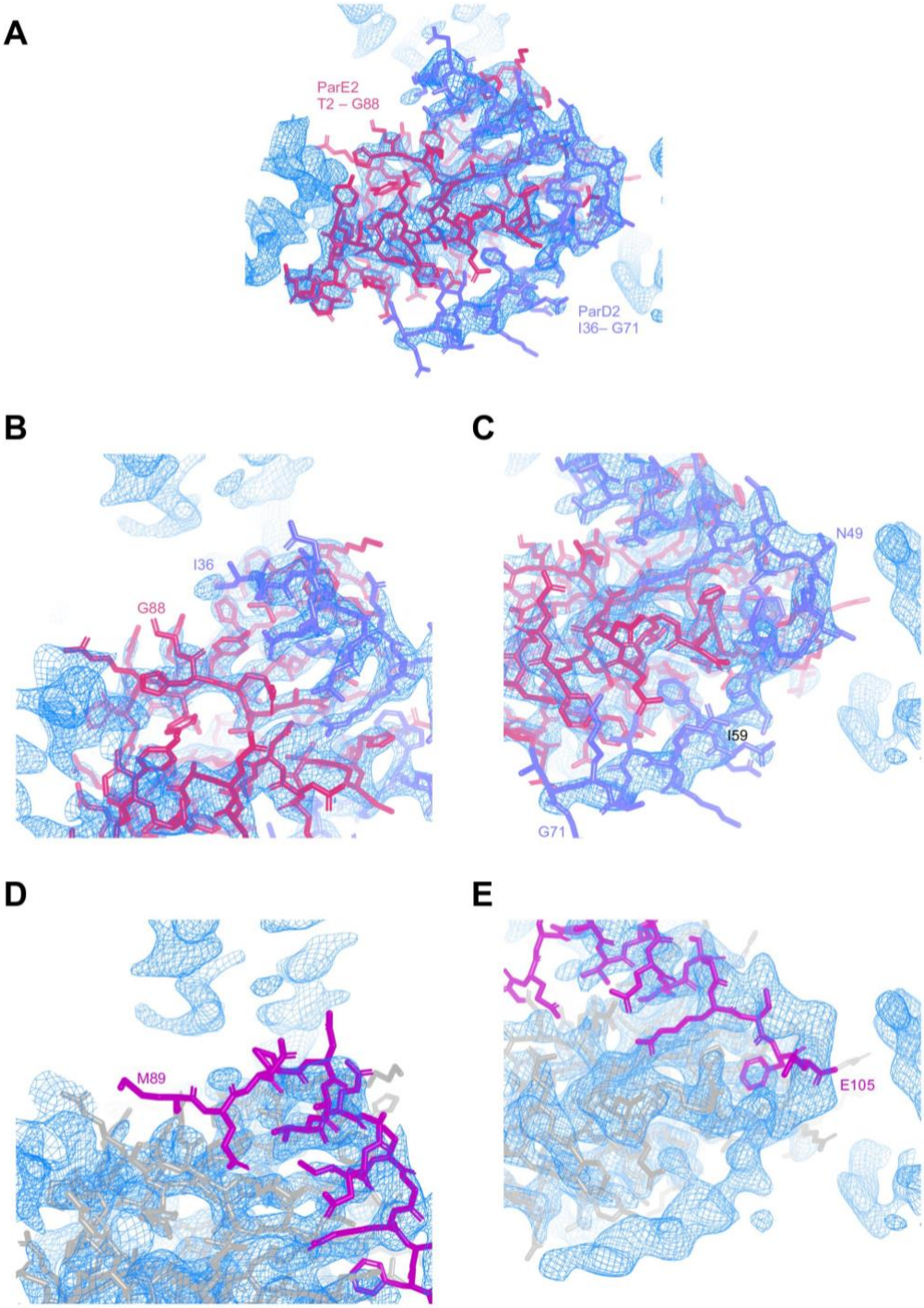


Figure legend overleaf

Figure 4.17 Electron density of the ParDE2 crystal structure

(**A**) ParDE2 crystal structure with F_o-F_c electron density map contoured at 2σ. Both ParE2 (magenta, T2 – G88) and ParD2 (slate blue, I36 – G71) are represented as stick models; (**B**) Enhanced view of the ParE2 C-terminus (G88) and ParD2 N-terminus (I36), from A, held by electron density; (**C**) Enhanced view of ParD2 N49 – G71 held by electron density. The linker region between the two resolved ParD2 helices is N49 – I59; (**D**) Enhanced view of AlphaFold ParE2 (T2 – E105) modelled into the electron density map from A. ParE2 residues T2 – G88 are coloured grey, with M89 – E105 (C-terminal linker and alpha helix) are coloured purple. No density was present at the linker region, from M89 – P92; (**E**) Reoriented view of D to show the AlphaFold ParE2 C-terminal residues (up to E105). These residues do not explain the density beyond E105, which remains unmodelled in this view.

4.6.3 AlphaFold multimer modelling of the ParDE2 complex

To further investigate the ParDE2 complex structure and validate AlphaFold predictions, the multimer mode (Evans *et al.*, 2022) was employed to generate a series of ParDE2 complexes. It was expected that, given the crystal structure data indicating the presence of the ParD2 α1 helix across the β sheet hydrophobic region, AlphaFold would generate a multimer model demonstrating preference for this interaction over the ParE2 α4 helix. This was not the case when the full ParE2 sequence was entered alongside the full ParD2 sequence; the full ParE2 model was generated as seen in **Figure 4.16 B**, with a new fold for ParD2 (**Figure 4.18 A and B**).

When a 1:1 stoichiometry was entered, AlphaFold generated a complex with a similar ParD2 α3/α4-ParE2 interaction (at the hydrophobic site under the ParE2 hairpin structure). The remainder of the antitoxin formed three further helices which are apparently displaced by the ParE2 α4 (**Figure 4.18 A**). A similar model is seen when a 2:2 stoichiometry is analysed, with the heterodimers interacting through the bundled helices of ParD2 at loop regions (Model 1: **Figure 4.18 B**). PISA analysis cannot be completed using an AlphaFold model due to the lack of crystallographic data, however, interactions between loop regions are unlikely to create large interface areas and are therefore likely to be biologically irrelevant. The highest scoring AlphaFold solution places the C-terminal helix of ParE2 at the β sheet hydrophobic region, rather than this region being occupied by any ParD2 residues. We do not believe this to be correct based on the crystal structure (**Figure 4.13**); rather, the C-terminal ParE2 helix is displaced by ParD2, similar to other RelE/ParE superfamily systems (Bøggild et al., 2012).

Given the crystal structure, and our confidence in the model due to the presence of electron density in the ParD2 alpha helix linker region (**Figure 4.17 C**), we investigated whether truncating the ParE2 sequence prior to input would present alternative solutions. Indeed, when the C-terminal M88 – E105 residues (ParE2$^{\Delta 88\text{-}105}$) (unresolved in the crystal) are removed from the sequence, two alternative complex solutions with a more biologically likely quaternary structure can be generated (Model 2: **Figure 4.18 C,** and Model 3**: Figure 4.18 E**). These solutions now show the antitoxin occupying both hydrophobic patches across the β sheet core and underside of the α helical hairpin, extending upward to form two further N-terminal helices. Given that this is the conserved mechanism of interaction within the entire RelE/ParE superfamily, and our crystal structure, we expect the placement of the antitoxin chain here to be correct (Anantharaman and Aravind, 2003; Dalton and Crosson, 2010). Additionally, the ParD2 monomers now interact via their N-terminal helices, which is typically seen for members of the RelE/ParE TA system family (Overgaard, Borch and Gerdes, 2009; Dalton and Crosson, 2010; Bøggild et al., 2012; Muthuramalingam et al., 2018; Kamruzzaman and Iredell, 2019).

Figure 4.18 AlphaFold multimer modelling of the ParDE2 complex

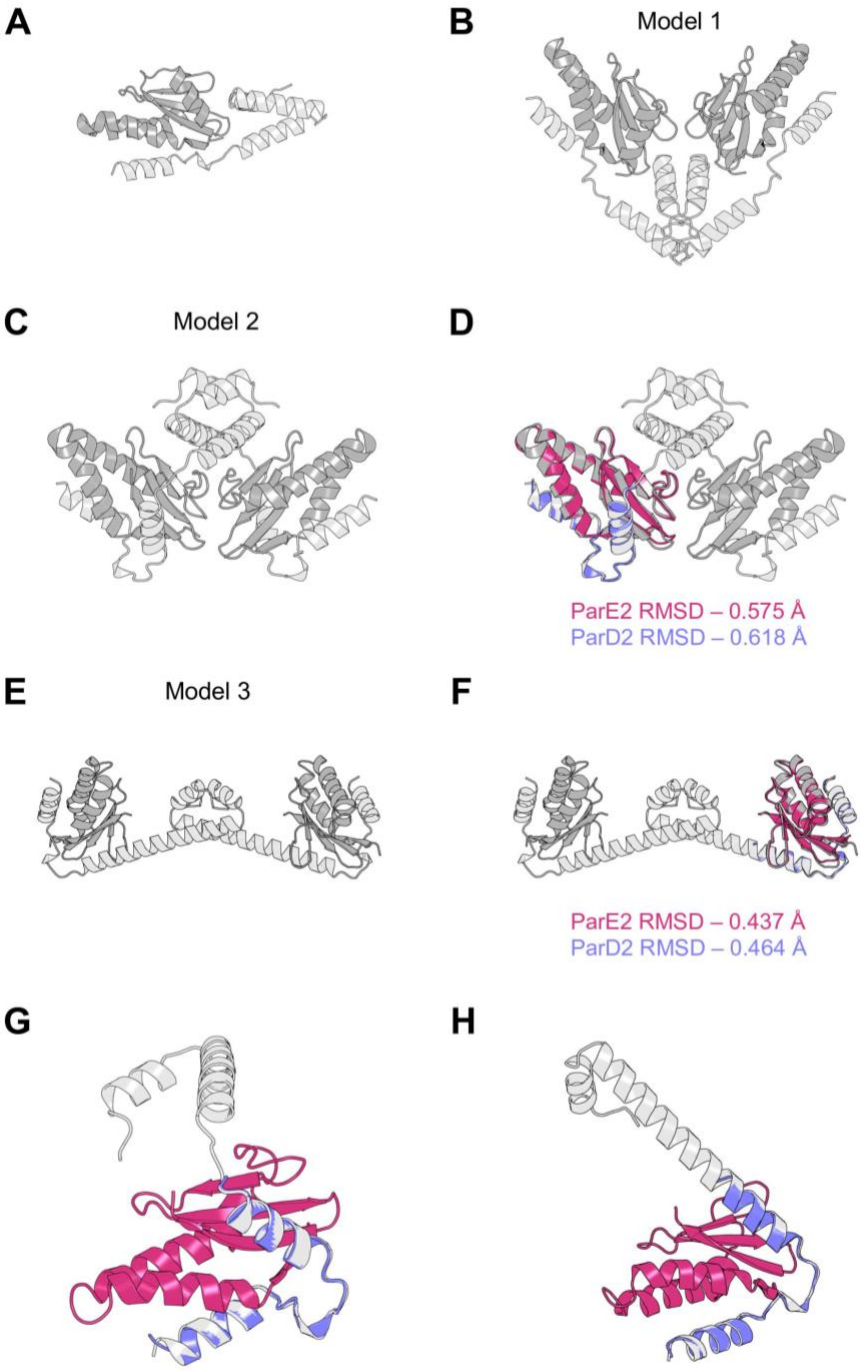


Figure legend overleaf

Figure 4.18 AlphaFold multimer modelling of the ParDE2 complex

(**A**) AlphaFold multimer model of $ParD2_1ParE2_1$; (**B**) AlphaFold multimer 'Model 1' of $ParD2_2ParE2_2$; (**C**) AlphaFold multimer 'Model 2' of $ParD2_2ParE2^{\Delta 88\text{-}105}{}_2$; (**D**) ParDE2 crystal structure aligned to the AlphaFold model in C. RMSD values for the individual ParE2/ParD2 alignments are presented below (Å); (**E**) AlphaFold multimer 'Model 3' of $ParD2_2ParE2^{\Delta 88\text{-}105}{}_2$; (**F**) ParDE2 crystal structure aligned to the AlphaFold model in E. RMSD values for the individual ParE2/ParD2 alignments are presented below (Å); (**G**) Enhanced view of ParD2 from C aligned to the ParDE2 crystal structure; (**H**) Enhanced view of ParD2 from E aligned to the ParDE2 crystal structure. Crystal structure colouring is as previously presented. AlphaFold models are coloured as ParD2 (light grey) and ParE2 (dark grey).

This N-terminal interaction, which is now consistent between the models (**Figure 4.18 C and E**), appears to form a helix-turn-helix like motif and Dali searches using the isolated ParD2 N-terminal dimer identified the TP901-1 Clear repressor protein from a Lactococcus phage to possess a highly similar dimerization and DNA-binding domain (PDB: 6XFA, RMSD 1.7 Å, (Rasmussen *et al.*, 2018)).

Importantly, when the crystal structures of the ParD2 and ParE2 monomers are aligned to the AlphaFold models of **Figure 4.18 C and E**, low RMSD values are returned indicating high structural similarity (**Figure 4.18 D and F**, respectively). A higher degree of structural similarity is seen for alignments to Model 3 (**Figure 4.18 F**), with the ParE2 toxin aligning to 0.437 Å, and the ParD2 antitoxin aligning to 0.464 Å (RMSD), compared to Model 2 (Figure 4.18 D) with the ParE2 toxin aligning to 0.575 Å, and the ParD2 antitoxin aligning to 0.618 Å (RMSD). Isolating the alignments as ParDE dimers of Model 2 and Model 3 demonstrates how the ParD2 antitoxin unresolved region could extend beyond the crystal structure (**Figure 4.18 G and H**, respectively). The N-terminal region of ParD2 from Model 3 is more similar to the ParD1 antitoxin structure (**Figure 4.3 E**) with a long central alpha helix. In order to determine the most likely model for ParDE2 (between Models 2 and 3), beyond the RMSD comparisons, we used the analytical SEC data and Stokes radius calculations (**Chapter 3**, **Figure 3.16 D**). The Stokes radius for Model 2 (**Figure 4.18 C**) was calculated to be 26.10 Å, and the Stokes radius for Model 3 (**Figure 4.18 E**) was calculated to be 30.50 Å. As previously presented in **Figure 3.17 D**, Model 3 provided an extremely close value to the observed from analytical SEC (30.27 Å) indicating that this is our best model for the ParDE2 heterotetrameric complex thus far. This also indicates the necessity for AlphaFold validation through a variety of techniques.

The crystal structure data played a significant role in this modelling process as it allowed us to understand that the β sheet hydrophobic patch of ParE2 somehow has preference for the ParD2 antitoxin region, rather than the ParE2 α4, and the ParDE interaction likely prevents full folding of ParE2.

4.7 Discussion

In solving both the ParDE1 complex (**Figure 4.1**) and ParDE2 complex (**Figure 4.13**) structures we can highlight several similarities and differences between the structures of the toxin and antitoxin constituents, alongside the complex structures.

Comparisons of the ParE toxins as isolated monomers from their respective crystal structures indicates that secondary structure is largely conserved; the superfamily structure βααββββα (Anantharaman and Aravind, 2003) is largely present in both toxins, however, with variation in lengths of secondary structure units and sequence variation throughout (further explored in **Chapter 5**). Sequence-based alignment of ParE1 and ParE2 returned an RMSD of 6.175 Å indicating low level sequence similarity, however, sequence-independent superposition returns a greatly improved RMSD value of 2.163 Å with strong alignment of the alpha helix hairpin and central, beta sheet core. This indicates that structure is more greatly conserved than sequence in the ParE toxins.

As we have previously shown, the ParE toxins appear to exhibit differing levels of affinity for *M. tuberculosis* gyrase in cleavage assays (**Figure 3.5 and Figure 3.14**). It was apparent that the ParE2 toxin had a lower affinity for DNA gyrase however was still capable of stabilising the cleavage complex and generating high levels of linear species DNA in our assays; a notable difference between the toxin structures is evident at the C-termini (**Figure 4.1 And Figure 4.13**). While both C-termini for the respective toxins were partly unstructured, the ParE1 toxin had a consistent number of residues resolved, indicating that while unstructured, it may be stabilised with the single turn of Q90 – M92. In contrast, the 17 C-terminal residues of ParE2 could not be resolved; P92 – G101 of which is predicted to form an alpha helix by AlphaFold (**Figure 4.16 B**) and occupy the superfamily hydrophobic surface across the beta sheet core. This surface is occupied in the crystal by ParD2 (and corresponds to the site occupied by ParD1 in the ParDE1 structure) (**Figure 4.14 C**). The importance for toxicity of these residues in ParE2 has previously been demonstrated as removing E95 – E105 (C-terminal 10 amino acids), or making mutants E98A or R102A, renders the toxin ineffective (Gupta *et al.*, 2016). When considering the RelE/ParE superfamily (Anantharaman and Aravind, 2003), re-organisation of this helix to be positioned across the beta sheet core is of significant importance in RelE toxins for the positioning of essential catalytic residues (Bøggild et al., 2012) and thus, ribonuclease activity when bound to the ribosome. ParE2 does not possess the canonical RelE catalytic core residues (presented and discussed later in **Chapter 5**), therefore, the significance of the C-terminal helix requires further investigation. It is possible that the refolding of these amino acids is required to create a more compact structure, or position externally facing residues for target interaction.

The full ParDE2 complex is yet to be fully elucidated as the N-terminus of the antitoxin has been cleaved (**Figure 4.13 C and D**), likely during the crystallisation process which took 3 months, and therefore has lost the ParD2 N-terminal dimerisation domain. AlphaFold consistently generated solutions whereby the conserved surface across the beta sheet core of ParE2 was occupied by the ParE2 C-terminal helix residues (**Figure 4.17 A and B**); this would contrast what is seen for ParE, and indeed RelE, structures in the PDB (**Table 1.4**) ((PDB: 3KXE (Dalton and Crosson, 2010); 5CEG (Aakre *et al.*, 2013); 60XA (Lite *et al.*, 2020); 7R5A (Garcia-Rodriguez et al., 2021); 6XRW (Snead, Moore and Bourne, 2022)). Our crystal data aligns well with the canonical mechanism of ParDE/RelBE TA interaction (Dalton and Crosson, 2010); a ParD2 antitoxin helix occupied the ParE2 surface and is confirmed by density in the linker region, N49 – I59, and through to the C-terminal G71 residue (**Figure 4.17 C**). Complementing this was the lack of density for the predicted ParE2 linker region to the C-terminal residues (**Figure 4.17 D**). While AlphaFold is highly confident of the ParE2 monomer solution, it has struggled with the structure of the ParD2 antitoxin as a monomer (**Figure 4.16 A**) and in complex with ParE2 (**Figure 4.17**). This is likely due to the lack of similar sequences during MSA in structure prediction (Jumper *et al.*, 2021). Our current best model for the ParDE2 system is the $ParD2_2ParE2_2$ heterotetramer presented in **Figure 4.18 E**. This model is supported by good alignment to the crystal structure (**Figure 4.18 F and E),** and analytical SEC data in **Chapter 3** (**Figure 3.16**) as it possesses a highly similar Stokes radius (ratio of 1.05). Interestingly, the heterotetramer models in **Figure 4.18 C and E** possess a ParD2 N-terminal dimerisation domain with structural similarity to the dimerisation domain of the *Lactococcus* phage TP901-1 Clear 1 repressor (PDB: 6FXA (Rasmussen *et al.*, 2018)), indicating DNA-binding capabilities of this N-terminal motif. Models for ParDE2 need to be developed further as this system is clearly peculiar in its structural biology, especially in the ParD2 antitoxin and how it may compete with the ParE2 C-terminal helix. Further crystal optimisation to more rapidly generate crystals could also be explored in the hope that a more favourable condition will decrease crystallisation time, prevent ParD2 cleavage, and produce higher resolution data to provide a full ParDE complex structure to validate our current best model (**Figure 4.18 E**). Further to this, an open reading frame is identifiable upstream of the ParDE2 operon, the translated product of which shares 27 % sequence with ParD1. It is plausible that this is the third member of a tripartite style ParDE system, similar to that seen for ParD2-PaaA2-ParE2 (Hallez *et al.*, 2010; Sterckx *et al.*, 2016). No model structure is available for the product of this ORF, if one is transcribed and translated. It could be interesting to explore the potential role of this putative protein in the biology of the ParDE2 TA system.

Similar to the ParDE2 system, the structure of ParDE1 was solved in an unexpected stoichiometry as a $ParD1_4ParE1_2$ heterohexamer. As previously stated, the toxin molecules have a high degree of structural similarity, however, the ParD antitoxins are where we see the greatest structural

differences. ParD1 forms a dimeric CopG ribbon-helix-helix motif through its N-terminal region (**Figure 1.7, Figure 4.1 and Figure 4.2 B and C**), followed by the toxin interaction helices that are also seen in ParDE2. When comparing the crystal structure ParE1 and ParD1 chains to AlphaFold predictions we see a high degree of structural similarity, largely validating the models (**Figure 4.11**). It is worth noting that AlphaFold generates a full-length structure for the ParD1 antitoxin, indicating structured helices throughout; our crystallographic data contrasts this and indicates that the C-terminal 24 residues are unstructured in solution, emphasised by the presence of two distinct ParD antitoxin chains. ParD antitoxins have also previously been shown to have unstructured C-terminal domains in solution (PDB: 2AN7 (Oberer *et al.*, 2007)) that, like many other type II TA system antitoxins, become structure on toxin binding (Madl *et al.*, 2006; Garcia-Pino *et al.*, 2016). It is peculiar, then, that AlphaFold generated a structured solution for the antitoxin. However, when models for the ParDE complex in the 4:2 stoichiometry were generated using the Colaboratory (Mirdita *et al.*, 2022), the appropriate residues appear to be modelled unstructured, as two ParD1 antitoxin C-termini compete for the same ParE1 interface (**Figure 4.12 A**).

Analyses of the ParDE1 complex highlight that the ParD1 – ParE1 interaction is highly specific, not only interacting via the conserved superfamily interface as expected (**Figure 4.5**), but also via several tuned ParD – ParE interactions in regions of lower sequence conservation (**Figure 4.10**). The heterohexameric complex structure is apparently stabilised by several ionic bonds between the full-length ParD1 chains that abut each other in the complex (**Table 4.2**). Investigating the structure of the CopG domain, of which there is one fully resolved in the structure, between what we have termed the primary and auxiliary antitoxin chains, indicates a DNA-binding role as seen in the FitAB system (Mattison *et al.*, 2006). Interestingly, this domain appears to permit interactions with elongated operators within promoter regions; the FitAB system employs two CopG domain that interact with operator sites ~ 15 bp apart (PDB: 2BSQ (Mattison *et al.*, 2006)). In the *V. cholera* ParDE system, three CopG domains exist back-to-back and create three DNA-binding sites for proposed enhanced DNA-binding through operator site interactions (PDB: 7R5A (Garcia-Rodriguez et al., 2021)). In contrast, our analyses suggest that ParDE1 has no DNA-binding capabilities due to the loss of essentially positioned arginine residues in the RHH/CopG domain (**Figure 4.2 B and C**), alongside the lack of identified operator sites. *M. tuberculosis* promoter regions are, however, known to be complex (Cortes *et al.*, 2013) and operator sites within the ParDE1 promoter may exist further upstream or as obscure sequences.

The ParDE1 heterohexamer structure, in part, supports our proposed model of ParDE1 complex remodelling for toxin release, based on analytical SEC experiments coupled with biochemistry (**Chapter 3**, **Figure 3.14**). AlphaFold has been used to successfully generate a heterotetramer

model, which we predict to be the initial complex state for ParDE1 (**Figure 4.12 A**). Interface analysis of the complex seen in the crystal indicates that four interfaces are relevant in the formation of the heterohexamer (**Figure 4.4**). Comparison to the search model (PDB: 3KXE (Dalton and Crosson, 2010)) and highly similar ParDE1 heterotetramer AlphaFold prediction, alongside the PISA analysis (**Figure 4.4**) and subsequent investigation of polar contacts (**Table 4.2**), indicated that the heterohexamer structure is in fact formed of $ParD1_2ParE1$ trimeric structures that interact mainly through their ParD1 CopG domains. This is supported by the structures of *V. cholera* ParDE (PDB: 7R5A (Garcia-Rodriguez et al., 2021)) and *V. cholera* ParD (PDB: 7B22 (Garcia-Rodriguez et al., 2021)) whereby the highly similar CopG domain multimerises in the same orientation seen for ParDE1. This observation has allowed us to develop our model for ParDE1 complex remodelling (**Figure 4.19**) (each chain is coloured independently for clarity), building on from what is proposed in Chapter 3.

We propose that in complex remodelling, the ParD1 CopG domain remains intact as a highly stable and conserved dimerisation domain. We propose that the CopG domains from two independent complexes interact through central ParD1 chains (blue and teal) (**Figure 4.19 A and B**). Displacement occurs at the relatively weak polar ParE1 – ParE1 interface (**Figure 4.4 A iii** and **Table 4.2**), and due to space, the toxin pairs are reorganised. The cognate light orange and purple toxins, for the abutting blue and teal ParD1 chains respectively, now form a pair at the centre of the structure as the red and brown ParE1 molecules are pushed out. The central blue and teal antitoxin chains remain full-length (previously, primary chains) and interfaced with the light orange and purple ParE1 toxins, respectively (**Figure 4.19 C**). The blue/teal ParD1 interface is stabilised heavily by ionic interactions (**Table 4.2**), and the external light green and pink ParD1 chains release their cognate toxins (red and brown, respectively) (**Figure 4.19 D**). The light green and pink ParD1 chains subsequently revert to a disordered C-terminus, and as such are unresolved in the crystal structure. Why this process is driven by increasing temperature, and why the increased ionic bonding between ParD CopG domains is preferred at higher temperatures (**Table 4.2 and Chapter3, Figure 3.12**), is yet to be fully understood. Additionally, further work needs to be performed on elucidating the higher-order species of ParDE1 that evolves in the later stages of sample incubation (**Chapter 3, Figure 3.12**). Our biophysical studies indicate that the ParE1 toxins are reabsorbed, potentially re-associating with the partially unstructured ParD1 chains (light green and pink) (**Figure 4.19 B**).

Figure 4.19 Updated model for ParDE1 complex remodelling and ParE1 toxin release

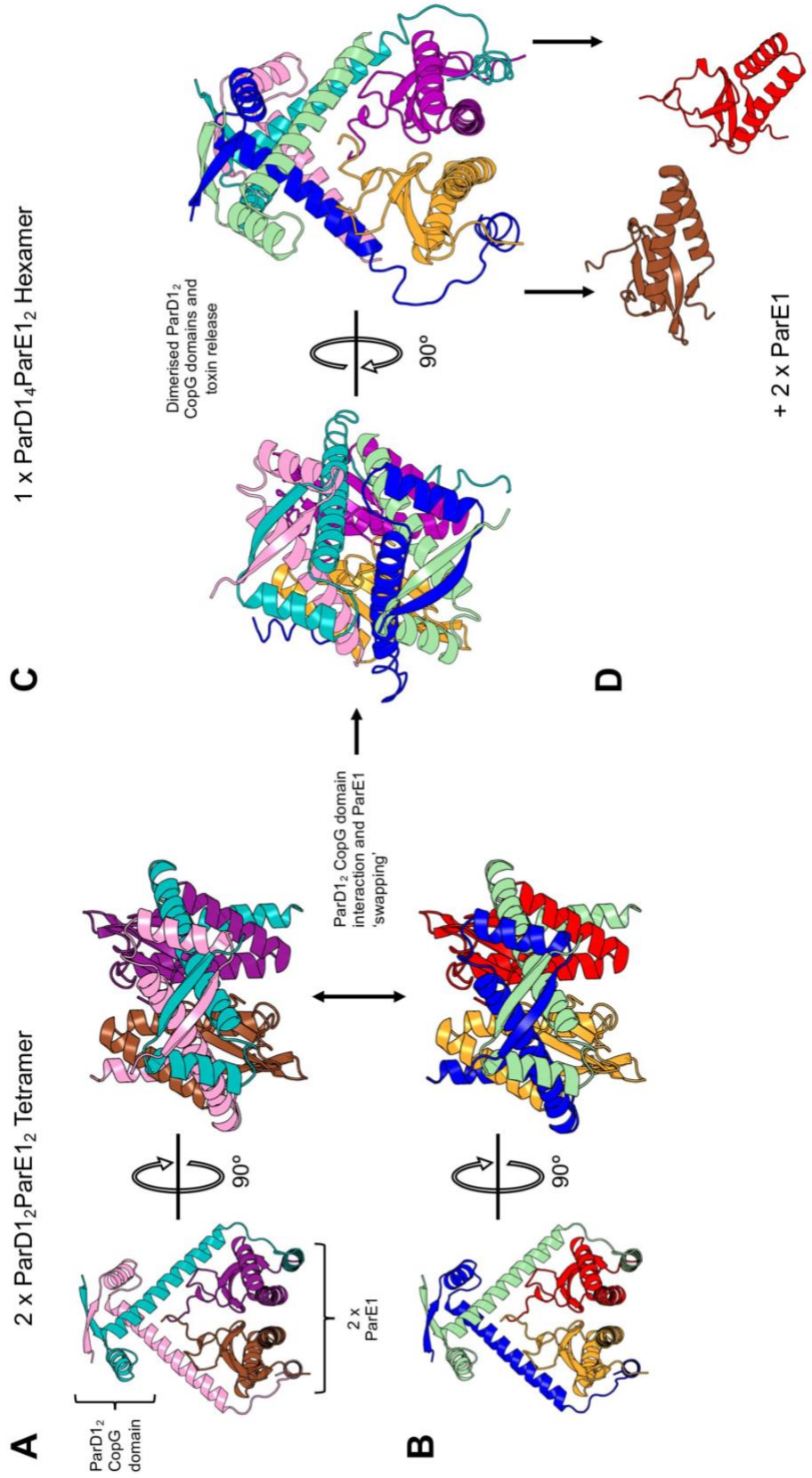


Figure legend overleaf

Figure 4.19 Updated model for ParDE1 complex remodelling and ParE1 toxin release

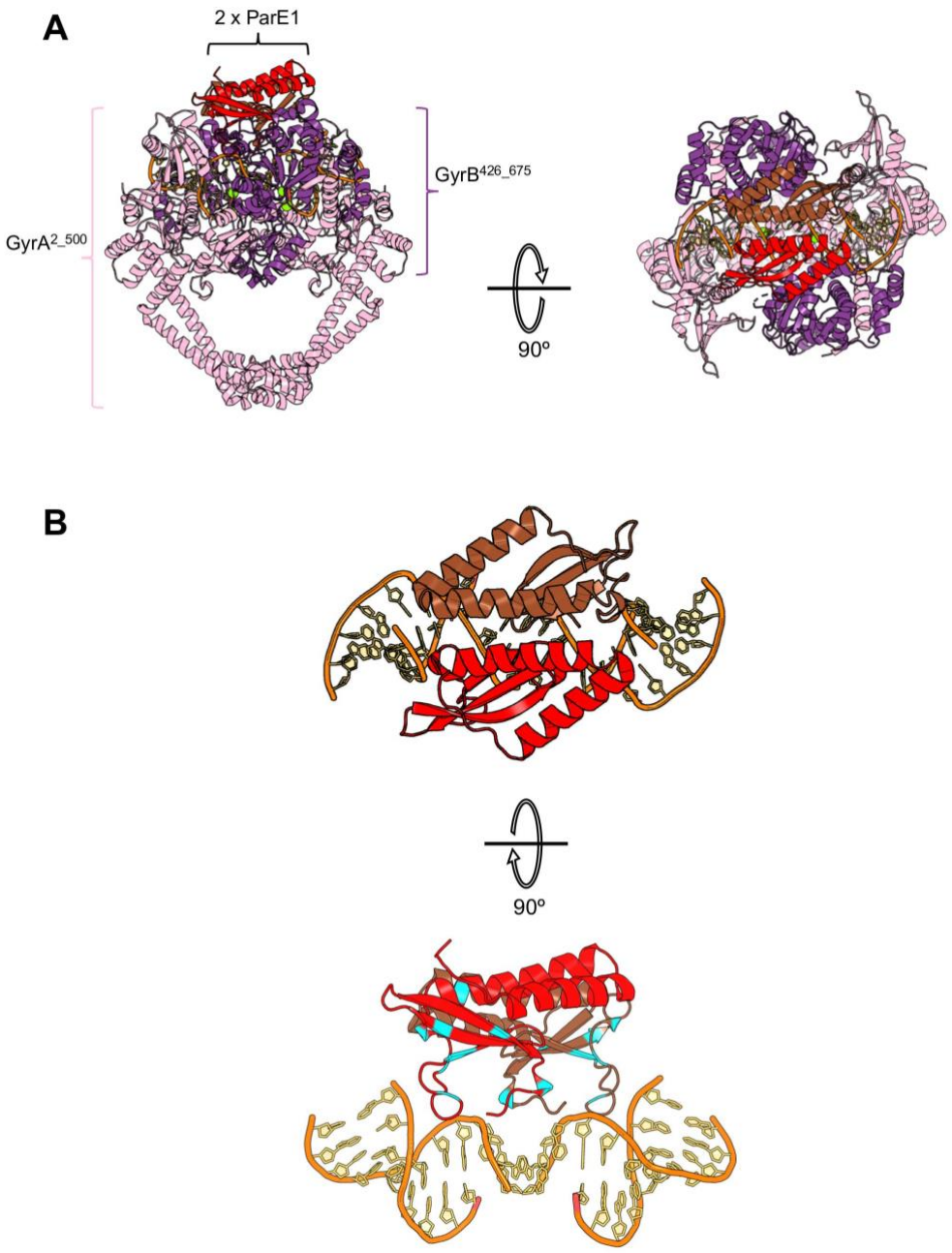
(**A**) AlphaFold generated $ParD1_2ParE1_2$ heterotetramer with ParD chains coloured teal and pink and ParE chains coloured brown and purple. The ParD1 dimer CopG domain is labelled appropriately; (**B**) AlphaFold generated $ParD1_2ParE1_2$ heterotetramer with ParD chains coloured light green and blue and ParE chains coloured light orange and red; (**C**) AlphaFold generated $ParD1_4ParE1_2$ heterohexamer after complex remodelling has occurred through dimerisation of the tetramer CopG domains and re-pairing of ParE1 toxins (now purple and light orange, from A and B, respectively). Internal facing ParD1 chains (teal and blue, from A and B, respectively) remain fully structured in complex with toxins; (**D**) Liberated brown and red ParE1 toxins, from A and B, respectively. Toxins are released, resulting in the transition to disordered/unstructured ParD1 C-termini for external pink and light green ParD1 chains (from A and B, respectively).

Using this model of the liberation of two ParE1 toxins, alongside the PISA information suggesting the existence of a potentially stabilising information between monomers, we propose a mechanism of ParE1 induced gyrase inhibition (**Figure 4.20**). While our biophysical data suggests that ParE1 exists in solution as a monomer, ParE1 may be stabilised as a dimer through the previously highlighted interface (**Figure 4.4 A iii**) when interacting with gyrase. Biochemical data indicates that the ParE – gyrase interaction would occur at the G-gate and could be supported by the ATPase domain of GyrB (**Chapter 3**, **Figure 3.8**). Using the $GyrB^{426\_675}A^{2\_500}$ crystal structure whereby the cleavage complex is stabilised with moxifloxacin (PDB: 5BS8 (Blower *et al.*, 2016)), we can manually model (PyMol) the ParE1 toxin pair to interact with both enzyme and DNA (**Figure 4.20 A**). This interaction may be further supported by the ATPase dimerising. When the ParE1 pair is isolated from the complex, the toxic C-terminal tails align to the asymmetric nick sites of DNA in the gyrase cleavage complex (**Figure 4.20 B**). Interestingly, when positioned this way the ParE1 toxin has several positively charged arginine and lysine residues that face toward the negatively charged DNA; the C-terminal tail also has an additional arginine (R96) that is unmodelled due to flexibility in our structure. Through this model, we propose that the C-terminal tail of the ParE1 toxin could sterically hinder the resealing of DNA at the G-gate, thus trapping the cleavage complex. This model does, however, rely on the stabilised ParE1 dimer at G-gate interface. One would expect to have seen increasing nicked form DNA in cleavage assays (**Chapter 3, Figure 3.14 B**) if this model were accurate as a single ParE1 could occupy the site unless both are somehow required to stabilise the interaction. This mechanism may potentially differ in ParE2, also monomeric in solution (**Chapter 3, Figure 3.16 B and D**) as the structured alpha helical C-terminus may not be able to interact with DNA in the same manner. The presence of this secondary structure may, however, explain the lowered potency demonstrated in our biochemistry compared to ParE1 (**Chapter 3, Figure 3.5 and Figure 3.14 B**). The ParE – gyrase interaction could be modelled with AlphaFold in future to either support or dispute this proposal as we build toward structural studies in cryo-EM or crystallography.

While we have observed the proposed ParDE1 remodelling process in analytical SEC, and coupled this with biochemistry to support the model (**Chapter 3**, **Figure 3.14**), we must further explore these interactions in future work. Importantly, we need to better understand the number of copies of ParD1 and ParE1 expressed within the cell; previous studies on RelBE systems indicate that the antitoxin RelB exists in 10-fold excess of the RelE toxin (Overgaard, Borch and Gerdes, 2009). Additionally, the *V. cholera* ParD structure indicates that CopG domains will multimerise into higher-order structures without the presence of the C-terminal domain (PDB: 7B22 (Garcia-Rodriguez et al., 2021)). It will be important to test whether increasing the amount of ParD1 in the remodelling process will prevent toxin release while still remodelling the complex; essentially, the incoming ParD1 dimer may simply adopt a ParE1 molecule and result in the same

heterohexamer structure. The use of single molecule FRET (Qiao *et al.*, 2021) may be an excellent method to further investigate the remodelling process. If the initially expressed heterotetramer ParDE1 complex sample can be divided, and the toxin components subsequently labelled with the appropriate dyes, mixing of the labelled samples may permit visualisation of the proposed toxin-swapping process previously proposed as the dyes come into close proximity with each other.

Figure 4.20 Model for ParE1 induced DNA cleavage



(**A**) Model for the interaction of ParE1 with the *M. tuberculosis* gyrase G-gate. The two ParE1 molecules (red and brown) can be manually positioned to interact with the $GyrB^{426\_675}A^{2\_500}$ fusion protein (PDB: 5BSQ, moxifloxacin removed). The GyrA domains are coloured pink and GyrB domains coloured purple; (**B**) Isolated ParE1 – DNA interaction from A. The C-terminal tails of the ParE1 pair align with the asymmetric nick sites created by gyrase during the reaction cycle. ParE1 arginine (R43, R49, R52, R58, R62, R74, R85, and R90) and lysine (K63) residues are coloured cyan.

Chapter 5. Phylogenetic and structural analyses of the *Mycobacterium tuberculosis* RelE/ParE superfamily

Chapter 5. Introduction

Mycobacterium tuberculosis harbours three RelBE TA systems, alongside the two ParDE systems studied in this work so far, with RelBE1 being studied by colleagues in the Genevaux Lab, Toulouse, France. Via personal correspondence we had learned that the RelE1 toxin does not appear to be catalytically active in a ribosome-dependent mRNA cleavage assay, as would be expected for a RelE toxin (Bøggild *et al*., 2012; Griffin, Davis and Strobel, 2013). Working with the Genevaux lab, we decided to investigate RelE1 further as a member of the RelE/ParE superfamily (Anantharaman and Aravind, 2003). Given the biochemical and structural data gathered thus far on the ParDE1 and ParDE2 systems, we have a good foundation for comparisons.

Initial study was performed via multiple sequence alignment (MSA) analysis and phylogenetics, followed by a structural characterisation of the RelBE1 complex. MSA of RelE and ParE toxins will elucidate the evolutionary relationships and diversity within the respective families. Given the supposed lack of canonical catalytic activity seen for RelE1 these analyses may highlight an evolutionary difference between it and the rest of the RelE toxins. This will also supplement the structural data gathered so far for the *M. tuberculosis* ParDE systems as the relationship between the ParE1 and ParE2 toxins can be further analysed. Following on from this, examination of the RelE/ParE superfamily via MSA and phylogenetics may allow us to better understand the relationships between the constituent families and whether there are clear sequence-level differences in the RelE and ParE toxins of interest in this study. It is possible, given the superfamily and data gathered thus far on RelE1, that the RelE1 toxin may have a close relationship with the catalytically inactive ParE family. Building on from this, structural study of the RelBE1 system will highlight potential structural abnormalities that may influence its function. Usefully, the structures of the RelBE2 and RelBE3 TA systems had been solved as part of the TB Structural Genomics Consortium (**Table 1.4**, PDB: 3G50 and 3OEI respectively) (Miallau *et al.,* 2013).

These analyses allow us to better understand the observation from the Genevaux lab (unpublished), and guide future studies through comparison of the RelBE and ParDE systems via sequence and structure.

5.1 Phylogenetic analyses of the RelE/ParE superfamily

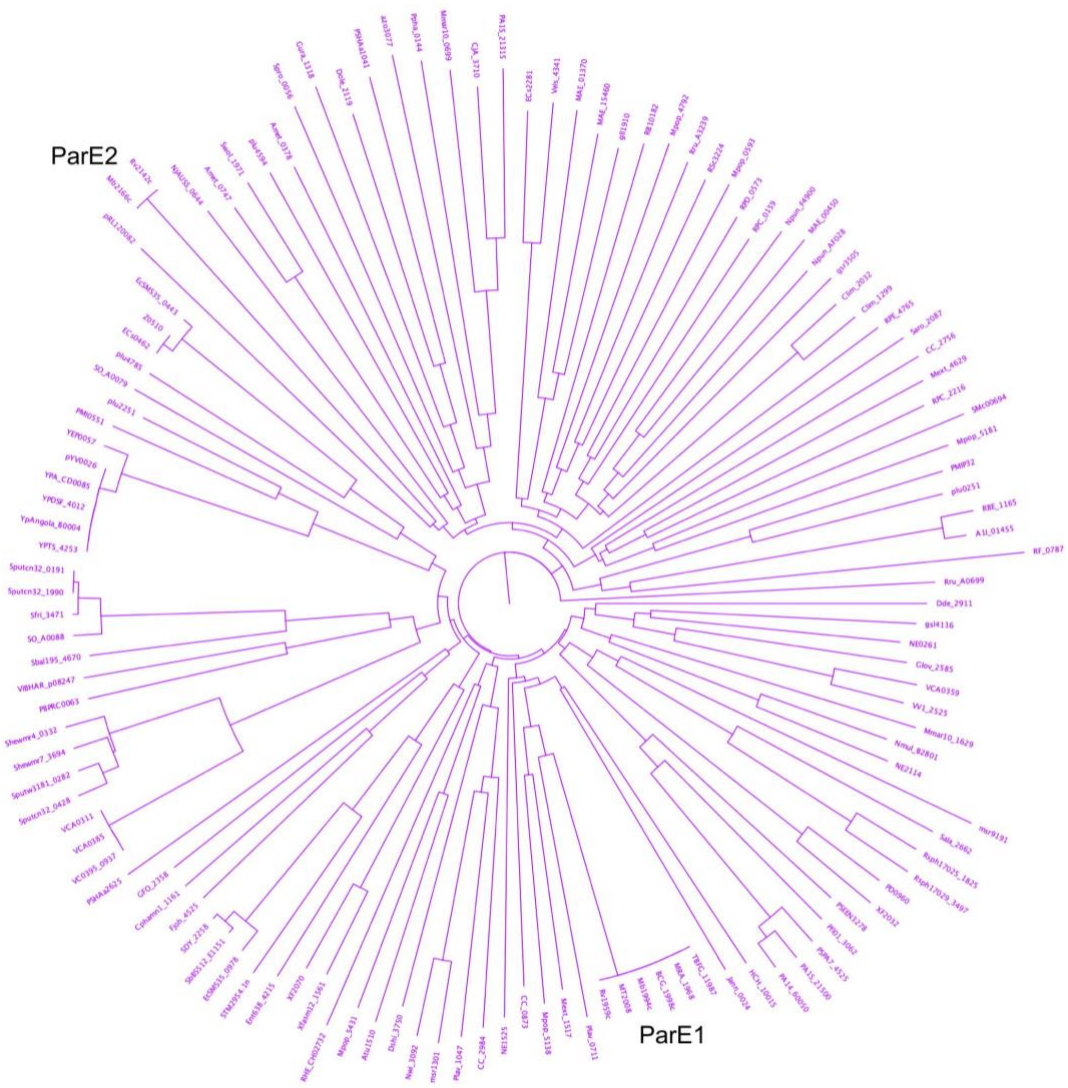
5.1.1 Independent alignment and analysis of RelE and ParE toxins

The first goal was to compare sequences within the RelE/ParE superfamily (Anantharaman and Aravind, 2003). Initially, the relevant protein sequences were collected from the TA system database, TADB2.0 (available at: https://bioinfo-mml.sjtu.edu.cn/TADB2/index.php) (Xie *et al.,* 2018) and inspected for potential mis-annotation. Before combining the datasets, protein sequence alignments were created for the respective RelE and ParE families using Clustal Omega (Sievers and Higgins, 2014). Based on these alignments, independent phylogenetic trees for ParE and RelE toxins were generated as per **Materials & Methods 2.11** (**Figure 5.1** and **Figure 5.2**, respectively), using FigTree (Cummings, 2004) (midpoint rooted due to lack of confirmed common ancestor) (Hess and De Moraes Russo, 2007). This allowed us to visualise the evolutionary relationships between the individual toxin sequences collected from TADB2.0 (Xie *et al.,* 2018).

The ParE phylogenetic tree (**Figure 5.1**) shows that the ParE sequences available in the TA database are potentially quite diverse, indicated by the large number of nodes close to the root, subsequently creating numerous branches and clades. This is contrasted by the RelE phylogenetic tree (**Figure 5.2**) which has far fewer early branch points, resulting in higher clustering of protein sequences at the leaves. This is unsurprising considering the requirement for higher RelE sequence conservation to maintain not only a mechanism of protein recognition (common to the RelE/ParE superfamily) (Dalton and Crosson, 2010), but also in maintaining essential residues for catalytic activity and toxin function (Griffin, Davis and Strobel, 2013).

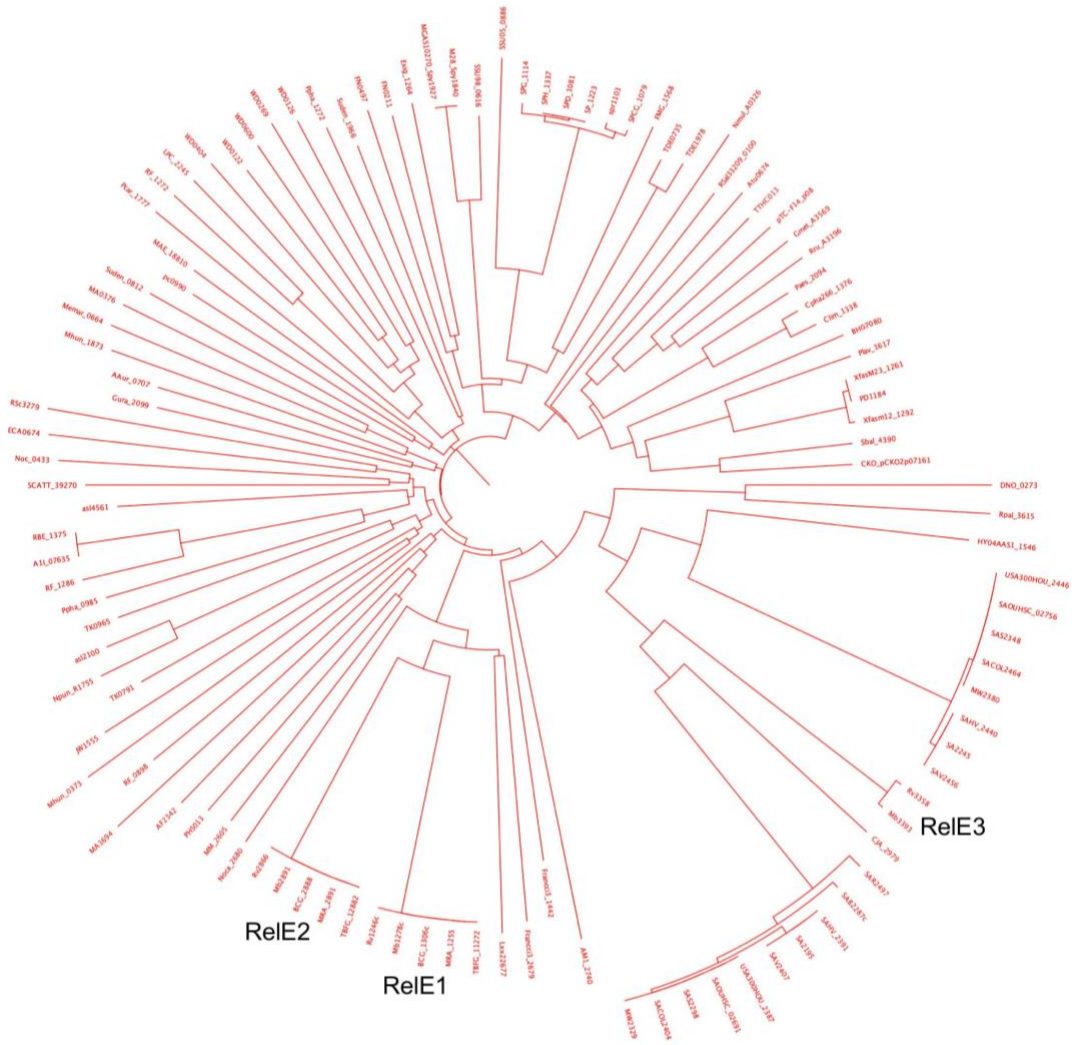
ParE1 and ParE2 (alongside their orthologous sequences in the Mycobacteria) are positioned distantly from each other, indicating a significantly different ancestor (**Figure 5.1**). Contrastingly, RelE1 and RelE2 (plus respective orthologues) share a more recent common ancestor, with RelE3 (and its single orthologue) then being more distantly related (**Figure 5.2**).

Figure 5.1 ParE toxin family phylogenetic tree



Phylogenetic tree of ParE protein sequences from the TA system database TADB 2.0, generated in FigTree with midpoint rooting. Initial sequence alignment was generated with Clustal Omega. Mycobacterial ParE clades are labelled as 'ParE1' and 'ParE2', respectively.

Figure 5.2 RelE toxin family phylogenetic tree



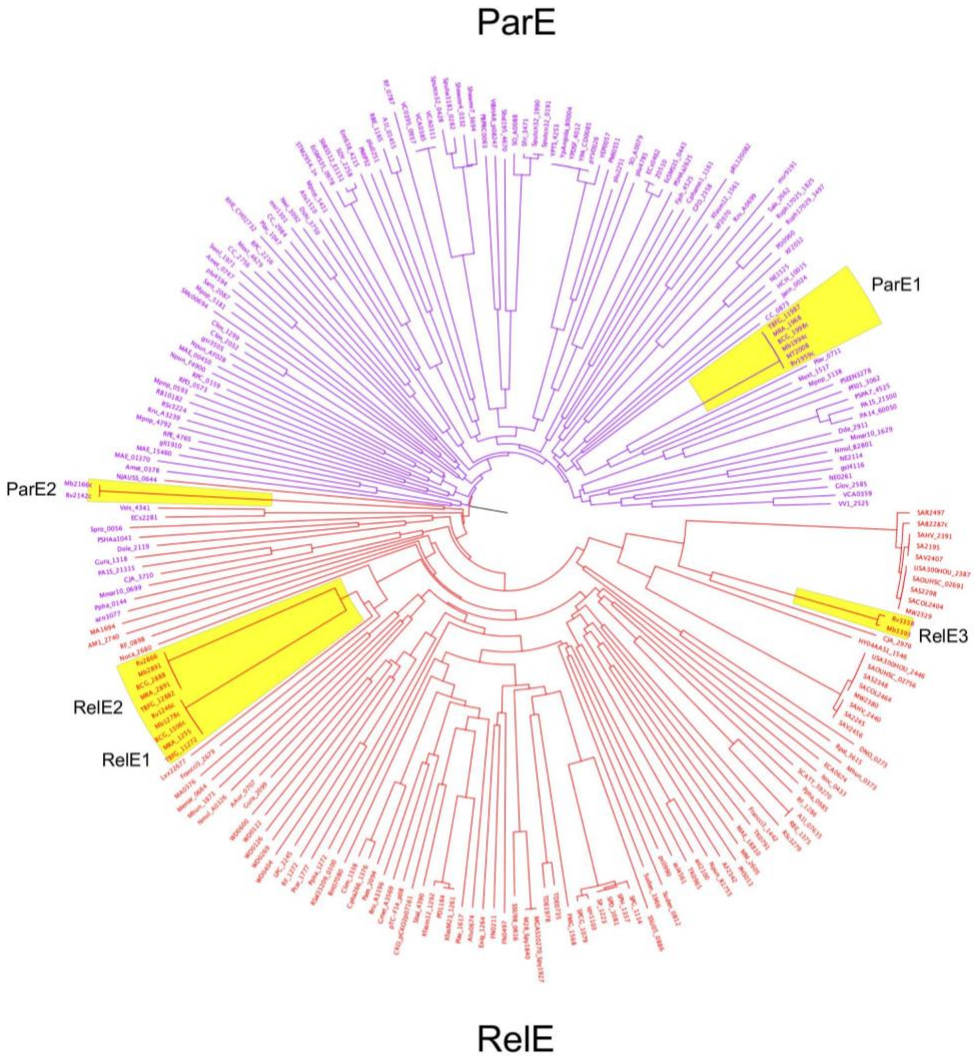
Phylogenetic tree of RelE protein sequences from the TA system database TADB 2.0, generated in FigTree with midpoint rooting. Initial sequence alignment was generated with Clustal Omega. Mycobacterial RelE clades are highlighted as 'RelE1', 'RelE2', and 'RelE3', respectively.

5.1.2 Alignment and analysis of the RelE/ParE toxin superfamily

Having visualised the evolutionary relationships within the RelE and ParE toxin families, highlighting the generally higher, but unsurprising, conservation of sequence through the RelE family, we can now investigate how the independent families are related. Performing the same analysis, using a Clustal Omega (Sievers and Higgins, 2014) generated protein sequence alignment to analyse sequence conservation in the superfamily, we can create a superfamily phylogenetic tree (**Materials & Methods 2.11**) (**Figure 5.3**). This allows us to visualise the evolutionary relationships between the RelE and ParE toxins, highlighting the most and least related between the respective families. By keeping the sequence identifiers (leaves) coloured as per **Figure 5.1** and **Figure 5.2** we see that the tree splits into hemispheres indicated by 'ParE' and 'RelE' family labels. However, by colouring the calculated ParE and RelE clades by their respective colour schemes, we highlight a number of ParE toxin sequences that now belong to a RelE clade. This small group interestingly contains the sequence for *M. tuberculosi*s ParE2 (highlighted) (**Figure 5.3**). Even more interestingly is that on further inspection, while ParE2 appears to be more related to the RelE clade, it becomes the first branch away from the remaining RelE sequences. Further to this, the mid-point rooted superfamily phylogenetic tree indicates that the *M. tuberculosis* RelE1 toxin (and neighbouring RelE2) are likely some of the most related in protein sequence to the ParE toxin family (low number of nodal points from that separating RelE and ParE families) (**Figure 5.3**). Given the surrounding ambiguous ParE sequences (purple leaves, red branches), these taxa may represent a transitory state between the RelE and ParE families. While the evolutionary relationships are not properly understood for this superfamily, it is believed that RNase T1 is a likely ancestor of RelE toxins given high similarities in the structure of the catalytic core (Neubauer *et al.*, 2009; Zhang *et al.*, 2012; Gucinski *et al.*, 2019). RNase T1 may therefore provide a good outgroup for re-rooting this tree and potentially providing insights into how individual systems emerged and to check the validity of the position of ambiguous ParE sequences.

Somewhat unsurprisingly, the overall structure of the separate ParE and RelE clades remained the same (**Figure 5.1 – Figure 5.3**) as the majority of the sequences are correctly classified within their respective ParE or RelE family. The positioning of the midpoint also highlights how diverged the sequences of ParE1 and RelE3 are (highlighted, **Figure 5.3**).

Figure 5.3 RelE/ParE toxin superfamily phylogenetic tree



Phylogenetic tree of ParE and RelE protein sequences from the TA system database TADB 2.0, generated in FigTree with midpoint rooting. ParE and RelE input sequences are indicated by purple and red leaf (label) colouring, respectively. ParE and RelE general clades are indicated by purple are red branches (connecting lines), respectively. Initial sequence alignment was generated with Clustal Omega. Mycobacterial ParE and RelE protein sequence clades are highlighted in yellow and labelled appropriately.

Re-aligning the RelE/ParE superfamily sequences with the putative ancestor sequence of RNase T1 (*Aspergillus oryzae*) allows us to plot a new phylogenetic tree (**Figure 5.4**) (fanned out rather than circularised for visualisation purposes). Midpoint rooting of this phylogenetic tree positioned the RNase T1 sequence between the RelE and ParE clades (not shown), and when rooting to the RNase sequence as the outgroup the overall structure of the tree appears strikingly similar to that seen in **Figure 5.3**. We initially hypothesised that, due to the conserved catalytic residues, RNase would be the ancestor of RelE, from which the ParE proteins evolved. This analysis suggests that both descend from RNase T1 divergently, rather than one family evolving from the other.

The initial noticeable difference in this analysis is that the previously identified 'ambiguous' ParE sequence clades have been resolved (**Figure 5.3 and Figure 5.4**). These are now positioned back amidst the ParE taxa, likely due to the addition of catalytic residues creating better definition between the RelE and ParE sequences. Instead of this, three distinct initial clades are generated from the RNase root. Interestingly, a previously identified ParE sequence (msr9191 – *Mesorhizobium japonicum* toxin) has become the initial branch and evidently unrelated to any of the other RelE or ParE sequences (**Figure 5.4**). The separate clade diverges into ParE and RelE sequences, indicated by the red (RelE) and purple (ParE) branches; what we see again is the RelE clade then splitting into the clade for the *M. tuberculosis* ParE2 sequence and the remaining RelE sequences. The ParE2 sequence is peculiar in that, by these phylogenetic analyses, it is apparently not a ParE at all, rather it may be an early divergent from the RelE family (**Figure 5.4**, highlighted yellow with ParE2 label, black branching).

Alongside this observation we maintain that the *M. tuberculosis* RelE1 and RelE2 sequences are not only closely related to each other but are also some of the closest in relation to the ParE family (**Figure 5.4**, highlighted yellow with RelE1/RelE2 label). The reasons for this are as yet unknown. Taking this into account alongside the correspondence with the Genevaux group, there is the potential that *M. tuberculosis* RelE1 (and RelE2) has a close relationship with the ParE family and may have an altered protein target, or simply loss of catalytic activity. As sequence analysis has provided little indication of this, with the catalytic residues being present and aligned within the RelE family, structural analysis of the RelBE1 system may offer better insights. Combining this with comparisons to known catalytically active RelE structures, and those from *M. tuberculosis*, may provide guidance as to why this system has not shown ribosome-dependent mRNAse activity (unpublished, personal correspondence). Additionally, comparison with the crystal structures of ParE1 and ParE2 may offer insights into the peculiar positioning of the ParE2 sequence and its potential as a closer relative to the RelE toxins.

Figure 5.4 RNase T1-rooted RelE/ParE phylogenetic tree

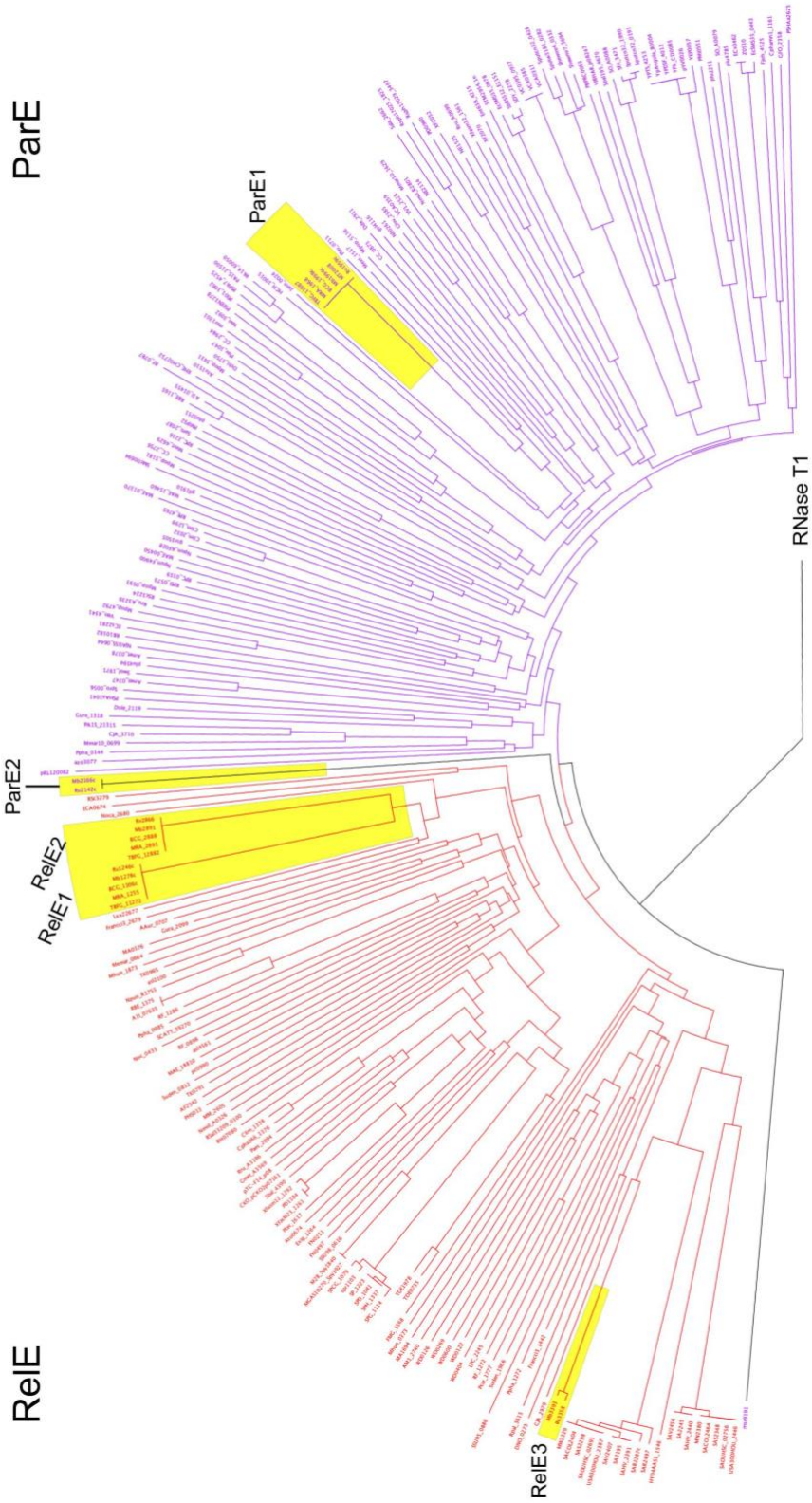


Figure legend overleaf

Figure 5.4 RNase T1-rooted RelE/ParE phylogenetic tree

Phylogenetic tree of RNase T1, ParE and RelE protein sequences from the TA system database TADB 2.0, generated in FigTree. Tree is rooted to the RNase T1 sequence identified as the outgroup. ParE and RelE input sequences are indicated by purple and red leaf (label) colouring, respectively. ParE and RelE general clades are indicated by purple are red branches (connecting lines), respectively. Branches for the sequences of *M. tuberculosis* ParE2 and *Mesorhizobium japonicum* msr9191 are colour black for identification. Initial sequence alignment was generated with Clustal Omega. Mycobacterial ParE and RelE protein sequence clades are highlighted in yellow and labelled appropriately.

5.2 RelBE1 complex crystal structure

Protein expression and purification for RelBE1 occurred as generally shown in **Chapter 3.1**, with *E. coli* Rosetta™ 2(DE3) pLysS being transformed with pTRB638. Details of expression, purification, crystallisation, and X-ray crystallography experiments can be found in **Materials & Methods 2.4.2, 2.5.10, 2.8.2.1, and 2.9**, respectively. Hexagonal bifrustum shaped crystals formed in 4 M ammonium acetate, 0.1 M Bis-Tris propane which could be transformed into needle morphology on reducing the ammonium acetate concentration. Importantly, the crystal buffer for these experiments had to contain 5 % glycerol to reduce the amount of precipitation and micro-crystal formation. Both crystal forms diffracted well providing datasets refining in spacegroups $I2_12_12_1$ and $P6_522$, respectively. The hexagonal bifrustum form was more resilient and allowed for the collection of a greater number of datasets; 13 datasets from a single crystal were merged to a final resolution of 2.20 Å in space group $I2_12_12_1$. The RelBE1 structure was determined by molecular replacement using PDB: 3G5O as a search model (**Materials & Methods 2.9.5**, **and Table 5.1**).

The asymmetric unit contained a heterotetrameric $RelB1_2RelE1_2$ protein complex (**Figure 5.5 A**). The quaternary structure resembled that of the RelBE2 search model (PDB: 3G50) and RelBE3 (PDB: 3OEI), but not of others in the PDB; this indicates a common complex structure specific to Mycobacteria for RelBE TA systems. The toxin-antitoxin interaction somewhat resembled that seen for the ParDE1 complex (**Figure 4.1 A**), however this apparently occurs via three antitoxin α helices in RelBE1. Overall, the composition creates a triangular (or V-shaped complex) (Bøggild *et al.,* 2012) with the RelE1 toxin molecules positioned away from each other, and the RelB1 antitoxin running up the middle, complexing at the top (**Figure 5.5 A**). The complex can be broken down into RelB – RelE dimers which appear to interact through the RelE toxin conserved β sheet core and three α helices of the RelB antitoxin (**Figure 5.5 B**). A single non-hydrogen ligand atom is built in (**Table 5.1**) as a chloride ion and is not believed to contribute to the stability or function of the complex.

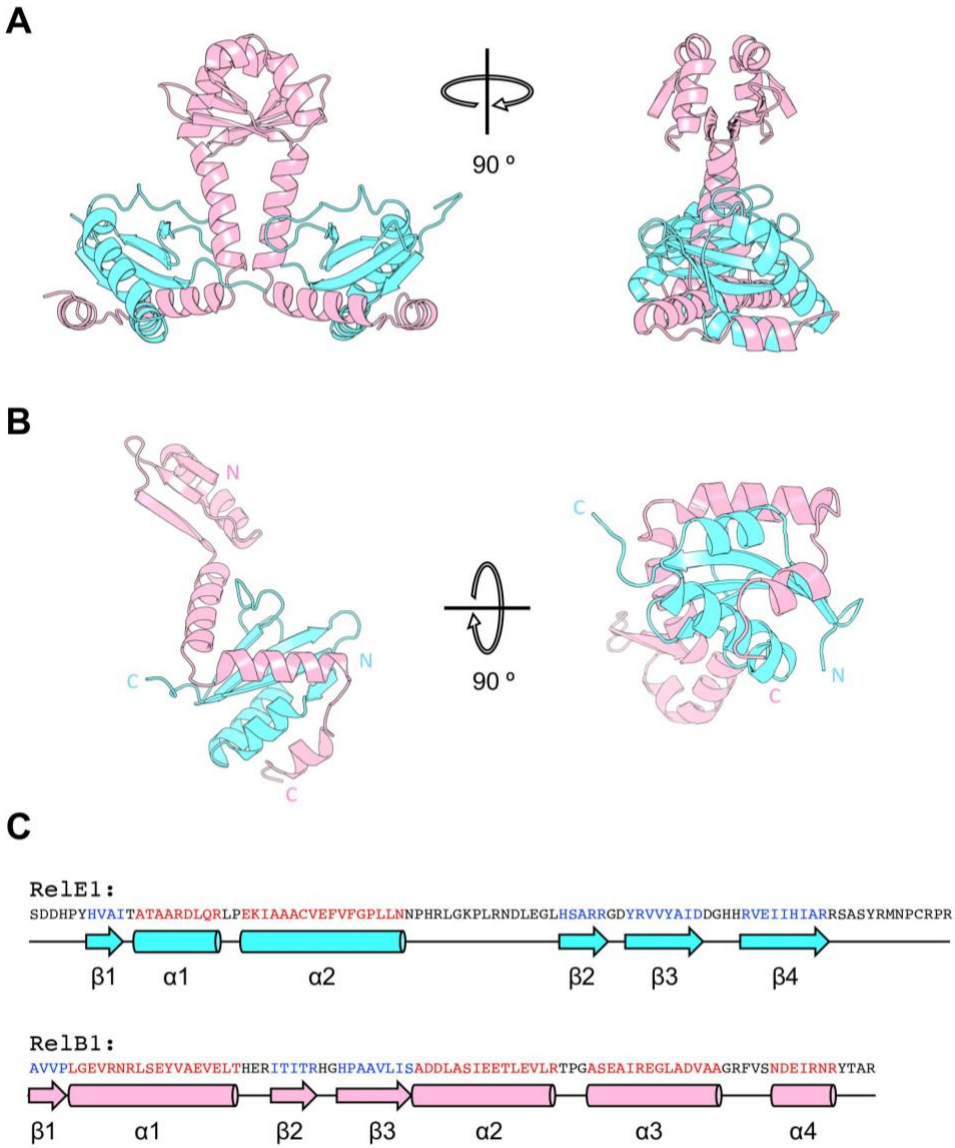
The RelE1 toxin is comprised of β1 (H8 – I11), α1 (A13 – R21), α2 (E24 – N40), β2 (H57 – R61), β3 (Y64 – D71), and β4 (H75 – R84) (**Figure 5.5 C**). The C-terminal 11 residues (A87 – R97) were not resolved in the crystal structure, likely due to flexibility in this region. The RelB1 antitoxin is comprised of β1 (A2 – P5), α1 (L6 – T23), β2 (I27 – R31), β3 (H34 – S41), α2 (A42 – R56), α3 (A60 – A73), and α4 (N79 – R85) (**Figure 5.5 C**). The C-terminal R86 was not resolved in the crystal structure, likely due to flexibility in this region.

Table 5.1 Data collection and refinement statistics for RelBE1

	RelBE1
PDB ID Code	8C22
Number of crystals	12
Beamline	Diamond I04
Wavelength, Å	0.9793
Resolution range, Å	65.68 – 2.20 (2.279 – 2.20)
Space group	P 21 21 21
Unit cell, *a b c* (Å), α β γ (°)	88.99 98.54 97.35, 90 90 90
Total reflections	22105
Unique reflections	22101 (2193)
Multiplicity	1.0 (1.0)
Completeness (%)	100 (100)
Mean I/sigma(I)	20.9 (0.4)
R_{merge}	0.016 (0.849)
R_{meas}	0.023 (1.201)
$CC_{1/2}$	1.0 (0.536)
R_{work}	0.237 (0.368)
R_{free}	0.271 (0.373)
No. of non-hydrogen atoms	2676
Macromolecules	2671
Ligands	1
Solvent	4
Protein Residues	342
RMSD (bonds, Å)	0.007
RMSD (angles, °)	0.98
Ramachandran favoured (%)	95.81
Ramachandran allowed (%)	4.19
Ramachandran outliers (%)	0.00
Average B-factor	80.13
Macromolecules	80.11
Ligands	105.88
Solvent	83.42

Values in parenthesis are for the highest resolution shell. R_{free} was calculated with 5% of the reflections selected.

Figure 5.5 RelBE1 crystal structure and secondary structures of the component RelE1 and RelB1 proteins



(**A**) Cartoon representation of the heterotetrameric $RelB1_2RelE1_2$ complex crystal structure. RelB1 antitoxins are coloured light pink. RelE1 toxins are coloured cyan; (**B**) Rotated views of the isolated RelBE dimer with N and C -terminal labels; (**C**) Secondary structure schematics of RelE1 and RelB1. Secondary structure is displayed underneath the protein sequence, whereby amino acids constituting α helices are coloured red, and those constituting β strands are coloured blue. Secondary structure elements are coloured as per their crystal structure counterpart and represented by tubes for α helices and arrows for β strands and are labelled accordingly.

The tertiary fold of RelE1 displays the generally conserved structure for the superfamily; the N-terminal β1/α1 and α2 form a hairpin, α2 connects via a loop to the antiparallel β sheet core of β2, β3, and β4, which sits on top of the hairpin. A parallel β strand interaction is formed between β1 and β2 (**Figure 5.6 A**). The C-terminus remains unstructured in RelE1, likely due to displacement by the RelB1 interaction, as this region would usually form an α helix in RelE toxins (Overgaard, Borch and Gerdes, 2009; Bøggild *et al.,* 2012). Similarly, both ParE1 and ParE2, within their respective crystal structures, had unstructured C-terminal regions resolved to a varying number of amino acids (**Figure 4.1 and Figure 4.13**) The two RelE1 toxin molecules present in the crystal structure are highly similar in fold; aligning the two protomers returned an RMSD of 0.390 Å indicating an extremely stable tertiary structure (**Figure 5.6 A**).

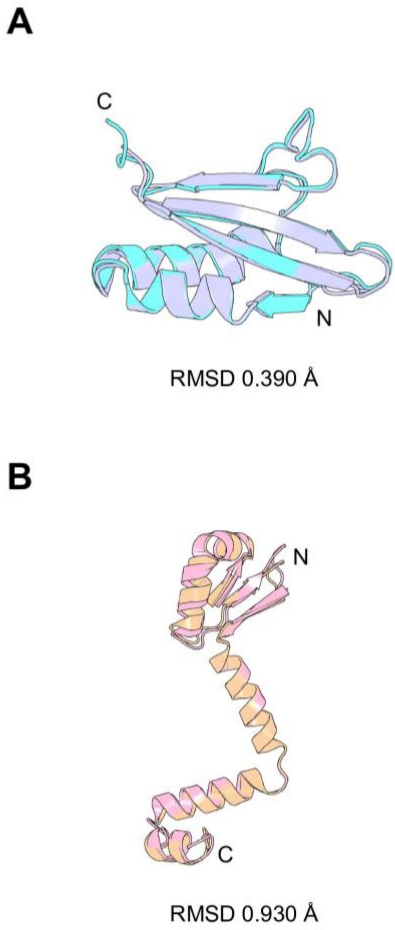
The tertiary structure of RelB1 is composed of β1 and β2 interacting in parallel, joined by the arched helix of α1. β2 and β3 form an anti-parallel sheet followed directly by α2, which connects to α3 via a two amino acid linker. α3 connects to the shorter α4 via a 5 amino acid loop, with the C-terminal residues of Y96 – A98 being unstructured (**Figure 5.6 B**). RelB1 is slightly more flexible in structure as protomer alignment returned an RMSD of 0.930 Å with the highest degree of alignment being concentrated through α2 – α4 (the region forming the toxin interface) (**Figure 5.6 B**).

5.2.1 RelBE1 is a probable autoregulatory DNA-binding complex

Investigation of the *relB* upstream region using einverted (EMBOSS) (Rice, Longden and Bleasby, 2000) highlighted a perfect, 13 base-pair inverted repeat sequence (**Figure 5.7 A**), a common operator motif found in TA system promoter regions (Fraikin, Goormaghtigh and van Melderen, 2020). Rendering the RelBE1 complex surface using APBS electrostatics (PyMol) further supports this, as an electropositive surface is concentrated at the antitoxin complexing region (**Figure 5.7 B**); a typical feature seen in canonical type II TA systems which autoregulate via conditional cooperativity (Garcia-Pino *et al.,* 2010). Dali searches (Holm, 2020) using a dimer of RelB (A2 – R56) (N-terminus to C-terminus of α2, the likely DNA-binding structure) returned several Rel, Phd, Vap, and Yef system antitoxins indicating a widespread mechanism of DNA-binding. The crystal structure of a highly similar Phd N-terminal DNA-binding region complexed with its operator inverted repeat sequence (PDB: 4ZM2) (Garcia-Pino *et al.*, 2016) was used to create a DNA-binding model (**Figure 5.7 C**). The RelB1 dimer interface is created by a six-stranded β sheet, with an antiparallel interaction occurring between β3 of each monomer, forming the 'head' of the Phd/YefM DNA-binding structure (**Figure 1.7**) (Garcia-Pino *et al.*, 2016). The loop between β3 and α2 creates a wing on either side (**Figure 5.7 C**). The 'head' of the RelB1 dimer contains R10 and R12 in α1, while the wing contains R31; it is likely that all play a role in the operator DNA interaction with R10 and R12 inserting into the major groove, while R31 interacts with the DNA-

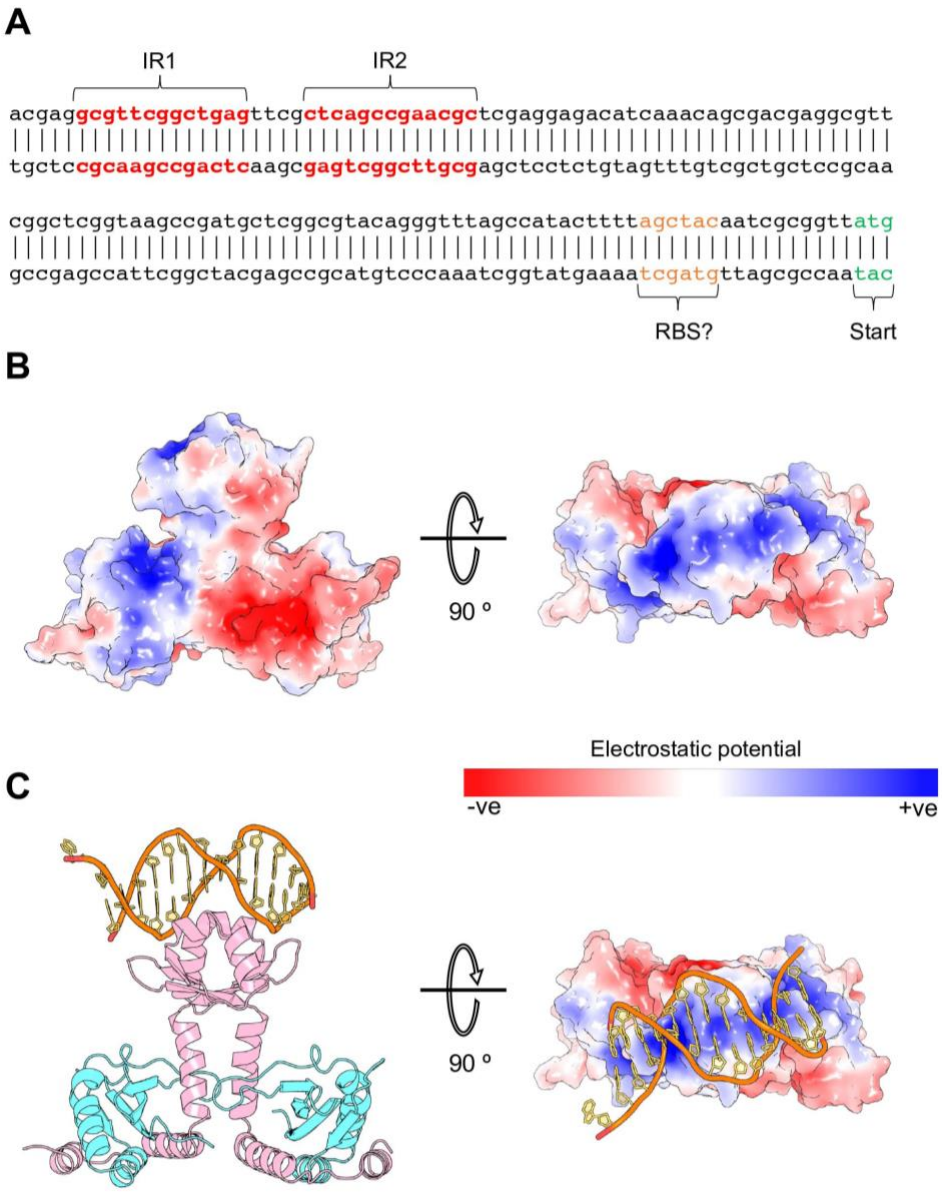
backbone (Garcia-Pino *et al.*, 2016). The presence of the inverted repeat indicates that two RelBE1 complexes will interact with the operator site, and RelB1 – RelE1 stoichiometry is likely to play a role in governing the autoregulatory capacity (Garcia-Pino *et al.,* 2010; Garcia-Pino *et al.,* 2016).

Figure 5.6 RelE/RelB protomer comparison



(**A**) Alignment of the RelE1 monomers with one RelE1 recoloured to slate; (**B**) Alignment of the RelB1 monomers with one RelB1 recoloured to wheat. N and C -terminal labels are included. RMSD values obtained from the respective alignments are included below each structure.

Figure 5.7 RelBE1 is a probably DNA-binding protein complex



(**A**) Upstream region of the *relBE* open reading frame with 13 bp inverted repeat highlighted in red text and brackets showing inverted repeat (IR) 1 and 2. Putative ribosome binding site (RBS) is highlighted in orange alongside the *relBE* start codon in green; (**B**) Surface rendering of the RelBE1 complex crystal structure coloured by electrostatic potential using the APBS plugin (PyMol); (**C**) DNA-bound model for a single RelBE1 complex bound to a single inverted repeat with rotated view to show the electrostatic potential as shown in B. DNA inverted repeat sequence is taken from PDB: 4ZM0, antitoxin Phd from phage P1 bound to operator DNA.

5.3 RelBE1 complex assembly and interfaces

As part of the RelE/ParE superfamily, previous analyses indicate that a conserved mechanism of protein recognition would be demonstrated within the complex structure. Amino acid conservation within the RelE toxin should form a largely hydrophobic groove (Anantharaman and Aravind, 2003; Dalton and Crosson, 2010), spanning the beta sheet core and toxin underside, between the alpha helix hairpin (as has also been shown in the **Chapter 4** ParDE system structures). Analysis of the interfaces within the complex structure will allow us to better understand how the toxin and antitoxin come together to form the quaternary structure.

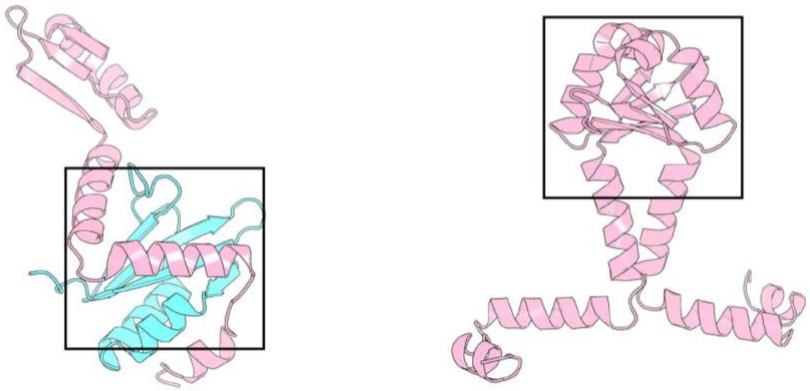
5.3.1 PISA analysis of the RelBE1 complex assembly

Using PISA (Chen and Zhou, 2005) to analyse the interfaces essential for RelBE1 complex assembly highlights only two unique interfaces of sufficient surface area (Å^2) to be potentially complex forming (**Figure 5.8 A**). Both interfaces within the complex are calculated to score the maximum of 1 in complexing significance (CSS) (**Figure 5.8 B**) indicating they are both essential to the formation of the quaternary structure seen in the crystal. The interface formed between the RelB1 antitoxin and RelE1 toxin (**Figure 5.8 A i**) appears twice within the structure (values from only one have been reported as scored were highly similar), and the interface between the RelB1 monomers, that which forms the predicted DNA-binding domain (**Figure 5.8 A ii**), appears only once.

Interestingly, while both interfaces are of similar surface areas and stabilised by comparable numbers of hydrogen bonds and salt bridges (with no disulphide bonds present), the RelB1-RelE1 interface was calculated to have a higher free energy gain (ΔiG kcal/mol) on formation, indicating that it would form more readily than that of the RelB1-RelB1 interface (**Figure 5.8 B**). Electron density at this essential interface is demonstrated in **Figure S2 C**, showing the 2Fo-Fc map for the relevant residues of the RelB1 antitoxin. The RelB1-RelB1 interface free energy gain value and CSS score indicate that in isolation RelB1 would form a stable dimer, likely capable of autoregulation due to the formation of the globular N-terminal DNA-binding domain.

Figure 5.8 PISA analysis of the RelBE1 complex

A



i ii

B

Interface	Interface area (Å²)	Δ^iG kcal/mol	Δ^iG P-value	N^{HB}	N^{SB}	N^{DS}	CSS
i	1378.0	-16.0	0.183	15	6	0	1.000
ii	1308.5	-14.7	0.319	18	7	0	1.000

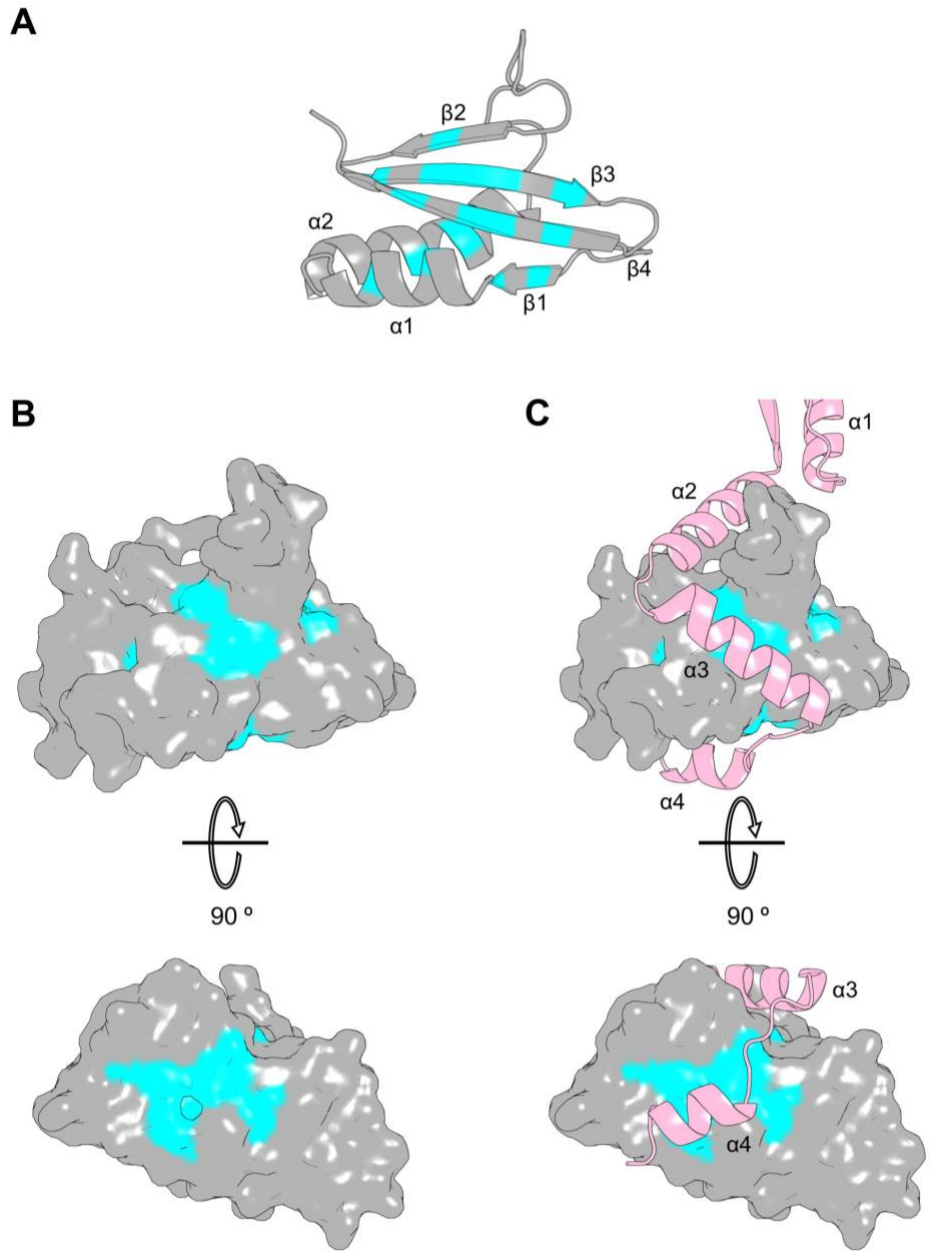
(**A**) Interfaces of the RelBE1 complex analysed by PISA. i: RelB1 – RelE1 interface formed between the RelE1 beta sheet core and the c-terminal helices of RelB1; ii: RelB1 – RelB1 interface formed at the RelB1 N-terminal domains; (**B**) PISA analysis of interfaces i – ii. Interface area is calculated in square angstroms (Å²); Δ^iG – solvation energy on folding (free energy gain); N^{HB} – Number of hydrogen bonds within the respective interface; N^{SB} – Number of salt bridges within the respective interface; N^{DS} – Number of disulphide bonds formed within the respective interface; CSS – Complex Significance Score ranges between 0 – 1 as interface relevance to complex formation increases.

5.3.2 RelBE1 conserved mechanism of protein recognition

Multiple sequence alignment of the RelE/ParE superfamily (Anantharaman and Aravind, 2003) highlighted the conservation of hydrophobic residues throughout the individual protein sequences. Dalton and Crosson ((Dalton and Crosson, 2010) used this superfamily analysis to demonstrate that these conserved hydrophobic amino acids constitute a structurally conserved hydrophobic groove for antitoxin recognition and binding using the *C. crescentus* ParE toxin structure (PDB: 3KXE).

As the *M. tuberculosis* RelE1 sequence was included in the superfamily analysis within the RelE family, we can highlight the highly conserved residues in the crystal structure of the toxin (isolated from the RelBE1 complex) (**Figure 5.9 A**). Visualising these conserved residues (**Figure 5.9 A**) demonstrates that they are concentrated across the β sheet core (β1 – β4), as is seen for *C. crescentus* ParE (Dalton and Crosson, 2010) and the *M. tuberculosis* ParE1 and ParE2 structures (**Chapter 4**, **Figure 4.5 and Figure 4.14)**. Rendering the surface of RelE1 while maintaining the highlighted residues shows us that these regularly spaced hydrophobic amino acids creates specifically positioned patches at the surface (**Figure 5.9 B**) to which the antitoxin, RelB1, binds (**Figure 5.9 C**). RelB1 interacts with the conserved hydrophobic patches via α3, α4, and the connecting loop, while α2 does not appear to interact with the RelE1 toxin at a highly conserved surface by this analysis (**Figure 5.9 C**). Taken together, with the presence of the RelB1 DNA-binding domain, this suggests that RelB1 A2 – R56 (β1 – α2, 'head' and 'shaft') constitutes an autoregulatory domain, while T57 – R89 (α3 – α4) constitute a specific RelE toxin-binding domain.

Figure 5.9 RelBE1 conserved mechanism of protein recognition



(**A**) Cartoon representation of the RelE1 toxin, coloured grey, with conserved residues from the RelE/ParE superfamily highlighted in cyan; (**B**) Surface rendered RelE1 toxin with conserved residues highlighted in cyan from A; (**C**) Surface rendered RelE1 with conserved residues in cyan complexed with RelB1 (coloured light pink). Secondary structure elements are labelled where appropriate throughout.

5.4 Amino acid conservation within the RelBE1 system

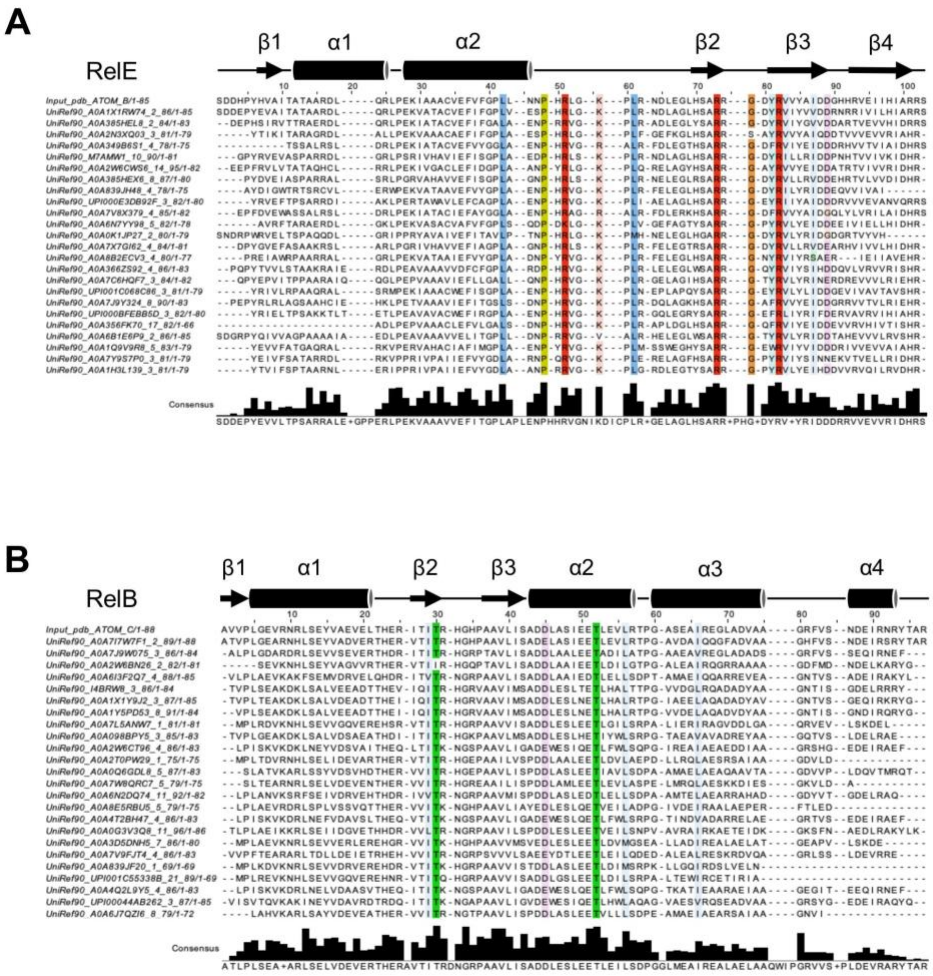
Amino acid conservation within the RelBE/ParDE system expands beyond a mechanism of protein recognition, as demonstrated through superfamily sequence alignment and identification of hydrophobic interfaces on the toxin components (Anantharaman and Aravind, 2003; Dalton and Crosson, 2010). Analysis of the *M. tuberculosis* RelBE1 system components, RelB1 and RelE1, using ConSurf allows us to demonstrate the conservation of amino acids determined by multiple sequence alignment (Ashkenazy *et al.*, 2016) (**Figure 5.10**), followed by rendering of the crystal structures using a spectrum of conservation scores to highlight the highest and lowest conserved residues.

5.4.1 Multiple sequence alignment of RelE1 and RelB1

Using the ConSurf generated multiple sequence alignments (Ashkenazy *et al.*, 2016) from the isolated RelE1 and RelB1 structures, amino acids can be coloured in the Clustal format in Jalview (Waterhouse *et al.*, 2009) and visualised by conservation threshold set to 80% (**Figure 5.10**) (25 of 125 total sequences shown for both). Plotting this alongside the secondary structure diagram demonstrates where the highest conserved amino acids are concentrated; a number of highly conserved residues are present through the RelE1 structure (as has been previously shown though superfamily alignment (**Figure 5.9**) (Anantharaman and Aravind, 2003). However, in contrast, using the RelE1 alignment we show that the highest conserved residues are a mixture of hydrophobic and charged amino acids positioned from the C-terminus of α2 through to the C-terminus of β2 (**Figure 5.10 A**). As previously commented on, RelE1 proteins must conserve not only a mechanism of protein recognition in order to interact with their counterpart antitoxin, RelB, but also conserve a number of residues such as arginine, lysine, and tyrosine, essential to mRNase action when positioned within the ribosome (Neubauer *et al.*, 2009).

The ConSurf analysis for RelB1 shows a very low level of sequence conservation when set to a threshold of 80%, demonstrated by the lack of strong Clustal colouring throughout the sequence (**Figure 5.10 B**). The highest conserved amino acids are apparently two threonine residues, one positioned in β2 and the other in α2, with conservation of a negatively charged amino acid (D/E) in α2 also (**Figure 5.10 B**). Comparing the consensus sequence graphs for both RelE1 and RelB1 demonstrates the generally more highly conserved sequence spanning the entire length of the sequence for RelE1, while RelB1 appears more variable and to decrease significantly by consensus toward the C-terminus (**Figure 5.10 A and B**). Plotting the spectrum of conservation scores on the crystal structures will allow for better understanding of the positioning of essential residues and how the TA system components have evolved to have a specific interaction while maintaining a conserved toxin function.

Figure 5.10 ConSurf multiple sequence alignment of RelE1 and RelB1



(**A**) RelE1; (**B**) RelB1. Both alignments are visualised in JalView, with Clustal colouring shown at 80% conservation. Matching secondary structure diagrams are presented above each alignment (cylinders represent α helices and arrows represent β strands). Consensus sequences and corresponding graphs are shown below each alignment. Multiple sequence alignment generated by ConSurf for both RelE1 and RelB1 using Clustal; top 25 sequences out of 125 are shown.

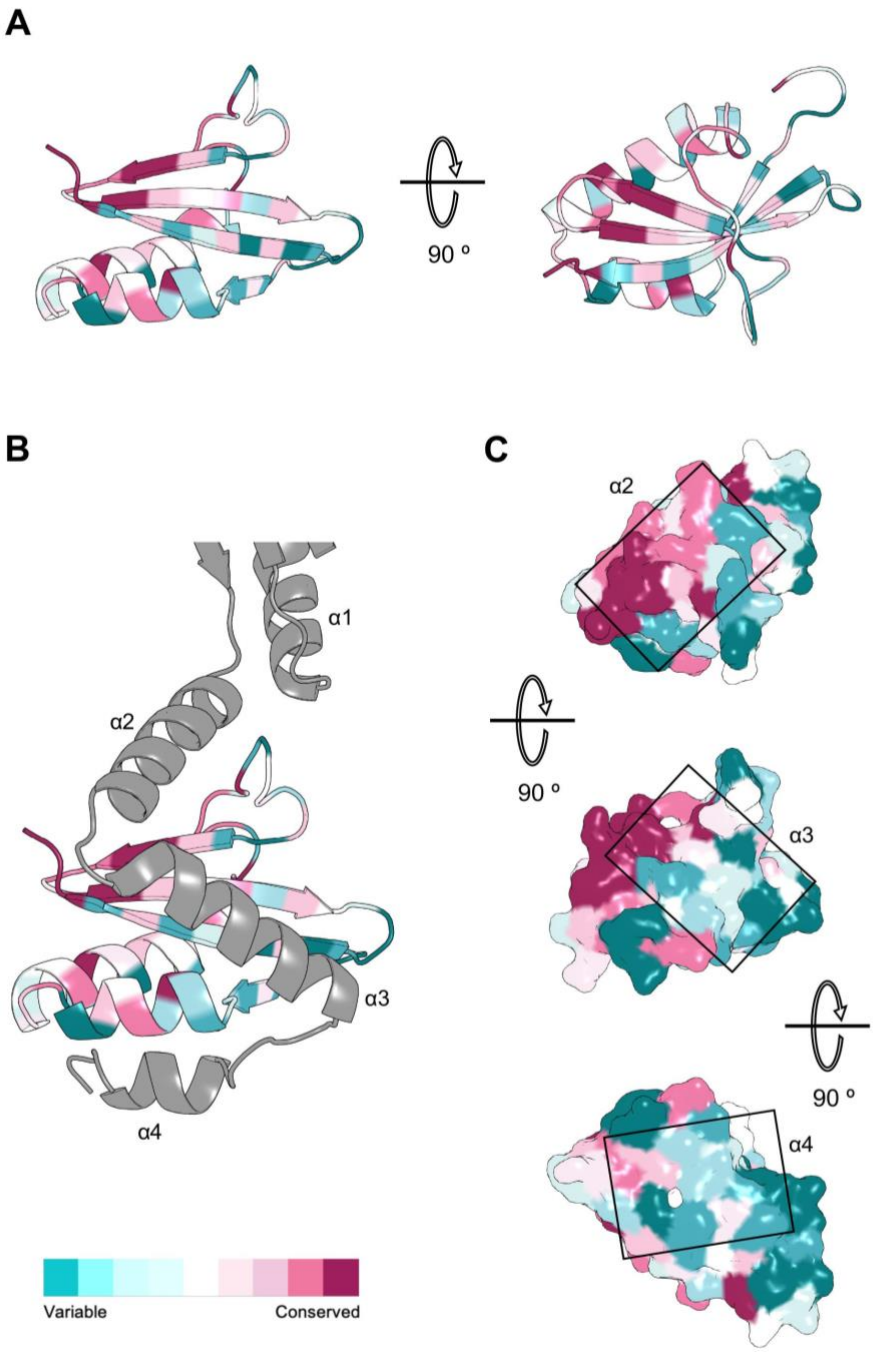
5.4.2 ConSurf analysis of the RelE1 toxin

Rendering the RelE1 crystal structure, isolated from the RelBE1 complex, by conservation score colouring (Ashkenazy *et al.*, 2016) shows that the highest conserved structural elements are β2 and β3, alongside the loop connecting α2 to β2; essentially the top half of the β sheet core (**Figure 5.11 A**). β1, α1, α2, and β4 have reasonably highly conserved residues interspersed with positions of very low amino acid conservation, potentially indicating a number of positions important for maintaining the tertiary structure alongside some tuning for the specific RelB1 interaction. Considering the positioning of these residues in the context of the RelB1 antitoxin structure, it is apparent that three separate interfaces are created by differing RelE1 amino acid conservation levels (**Figure 5.11 B**). RelB1 α2 appears to align to the highest conserved region of RelE1 at β2 and β3 (magenta to white colouring), which creates a highly conserved surface of several hydrophobic and charged amino acids (those that create part of the catalytic core) (**Figure 5.11 B and C**). RelB1 α3 and α4 interact at RelE1 regions of lower conservation (green to white colouring), created by mainly hydrophobic amino acids (**Figure 5.11 B and C**). This indicates that the RelBE1 interaction is more complex than a set of highly conserved superfamily hydrophobic residues.

5.4.3 ConSurf analysis of the RelB1 antitoxin

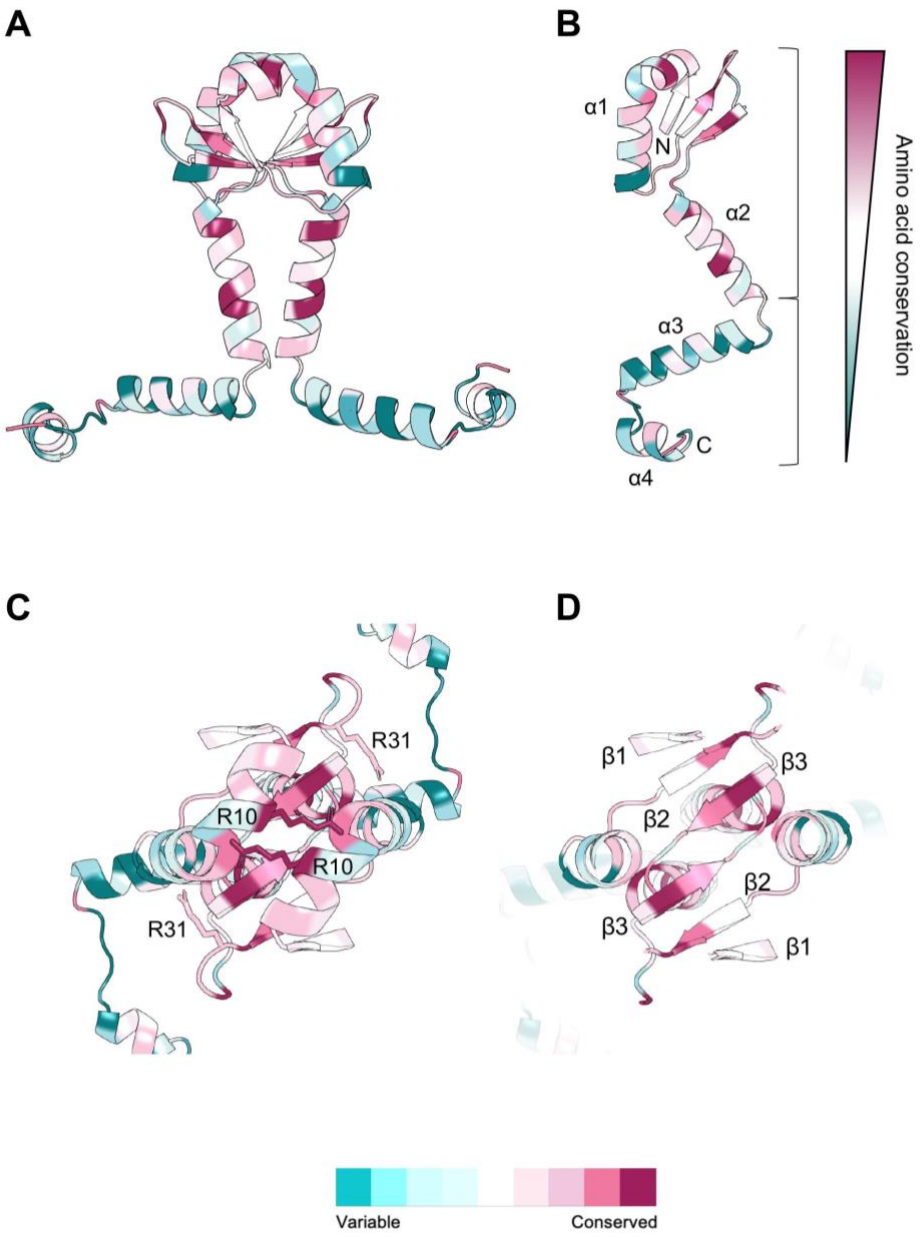
Visualising the isolated RelB1 crystal structure using ConSurf conservation score colouring (Ashkenazy *et al.*, 2016) demonstrates a similar pattern of localised, or concentrated, areas of conservation. The most conserved amino acid sites are almost entirely limited to the N-terminal 'DNA-binding' region (**Figure 5.12 A and Figure 5.7 C**), spanning from the N-terminus to R56 (C-terminus of α2). This leaves α3 and α4 to be distinctly more variable in sequence, indicated by the generally green to white colouring (**Figure 5.12 A**) and creates two domains of varied conservation for the antitoxin (**Figure 5.12 B**). Within the N-terminal domain, likely DNA-binding domain, are the highly conserved R10 and R31 residues (**Figure 5.12 C**), with the overall structure of this domain being maintained by the 6 β strands at the RelB1 dimerisation interface (**Figure 5.12 D and Figure 5.8 A ii**). The C-terminal, likely RelE1-toxin specific region, is comprised of variable amino acids (**Figure 5.12 B**); taken together, this analysis indicates that the N-terminal 'conserved' domain is a structured region, while the two C-terminal alpha helices are unstructured and RelE1 toxin specific, folding on interacting with the corresponding RelE1 interfaces.

Figure 5.11 ConSurf analysis of RelE1



(**A**) Rotated views of the RelE1 monomer cartoon; (**B**) RelBE dimer with RelE as in A and RelB1 recoloured grey with secondary structure labels; (**C**) Surface rendered RelE1 with black boxes indicating RelB1 α helix binding grooves as labelled in B. Cartoon and surface representations of RelE1 from the RelBE1 crystal structure are rendered in ConSurf colouring scheme (shown as a gradient below C) to represent amino acid conservation.

5.12 ConSurf analysis of the RelB1 antitoxin



(**A**) RelB1 dimer cartoon; (**B**) RelB1 monomer cartoon with secondary structure α helices labelled and N and C- terminal labels. A generalised amino acid conservation gradient from high (magenta) to low (teal) is presented alongside; (**C**) View from above the RelB1 N-terminal dimerisation/DNA-binding domain with conserved arginine residues labelled (single letter code R and numbering); (**D**) Visual cut through of C showing the N-terminal dimerised β strands with secondary structure labels. Cartoon and of the isolated RelB1 structures from the RelBE1 crystal structure are rendered in ConSurf colouring scheme (shown as a gradient below C and D) to represent amino acid conservation.

5.4.4 RelBE1 complex interactions are complementary in conservation

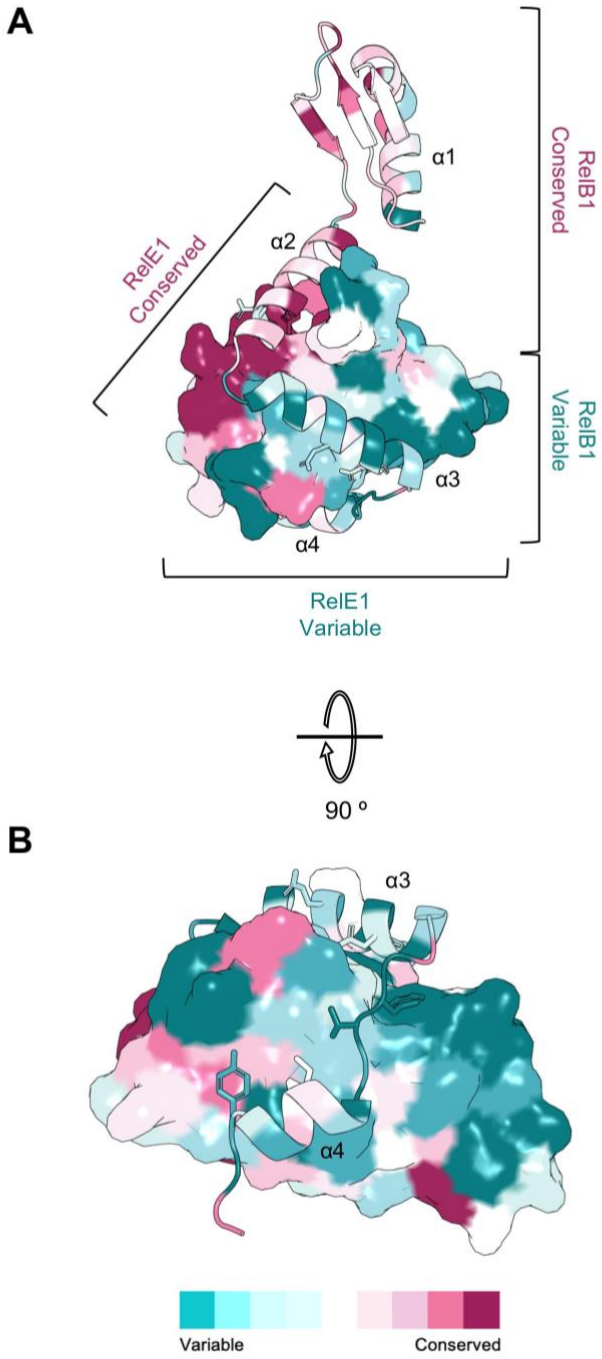
As indicated in **Figures 5.11** and **5.12**, the interactions between RelE1 and RelB1 occur at both conserved and variable positions. Interestingly, the regions of conservation and variability appear to align well in terms of interfacing regions. While the entire RelB1 α2 – α4 region contributed to the essential interface identified for complex formation (**Figure 5.8 A i**), ConSurf allows us to better understand the relationships between the RelE1 and RelB1 amino acid sequences creating the complex.

Both RelB1 and RelE1 can be generalised into two hemispheres of conservation; one being more highly conserved (indicated by concentrated magenta colouring), with the other being more variable (indicated by concentrated green colouring (**Figure 5.13**). In the RelBE1 system dimer, the 'top half' can be considered to be the complementary conserved portion, while the 'bottom half' is the complementary variable portion (**Figure 5.13 A**). In RelE1, the binding groove for RelB1 a2 (**Figure 5.11 C**) is conserved in both superfamily hydrophobicity (**Figure 5.9**) (Anantharaman and Aravind, 2003), and RelE toxin specific catalytic residues (Neubauer *et al.*, 2009); this creates both a hydrophobic and polar microenvironment of high conservation. This is complemented by the highly conserved N-terminal region of RelB1; while the N-terminal DNA-binding motif is well conserved for antitoxin dimerisation and autoregulation, α2 is also conserved to interact specifically with the RelE1 conserved patch (**Figure 5.13 A**). Notably, RelB1 has hijacked the catalytic core amino acids by forming a number of polar contacts and ionic bonds, notably between the highly conserved RelB1 E50 and RelE1 R65.

In contrast to this, while the superfamily RelE hydrophobic antitoxin recognition surface is maintained through the RelB α3 and α4 binding grooves, these amino acids are apparently more variable when considering the RelE ConSurf analysis (**Figure 5.13 B**). Here we find a greater number of tuned hydrophobic interactions between the RelB1 and RelE1 variable regions, with the C-terminus of RelB1 becoming rich in hydrophobic residues (**Figure 5.13 B and Figure 5.10 B**).

Overall, the level of amino acid conservation throughout the RelBE1 dimer generally matches the increasing variability from N to C-terminus of RelB1; this indicates the interactions have become highly complementary in supporting the superfamily recognition surface (Dalton and Crosson, 2010). This slightly contrasts the pattern of complementation seen for the ParDE1 complex; while the ParE1 – C-terminal ParD1 interactions were highly variable, they appeared to be tuned via polar contacts rather than increased hydrophobicity (**Chapter 4**, **Figure 4.10**).

Figure 5.13 RelB1 – RelE1 interaction is specific and tuned



(**A**) Surface rendered RelE1 dimerised with RelB1 cartoon. Brackets indicate localised and generalised levels of conservation for both RelE1 and RelB1; (**B**) Rotated and enhanced view of A. Secondary structure labels for RelB1 α helices are included throughout. Cartoon and surface representations of the RelBE1 dimer from the RelBE1 crystal structure are rendered in ConSurf colouring scheme (shown as a gradient below B) to represent amino acid conservation.

5.5 Biophysical analyses of the *M. tuberculosis* RelBE1 complex

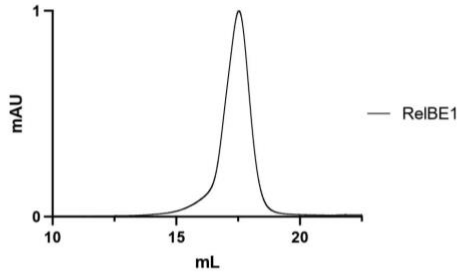
The RelBE1 complex was analysed by SEC, circular dichroism spectroscopy, and thermal denaturation (**Figure 5.14**) as previously shown for the ParDE system complexes (**Figure 3.11 and Figure 3.16**). The RelBE1 complex elutes in a single peak (**Figure 5.14 A**), similar to the ParDE2 complex (**Figure 3.16 A**), without the observation of multiple complex peaks as seen for ParDE1 (**Figure 3.11 A**). This indicates the RelBE1 complex is stable in solution which is further supported by thermal denaturation experiments; the complex did not reach its melting temperature, even when temperature increased to 80 °C (**Figure 5.14 B**).

CD scans taken before thermal denaturation (**Figure 5.14 C**, solid black) indicate a high degree of helicity in the structure which is supported by the crystal structure (**Figure 5.5**). Unsurprisingly, as T_m was not reached, CD scans taken after thermal denaturation indicate the retention of a high degree of the largely helical structure (**Figure 5.14 C**, dashed black).

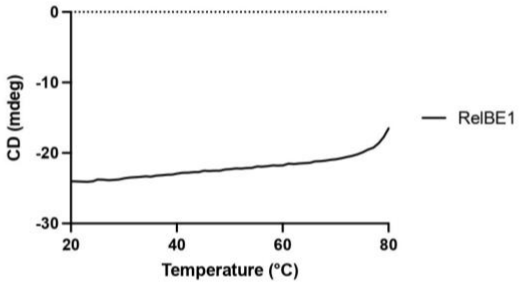
Using the elution volume for RelBE1 obtained in analytical SEC (**Figure 5.14 A**), we can generate observed molecular weights (M_r) and Stokes radii (R_{st}) values in kilodaltons (kDa) and angstroms (Å), respectively (**Figure 5.14 D**). Previous experiments have shown the R_{st} predictions to be more accurate, indicating that the complexes studied thus far are less globular in their overall structure. The RelBE1 complex molecular weight was calculated to be 41.14 kDa as per the 2:2 stoichiometry observed in the crystal structure (**Figure 5.5**), and the observed M_r for the complex from SEC was only 1.06 times this, at 43.45 kDa (**Figure 5.14 D**). This indicates that, unlike the ParDE TA systems, RelBE1 runs close to true to its molecular weight via SEC. Unsurprisingly, when comparing the R_{st} for the crystal structure of RelBE1 (28.90 Å) to the observed value from SEC (27.75 Å), we see little discrepancy with a ratio of 0.96 (**Figure 5.14 D**). For RelBE1, both M_r and R_{st} values support that the RelBE1 complex exists in solution as the heterotetramer seen in the crystal structure (**Figure 5.5**). Further to this, we have shown that the complex is incredibly thermostable up to temperatures of 80 °C, indicating a very strong interaction between the toxin and antitoxin.

Figure 5.14 Biophysical studies of the RelBE1 complex

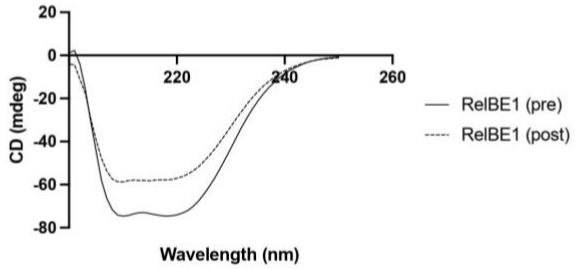
A



B



C



D

Sample	Calculated M_r (kDa)	Observed M_r (kDa)	Observed M_r/ Calculated M_r	Calculated R_{st} (Å)	Observed R_{st} (Å)	Observed R_{st}/ Calculated R_{st}
RelBE1	41.14	43.45	1.06	28.90	27.75	0.96

(**A**) Analytical SEC trace for RelBE1; (**B**) Thermal denaturation curve for RelBE1; (**C**) Circular dichroism spectroscopy scans for RelBE1 before (solid black, pre) and after (dashed black, post) melting; (**D**) Table of RelBE1 protein molecular weight (M_r) and Stokes Radius (R_{st}) calculations, observations, and comparisons. R_{st} value was generated from the RelBE1 crystal structure using HullRad (Fleming and Fleming, 2018). Comparison of observed/calculated is coloured green if within 10 % of the predicted ratio, yellow if > 10 ≥ 25 %, and red if > 25 %. Chromatograms are representative of duplicate data and are normalised between 0 – 1 for presentation and comparison. Graphs are cropped to the appropriate scale (10 – 22.5 mL).

5.6 RelE toxin structural conservation

From the analyses so far, we have seen that the sequence of RelE toxins is apparently highly conserved through the core of the protein, maintaining a mechanism of interaction and residues required for toxin function. Variable sequence regions exist within the toxin in order to tune the interactions to the cognate RelB1 toxin. Both conserved and variable regions must maintain a common tertiary structure to the RelE family in order to be ribosome-specific and catalytically active (Bøggild *et al.,* 2012; Griffin, Davis and Strobel, 2013).

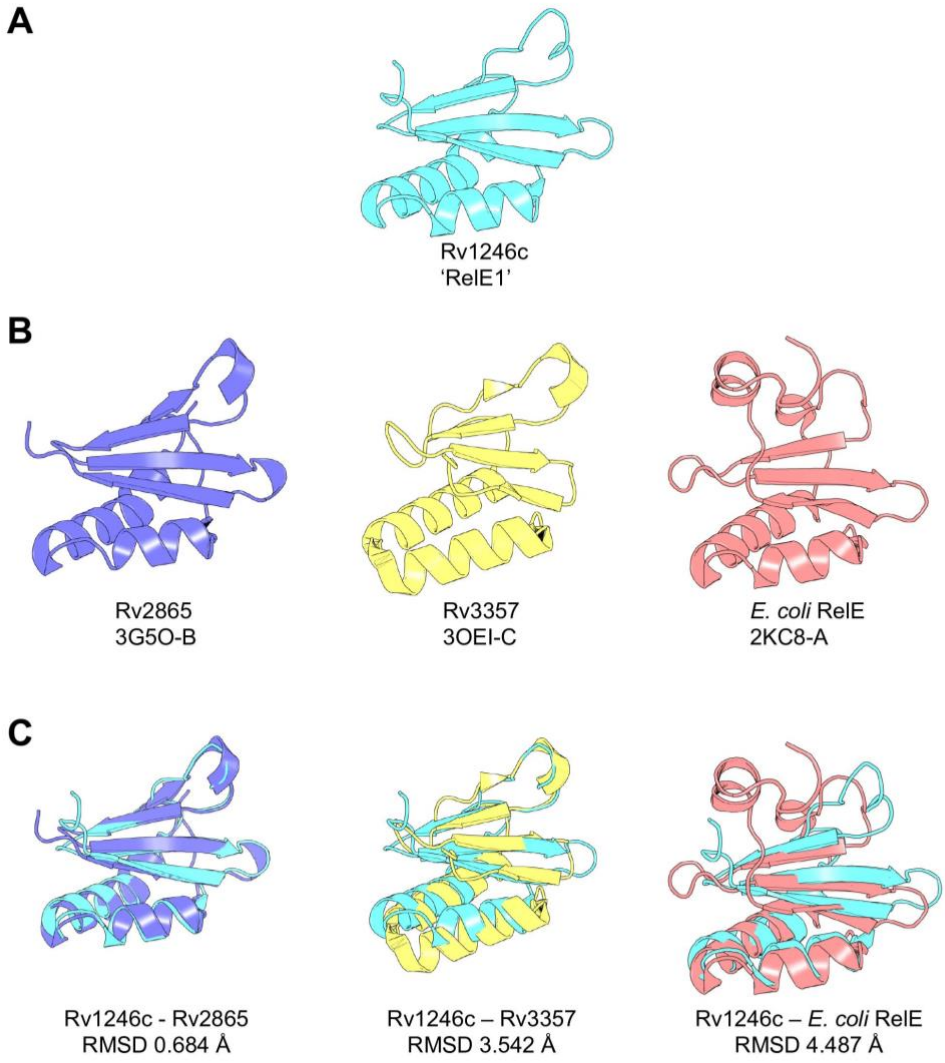
5.6.1 RelE toxins have a conserved tertiary globular structure

Communications with the Genevaux group, stated previously, indicated that the RelE1 toxin does not show activity *in vitro* during ribosome-dependent mRNase assays (but is toxic *in vivo*), while the *E. coli* RelE toxin acted as a positive control demonstrating activity. Comparing the structures of the *M. tuberculosis* and *E. coli* RelE toxins may provide insights as to why there is a lack of typical RelE1 activity in the standard RelE biochemical toxicity assay.

M. tuberculosis RelBE systems have an overall similar complex quaternary structure sharing the RelB1 antitoxin structure and mechanism of complexing. The RelE1 (Rv1246c) toxin (**Figure 5.15 A**) and toxins from the two previously solved RelBE systems, Rv2865 (RelE2, PDB: 3G5O-B) and Rv3357 (RelE3, PDB: 3OEI-C) (**Figure 5.15 B**), possess a visibly similar tertiary structure; Rv2865 was used as the MR search model. Performing a sequence-independent structural alignment using PyMol we can assess the degree of tertiary structure conservation across these toxins irrespective of primary sequence. RelE1 and RelE2 are highly similar, returning an RMSD of 0.684 Å, with RelE1 and RelE3 still being reasonably similar, returning an RMSD of 3.542 Å. All three *M. tuberculosis* RelE toxins present the same α helical hairpin and β sheet core as is common to the RelE/ParE superfamily (Gucinski *et al.*, 2019), with minor variations in α helix and β strand length. It is worth noting the RelE C-terminal residues were unresolved in each of the three *M. tuberculosis* RelBE crystal structures.

This contrasts the structural characterisation of the *E. coli* RelE toxin (**Figure 5.15 B**) in isolation from the complex (PDB: 2KC8-A) (Li *et al.*, 2009). This toxin structure presents the C-terminal residues as an α helix bent back over the β sheet core. Sequence-independent alignment with RelE1 returns a respectable RMSD of 4.487 Å; the tertiary structure remains consistent however, with the addition of the *E. coli* RelE C-terminal α helix (**Figure 5.15 C**).

Figure 5.15 Structural conservation of RelE toxins



(**A**) *M. tuberculosis* RelE1 (cyan) from the RelBE1 crystal structure; (**B**) *M. tuberculosis* RelE2 (slate blue), *M. tuberculosis* RelE3 (yellow), *E. coli* RelE1 (salmon); (**C**) Sequence-based alignments of *M. tuberculosis* RelE1 against the structures directly above in B. RMSD values calculated in PyMol are presented below each alignment. *M. tuberculosis* proteins are labelled with their respective 'Rv' identifier.

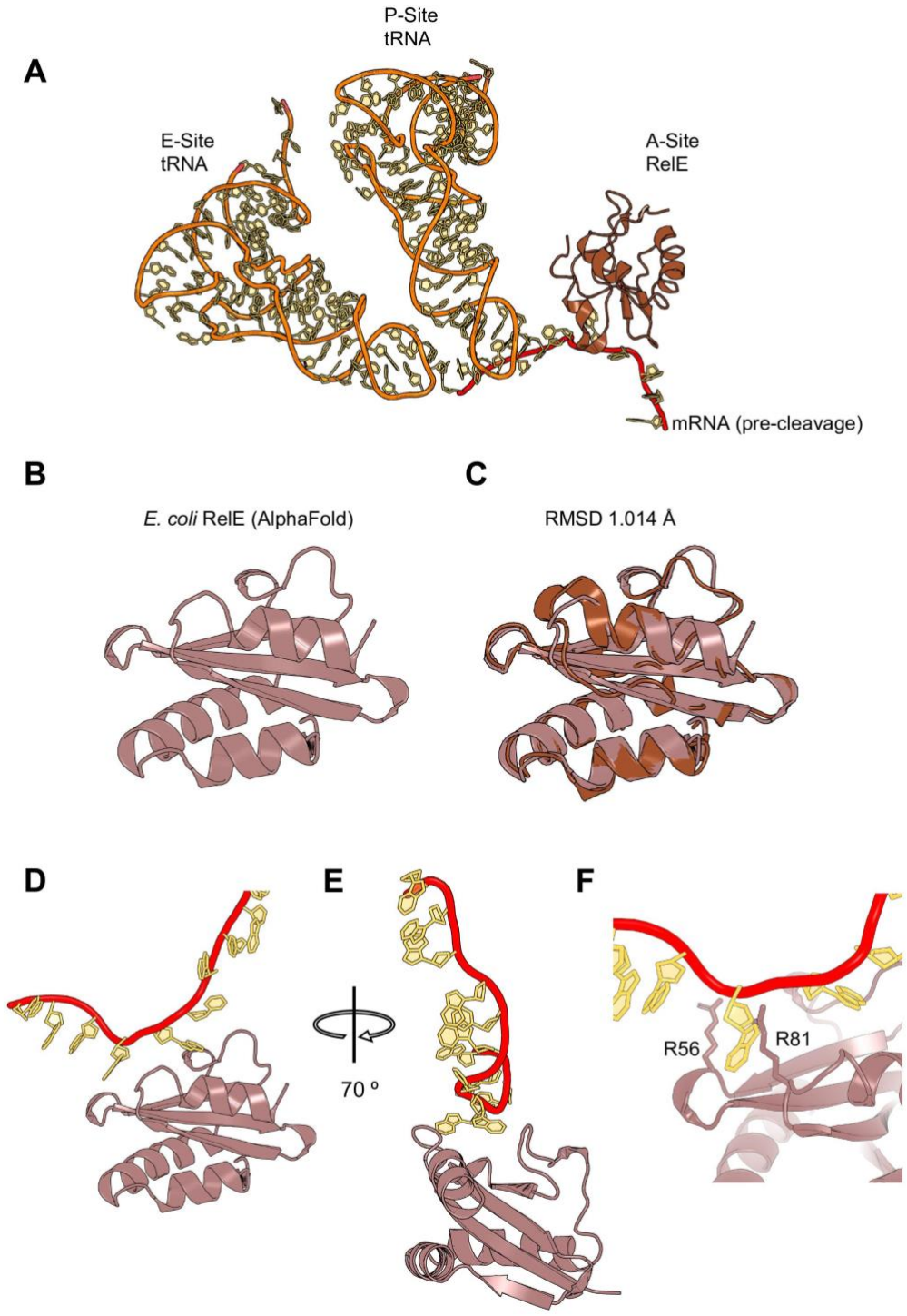
5.5.2 RelE toxins are ribosome-dependent and occupy the A-site

Crystallographic studies of the *E. coli* RelE1 toxin (Neubauer *et al.*, 2009) elucidated the mechanism by which mRNA cleavage occurs (**Figure 5.16**). Both wild-type and a catalytically inactive mutant of RelE1 were crystallised in complex with *T. thermophilus* 70S ribosomes and clearly demonstrate that RelE occupies the A-site of the ribosome, preventing the recruitment of subsequent aminoacyl-tRNA while also positioning to cleave mRNA (catalytically inactive RelE caused no mRNA cleavage) (**Figure 5.16 A**). The *E. coli* RelE – ribosome interaction is stabilised by contacts between RelE α1 R10, K13, K17, α2 K28 and K29, and α3 R93 with 16S rRNA (Neubauer *et al.*, 2009).

Both active and inactive RelE structures from Neubauer *et al.* (2009) can be replaced with the AlphaFold model (**Figure 5.16 B**) available from the online database (Varadi *et al.*, 2022) to not only compare the positioning of essential residue side chains (otherwise not present in the inactive mutant), but also to complete the RelE structure, as the crystal structure is of good resolution, but many α helices and β sheets are not fully resolved (**Figure 5.16 A**). Sequence-based structural alignment of the *E. coli* AlphaFold RelE1 with the crystal structure RelE1 returned an RMSD of 1.014 Å (**Figure 5.16 C**). This indicates a high degree of structural similarity with key structural components positioned extremely well (α helical hairpin, β sheet core, C-terminal helix). Subsequent positioning of the AlphaFold RelE1 model into the A-site via alignment to the crystal structure RelE1 allows us to visualise the importance of the C-terminal helix position and potentially the R45 and R81 residues mutated to obtain the pre-cleavage structure (**Figure 5.16 D and E**).

RelE (and potentially ParE) proteins re-organise the positioning of the C-terminal residues (Bøggild *et al.,* 2012) into this helix, likely 'switching on' the toxin via repositioning key residues within the RelE catalytic core. Essential RelE residues, such as the basic R56 and R81, are positioned toward the mRNA in the pre-cleavage structure, with Y87 stabilising the orientation via base stacking with the substrate (PDB: 4V7J) (Neubauer *et al.*, 2009) at the cleavage site (**Figure 5.16 F**). Understanding this mechanism of action and demonstrating the positioning of key structural features, such as the C-terminal helix, in the *E. coli* RelE1 toxin allows us to compare the *M. tuberculosis* RelE1 structure and potentially highlight key differences rendering the toxin catalytically inactive.

Figure 5.16 *E. coli* RelE1 occupies the A-Site of the ribosome



(**A**) *E. coli* RelE1$^{(R45A/R81A)}$ occupying the ribosomal A site with pre-cleavage mRNA; (B) AlphaFold model of the E. coli RelE1 toxin; (C) Sequence-based alignment of the RelE1 AlphaFold (light brown) model in B with the RelE1 crystal structure (brown) from A. RMSD value calculated in PyMol is presented above; (**D** and **E**) AlphaFold RelE1 – mRNA interaction created via alignment with RelE1 from A and editing to isolate the RelE1 – mRNA complex; (**F**) Coordination of arginine residues from AlphaFold RelE1 (notably R81 – mutated in A for pre-cleavage structure) toward mRNA.

5.6.3 The RelE catalytic core

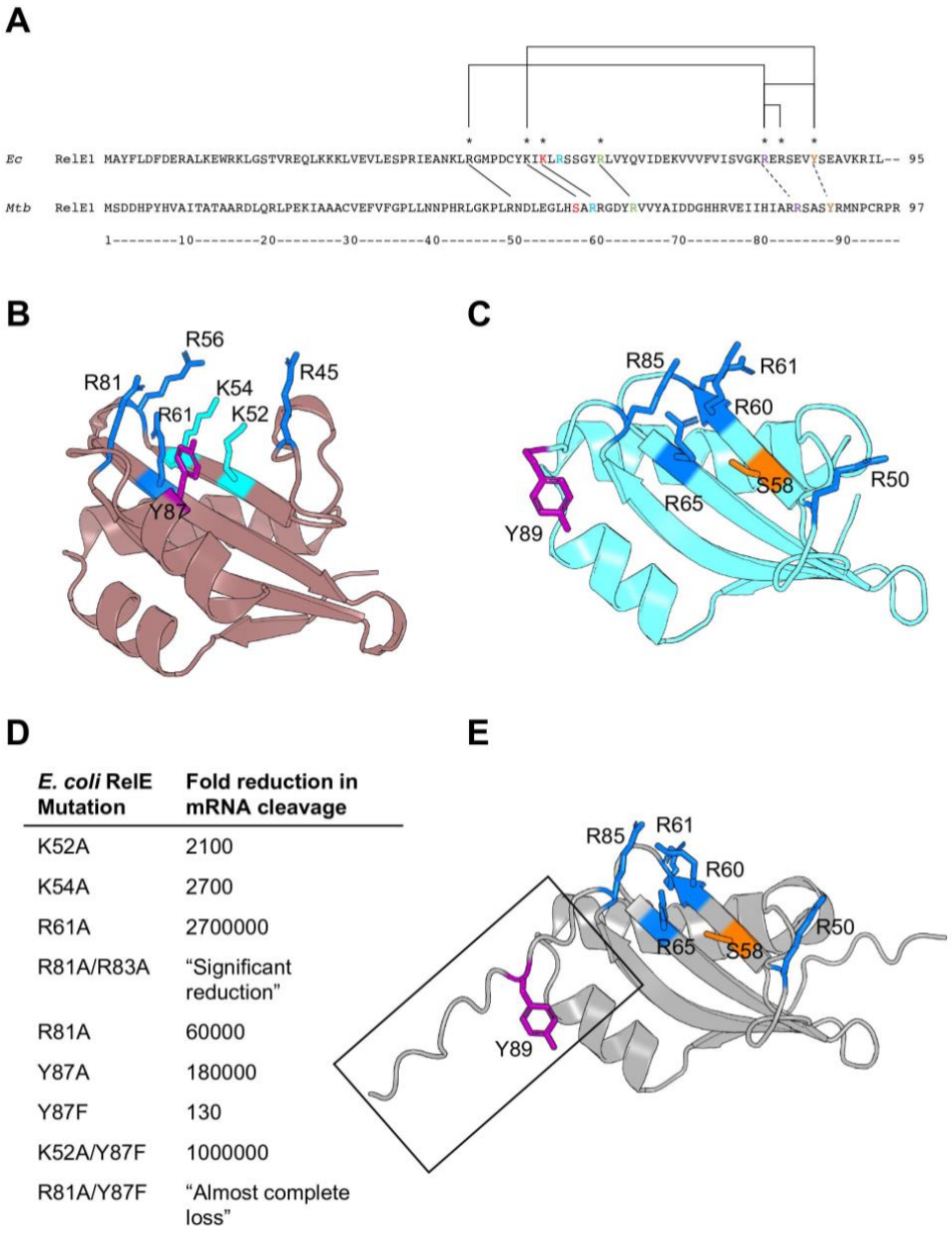
As previously shown via the ConSurf generated MSA (**Figure 5.10 A**), several of the most conserved amino acids in RelE toxin sequences are charged, align to the β sheet core of the secondary structure, and have previously been shown to be essential for catalytic activity (Griffin, Davis and Strobel, 2013).

As the tertiary fold of the RelE1 toxin appears to be similar to that of not only the other Mycobacterial RelE toxins, but also to the catalytically active *E. coli* RelE toxin (**Figure 5.15 C**), it is essential to further investigate the composition of the catalytic core. Using a combination of AlphaFold models (Varadi *et al.*, 2022), sequence alignment and sequence-independent structure alignment, we can demonstrate several differences between the catalytically active *E. coli* RelE core and the catalytically inactive (personal correspondence) *M. tuberculosis* RelE1 core (**Figure 5.17**).

Using the AlphaFold model for the *E. coli* RelE (validated through high similarity in alignment to solved structured) allows for the full structure to be present in the alignment. Comparison of the positions of catalytic residues is summarised in **Figure 5.17 A**, whereby essential *E. coli* residues are highlighted, coloured for alignment via MSA, and directly connected via lines for structural alignment. This reveals certain 'mis-matches': *E. coli* R45 and *M. tuberculosis* R50 are structurally aligned but not in sequence; *E. coli* K52 structurally aligns to *M. tuberculosis* S58, removing a basic residue; *E. coli* K54 aligns to the *M. tuberculosis* S58 in sequence, but to R60 in the crystal structure, while *M. tuberculosis* R60 aligns in sequence to *E. coli* R56 (which is apparently not important catalytically) (**Figure 5.17 A**). Both *E. coli* R61 and R81 align in sequence to *M. tuberculosis* R65 and R85 respectively, these amino acids also have the potential to align in structure as they are in the flexible C-terminal region of *M. tuberculosis* RelE1.

The important catalytic residues can be highlighted in the *E. coli* AlphaFold model, whereby a core of basic arginine and lysine residues is evident, supplemented by Y87 (**Figure 5.17 B**). This core catalytic structure is largely present in the *M. tuberculosis* RelE1 structure, with the substitution of the K52 position (*E. coli*) for S58, and the positions of R85/Y89 are flexible (**Figure 5.17 C**), likely due to the presence of the RelB1 antitoxin in the crystal displacing the C-terminal helix. The importance of the catalytic residues, investigated by mutagenesis, is summarised in **Figure 5.17 D** showing that *E. coli* R61, R81, K52, and Y87 in combinations have significant roles in mRNA cleavage rates. While these amino acids likely align with R65, R85, S58, and Y89 of the *M. tuberculosis* RelE1 catalytic core (**Figure 5.17 C**), the corresponding RelE1 AlphaFold model indicates that the flexible C-terminal region may not form a helix that reorganises across the β sheet core (**Figure 5.17 E**). This may explain some of the loss of catalytic activity as the highly important R85/Y89 combination may not be correctly positioned (**Figure 5.17 C – E**).

5.17 RelE catalytic core comparison



E. coli RelE Mutation	Fold reduction in mRNA cleavage
K52A	2100
K54A	2700
R61A	2700000
R81A/R83A	"Significant reduction"
R81A	60000
Y87A	180000
Y87F	130
K52A/Y87F	1000000
R81A/Y87F	"Almost complete loss"

(**A**) Structure-based sequence comparisons of *E. coli* (*Ec*) and *M. tuberculosis* (*Mtb*) RelE1 toxins. * *E. coli* residues that have been mutated to interfere with catalytic activity, connecting brackets indicate mutated pairs; Matched colouring of residues between sequences indicates sequence-based alignment; Solid inter-sequence connecting lines indicate structurally aligned residues; Dashed inter-sequence connecting lines indicate residues with the potential to structurally align; (**B**) AlphaFold *E. coli* RelE1 with experimentally validated catalytic residues highlighted and side-chains displayed; (**C**) *M. tuberculosis* RelE1 with both sequence and structurally conserved catalytic core residues highlighted; (**D**) Summary of experimentally validated catalytic mutations to *E. coli* RelE1; (**E**) AlphaFold *M. tuberculosis* RelE1 model with residues highlighted as per C. Black box indicates the unstructured and flexible C-terminal residues. Arginine (R – Blue); Lysine (K – Cyan); Tyrosine (Y – Purple); Serine (S – Orange).

5.7 RelE/ParE superfamily structure comparison

The comparisons of the RelE catalytic core to the crystal structure and AlphaFold model of *M. tuberculosis* RelE1 highlighted only a few differences (**Figure 5.17**) that may cause the lack of ribosome-dependent ribonuclease activity (personal correspondence, Genevaux group). Expanding this study of the catalytic core to the entire mycobacterial RelE/ParE toxin complement may not only help us to direct future studies on the mechanism of toxicity of the RelE1 toxin, but also to help us better understand the phylogenetic relationships between the systems initially demonstrated in **Figures 5.1 – 5.4**. This series of comparisons may also allow us to explain the biochemical results obtained earlier for the ParE toxins (**Chapter 3**), alongside developing theories surrounding the peculiar phylogenetic positioning of the ParE2 toxin and its clear relationship to the RelE systems (**Figure 5.4**).

Comparisons were performed using the now validated and highly accurate AlphaFold models (**Figure 5.18**) (Varadi *et al.*, 2022) of each toxin as the entire sequence is present, including the regularly unresolved C-terminal toxin residues in crystallographic structures. For clarity, the RMSD values obtained on sequence-based alignments of each crystal structure to its respective AlphaFold model are as follow (RMSD values in parentheses); ParE1 (0.673 Å), ParE2 (0.548 Å), RelE1 (0.540 Å), RelE2 (0.382 Å), RelE3 (0.254 Å). Structural homology comparisons were made between the toxins themselves and to the experimentally validated and well define catalytic core of *E. coli* RelE (**Figure 5.18 A**), as previously shown for *M. tuberculosis* RelE1 (**Figure 5.17**). Sequence-independent superpositions of the toxins reveals overall structure is generally very well conserved (as is expected for the superfamily), with the closest structural homologues being RelE1/RelE2 (RMSD 0.445 Å), *E. coli* RelE/RelE3 (RMSD 1.114 Å), and surprisingly ParE1/RelE2 (RMSD 1.233 Å) (**Figure 5.18 Å**). The least related appeared to be the ParE2 toxin when compared to the *M. tuberculosis* systems and the *E. coli* RelE toxin (**Figure 5.18 A**).

Given the generally high structural similarity between the toxins the positioning of catalytic residues was analysed, using the validated *E. coli* residues as a scaffold. From this, we see generally consistent conservation, and interestingly substitutions, between the *E. coli* RelE and *M. tuberculosis* RelE toxins (**Figure 5.18 B**). Notably, the *E. coli* K52 residue is substituted for serine, and the K54 position is substituted for arginine throughout the mycobacterial RelE toxins. The R61, R81, and Y87 positions are well conserved throughout, however RelE1 Y89 and RelE3 R82/Y85 appear is if they could structurally align and are simply in a flexible portion (**Figure 5.18 B**). This feature was previously shown for RelE1 in **Figure 5.17 E**. Unsurprisingly this catalytic core structure is almost completely lost in ParE1, as a distantly related toxin belonging to the ParE clade (**Figure 5.4**); the only residue that is structurally conserved is ParE1 K64, which aligns to *E. coli* RelE K52. When comparing the catalytic core residues of *E. coli* RelE to ParE2, we also see

very little conservation (**Figure 5.18 B**), only R65 aligns to *E. coli* RelE R56, and this residue only weakly contributes to ribonuclease activity (Neubauer *et al.*, 2009). This initially indicates that the ParE toxins are distinct from the RelE toxins, notably ParE2 which is the closer relative. However, the substitutions present in ParE2 at K52 (H62), K54 (Y63), and R61 (Y70) are all residues capable of having roles in the catalytic core of ribonuclease enzymes (Yang, 2011). The substitutions present at R81/Y87 (M89/P92), arguably the positions of highest catalytic importance, are not capable of fulfilling the same roles in the ribosome-dependent ribonuclease mechanism. Taken together with the phylogenetic data (**Figure 5.4**), this suggests that the ParE2 toxin is indeed distinct from not only the ParE1 toxin, but potentially also the RelE toxins and could possess a different, as yet unknown, catalytic mechanism, if any at all. Given the biochemical data previously presented (**Chapter 3**, **Figure 3.5**), this could tentatively indicate ParE2 to be a gyrase-dependent nuclease.

These distinct differences in the sequence and structure of ParE2, and indeed the similarities in the RelE toxins, are further exemplified through the RelE/ParE superfamily alignment (**Figure 5.18 C**). This alignment was used in generating the phylogenetic tree in **Figure 5.3**. Initially, we can see the highly conserved hydrophobic patches (amino acids coloured blue), however, combining the MSA with structural alignment comparisons (**Figure 5.18 B**) allows us to highlight the conserved/substituted positions between sequence and structure (**Figure 5.18 C**, red boxes). On inspecting the entire MSA, RelE sequences appeared to be divided into three sub-types at the *E. coli* K52 aligned position; one subset presented either arginine or lysine at this position (although arginine was far more frequent), the other two subsets had this substituted for serine as is the case for the mycobacterial RelE toxins. However, those that contained serine were either followed by an arginine residue, like in RelE3, or by an alanine residue, like in RelE1 and RelE2. These subsets were around equal in abundance in the total sequences analysed. Whether or not serine can functionally substitute in for the active proton acceptor at this position is yet to be discussed, however, for those with serine - arginine may be able to with the arginine simply fulfilling this role (**Figure 5.18 C**).

The alignment also allows us to visualise the large variation in C-terminal sequence length, essentially the residues following the conserved hydrophobic patch (**Figure 5.18 C**). Evidently, this is also an area of high sequence variation, however, there is a generally conserved hydrophobic amino acid that may assist in reforming the C-terminal helix. This process is likely essential for RelE activity as it positions the C-terminal tyrosine (*E. coli* Y87) to interact with substrate mRNA in the ribosome. ParE1 also maintains a hydrophobic residue aligning to this tyrosine, while ParE2 has V95 and I99 that appear to assist in the formation of the predicted C-terminal helix (neither align to Y87 in sequence or structure). The exact role of reforming a C-

terminal helix in ParE toxins is not yet understood; this may be involved in creating a more compact structure to permit interaction with the target, DNA gyrase. In the case of ParE2 this may actually be involved in creating a catalytic core, however, this is highly speculative at this point. Importantly, removal of the C-terminal helix does render the ParE2 toxin inactive against gyrase (Gupta *et al.*, 2017).

Figure 5.18 RelE/ParE structure and sequence comparisons

A

AlphaFold	ParE1	ParE2	RelE1	RelE2	RelE3
ParE1					
ParE2	1.844 (6.361)				
RelE1	1.615 (2.543)	2.832 (9.694)			
RelE2	1.233 (2.638)	2.959 (1.187)*	0.445 (0.446)		
RelE3	1.374 (8.506)	1.617 (0.837)**	1.293 (5.121)	1.327 (3.021)	
E. coli RelE	1.933 (6.330)	4.301 (3.965)	1.578 (3.665)	1.544 (6.300)	1.114 (2.545)

B

E. coli	***M. tuberculosis***				
RelE	RelE1	RelE2	RelE3	ParE1	ParE2
K52	S58	S54	S58	K64	H62
K54	R60	R56	R60	S66	Y63
R56	(R61) G62	(R57) G58	D63	G68	R65
R61	R65	R61	R66	T71	Y70
R81	R85	R81	R82*	Q90 (R91)	M89
Y87	Y89*	Y85	Y85*	V94	P92

C

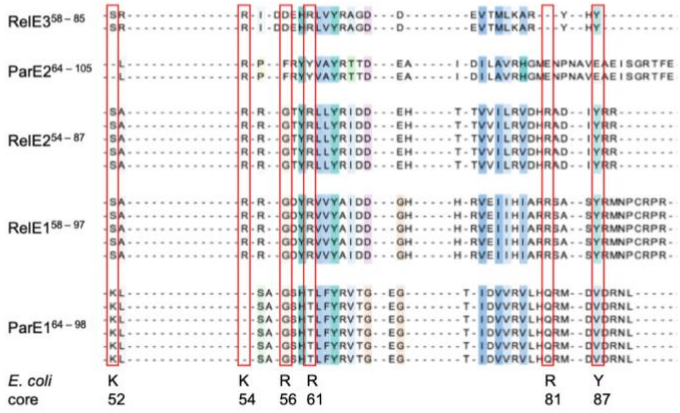


Figure legend overleaf

Figure 5.18 RelE/ParE structure and sequence comparisons

(**A**) Summary of ParE and RelE toxin structural comparison via sequence-independent superposition. AlphaFold models were used in the analysis and values are RMSD in Ångstroms (Å). Values in parentheses are for sequence-dependent alignments. *120 atoms only used in alignment, **106 atoms only used in alignment; (**B**) Comparison of residues aligned to the catalytic core positions of the reference structure, *E. coli* RelE. *Residue in flexible position and may align, residues in parentheses indicate neighbouring amino acid of appropriate species; (**C**) Cropped MSA of RelE/ParE superfamily to show C-terminal portion of the RelE/ParE toxins of interest. Catalytic core aligning positions are highlighted by red boxes. Alignment was generated in Clutal Omega and amino acids are coloured in the Clustal colour scheme. Mycobacterial RelE/ParE are labelled with appropriate amino acid numbering for presented sequences *E. coli* reference structure catalytic residues are presented below. Amino acids are represented by their single-letter code, throughout.

5.8 Discussion

To develop our hypotheses regarding the observed lack of cleavage/degradation type toxicity by the Mtb RelE toxin, we have considered its sequence and structure as part of the RelE/ParE superfamily (Anantharaman and Aravind, 2003). Phylogenetic analyses of the ParE (**Figure 5.1**) and RelE (**Figure 5.2**) families initially highlight the general levels of sequence conservation within the respective families; the presence of a catalytic core in RelE toxins adds a higher degree of sequence conservation within this family of toxins. The ParE1, RelE1 and RelE2 toxins are well conserved throughout Mycobacteria, while the ParE2 and RelE3 toxins only have single orthologues in Mycobacteria. Whether this is an indication of importance in virulence due to the association of TA systems with *M. tuberculosis* pathogenicity (Ramage, Connolly and Cox, 2009; Sala, Bordes and Genevaux, 2014) requires further study. Our methods of phylogenetic analysis are largely limited by our method of sequence identification, using TADB 2.0 (Xie et al., 2018); this heavily relies on the accuracy and maintenance of the database and does not include many sequences. In future, expansion of our MSA would prove incredibly useful in better positioning individual sequences.

In comparing the RelE/ParE superfamily with and without the RNase T1 rooting (**Figure 5.3** and **Figure 5.4**), we highlight that ParE2 is distinct in its protein sequence. Despite our biochemical analyses confirming that ParE2 is indeed a gyrase-targeting toxin (**Chapter 3**, **Figure 5.5**), the protein sequence is better aligned to the RelE family (**Figure 5.3 and Figure 5.4**). Interestingly, while the toxin does appear to cluster with the RelE clades, it appears as an early divergent from the group (**Figure 5.4**). Further analyses of the sequences highlight that the RelE canonical catalytic core (Neubauer *et al.*, 2009; Griffin, Davis and Strobel, 2013) is not present, at least in full, in the ParE2 sequence (**Figure 5.18**). Here, we identify a potential limitation in our methods; the RelE family can be further divided into several toxin groups with distinct catalytic mechanism, while maintaining ribosome specificity (Liang *et al.*, 2014). It may, therefore, be important to diversify the MSA beyond just the RelE toxins and include sequences such as YafQ, HigB, and YoeB. This may create distinct clades depending on conserved residues at the catalytic core, to which ParE2 may be better aligned. Alternatively, it may cause ParE2 to become distinct from the RelE family altogether. Further to this, it is important to highlight that analyses presented here are dependent on RNase T1 being the true ancestor of the ParE/RelE superfamily, however, this is generally accepted to be the case due to structural homology and good alignment of catalytic residues in the catalytic sub-families (Overgaard *et al.,* 2008; Neubauer *et al.,* 2009; Bøggild *et al.,* 2012; Bechhofer and Deutscher, 2019). Altogether, the peculiar evolutionary relationship of ParE2, alongside the biochemistry and structural biology presented in **Chapters 3 and 4**, respectively, highlights the need for further study. ParE2 may indeed have catalytic capabilities,

albeit with the altered target of DNA gyrase; further structural work and biochemistry, including testing the ParE2 toxin in ribosome cleavage assays, will assist us in better understanding this TA system.

The RelE1 and RelE2 toxins have a high degree of sequence similarity and a close evolutionary relationship (**Figure 5.2**). The initial purpose of these analyses was to highlight evolutionary and structural differences that might explain the lack of ribosome-dependent mRNase activity seen for the RelE1 toxin (personal correspondence, Genevaux lab). Superfamily analysis initially indicated that the RelE1 toxin is indeed a RelE toxin by sequence, however, it appears to share a closer common ancestor with the ParE toxins than the majority of the RelE sequences analysed (**Figure 5.3 and Figure 5.4**). Structural studies of the RelBE1 system assisted in highlighting potentially important differences at the catalytic core of RelE1 (**Figure 5.17**). As previously mentioned, many RelE family core structures exist, however RelE1 aligns well with the canonical RelE catalytic core as seen in *E. coli* with K52, K54, R61, R81, and Y87 (Neubauer *et al.*, 2009), however, there are subtle differences (**Figure 5.17 and Figure 5.18**).

Importantly, the mechanism of *E. coli* RelE induced mRNA cleavage involves Y87 base stacking with the second nucleotide in the A-site to correctly position the substrate mRNA for nucleophilic attack; the positively charged microenvironment activates K52 as a base, while K54 and R61 stabilise the transition state and R81 acts as an acid (Griffin, Davis and Strobel, 2013). We highlight that K52, acting as the proton acceptor in the reaction, is substituted for S58 in RelE1 (and interestingly in RelE2 and RelE3, indicating this is a conserved change in mycobacterial RelE toxins). Mutational analyses of the *E. coli* RelE toxin has shown K52 to be largely dispensable for catalytic activity, albeit rate-reducing (Neubauer *et al.*, 2009; Griffin, Davis and Strobel, 2013); this is potentially due to R61 and K54 compensating for the loss of a basic residue. For S58 to act as the proton acceptor it must initially be deprotonated or can alternatively be phosphorylated as a mechanism of catalytic control (Gone *et al.*, 2016). Phosphorylation of the toxin adds another layer of control, which is emerging as a theme in some TA systems (Cai *et al.*, 2020; Jurėnas and van Melderen, 2020; Yu *et al.*, 2020); given that the RelE1 toxin has been shown to inhibit the growth of not only *M. tuberculosis*, but also *M. smegmatis* and *E. coli*, if phosphorylation of S58 were to be an essential step for activity, the kinase responsible must be highly conserved across bacteria and not part of an organism specific pathway or system. Further to this, through our crystal structure (**Figure 5.5**) and AlphaFold models (**Figure 5.17 C**), we have highlighted that the C-terminal residues of RelE1 may not form an alpha helix as is expected for a RelE toxin. The refolding of this motif has been previously proposed to be essential for toxicity as it positions the tyrosine residue in the RelE active site for interaction with, and orientation of, the mRNA substrate (Neubauer *et al.,* 2009; Bøggild *et al.,* 2012). Refolding may also contribute to forming

a compact fold permitting interaction with the ribosome A-site. The C-terminal residues (**Figure 5.17 C and Figure 5.18 C**) are routinely displaced when the RelB antitoxin is bound, and hence are usually unstructured in RelBE complex structures (**Figure 5.5**). If these residues do not subsequently refold to position catalytic core residues, RelE1 may not have ribosome-dependent enzymatic activity. This would mean that RelE1 may have an altered target, interact with the ribosome at a different site or via a novel mechanism. Interestingly, we see sequence variability in α1 and α2 of the *M. tuberculosis* RelE toxins (indicated in **Figure 5.11 A**) and especially in the positioning of the 16S rRNA interacting basic residues highlighted in the *E. coli* structure (Neubauer *et al.*, 2009); this may indicate that these contacts are not conserved across RelE toxins, or support that *M. tuberculosis* RelE1 has altered target specificity. Together, our structural and phylogenetic analyses of RelE1 indicate that S58 is a potential residue of interest, along with the potential that the C-terminal residues may not form an alpha helix to correctly reposition the highly important Y89 residue for substrate mRNA orientation. Sequence variability in stabilising basic amino acids on the RelE alpha helix hairpin may also result in target variation, or a slightly different mechanism of stabilising the RelE-target interaction.

The structural work across **Chapter 4** (ParDE) and here, in **Chapter 5** (RelBE), highlights that antitoxin chains generally demonstrate decreasing levels of conservation from N – C termini (**Chapter 4, Figure 4.8 and Chapter 5, Figure 5.12**). This is due to the need for a stable N-terminal antitoxin dimerisation domain in many type II TA systems, in most systems for the purposes of autoregulation (Chan, Espinosa and Yeo, 2016). C-terminal domains of these antitoxins are all likely unstructured and transition from disorder to order to cognate toxin binding (Garcia-Pino *et al.*, 2010; Chan, Espinosa and Yeo, 2016; Garcia-Rodriguez *et al.,* 2021). We demonstrate here, for RelB, that these C-terminal domains are toxin specific through a region of less conserved amino acids that are tuned to their cognate toxin for both hydrophobic and ionic interactions (**Figure 5.13**). Contrasting what we have shown for ParDE1, whereby these interactions are hydrophobic across the conserved superfamily surface and tuned ionic at the antitoxin C-terminus (**Chapter 4, Figure 4.10**), the RelBE system demonstrates increased ionic interaction across the RelE beta sheet core (hijacking the catalytic core residues) and increased tuned hydrophobic interactions at the RelB C-terminus (**Figure 5.13**). The N-terminal RelB1 dimerisation domain (**Figure 5.8 A ii**) is likely important in a conditionally cooperative autoregulatory mechanism in a similar manner to that seen for the Phd-Doc system (PDB: 3K33, (Garcia-Pino *et al.,* 2010). This N-terminal structure is well conserved across a number of different TA system antitoxin such as YefM (Nieto *et al.*, 2007; Kumar *et al.*, 2008) and Phd (Garcia-Pino *et al.,* 2010), similar to the widespread dissemination of the CopG domain in antitoxins (**Figure 1.7**). Together with the structural data on ParDE system structures (**Chapter 4, Figure 4.2**) it is becoming apparent that a limited number of antitoxin structures within type II systems exist. In future, it

would be interesting to explore the phylogenetic and structural relationships of the antitoxins that have conserved structures between systems. While clear TA system partners exists, it may be possible to highlight distinct clades of antitoxins grouped through their N-terminal domains, to which they can be better clustered in future. Antitoxins of type II systems may, therefore, require redefining, not based on their cognate toxin, rather, based on their own structural characteristics.

Having identified two potential operator sites within the *M. tuberculosis relBE1* promoter region (**Figure 5.7 A**), alongside previous studies on structurally similar N-terminal domains (A Garcia-Pino *et al.*, 2010), it is highly likely that RelBE1 has an autoregulatory capacity which should be explored in future. Initial studies via EMSA and promoter assays (Hampton *et al.*, 2018; Beck *et al.*, 2020) would be an excellent starting point, followed by a more thorough investigation into the effects of varying RelB – RelE stoichiometries. This would require the optimisation of purification methods for both RelB1 and RelE1; the expression of RelE1 is expected to be challenging due to its reported toxicity in *E. coli* (Sala, Bordes and Genevaux, 2014). In order to avoid the difficulties with expression, and often cloning, the toxins in this study were co-expressed with their cognate antitoxins and we have managed to discover methods for ParE toxin release during purification (**Materials & Methods 2.6.4.4**) or through biophysical studies (**Chapter 3, Figure 3.14**). The apparent thermostability of the RelBE1 complex (**Figure 5.14**) may be an indication that this complex is unlikely to dissociate under natural conditions. Refolding methods have been reported for a number of TA systems whereby the complex is denatured chemically, and the toxin refolded by specific methods (Sterckx *et al.*, 2015). Alternately, the role of Clp proteases in antitoxin degradation across multiple Mycobacterial TA systems has recently been reviewed (Bordes and Genevaux, 2021), and interestingly, RelB1 was highlighted as a specific substrate for the ClpC1P1P2 protease (Ziemski *et al.*, 2021). However, the affinity of the protease for RelB1 when in complex with RelE1 was not tested, and the role of proteases in preferential antitoxin degradation to 'activate' toxins is heavily disputed (Song and Wood, 2020). These may, however, be useful avenues to explore in future to assist in isolating RelE1 for use in a variety of biochemical studies.

AlphaFold multimer (Evans *et al.*, 2022) may assist in predicting the ability of the RelBE1 system to form complexes involving non-equimolar stoichiometries; previous studies of RelBE systems have highlighted that RelB antitoxins may exist in 10-fold excess of RelE toxins (Bøggild *et al.,* 2012), favouring complexes that are negatively autoregulatory (Marianovsky *et al.*, 2001; A Garcia-Pino *et al.*, 2010). As part of this autoregulatory process, it is likely that dimerisation at the N-terminal domain between two separate RelBE1 complexes will occur and form a structure similar to that seen in the *S. aureus* YoeB-YefM TA system bound to similarly spaced operator

sites (PDB: 6L8E, (Xue *et al.*, 2020)). We currently have no evidence to support that this dimerisation process will, similar to ParDE1 (**Chapter 4, Figure 4.19**), initiate complex remodelling and toxin release. This is worth further study; however, we have not observed anything other than a homogenous sample thus far during purification and analytical SEC (**Figure 5.14 A**) alongside the apparent thermostability of the complex (**Figure 5.14 B**).

Altogether, protein sequence-based analyses indicate that the RelE1 toxin aligns well to the RelE class of toxins (**Figure 5.2**). Combining this with our crystal structure (**Figure 5.5**), and structure of the well characterised *E. coli* RelE toxin bound to the ribosome (**Figure 5.16**), we identify several potentially key differences (**Figure 5.17**). RelE toxins have been shown to undergo replacement of a C-terminal helix to correctly position an essential catalytic tyrosine (Li *et al.*, 2009; Neubauer *et al.*, 2009); Mtb RelE1 may not be capable of this (**Figure 5.17 E**), subsequently not reconstituting a catalytic core and compact fold for ribosomal interaction. Additionally, a serine residue exists at the site of the typical proton acceptor lysine (**Figure 5.17 C**), therefore this residue may require modification to fulfil this role contributing to the proposed RelE catalytic mechanism of mRNA cleavage (Gone *et al.*, 2016). Interestingly, this substitution is consistent throughout the Mtb RelE toxins (Miallau *et al.*, 2013), indicating a potentially conserved mechanism of toxicity and activation. Further to this, we identify ParE2 as a divergent of both the ParE and RelE toxin families through phylogenetic analyses (**Figure 5.4**), something which requires further study in the elucidation of potentially catalytic activity in light of the clear gyrase-dependent cleavage we present in **Chapter 3**.

Chapter 6. Conclusion

6.1 Summary of results

M. tuberculosis DNA gyrase has been central to this project and a range of gyrase constructs have been used to allow the characterisation of proteinaceous inhibitors. The gyrase holoenzyme was expressed and purified as separate, stable, dimeric subunits (GyrA and GyrB) (**Figure 3.10 A**) and shown to be active in ATP-independent relaxation, ATP-dependent supercoiling, and was susceptible to poisoning with well-studied inhibitors in cleavage assays (**Figure 3.2**). We expressed three gyrase fusion proteins with the intention of using them in structural studies, and so also characterised their biochemical activities (**Figure 3.6**). A full-length GyrBA fusion protein was shown to exists as a stable dimer in solution (**Figure 3.10 B**) and was comparably active in ATP-independent relaxation, albeit in a contrastingly processive manner (**Figure 3.7 A**). Two truncated fusion proteins were shown to be stable in solution as dimers (**Figure 3.10 B**) but only capable of variable levels of linearisation and nicking in an ATP-independent manner (**Figure 3.7 B and C**). Nevertheless, each of the gyrase proteins subsequently participated in experiments characterising the effects of ParE toxins and provided guidance regarding future structural work.

The two ParDE toxin-antitoxin systems found on the chromosome of *M. tuberculosis* H37Rv have been characterised in this body of work using a complement of biochemical, biophysical, structural, and phylogenetic analyses. The RelBE1 system (one of three on the chromosome of *M. tuberculosis* H37Rv) has also been structurally characterised and phylogenetically analysed as part of its own TA system family, and part of the RelE/ParE superfamily alongside the ParDE1 and ParDE2 toxins, ParE1 and ParE2, respectively.

A broad range of techniques has been used to understand the TA system complex structures. Biophysical studies of the ParDE1 system indicated that this complex exists as multiple species in solution (**Figure 3.1 B** and **Figure 3.11 A**); size-exclusion chromatography allowed us to resolve two species (**Figure 3.11 B**), one predicted to be a heterotetramer and one predicted to be a heterohexamer through multimeric AlphaFold models (**Figure 3.11 F** and **Figure 4.12**). The $ParD1_2ParE1_2$ heterotetramer model (**Figure 4.12 A**) resembled the *C. crescentus* ParDE system complex crystal structure (PDB: 3KXE (Dalton and Crosson, 2010)) used as the search model in MR for crystallographic studies of ParDE1; the resulting solution, in contrast, confirmed the existence of the $ParD1_4ParE1_2$ heterohexamer species (**Figure 4.1 A** and **Figure 4.12 A**). Interestingly, further analytical SEC experiments indicated that the heterotetramer could change to the heterohexamer over time (**Figure 3.12 A**), a process that was observed to be accelerated by increased temperature (**Figure 3.12 A and D**). Using this information, we developed a hypothesis for ParE1 toxin release due to the resulting imbalance in toxin:antitoxin stoichiometries between starting heterotetramer and resulting heterohexamer. Not only were we able to observe a chromatographic peak for free ParE1 during the transition of

heterotetramers to heterohexameric complexes (**Figure 3.12 C**), we were also able to scale-up the process for purification (albeit with precipitated ParE1 product) (**Figure 3.13**). Importantly, this process could be associated with enhanced gyrase-dependent cleavage as incubated ParDE1 underwent remodelling (**Figure 3.14**). These analyses demonstrated that ParE1 is a gyrase poisoning toxin that causes DNA damage due to non-specific gyrase-dependent cleavage of DNA substrates (**Figure 3.14 B**). From these data we were able to develop a model for ParE1 toxin release (**Figure 4.19**) and subsequent inhibition of DNA gyrase (**Figure 4.20**). We hypothesise that the ParD1 N-terminal CopG domain facilitates dimerisation of multiple ParDE1 complexes and subsequent steric clashes between neighbouring ParE1 monomers forces toxin exchange between complexes (**Figure 4.19 A - C**). As a result, two ParE1 toxins are expelled and their cognate ParD1 interface reverts to an unstructured domain (**Figure 4.19 C** and **Figure 4.1 B**). Building on from this, we propose that the released ParE1 monomers, which likely do not form a dimer in solution (**Figure 3.12 G**) but do have an interface within the ParDE1 complex (**Figure 4.4 A iii**), may interact with DNA gyrase at the G-gate (**Figure 4.20 A**). The ParE1-ParE1 interface seen in the complex may reform and subsequently position the toxic C-terminal residues in line with the asymmetric nick sites at the G-gate, preventing resealing (**Figure 4.20 B**).

Altogether, we have shown that the ParDE1 complex undergoes a post-translational remodelling process which releases ParE1 toxins; this would be the first reported mechanism of this sort in TA systems. The overall process may have evolved from mechanisms seen in conditional cooperativity; whilst the dimerisation of CopG domains is a recurring theme in autoregulation (Costa *et al.*, 2001; Mattison *et al.*, 2006), when at equimolar stoichiometries steric occlusion usually forces complexes to dissociate from cognate operator DNA. While a similar CopG dimerisation event has occurred in our crystal structure of ParDE1, we have no further evidence to suggest that the complex has any autoregulatory capacity (**Figure 4.2**). Instead, we argue that the ParE – ParE (**Figure 4.4 A iii**) and ParD – ParE (**Figure 4.4 A ii**) interfaces have increased plasticity as toxins are liberated from ParDE1 complexes. This is a reasonable conclusion as rather than existing in an equilibrium between equimolar and non-equimolar stoichiometries, ParDE1 complexes appear to develop toward ever higher order states (**Figure 3.12 C**). If ParDE1 did indeed act as per conditionally cooperative systems, once expressed in the equimolar heterotetramer (**Figure 3.12 A**, black), the complex would not form higher order stoichiometries due to the CopG domains failing to form stable dimers between complexes due to steric clashes between the saturated antitoxin chains (Garcia-Pino *et al.*, 2010).

In contrast with ParDE1, the ParDE2 complexes proved more peculiar throughout their study. Biophysical studies indicated that, unlike ParDE1, this system exists as a relatively stable complex in solution (**Figure 3.16 A**, black); we predicted the stoichiometry to be 2:2 as per most TA

complexes structures in the RelE/ParE superfamily (**Chapter 1, Table 1.4**) (Anantharaman and Aravind, 2003; Francuski and Saenger, 2009; Dalton and Crosson, 2010; Blower, Salmond and Luisi, 2011). However, AlphaFold consistently struggled in generating models forming biologically relevant ParD dimerisation domains (**Figure 4.16 A** and **Figure 4.18 A and B**), alongside Stokes radii not matching that observed in analytical SEC (**Figure 3.16 D**). The generation of poor ParDE2 complex models was due to high preference for the formation of an alpha helix by the ParE2 C-terminal residues across the conserved hydrophobic RelE/ParE antitoxin recognition surface, both in monomeric solutions (**Figure 4.16 B**) and multimeric in the context of ParD2 (**Figure 4.18 A and B**). The crystal structure of this system was unfortunately incomplete with respect to a heterotetrametric structure (**Figure 4.13**), likely due to the loss of the ParD2 N-terminal domain (**Figure 4.13 D**) during the long crystallisation process. However, the C-terminal ParD2 residues occupied the conserved antitoxin recognition surface (**Figure 4.13 A**), and this was confirmed with electron density maps (**Figure 4.17**). This allowed us to remodel the ParDE2 complex using AlphaFold by removing the ParE2 C-terminal 17 residues; subsequently two reasonable models could be generated (**Figure 4.18 C and E**), but only one was supported by the SEC data and crystal structure ('Model 3', **Figure 4.18 E and F**). Interestingly, the global structure of the antitoxin resembled the ParD1 antitoxin (**Figure 4.18 H** and **Figure 4.3 E**), however, the N-terminal domain did not include a beta sheet and dimerised through a helix-turn-helix domain instead. Additionally, our best model for ParDE2 positions the ParE2 toxin molecules at opposite sides of a widened V-shaped complex (**Figure 4.18 E**), rather than positioned centrally with a ParE – ParE interface (**Figure 4.1 A**).

Rather than isolating the ParE2 toxin through remodelling, we were able to purify this toxin through a specific purification method that uses its distinct isoelectric point (**Figure 4.18 A and B**). We show that the ParE2 toxin is also a gyrase inhibiting toxin, albeit with potentially lower affinity than ParE1 (**Figure 3.5 B** and **Figure 3.14 B**). ParE2 was shown to stabilise high proportions of cleavage complexes when tested against $GyrB_2A_2$ and GyrBA; a more rapid increase in linear species in seen with the GyrBA fusion (**Figure 3.8 A**), potentially due to the ParE2 toxin interacting with the enzyme at GyrB-GyrA interfaces. This is also supported by the decreasing levels of linearisation seen with both fusion proteins, $GyrBA^{56}$ and $GyrB^{28}A^{56}$ (**Figure 3.8 B and C**). The removal of the GyrA-CTD reduced linearisation from ~60 % to ~25 % (**Figure 3.8 A and B**), and the subsequent removal of the GyrB-ATPase reduced linearisation from ~25 % to ~8 % (**Figure 3.8 B and C**). Though it may be argued it is unlikely that the toxin requires interfaces formed by these domains, rather, efficient DNA processing and gyrase function may be required for more effective ParE2 induced inhibition. These studies indicated that the GyrBA and $GyrBA^{56}$ fusion proteins may be very useful in future structural studies.

While the ParE2 toxin performed as expected in biochemical assays, inhibiting gyrase by trapping cleavage complexes (**Figure 3.5 B**), the phylogeny of the toxin sequence and the mechanism by which poisoning occurs is now open for debate. Our phylogenetic analyses position the ParE2 toxin in an early divergent clade of the RelE toxin family (**Figure 5.3** and **Figure 5.4**), rather than a true member of the ParE family devoid of catalytic residues. The ParE2 toxin may therefore contain a catalytic core of some activity and may require DNA-gyrase for correct positioning to its substrate. This would be an interesting mechanism of action and contrasts the model that we have built for ParE1 (**Figure 4.20**). The predicted existence of the C-terminal helix in ParE2 (**Figure 4.16 B**) further supports a potential catalytic role; this is a feature of RelE toxins (Bøggild *et al.*, 2012) and is a mechanism of toxin activation through the reposition of an essential tyrosine residue with its mRNA substrate within the ribosome (Griffin, Davis and Strobel, 2013). No tyrosine residue is present in the C-terminal helix of ParE2 however (**Figure 5.18 C**), therefore, the hypothesised mechanism of catalysis will be distinct from that seen for RelE toxins. As we have shown this toxin to induce significant linearisation in a gyrase-dependent manner (**Figure 3.5 B**), gyrase may somehow stabilise the ParE2 toxin for nuclease function. This is highly speculative at this point, however.

Phylogenetic analyses positioning ParE2 in the RelE family also positioned the RelE1 toxin as a comparably close relative to the ParE toxins (**Figure 5.3** and **Figure 5.4**). Our work on this system was in collaboration with the Genevaux group who observed that RelE1 does not possess ribosome-dependent mRNase activity (personal communication). Our structural characterisation of the RelBE1 system complex (**Figure 5.5**) became a precursor to identifying potential differences in the RelE1 toxin that may contribute to the lack of mRNase function. Investigation of the RelE1 catalytic core indicated a high degree of structural and sequence similarity to the well characterised *E. coli* RelE core (**Figure 5.17**) (Neubauer *et al.*, 2009; Bøggild *et al.*, 2012; Griffin, Davis and Strobel, 2013); however, the *E. coli* catalytic residue K52 is substituted for S58 in *M. tuberculosis* RelE1 (**Figure 5.17 B and C**) and the C-terminal tail did not appear to be modelled as folded into the 'activated' position by AlphaFold (**Figure 5.17 E**). While the K52 – S58 mutation could be redundant due to the supplementary role of K52 in catalysis (Griffin, Davis and Strobel, 2013), the improper repositioning of *M. tuberculosis* RelE1 Y89 would be highly rate-reducing (Neubauer *et al.*, 2009; Griffin, Davis and Strobel, 2013); in combination, these two differences could render the toxin completely ineffective in ribosome-dependent mRNA cleavage. Alternatively, *M. tuberculosis* RelE1 has a varied sequence through α2 and does not contain the residues identified to stabilise the *E. coli* RelE-ribosome interaction (Neubauer *et al.*, 2009); whether the RelE1 toxin is even capable of interacting with the ribosome is not yet understood, however, it may simply form interactions through several different residues. Both

phylogenetic (**Figure 5.2** and **Figure 5.3**) and structural analyses (**Figure 5.5**) indicate that RelE1 is indeed a RelE toxin, after all.

The complex structure of the RelBE1 system presented the third different antitoxin dimerisation domain in this study. This domain, made up of two arched helices and 6 beta sheets, is present in numerous different TA systems, similar to the CopG domain. We predict that this has an autoregulatory function due to the presence and position of essential basic residues and the identification of two operator sites (inverted repeats) within the promoter for *relBE1.*

6.2 Future research

6.2.1 Protein expression, purification, and structural studies

The main goal of future work with regards to what has been presented here will be to structurally characterise the interaction of both ParE1 and ParE2 with *M. tuberculosis* gyrase. An essential step toward this is the development of more efficient methods of purification for both gyrase and ParDE proteins, specifically the ParE toxins.

With regards to protein expression and purification methods, the addition of specific recombinant tags may be a useful tool to improve purification of notoriously 'sticky' gyrase proteins, albeit one that requires further gene synthesis and cloning. Considering the gyrase proteins used thus far, highly efficient methods have previously been published in the study of the *E. coli* gyrase by Cryo-EM; N-terminal His-TEV tags were used in combination with C-terminal strep tags to improve protein purification. Subsequent methods of gyrase reconstitution from the GyrA and GyrB subunits were also published and utilise very low salt buffers in contrast to our methods. These methods of purification will need to be explored in future to check for protein stability and the effective reconstitution of the $GyrB_2A_2$ holoenzyme; this can be performed using our already optimised methods of analytical SEC. Once these methods are optimised, the addition of specific DNA substrates, such as the 180 bp substrate for $GyrB_2A_2$ and GyrBA (vanden Broeck *et al.*, 2019), or the 24 bp substrate for $GyrBA^{56}$ and $GyrB^{28}A^{56}$ (Blower *et al.*, 2016) can be studied alongside the addition of ATP analogues to promote complex stability. Following this, the ParE toxins can be used to stabilise the cleavage complexes of selected gyrase enzymes for subsequent structural characterisation by Cryo-EM or crystallography. While our methods in crystallography are currently sound, we must develop practices in Cryo-EM, likely starting with negative staining once protein purification methods and complexing is optimised.

As highlighted, an essential step on forming the cleavage complexes is to have high purity ParE toxins. Like the gyrase enzyme proteins previously mentioned, altering recombinant tags may be a useful avenue to explore; our current hypotheses regarding the toxic mechanisms for both

ParE1 and ParE2 rely on C-terminal residues interacting with gyrase to some extent. Our model for ParE1 involves the C-terminal residues being able to access the DNA cleavage sites at the G-gate during the gyrase reaction cycle. For ParE2, we currently believe that the refolding of the C-terminal helix is an essential feature of this protein in potentially reforming a catalytic core. If interactions with the C-terminal ParE residues can be perturbed this may render the toxins non-toxic, or at least tolerable within the cell, therefore the use a range of C-terminal tags should be explored to allow for expression of the toxins in the absence of their cognate ParD antitoxin. We may be able to prevent the ParE – gyrase interaction by fusing the C-terminus with a cleavable tag which can be removed during purification.

We have previously been able to purify the ParE2 toxin away from the complex in high purity, however this was apparently sporadic and varied greatly in yield. This process may simply require optimisation in terms of buffers, protein concentration, time, and temperature used in the N-terminal 6His-SUMO cleavage step. In this study we have built a model for ParDE1 complex remodelling and ParE1 toxin release. Attempts were made to purify the ParE1 toxin from a 'large-scale' protein expression, however, while a good yield of ParE1 toxin was obtained the fraction had precipitated. The remodelling process could be optimised in future for ParE1 purification through buffers, protein concentration, time, and temperature used during ParDE1 incubation.

As an alternative to recombinant tags and existing methods that require optimisation, both complexes could be subjected to chemical denaturation and specific refolding of the toxins. Methods have been published for the efficient refolding of TA system toxins (Sterckx *et al.*, 2015); however, refolding is likely protein specific and would require significant optimisation not only in the refolding process but potentially in the unfolding/denaturation process also. Importantly, our expression vectors and methods are largely optimised for these processes; denaturation of the protein complexes in Gn.HCl or urea can occur on-column (Ni-NTA) as the ParE toxins are tagged. Several different methods can subsequently be attempted to refold the protein on-column, or the protein can be eluted and refolded using several dialysis steps. With regards to ParE1, the precipitated fraction from our high-scale complex remodelling attempt could be resolubilised in denaturant and refolding attempts made on column or by dialysis in a similar manner. These processes can, however, be extremely time-consuming and may require considerable optimisation for potentially low yields. The same techniques discussed here for the purification of ParE toxins will also be applicable to the RelE1 toxin for future study. Exploring alternate expression systems could also prove fruitful; the use of eukaryotic insect or yeast expression may be possible due to divergence in their type IIA topos and nuclear localisation.

Once efficient methods of toxin isolation are developed, as previously mentioned the ParE proteins can be used to trap cleavage complexes of DNA gyrase, followed by structural studies

through Cryo-EM and crystallography. The selection of the appropriate gyrase construct can be informed by a combination of the biochemistry gathered in this thesis, and future biophysical analyses of the ParE-gyrase interaction; potentially via our established analytical SEC methods. Currently, the full-length GyrBA fusions appears to be the best candidate having demonstrated the highest susceptibility to ParE2 poisoning (**Figure 3.8**), purifies well, and existed in solution as a dimer (**Figure 3.10**). Additionally, further structural work should be pursued with regards to both ParDE1 and ParDE2 complexes. To better support our theories of ParDE1 remodelling, validation of the AlphaFold $ParD1_2ParE1_2$ heterotetramer model through crystallographic studies would be useful. This may prove challenging given that the remodelling process appears to occur naturally over time. If the complex can be expressed and purified as the heterotetramer species, crystallisation experiments could be conducted at low temperatures to slow down the remodelling process; if a condition can be identified that promotes rapid crystallisation of the sample at low temperature, we may be able to capture the heterotetramer and confirm the existence of this species. The ParDE2 structure is currently incomplete, however, we have an AlphaFold model supported by both SEC and crystal structure data. The apparent stability of the purified ParDE2 complex in solution was problematic in crystallisation trials; even when the starting concentration was 20 mg/mL crystallisation took approximately 3 months (with very few conditions precipitating, and only 3 producing crystals). Even higher concentrations could be explored in future, or the crystal buffer optimised to promote more rapid crystallisation. With this, we may be able to avoid the loss of the ParD2 N-terminal dimerisation domain and elucidate the whole complex structure. It is possible, however, that the removal of the N-terminal domain is essential for packing of this structure, and therefore crystals may not form for the full-length complex.

6.2.2 Biochemistry and biophysical studies

An essential part of future study will be in advancing our understanding of the proposed ParDE1 remodelling process using a variety of techniques. It will be interesting to use single molecule FRET to potentially visualise the re-arrangement we propose to occur at the ParE1 – ParE1 toxin interface (Roy, Hohng and Ha, 2008); labelling separate samples with the appropriate fluorophores may prove that ParE1 toxins are exchanged between heterotetramers as they interact to form a heterohexamer and liberate toxins. Initially, we expect that one dye would dominate the emission as separately labelled samples remain diffuse (Mazal and Haran, 2019). In remodelling, the fluorophores would be forced into close proximity as toxins are exchanged between heterotetramers as they interact via the ParD1 CopG domains; contrastingly labelled toxin molecules would result in the dominance of the second fluorophores (Roy, Hohng and Ha, 2008; Mazal and Haran, 2019). Data from such an experiment may be obscured by noise if the liberated toxins subsequently exist in solution as monomers, as predicted. These molecules

would theoretically not interact resulting in stronger output from the initial signal. However, we predict that this process eventually reabsorbs the liberated toxin molecules, indicate by SEC data, which would result in the absolute preference for the second fluorophores, indicating mixing between the samples (Mazal and Haran, 2019). As a whole process ParDE1 remodelling could return some interesting and highly informative FRET data to support or disprove our model.

We must also better explore the endpoint of this process; as mentioned, our biophysical data suggest the evolution of a higher order species beyond the heterohexamer that reabsorbs the liberated toxin molecules (potentially an equimolar hetero-octamer). AlphaFold modelling and potentially crystallographic studies may assist in this study (Evans *et al.*, 2022), however, we currently do not know when the process stops therefore ParDE1 incubations must be elongated until the entire fraction is stable. The potential thermodynamic nature of the process should also be explored in future; once the final species is stable, we should explore whether lowering temperatures reverses the process. Currently, we do not believe this would occur as the process occurred naturally at 4 °C, indicating a proclivity of the system to remodel kinetically, rather than thermodynamically.

Supplementing the ParDE1 remodelling process with ParD antitoxin will also be an important avenue to explore. Given our model, we propose it may be possible that rather than ParDE1 heterotetramers interacting, an interaction of a ParDE1 heterotetramer with a ParD1 dimer may be preferred and could result in the heterohexamer species if a ParE1 toxin can be 'adopted' by the incoming ParD1 dimer. This study should be supplemented by developing our understanding of the ratio of toxin:antitoxin translated in the cell under a variety of conditions. Many studies indicate that antitoxins are translated in far greater amounts and that upregulation of toxins is a mechanism of TA system activation. Indeed, ParD1 abundance would prevent the proposed ParE1 liberation during remodelling, however, if under specific conditions the translation of component is found to be equimolar, the complex would be able to become toxic and the system 'activated', at least for a period of time. Better understanding of the affinities of the individual TA system components for one another will be interesting data and should be explored through techniques such as SEC-MALS. In the context of ParDE1, this may highlight why ParE1 toxins can be separated from the complex. It will also be useful to understand the oligomeric state of ParD1 in solution and to explore the affinity of the CopG domain for other ParD1 dimers, as well as for a ParDE1 heterotetramer as this will influence remodelling. Beyond these studies, the affinity of ParE1 and ParE2 for gyrase should be explored.

We must also develop our ParE1 induced gyrase inhibition studies. This can occur in two ways; one series of experiments will involve testing the purified ParE1 toxin more thoroughly against our range of gyrase proteins. The second series would involve more developed ParDE1

remodelling coupled with gyrase cleavage assays, whereby the components are incubated together throughout, rather than ParDE1 pre-incubation. This may reveal whether the development beyond the heterohexamer (and predicted reabsorption of ParE toxins) is preferred to the gyrase interaction. This would indicate that the process may be able to stall gyrase and slow cellular processes relying on the enzyme for a specific period of time, and gyrase could subsequently be returned to full function once the toxins are removed by the evolving ParDE1 complex. Further to this, growth-curves of *E. coli* expressing the ParDE1 system at a range of temperatures and over a range of times will hopefully confirm the toxic effects predicted to be associated with complex remodelling.

On a similar note, both ParDE1 and ParDE2 need further study regarding the antitoxicity of their cognate antitoxins. It has been reported that ParD2 can prevent the toxic effects of ParE2, but not rescue the ParE2 poisoned gyrase (Gupta *et al.*, 2016). We must confirm these findings. Furthermore, the biochemistry of both systems should be explored in the full range of gyrase activity assays, including supercoiling and decatenation to explore whether the toxins have an inhibitory preference. While this doesn't assist our main objective of structural classification of the interaction, it will develop our understanding of the effects of different ParE toxins.

With regards to the RelBE1 system, if the RelE1 toxin can be isolated further study of the target should be conducted. This could be performed with pull-downs using the tagged RelE1 toxin as 'bait'. It will be important to confirm that the ribosome, or specifically one subunit, is the target of this toxin. Once confirmed, SEC-MALS could be used to analyse the dissociation constant of the interaction. Regarding the yet unproven toxicity of RelE1 *in vitro*, several studies could be performed including attempts to phosphorylate S58 or mutation to lysine as seen in the *E. coli* RelE catalytic core.

Ultimately, studies on RelE1 may also progress toward structural characterisation bound to its target; if this is confirmed to be the ribosome there is precedent for such study (Neubauer *et al.*, 2009), and we expect to see a novel mechanism of action which would be interesting. Furthermore, we have highlighted the potential for autoregulation, and this can be explored biochemically and microbiologically through EMSAs and promoter assays (Beck *et al.*, 2020). Isolating the individual components will assist in bettering our understanding of the potential for conditional cooperativity in the RelBE1 system as the toxin and antitoxin stoichiometries can be altered in vitro, prior to binding promoter DNA.

6.2.3 Phylogenetic analyses

To better understand the evolutionary relationships of toxins studied in this thesis we should diversify the multiple sequence analysis aspect of our phylogenetic analyses. Specifically, we

need more information regarding the sequence of the ParE2 toxin and the potential existence of a catalytic core. As the ParE2 protein apparently aligns better with the RelE family of toxins, expanding our library of RelE family toxin sequences may assist us in identifying more similar proteins and highlight essential residues for catalytic activity. While catalytic activity would be surprising for the ParE2 toxin, it is worth further study and these analyses may highlight residues beyond the C-terminal tail for mutational analysis.

Beyond the toxins in this study, we believe that phylogenetic analyses of the antitoxin components of TA systems could be an interesting avenue of research. In this body of work, we present three TA system complex structures with individual, but not unique, antitoxin dimerisation domains. It has indeed previously been suggested to classify the antitoxins by their respective, likely structured, dimerisation domains (Blower, Salmond and Luisi, 2011; Chan, Espinosa and Yeo, 2016). Not only could this be useful in the categorisation of systems regarding their autoregulatory capacities, it may also highlight other systems capable of remodelling. We predict this process to be unique to the CopG domain, however, it will likely require the appropriate combination of CopG domain and cognate toxins.

6.3 Concluding remarks

In this thesis we present the crystal structures of the three remaining ParE/RelE TA systems of *M. tuberculosis* (**Figure 4.1, Figure 4.13, Figure 5.5, Figure 1.8, and Table 1.4**). Both ParDE structures are distinct not only from each other, but from other ParDE structures in the PDB (**Figure 4.1, Figure 4.13, and Table 1.4**). Combining structural, biochemical, and biophysical analyses we propose a novel mechanism of post-translational liberation of ParE1 toxin from its neutralised complex via a remodelling process (**Figure 3.14**), involving multimerisation of ParD1 dimerisation domains and ParE1 toxin exchange/release (**Figure 4.19**).

Biochemical studies indicate that both ParE toxins inhibit Mtb gyrase, trapping cleavage complexes to induce dsDNA breaks via a novel mechanism to the CcdB toxin. We propose two possible mechanisms for this based on the structure of the ParDE1 complex (**Figure 4.1**) and the biochemical analysis for ParE2 (**Figure 3.5 and Figure 3.8**), respectively; ParE toxins may bind at the gyrase DNA-gate to intercalate residues in their flexible C-terminal helices into gyrase-cleaved substrate DNA (**Figure 4.20**), or, they may bind allosterically at a site relying on interaction between gyrase subunits to interfere with gate dynamics, decoupling processes to prevent re-ligation (**Figure 3.8**). Indeed, both toxins may operate differently as phylogenetic analyses of the ParE/RelE superfamily highlights ParE2 as a distinct from both ParE and RelE toxins (**Figure 5.4**). This study has laid foundations for future research to untangle the molecular mechanisms of gyrase inhibition by ParE toxins; biophysical, biochemical, and mutagenesis studies will each play

a role in advancing this work, with structural biology via crystallography or cryo-EM, exploring the ParE-gyrase interaction, remaining the ultimate goal.

The same series of phylogenetic studies assisted in identifying subtle differences in the RelE1 sequence that may contribute to the reported lack of ribosome-dependent mRNase activity (personal communication; Genevaux group, Toulouse) (**Figure 5.3 and Figure 5.18**). Subsequent structural analyses highlighted several differences in the catalytic core and C-terminal residues of RelE1 that may render the toxin catalytically inactive (**Figure 5.15 and Figure 5.17**), simply inhibiting translation through direct inhibition of the ribosome, or requiring a post-translational modification for catalytic activation.

These findings advance our understanding of the type II complement of Mtb, notably systems which have been implicated in several roles contributing to the virulence and adaptation of Mtb during infection, and present ParE toxins as potentially potent gyrase inhibitors worthy of further study.

References

Aakre, C.D. *et al.* (2013) 'A bacterial toxin inhibits DNA replication elongation through a direct interaction with the β sliding clamp', *Molecular Cell*, 52(5). Available at: https://doi.org/10.1016/j.molcel.2013.10.014.

Aakre, C.D. *et al.* (2015) 'Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates', *Cell*, 163(3). Available at: https://doi.org/10.1016/j.cell.2015.09.055.

Abou-Nader, M. and Benedik, M.J. (2010) 'Rapid generation of random mutant libraries', *Bioengineered Bugs*, 1(5). Available at: https://doi.org/10.4161/bbug.1.5.12942.

Adams, P.D. *et al.* (2010) 'PHENIX: a comprehensive Python-based system for macromolecular structure solution.', *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp. 213–21. Available at: https://doi.org/10.1107/S0907444909052925.

Agarwal, N. and Tyagi, A.K. (2006) 'Mycobacterial transcriptional signals: requirements for recognition by RNA polymerase and optimal transcriptional activity.', *Nucleic acids research*, 34(15), pp. 4245–57. Available at: https://doi.org/10.1093/nar/gkl521.

Aghera, N.K. *et al.* (2020) 'Mechanism of CcdA-Mediated Rejuvenation of DNA Gyrase', *Structure*, 28(5), pp. 562-572.e4. Available at: https://doi.org/10.1016/J.STR.2020.03.006.

Agrawal, A. *et al.* (2013) 'Mycobacterium tuberculosis DNA gyrase ATPase domain structures suggest a dissociative mechanism that explains how ATP hydrolysis is coupled to domain motion', *Biochemical Journal*, 456(2). Available at: https://doi.org/10.1042/BJ20130538.

Agrawal, A. *et al.* (2013) 'Mycobacterium tuberculosis DNA gyrase ATPase domain structures suggest a dissociative mechanism that explains how ATP hydrolysis is coupled to domain motion', *Biochemical Journal*, 456(2). Available at: https://doi.org/10.1042/BJ20130538.

Ahmed, W. *et al.* (2014) 'Conditional silencing of topoisomerase I gene of Mycobacterium tuberculosis validates its essentiality for cell survival', *FEMS Microbiology Letters*, 353(2). Available at: https://doi.org/10.1111/1574-6968.12412.

Ahn, D.H. *et al.* (2017) 'Structural analyses of the MazEF4 toxin-antitoxin pair in Mycobacterium tuberculosis provide evidence for a unique extracellular death factor', *Journal of Biological Chemistry*, 292(46). Available at: https://doi.org/10.1074/jbc.M117.807974.

Akbergenov, R. *et al.* (2011) 'Molecular basis for the selectivity of antituberculosis compounds capreomycin and viomycin', *Antimicrobial Agents and Chemotherapy*, 55(10). Available at: https://doi.org/10.1128/AAC.00628-11.

Alangaden, G.J. *et al.* (1998) 'Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 42(5). Available at: https://doi.org/10.1128/aac.42.5.1295.

Albrethsen, J. *et al.* (2013) 'Proteomic profiling of Mycobacterium tuberculosis identifies nutrient-starvation-responsive toxin-antitoxin systems', *Molecular and Cellular Proteomics*, 12(5). Available at: https://doi.org/10.1074/mcp.M112.018846.

Alderwick, L.J. *et al.* (2015) 'The mycobacterial cell wall—peptidoglycan and arabinogalactan', *Cold Spring Harbor Perspectives in Medicine*, 5(8). Available at: https://doi.org/10.1101/cshperspect.a021113.

Aldred, K.J. *et al.* (2013) 'Topoisomerase IV-quinolone interactions are mediated through a water-metal ion bridge: Mechanistic basis of quinolone resistance', *Nucleic Acids Research*, 41(8). Available at: https://doi.org/10.1093/nar/gkt124.

Aldred, K.J. *et al.* (2016) 'Fluoroquinolone interactions with Mycobacterium tuberculosis gyrase: Enhancing drug activity against wild-type and resistant gyrase', *Proceedings of the National Academy of Sciences of the United States of America*, 113(7). Available at: https://doi.org/10.1073/pnas.1525055113.

Allué-Guardia, A. *et al.* (2021) 'Mycobacteriophages as potential therapeutic agents against drug-resistant tuberculosis', *International Journal of Molecular Sciences*. Available at: https://doi.org/10.3390/ijms22020735.

Ames, J.R. *et al.* (2019) 'Expression of different ParE toxins results in conserved phenotypes with distinguishable classes of toxicity', *MicrobiologyOpen*. Available at: https://doi.org/10.1002/mbo3.902.

Anantharaman, V. and Aravind, L. (2003) 'New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system', *Genome Biology*, 4(12), p. R81. Available at: https://doi.org/10.1186/gb-2003-4-12-r81.

Aravind, L., Leipe, D.D. and Koonin, E. v. (1998) 'Toprim - A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins', *Nucleic Acids Research*, 26(18). Available at: https://doi.org/10.1093/nar/26.18.4205.

Arcus, V.L. *et al.* (2011) 'The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array', *Protein Engineering, Design and Selection*. Available at: https://doi.org/10.1093/protein/gzq081.

Ashkenazy, H. *et al.* (2016) 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules', *Nucleic Acids Research*, 44(W1). Available at: https://doi.org/10.1093/NAR/GKW408.

Ashley, R.E. *et al.* (2017) 'Recognition of DNA Supercoil Geometry by Mycobacterium tuberculosis Gyrase', *Biochemistry*, 56(40). Available at: https://doi.org/10.1021/acs.biochem.7b00681.

Aslanidis, C. and de Jong, P.J. (1990) 'Ligation-independent cloning of PCR products (LIC-PCR)', *Nucleic Acids Research*, 18(20), pp. 6069–6074. Available at: https://doi.org/10.1093/nar/18.20.6069.

Aubry, A. *et al.* (2006) 'First functional characterization of a singly expressed bacterial type II topoisomerase: The enzyme from Mycobacterium tuberculosis', *Biochemical and Biophysical Research Communications*, 348(1). Available at: https://doi.org/10.1016/j.bbrc.2006.07.017.

Balasubramani, G.L. *et al.* (2020) 'Structure-based drug repurposing to inhibit the DNA gyrase of Mycobacterium tuberculosis', *Biochemical Journal*, 477(21). Available at: https://doi.org/10.1042/BCJ20200462.

Balfour, J.A., Bryson, H.M. and Brogden, R.N. (1996) 'Imipenem/cilastatin: An update of its antibacterial activity, pharmacokinetics and therapeutic efficacy in the treatment of serious infections', *Drugs*. Available at: https://doi.org/10.2165/00003495-199651010-00008.

Baptista, R. *et al.* (2018) 'Untargeted metabolomics reveals a new mode of action of pretomanid (PA-824)', *Scientific Reports*, 8(1). Available at: https://doi.org/10.1038/s41598-018-23110-1.

Baranello, L. *et al.* (2016) 'RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription', *Cell*, 165(2). Available at: https://doi.org/10.1016/j.cell.2016.02.036.

Barbosa, L.C.B. *et al.* (2012) 'Design and synthesis of peptides from bacterial ParE toxin as inhibitors of topoisomerases', *European Journal of Medicinal Chemistry*, 54. Available at: https://doi.org/10.1016/j.ejmech.2012.06.008.

Barth, V.C. and Woychik, N.A. (2020) 'The Sole Mycobacterium smegmatis MazF Toxin Targets tRNALys to Impart Highly Selective, Codon-Dependent Proteome Reprogramming', *Frontiers in Genetics*, 10. Available at: https://doi.org/10.3389/fgene.2019.01356.

Bates, A.D., Berger, J.M. and Maxwell, A. (2011) 'The ancestral role of ATP hydrolysis in type II topoisomerases: Prevention of DNA double-strand breaks', *Nucleic Acids Research*. Available at: https://doi.org/10.1093/nar/gkr258.

Bax, B.D. *et al.* (2010) 'Type IIA topoisomerase inhibition by a new class of antibacterial agents', *Nature*, 466(7309). Available at: https://doi.org/10.1038/nature09197.

Bax, B.D. *et al.* (2019) 'DNA Topoisomerase Inhibitors: Trapping a DNA-Cleaving Machine in Motion', *Journal of Molecular Biology* [. Available at: https://doi.org/10.1016/J.JMB.2019.07.008.

Baxter, J. and Diffley, J.F.X. (2008) 'Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast', *Molecular Cell*, 30(6). Available at: https://doi.org/10.1016/j.molcel.2008.04.019.

Bechhofer, D.H. and Deutscher, M.P. (2019) 'Bacterial ribonucleases and their roles in RNA metabolism', *Critical Reviews in Biochemistry and Molecular Biology*. Available at: https://doi.org/10.1080/10409238.2019.1651816.

Beck, I.N. *et al.* (2020) 'Antitoxin autoregulation of M. tuberculosis toxin-antitoxin expression through negative cooperativity arising from multiple inverted repeat sequences', *Biochemical Journal*, 477(12), pp. 2401–2419. Available at: https://doi.org/10.1042/BCJ20200368.

Belanger, A.E. *et al.* (1996) 'The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol', *Proceedings of the National Academy of Sciences of the United States of America*, 93(21). Available at: https://doi.org/10.1073/pnas.93.21.11919.

Bendtsen, K.L. *et al.* (2017) 'Toxin inhibition in C. crescentus VapBC1 is mediated by a flexible pseudo-palindromic protein motif and modulated by DNA binding', *Nucleic Acids Research*, 45(5). Available at: https://doi.org/10.1093/nar/gkw1266.

Berger, S.L. and Kimmel, A.R. (1987) 'Guide to molecular cloning techniques'. Academic Press Inc.,Orlando, FL.

Bermingham, A. and Derrick, J.P. (2002) 'The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery', *BioEssays*. Available at: https://doi.org/10.1002/bies.10114.

Betts, J.C. *et al.* (2002) 'Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling', *Molecular Microbiology*, 43(3). Available at: https://doi.org/10.1046/j.1365-2958.2002.02779.x.

Bizard, A.H. and Hickson, I.D. (2020) 'The many lives of type IA topoisomerases', *Journal of Biological Chemistry*. Available at: https://doi.org/10.1074/jbc.REV120.008286.

Black, G.F. *et al.* (2009) 'Immunogenicity of novel DosR regulon-encoded candidate antigens of mycobacterium tuberculosis in three high-burden populations in Africa', *Clinical and Vaccine Immunology*, 16(8). Available at: https://doi.org/10.1128/CVI.00111-09.

Blanco, D. *et al.* (2015) 'Mycobacterium tuberculosis gyrase inhibitors as a new class of antitubercular drugs', *Antimicrobial Agents and Chemotherapy*, 59(4). Available at: https://doi.org/10.1128/AAC.03913-14.

Blomgran, R. and Ernst, J.D. (2011) ' Lung Neutrophils Facilitate Activation of Naive Antigen-Specific CD4 + T Cells during Mycobacterium tuberculosis Infection ', *The Journal of Immunology*, 186(12). Available at: https://doi.org/10.4049/jimmunol.1100001.

Blomgran, R. *et al.* (2012) 'Mycobacterium tuberculosis inhibits neutrophil apoptosis, leading to delayed activation of naive CD4 T cells', *Cell Host and Microbe*, 11(1). Available at: https://doi.org/10.1016/j.chom.2011.11.012.

Blower, T.R. *et al.* (2009) 'Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of Erwinia', *Journal of Bacteriology*, 191(19), pp. 6029–6039. Available at: https://doi.org/10.1128/JB.00720-09.

Blower, T.R. *et al.* (2011) 'A processed noncoding RNA regulates an altruistic bacterial antiviral system', *Nature Structural & Molecular Biology*, 18(2), pp. 185–190. Available at: https://doi.org/10.1038/nsmb.1981.

Blower, T.R. *et al.* (2012) 'Identification and classification of bacterial Type III toxin–antitoxin systems encoded in chromosomal and plasmid genomes', *Nucleic Acids Research*, 40(13), pp. 6158–6173. Available at: https://doi.org/10.1093/nar/gks231.

Blower, T.R. *et al.* (2016) 'Crystal structure and stability of gyrase-fluoroquinolone cleaved complexes from Mycobacterium tuberculosis.', *Proceedings of the National Academy of Sciences of the United States of America*, 113(7), pp. 1706–13. Available at: https://doi.org/10.1073/pnas.1525047113.

Blower, T.R., Salmond, G.P. and Luisi, B.F. (2011) 'Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin–antitoxin partners', *Current Opinion in Structural Biology*, 21(1), pp. 109–118. Available at: https://doi.org/10.1016/J.SBI.2010.10.009.

Bøggild, A. *et al.* (2012) 'The Crystal Structure of the Intact E. coli RelBE Toxin-Antitoxin Complex Provides the Structural Basis for Conditional Cooperativity', *Structure*, 20(10), pp. 1641–1648. Available at: https://doi.org/10.1016/j.str.2012.08.017.

Bold, T.D. *et al.* (2011) 'Suboptimal activation of antigen-specific cD4+ effector cells enables persistence of M. tuberculosis in vivo', *PLoS Pathogens*, 7(5). Available at: https://doi.org/10.1371/journal.ppat.1002063.

Bolten, S.N., Rinas, U. and Scheper, T. (2018) 'Heparin: role in protein purification and substitution with animal-component free material', *Applied Microbiology and Biotechnology*. Available at: https://doi.org/10.1007/s00253-018-9263-3.

Borah, K. *et al.* (2021) ' Metabolic fluxes for nutritional flexibility of Mycobacterium tuberculosis ', *Molecular Systems Biology*, 17(5). Available at: https://doi.org/10.15252/msb.202110280.

Bordes, P. and Genevaux, P. (2021) 'Control of Toxin-Antitoxin Systems by Proteases in Mycobacterium Tuberculosis', *Frontiers in Molecular Biosciences*. Available at: https://doi.org/10.3389/fmolb.2021.691399.

Bouige, A. *et al.* (2013) 'Mycobacterium tuberculosis DNA gyrase possesses two functional GyrA-boxes', *Biochemical Journal*, 455(3). Available at: https://doi.org/10.1042/BJ20130430.

Brielle, R., Pinel-Marie, M.L. and Felden, B. (2016) 'Linking bacterial type I toxins with their actions', *Current Opinion in Microbiology*. Available at: https://doi.org/10.1016/j.mib.2016.01.009.

Briken, V. and Miller, J.L. (2008) 'Living on the edge: Inhibition of host cell apoptosis by Mycobacterium tuberculosis', *Future Microbiology*. Available at: https://doi.org/10.2217/17460913.3.4.415.

Brossier, F. *et al.* (2011) 'Molecular investigation of resistance to the antituberculous drug ethionamide in multidrug-resistant clinical isolates of Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 55(1). Available at: https://doi.org/10.1128/AAC.01030-10.

Bryant, P. *et al.* (2022) 'Predicting the structure of large protein complexes using AlphaFold and sequential assembly', *bioRxiv* [Preprint].

Cabral, J.H.M. *et al.* (1997) 'Crystal structure of the breakage–reunion domain of DNA gyrase', *Nature*, 388(6645), pp. 903–906. Available at: https://doi.org/10.1038/42294.

Cai, Y. *et al.* (2020) 'A nucleotidyltransferase toxin inhibits growth of Mycobacterium tuberculosis through inactivation of tRNA acceptor stems', *Science Advances*, 6(31). Available at: https://doi.org/10.1126/sciadv.abb6651.

Campbell, E.A. *et al.* (2001) 'Structural mechanism for rifampicin inhibition of bacterial RNA polymerase', *Cell*, 104(6). Available at: https://doi.org/10.1016/S0092-8674(01)00286-0.

Castro-Roa, D. *et al.* (2013) 'The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu', *Nature Chemical Biology*, 9(12). Available at: https://doi.org/10.1038/nchembio.1364.

Cataudella, I. *et al.* (2012) 'Conditional cooperativity in toxin–antitoxin regulation prevents random toxin activation and promotes fast translational recovery', *Nucleic Acids Research*, 40(14), pp. 6424–6434. Available at: https://doi.org/10.1093/nar/gks297.

Chackerian, A.A. *et al.* (2002) 'Dissemination of Mycobacterium tuberculosis is influenced by host factors and precedes the initiation of T-cell immunity', *Infection and Immunity*, 70(8). Available at: https://doi.org/10.1128/IAI.70.8.4501-4509.2002.

Chahine, E.B., Karaoui, L.R. and Mansour, H. (2014) 'Bedaquiline: A Novel Diarylquinoline for Multidrug-Resistant Tuberculosis', *Annals of Pharmacotherapy*. Available at: https://doi.org/10.1177/1060028013504087.

Champoux, J.J. (2001) 'DNA topoisomerases: Structure, function, and mechanism', *Annual Review of Biochemistry*. Available at: https://doi.org/10.1146/annurev.biochem.70.1.369.

Champoux, J.J. and Dulbecco, R. (1972) 'An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay).', *Proceedings of the National Academy of Sciences of the United States of America*, 69(1). Available at: https://doi.org/10.1073/pnas.69.1.143.

Chan, P.F. *et al.* (2017) 'Thiophene antibacterials that allosterically stabilize DNA-cleavage complexes with DNA gyrase', *Proceedings of the National Academy of Sciences of the United States of America*, 114(22). Available at: https://doi.org/10.1073/pnas.1700721114.

Chan, W.T., Espinosa, M. and Yeo, C.C. (2016) 'Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems', *Frontiers in Molecular Biosciences*, 3, p. 9. Available at: https://doi.org/10.3389/fmolb.2016.00009.

Chao, M.C. and Rubin, E.J. (2010) 'Letting sleeping dos lie: Does dormancy play a role in tuberculosis?', *Annual Review of Microbiology*. Available at: https://doi.org/10.1146/annurev.micro.112408.134043.

Chen, H. and Zhou, H.-X. (2005) 'Prediction of interface residues in protein-protein complexes by a consensus neural network method: Test against NMR data', *Proteins: Structure, Function, and Bioinformatics*, 61(1), pp. 21–35. Available at: https://doi.org/10.1002/prot.20514.

Chirehwa, M.T. *et al.* (2020) 'Population Pharmacokinetics of Cycloserine and Pharmacokinetic/Pharmacodynamic Target Attainment in Multidrug-Resistant Tuberculosis Patients Dosed with Terizidone', *Antimicrobial Agents and Chemotherapy*, 64(11). Available at: https://doi.org/10.1128/AAC.01381-20.

Cholo, M.C. *et al.* (2012) 'Clofazimine: Current status and future prospects', *Journal of Antimicrobial Chemotherapy*. Available at: https://doi.org/10.1093/jac/dkr444.

Cholo, M.C. *et al.* (2017) 'Mechanisms of action and therapeutic efficacies of the lipophilic antimycobacterial agents clofazimine and bedaquiline', *Journal of Antimicrobial Chemotherapy*, 72(2). Available at: https://doi.org/10.1093/jac/dkw426.

Chono, H. *et al.* (2011) 'Acquisition of HIV-1 resistance in T lymphocytes using an ACA-specific E. coli mRNA interferase', *Human Gene Therapy*, 22(1). Available at: https://doi.org/10.1089/hum.2010.001.

Chopra, S. *et al.* (2012) 'Evaluation of gyrase B as a drug target in Mycobacterium tuberculosis', *Journal of Antimicrobial Chemotherapy*, 67(2). Available at: https://doi.org/10.1093/jac/dkr449.

Christensen-Dalsgaard, M. and Gerdes, K. (2006) 'Two higBA loci in the Vibrio cholerae superintegron encode mRNA cleaving enzymes and can stabilize plasmids', *Molecular Microbiology*, 62(2). Available at: https://doi.org/10.1111/j.1365-2958.2006.05385.x.

Christensen, S.K. and Gerdes, K. (2003) 'RelE toxins from Bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA', *Molecular Microbiology*, 48(5). Available at: https://doi.org/10.1046/j.1365-2958.2003.03512.x.

Churchyard, G. *et al.* (2017) 'What We Know about Tuberculosis Transmission: An Overview', *Journal of Infectious Diseases*. Available at: https://doi.org/10.1093/infdis/jix362.

Clay, H., Volkman, H.E. and Ramakrishnan, L. (2008) 'Tumor Necrosis Factor Signalling Mediates Resistance to Mycobacteria by Inhibiting Bacterial Growth and Macrophage Death', *Immunity*, 29(2). Available at: https://doi.org/10.1016/j.immuni.2008.06.011.

Cole, S.T. *et al.* (1998) 'Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence', *Nature*. Available at: https://doi.org/10.1038/31159.

Collin, F. and Maxwell, A. (2019) 'The Microbial Toxin Microcin B17: Prospects for the Development of New Antibacterial Agents', *Journal of Molecular Biology*. Available at: https://doi.org/10.1016/j.jmb.2019.05.050.

Commandeur, S. *et al.* (2013) ' An Unbiased Genome-Wide Mycobacterium tuberculosis Gene Expression Approach To Discover Antigens Targeted by Human T Cells Expressed during Pulmonary Infection ', *The Journal of Immunology*, 190(4). Available at: https://doi.org/10.4049/jimmunol.1201593.

Corbett, K.D. (2006) ' DNA Topology . By Andrew D Bates and , Anthony Maxwell. Oxford and New York: Oxford University Press. xviii + 198 p; ill.; index. ISBN: 0–19–856709–X (hc); 0–19–850655–4 (pb). [Original version published in 1993, as part of the "In Focus" series.] 2005. ', *The Quarterly Review of Biology*, 81(1). Available at: https://doi.org/10.1086/503941.

Corbett, K.D. and Berger, J.M. (2003) 'Structure of the topoisomerase VI-B subunit: Implications for type II topoisomerase mechanism and evolution', *EMBO Journal*, 22(1). Available at: https://doi.org/10.1093/emboj/cdg008.

Corbett, K.D. and Berger, J.M. (2004) 'Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases', *Annual Review of Biophysics and Biomolecular Structure*. Available at: https://doi.org/10.1146/annurev.biophys.33.110502.140357.

Corbett, K.D. *et al.* (2005) 'The structural basis for substrate specificity in DNA topoisomerase IV', *Journal of Molecular Biology*, 351(3). Available at: https://doi.org/10.1016/j.jmb.2005.06.029.

Corbett, K.D., Shultzaberger, R.K. and Berger, J.M. (2004) 'The C-terminal domain of DNA gyrase A adopts a DNA-bending β-pinwheel fold', *Proceedings of the National Academy of Sciences of the United States of America*, 101(19). Available at: https://doi.org/10.1073/pnas.0401595101.

Cortes, T. *et al.* (2013) 'Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in Mycobacterium tuberculosis', *Cell Reports*, 5(4), pp. 1121–1131. Available at: https://doi.org/10.1016/J.CELREP.2013.10.031.

Costa, M. *et al.* (2001) 'Plasmid transcriptional repressor copG oligomerises to render helical superstructures unbound and in complexes with oligonucleotides', *Journal of Molecular Biology*, 310(2). Available at: https://doi.org/10.1006/jmbi.2001.4760.

Cowtan, K. and IUCr (2006) 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta Crystallographica Section D Biological Crystallography*, 62(9), pp. 1002–1011. Available at: https://doi.org/10.1107/S0907444906022116.

Crisona, N.J. *et al.* (2000) 'Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements', *Genes and Development*, 14(22). Available at: https://doi.org/10.1101/gad.838900.

Cruz, J.W. *et al.* (2014) 'Doc toxin is a kinase that inactivates elongation factor Tu', *Journal of Biological Chemistry*, 289(11). Available at: https://doi.org/10.1074/jbc.M113.544429.

Cruz, J.W. *et al.* (2015) 'Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase', *Nature Communications*, 6. Available at: https://doi.org/10.1038/ncomms8480.

Culviner, P.H. and Laub, M.T. (2018) 'Global Analysis of the E. coli Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis', *Molecular Cell*, 70(5). Available at: https://doi.org/10.1016/j.molcel.2018.04.026.

Cummings, M.P. (2004) 'FigTree', in *Dictionary of Bioinformatics and Computational Biology*. Available at: https://doi.org/10.1002/9780471650126.dob0904.

Dawson, C.C. *et al.* (2022) 'Discovery of a novel type IIb RelBE toxin-antitoxin system in Mycobacterium tuberculosis defined by co-regulated with an antisense RNA', *Molecular Micobiology*. Available at: https://doi.org/10.1111/mmi.14917.

Dahl, J.L. *et al.* (2003) 'The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 100(17). Available at: https://doi.org/10.1073/pnas.1631248100.

Dalton, K.M. and Crosson, S. (2010) 'A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex', *Biochemistry*, 49(10), p. 2205. Available at: https://doi.org/10.1021/BI902133S.

Daniel, J. *et al.* (2011) 'Mycobacterium tuberculosis uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages', *PLoS Pathogens*, 7(6). Available at: https://doi.org/10.1371/journal.ppat.1002093.

Danishuddin, M. and Khan, A.U. (2012) 'Molecular modelling and docking analysis of Beta-lactamases with inhibitors: A comparative study', *In Silico Biology*, 11(5). Available at: https://doi.org/10.3233/ISB-2012-0443.

Dao-Thi, M.H. *et al.* (2005) 'Molecular basis of gyrase poisoning by the addiction toxin CcdB', *Journal of Molecular Biology*, 348(5). Available at: https://doi.org/10.1016/j.jmb.2005.03.049.

Dar, A. *et al.* (2009) 'A unique 45-amino-acid region in the toprim domain of Plasmodium falciparum gyrase B is essential for its activity', *Eukaryotic Cell*, 8(11). Available at: https://doi.org/10.1128/EC.00149-09.

Das, U. *et al.* (2014) 'Crystal structure of the VapBc-15 complex from mycobacterium tuberculosis reveals a two-metal ion dependent pin-domain ribonuclease and a variable mode of toxin-antitoxin assembly', *Journal of Structural Biology*, 188(3). Available at: https://doi.org/10.1016/j.jsb.2014.10.002.

Davis, J.M. and Ramakrishnan, L. (2009) 'The Role of the Granuloma in Expansion and Dissemination of Early Tuberculous Infection', *Cell*, 136(1). Available at: https://doi.org/10.1016/j.cell.2008.11.014.

Deghorain, M. *et al.* (2013) 'Type II toxin-antitoxin loci: The ccdAB and parDE families', in *Prokaryotic Toxin-Antitoxins*. Available at: https://doi.org/10.1007/978-3-642-33253-1_4.

Dejesus, M.A. *et al.* (2017) 'Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis', *mBio*. Edited by C.L. Stallings, 8(1), pp. e02133-16. Available at: https://doi.org/10.1128/mBio.02133-16.

del Portillo, P. *et al.* (2019) 'Hypoxia is not a main stress when Mycobacterium tuberculosis is in a dormancy-like long-chain fatty acid environment', *Frontiers in Cellular and Infection Microbiology*, 9(JAN). Available at: https://doi.org/10.3389/fcimb.2018.00449.

Demirci, H. *et al.* (2013) 'A structural basis for streptomycin-induced misreading of the genetic code', *Nature Communications*, 4. Available at: https://doi.org/10.1038/ncomms2346.

Deng, J., Toledo, R.T. and Lillard, D.A. (1976) 'Effect of temperature and ph on protein-protein interaction in actomyosin solutions', *Journal of Food Science*, 41(2). Available at: https://doi.org/10.1111/j.1365-2621.1976.tb00599.x.

Dessen, A. *et al.* (1995) 'Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis', *Science*, 267(5204). Available at: https://doi.org/10.1126/science.7886450.

Dienemann, C. *et al.* (2011) 'Crystal structure of the VapBC toxin-antitoxin complex from Shigella flexneri reveals a hetero-octameric DNA-binding assembly.', *Journal of molecular biology*, 414(5), pp. 713–22. Available at: https://doi.org/10.1016/j.jmb.2011.10.024.

DiGate, R.J. and Marians, K.J. (1992) 'Escherichia coli topoisomerase III-catalyzed cleavage of RNA', *Journal of Biological Chemistry*, 267(29). Available at: https://doi.org/10.1016/s0021-9258(19)36714-6.

Divangahi, M. *et al.* (2010) 'Eicosanoid pathways regulate adaptive immunity to Mycobacterium tuberculosis', *Nature Immunology*, 11(8). Available at: https://doi.org/10.1038/ni.1904.

Dong, K.C. and Berger, J.M. (2007) 'Structural basis for gate-DNA recognition and bending by type IIA topoisomerases', *Nature*. Available at: https://doi.org/10.1038/nature06396.

Drlica, K. *et al.* (2008) 'Quinolone-mediated bacterial death', *Antimicrobial Agents and Chemotherapy*. Available at: https://doi.org/10.1128/AAC.01617-06.

Drolet, M. *et al.* (2003) 'The problem of hypernegative supercoiling and R-loop formation in transcription', *Frontiers in Bioscience* [Preprint]. Available at: https://doi.org/10.2741/970.

Ehrt, S. and Schnappinger, D. (2009) 'Mycobacterial survival strategies in the phagosome: Defence against host stresses', *Cellular Microbiology*. Available at: https://doi.org/10.1111/j.1462-5822.2009.01335.x.

el Sayyed, H. *et al.* (2016) 'Mapping Topoisomerase IV Binding and Activity Sites on the E. coli Genome', *PLoS Genetics*, 12(5). Available at: https://doi.org/10.1371/journal.pgen.1006025.

Emery, J.C. *et al.* (2021) 'Self-clearance of Mycobacterium tuberculosis infection: Implications for lifetime risk and population at-risk of tuberculosis disease', *Proceedings of the Royal Society B: Biological Sciences*, 288(1943). Available at: https://doi.org/10.1098/rspb.2020.1635.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta Crystallogr D Biol Crystallogr*. 2004/12/02, 60(Pt 12 Pt 1), pp. 2126–2132. Available at: https://doi.org/S0907444904019158 [pii] 10.1107/S0907444904019158.

Engelberg-Kulka, H., Hazan, R. and Amitai, S. (2005) 'mazEF: A chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria', *Journal of Cell Science*. Available at: https://doi.org/10.1242/jcs.02619.

Ernst, J.D. (2012) 'The immunological life cycle of tuberculosis', *Nature Reviews Immunology*. Available at: https://doi.org/10.1038/nri3259.

Eun, H.J. *et al.* (2020) 'Crystal structure of the YoeBSa1-YefMSa1 complex from Staphylococcus aureus', *Biochemical and Biophysical Research Communications*, 527(1). Available at: https://doi.org/10.1016/j.bbrc.2020.04.089.

Evans, P.R., Murshudov, G.N. and IUCr (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D Biological Crystallography*, 69(7), pp. 1204–1214. Available at: https://doi.org/10.1107/S0907444913000061.

Evans, R. *et al.* (2022) 'Protein complex prediction with AlphaFold-Multimer', *bioRxiv*.

Fauci, A.S. and Morens, D.M. (2012) 'The Perpetual Challenge of Infectious Diseases', *New England Journal of Medicine*, 366(5). Available at: https://doi.org/10.1056/nejmra1108296.

Feng, L. *et al.* (2021) 'The pentapeptide-repeat protein, MfpA, interacts with mycobacterial DNA gyrase as a DNA T-segment mimic', *Proceedings of the National Academy of Sciences of the United States of America*, 118(11). Available at: https://doi.org/10.1073/pnas.2016705118.

Fernández-García, L. *et al.* (2016) 'Toxin-Antitoxin Systems in Clinical Pathogens', *Toxins*, 8(7), p. 227. Available at: https://doi.org/10.3390/toxins8070227.

Fiebig, A. *et al.* (2010) 'Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems', *Molecular Microbiology*, 77(1). Available at: https://doi.org/10.1111/j.1365-2958.2010.07207.x.

Fineran, P.C. *et al.* (2009) 'The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair', *Proc Natl Acad Sci U S A*. 2009/01/07, 106(3), pp. 894–899. Available at: https://doi.org/0808832106 [pii] 10.1073/pnas.0808832106.

Flatman, R.H. *et al.* (2005) 'Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action', *Antimicrobial Agents and Chemotherapy*, 49(3). Available at: https://doi.org/10.1128/AAC.49.3.1093-1100.2005.

Fleming, P.J. and Fleming, K.G. (2018) 'HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties', *Biophysical Journal*, 114(4). Available at: https://doi.org/10.1016/j.bpj.2018.01.002.

Fogel, N. (2015) 'Tuberculosis: A disease without boundaries', *Tuberculosis*. Available at: https://doi.org/10.1016/j.tube.2015.05.017.

Ford, C.B. *et al.* (2011) 'Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection', *Nature Genetics*, 43(5). Available at: https://doi.org/10.1038/ng.811.

Forterre, P. (2002) 'A hot story from comparative genomics: Reverse gyrase is the only hyperthermophile-specific protein', *Trends in Genetics*. Available at: https://doi.org/10.1016/S0168-9525(02)02650-1.

Fozo, E.M. *et al.* (2010) 'Abundance of type I toxin-antitoxin systems in bacteria: Searches for new candidates and discovery of novel families', *Nucleic Acids Research*, 38(11). Available at: https://doi.org/10.1093/nar/gkq054.

Fraikin, N., Goormaghtigh, F. and van Melderen, L. (2020) 'Type II Toxin-Antitoxin Systems: Evolution and Revolutions.', *Journal of bacteriology*, 202(7). Available at: https://doi.org/10.1128/JB.00763-19.

Francuski, D. and Saenger, W. (2009) 'Crystal Structure of the Antitoxin-Toxin Protein Complex RelB-RelE from Methanococcus jannaschii', *Journal of Molecular Biology*, 393(4). Available at: https://doi.org/10.1016/j.jmb.2009.08.048.

Freire, D.M. *et al.* (2019) 'An NAD + Phosphorylase Toxin Triggers Mycobacterium tuberculosis Cell Death', *Molecular Cell*, 73(6). Available at: https://doi.org/10.1016/j.molcel.2019.01.028.

Gadelle, D. *et al.* (2014) 'DNA topoisomerase VIII: A novel subfamily of type IIB topoisomerases encoded by free or integrated plasmids in Archaea and Bacteria', *Nucleic Acids Research*, 42(13). Available at: https://doi.org/10.1093/nar/gku568.

Garces, A. *et al.* (2010) 'EspA acts as a critical mediator of ESX1-dependent virulence in Mycobacterium tuberculosis by affecting bacterial cell wall integrity', *PLoS Pathogens*, 6(6). Available at: https://doi.org/10.1371/journal.ppat.1000957.

Garcia-Pino, A *et al.* (2010) 'Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity', *Cell*. 2010/07/07, 142(1), pp. 101–111. Available at: https://doi.org/S0092-8674(10)00614-8 [pii] 10.1016/j.cell.2010.05.039.

Garcia-Pino, A. *et al.* (2016) 'An intrinsically disordered entropic switch determines allostery in Phd–Doc regulation', *Nature Chemical Biology*, 12(7), pp. 490–496. Available at: https://doi.org/10.1038/nchembio.2078.

Garcia-Rodriguez, G. *et al.* (2021) 'Entropic pressure controls the oligomerization of the Vibrio cholerae ParD2 antitoxin', *Acta Crystallographica Section D Structural Biology*, 77(7). Available at: https://doi.org/10.1107/s2059798321004873.

Garcia-Rodriguez, G. *et al.* (2021) 'Vibrio cholerae ParE2 toxin modulates its operon transcription by stabilization of an antitoxin DNA ruler', *bioRxiv*.

Gasteiger, E. *et al.* (1999) 'Protein identification and analysis tools in the ExPASy server.', *Methods in molecular biology (Clifton, N.J.)*, 112.

Gebreweld, F.H. *et al.* (2018) 'Factors influencing adherence to tuberculosis treatment in Asmara, Eritrea: A qualitative study', *Journal of Health, Population and Nutrition*, 37(1). Available at: https://doi.org/10.1186/s41043-017-0132-y.

Geldmacher, C. *et al.* (2010) 'Preferential infection and depletion of Mycobacterium tuberculosis-specific CD4 T cells after HIV-1 infection', *Journal of Experimental Medicine*, 207(13). Available at: https://doi.org/10.1084/jem.20100090.

Gellert, M. *et al.* (1976) 'DNA gyrase: an enzyme that introduces superhelical turns into DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 73(11). Available at: https://doi.org/10.1073/pnas.73.11.3872.

Gerdes, K. *et al.* (1986) 'Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon.', *The EMBO journal*, 5(8). Available at: https://doi.org/10.1002/j.1460-2075.1986.tb04459.x.

Gerdes, K., Rasmussen, P.B. and Molin, S. (1986) 'Unique type of plasmid maintenance function: Postsegregational killing of plasmid-free cells', *Proceedings of the National Academy of Sciences of the United States of America*, 83(10). Available at: https://doi.org/10.1073/pnas.83.10.3116.

Gibson, E.G. *et al.* (2018) 'Mechanism of Action of Mycobacterium tuberculosis Gyrase Inhibitors: A Novel Class of Gyrase Poisons', *ACS Infectious Diseases*, 4(8). Available at: https://doi.org/10.1021/acsinfecdis.8b00035.

Gone, S. *et al.* (2016) 'Mechanism of Ribonuclease III Catalytic Regulation by Serine Phosphorylation', *Scientific Reports*, 6. Available at: https://doi.org/10.1038/srep25448.

Gopal, P. *et al.* (2020) 'Pyrazinamide triggers degradation of its target aspartate decarboxylase', *Nature Communications*, 11(1). Available at: https://doi.org/10.1038/s41467-020-15516-1.

Gordon, S. v. and Parish, T. (2018) 'Microbe profile: Mycobacterium tuberculosis: Humanity's deadly microbial foe', *Microbiology (United Kingdom)*, 164(4). Available at: https://doi.org/10.1099/mic.0.000601.

Gore, S., Velankar, S. and Kleywegt, G.J. (2012) 'Implementing an X-ray validation pipeline for the Protein Data Bank', *Acta Crystallographica Section D Biological Crystallography*, 68(4), pp. 478–483. Available at: https://doi.org/10.1107/S0907444911050359.

Gotfredsen, M. and Gerdes, K. (1998) 'The *Escherichia coli relBE* genes belong to a new toxin-antitoxin gene family', *Molecular Microbiology*, 29(4), pp. 1065–1076. Available at: https://doi.org/10.1046/j.1365-2958.1998.00993.x.

Goto, T. and Wang, J.C. (1982) 'Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings.', *Journal of Biological Chemistry*, 257(10). Available at: https://doi.org/10.1016/s0021-9258(19)83859-0.

Goude, R. *et al.* (2009) 'The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 53(10). Available at: https://doi.org/10.1128/AAC.00162-09.

Graille, M. *et al.* (2008) 'Crystal Structure of an Intact Type II DNA Topoisomerase: Insights into DNA Transfer Mechanisms', *Structure*, 16(3). Available at: https://doi.org/10.1016/j.str.2007.12.020.

Greenfield, N.J. (2007) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols*, 1(6). Available at: https://doi.org/10.1038/nprot.2006.202.

Griffin, J.E. *et al.* (2011) 'High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.', *PLoS pathogens*, 7(9), p. e1002251. Available at: https://doi.org/10.1371/journal.ppat.1002251.

Grue, P. *et al.* (1998) 'Essential mitotic functions of DNA topoisomerase IIα are not adopted by topoisomerase IIβ in human H69 cells', *Journal of Biological Chemistry*, 273(50). Available at: https://doi.org/10.1074/jbc.273.50.33660.

GSK. (2019). GSK starts a phase III clinical programme for a potential first-in-class antibiotic, gepotidacin. [Online]. GlaxoSmithKline. Last Updated: October 2019. Available at: https://www.gsk.com/en-gb/media/press-releases/gsk-starts-a-phase-iii-clinical-programme-for-a-poten [Accessed 15 November 2022].

GSK. (2022). EAGLE-2 and EAGLE-3 phase III trials for gepotidacin stopped early for efficacy following pre-planned interim analysis b. [Online]. GlaxoSmithKline. Last Updated: November 2022. Available at: https://www.gsk.com/en-gb/media/press-releases/gsk-announces-phase-iii-trials-for-gepotidacin/ [Accessed 15 November 2022].

Gucinski, G.C. *et al.* (2019) 'Convergent Evolution of the Barnase/EndoU/Colicin/RelE (BECR) Fold in Antibacterial tRNase Toxins', *Structure*, 27(11). Available at: https://doi.org/10.1016/j.str.2019.08.010.

Guglielmini, J. and van Melderen, L. (2011) 'Bacterial toxin-antitoxin systems', *Mobile Genetic Elements*, 1(4), pp. 283–306. Available at: https://doi.org/10.4161/mge.18477.

Gundry, R.L. *et al.* (2009) 'Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow', *Current Protocols in Molecular Biology* [Preprint], (SUPPL. 88). Available at: https://doi.org/10.1002/0471142727.mb1025s88.

Guo, Y. *et al.* (2014) 'RalR (a DNase) and RalA (a small RNA) form a type I toxin–antitoxin system in Escherichia coli', *Nucleic Acids Research*, 42(10), pp. 6448–6462. Available at: https://doi.org/10.1093/nar/gku279.

Gupta, A. *et al.* (2017) 'Co-expression network analysis of toxin-antitoxin loci in Mycobacterium tuberculosis reveals key modulators of cellular stress', *Scientific Reports*, 7(1), p. 5868. Available at: https://doi.org/10.1038/s41598-017-06003-7.

Gupta, M. *et al.* (2016) 'The chromosomal parDE2 toxin-antitoxin system of Mycobacterium tuberculosis H37Rv: Genetic and functional characterization', *Frontiers in Microbiology*, 7(JUN). Available at: https://doi.org/10.3389/fmicb.2016.00886.

Gurnev, P.A. *et al.* (2012) 'Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers', *FEBS Letters*, 586(16). Available at: https://doi.org/10.1016/j.febslet.2012.06.021.

Hallez, R. *et al.* (2010) 'New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in Escherichia coli O157:H7', *Molecular Microbiology*, 76(3). Available at: https://doi.org/10.1111/j.1365-2958.2010.07129.x.

Hampton, H.G. *et al.* (2018) 'AbiEi Binds Cooperatively to the Type IV abiE Toxin–Antitoxin Operator Via a Positively-Charged Surface and Causes DNA Bending and Negative Autoregulation', *Journal of Molecular Biology*, 430(8), pp. 1141–1156. Available at: https://doi.org/10.1016/j.jmb.2018.02.022.

Harms, A. *et al.* (2015) 'Adenylylation of Gyrase and Topo IV by FicT Toxins Disrupts Bacterial DNA Topology', *Cell Reports*, 12(9). Available at: https://doi.org/10.1016/j.celrep.2015.07.056.

Harms, A. *et al.* (2017) 'A bacterial toxin-antitoxin module is the origin of inter-bacterial and inter-kingdom effectors of Bartonella', *PLoS Genetics*, 13(10). Available at: https://doi.org/10.1371/journal.pgen.1007077.

Hashimi, S.M. (2019) 'Albicidin, a potent DNA gyrase inhibitor with clinical potential', *Journal of Antibiotics*. Available at: https://doi.org/10.1038/s41429-019-0228-2.

Hashimi, S.M. *et al.* (2007) 'The phytotoxin albicidin is a novel inhibitor of DNA gyrase', *Antimicrobial Agents and Chemotherapy*, 51(1). Available at: https://doi.org/10.1128/AAC.00918-06.

Hazan, R., Sat, B. and Engelberg-Kulka, H. (2004) 'Escherichia coli mazEF-Mediated Cell Death Is Triggered by Various Stressful Conditions', *Journal of Bacteriology*, 186(11), pp. 3663–3669. Available at: https://doi.org/10.1128/JB.186.11.3663-3669.2004.

Heddle, J.G. *et al.* (2001) 'The antibiotic microcin B17 is a DNA gyrase poison: Characterisation of the mode of inhibition', *Journal of Molecular Biology*, 307(5). Available at: https://doi.org/10.1006/jmbi.2001.4562.

Hegde, S.S. *et al.* (2005) 'Biochemistry: A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA', *Science*, 308(5727). Available at: https://doi.org/10.1126/science.1110699.

Herrero, M. and Moreno, F. (1986) 'Microcin B17 blocks DNA replication and induces the SOS system in Escherichia coli', *Journal of General Microbiology*, 132(2). Available at: https://doi.org/10.1099/00221287-132-2-393.

Hess, P.N. and De Moraes Russo, C.A. (2007) 'An empirical test of the midpoint rooting method', *Biological Journal of the Linnean Society*, 92(4). Available at: https://doi.org/10.1111/j.1095-8312.2007.00864.x.

Hiasa, H. and Marians, K.J. (1996) 'Two distinct modes of strand unlinking during θ-type DNA replication', *Journal of Biological Chemistry*, 271(35). Available at: https://doi.org/10.1074/jbc.271.35.21529.

Hinchey, J. *et al.* (2007) 'Enhanced priming of adaptive immunity by a proapoptotic mutant of Mycobacterium tuberculosis', *Journal of Clinical Investigation*, 117(8). Available at: https://doi.org/10.1172/JCI31947.

Hobson, M.J. and Berger, J.M. (2019) 'Caught in the Open: A Domain Insertion of M. tuberculosis Gyrase Suppresses ATPase Dimerization', *Structure*, 27(4), pp. 561–563. Available at: https://doi.org/10.1016/J.STR.2019.03.010.

Hobson, M.J., Bryant, Z. and Berger, J.M. (2020) 'Modulated control of DNA supercoiling balance by the DNA-wrapping domain of bacterial gyrase', *Nucleic Acids Research*, 48(4), pp. 2035–2049. Available at: https://doi.org/10.1093/nar/gkz1230.

Holm, C. *et al.* (1985) 'DNA topoisomerase II is required at the time of mitosis in yeast', *Cell*, 41(2). Available at: https://doi.org/10.1016/S0092-8674(85)80028-3.

Holm, L. (2020) 'Using Dali for Protein Structure Comparison', in *Methods in Molecular Biology*. Available at: https://doi.org/10.1007/978-1-0716-0270-6_3.

Holm, L. and Sander, C. (1993) 'Protein structure comparison by alignment of distance matrices', *J Mol Biol*. 1993/09/05, 233(1), pp. 123–138. Available at: https://doi.org/S0022-2836(83)71489-0 [pii] 10.1006/jmbi.1993.1489.

Hudock, T.A. *et al.* (2017) 'Hypoxia sensing and persistence genes are expressed during the intragranulomatous survival of mycobacterium tuberculosis', *American Journal of Respiratory Cell and Molecular Biology*, 56(5). Available at: https://doi.org/10.1165/rcmb.2016-0239OC.

Hugonnet, J.E. *et al.* (2009) 'Meropenem-clavulanate is effective against extensively drug-resistant Mycobacterium tuberculosis', *Science*, 323(5918). Available at: https://doi.org/10.1126/science.1167498.

Hurley, J.M. and Woychik, N.A. (2009) 'Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites', *Journal of Biological Chemistry*, 284(28). Available at: https://doi.org/10.1074/jbc.M109.008763.

Huse, M. *et al.* (2006) 'T cells use two directionally distinct pathways for cytokine secretion', *Nature Immunology*, 7(3). Available at: https://doi.org/10.1038/ni1304.

Imai, Y. *et al.* (2022) 'Evybactin is a DNA gyrase inhibitor that selectively kills Mycobacterium tuberculosis', *Nature Chemical Biology* [Preprint]. Available at: https://doi.org/10.1038/s41589-022-01102-7.

Irvine, G.B. (2000) 'Determination of Molecular Size by Size-Exclusion Chromatography (Gel Filtration)', *Current Protocols in Cell Biology*, 6(1). Available at: https://doi.org/10.1002/0471143030.cb0505s06.

Izudi, J., Tamwesigire, I.K. and Bajunirwe, F. (2020) 'Treatment success and mortality among adults with tuberculosis in rural eastern Uganda: A retrospective cohort study', *BMC Public Health*, 20(1). Available at: https://doi.org/10.1186/s12889-020-08646-0.

Jaffé, A., Ogura, T. and Hiraga, S. (1985) 'Effects of the ccd function of the F plasmid on bacterial growth.', *Journal of bacteriology*, 163(3), pp. 841–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3897195 (Accessed: 13 March 2019).

Jain, S.K. *et al.* (2007) 'Accelerated detection of Mycobacterium tuberculosis genes essential for bacterial survival in guinea pigs, compared with mice', *Journal of Infectious Diseases*, 195(11). Available at: https://doi.org/10.1086/517526.

Jankevicius, G. *et al.* (2016) 'The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA', *Molecular Cell*, 64(6). Available at: https://doi.org/10.1016/j.molcel.2016.11.014.

Janowski, R. *et al.* (2009) 'Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from Mycobacterium tuberculosis', *Journal of Structural and Functional Genomics*, 10(2), pp. 137–150. Available at: https://doi.org/10.1007/s10969-009-9060-4.

Jayakumar, D., Jacobs, W.R. and Narayanan, S. (2008) 'Protein kinase E of Mycobacterium tuberculosis has a role in the nitric oxide stress response and apoptosis in a human macrophage model of infection', *Cellular Microbiology*, 10(2). Available at: https://doi.org/10.1111/j.1462-5822.2007.01049.x.

Jiang, Y. *et al.* (2002) 'ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase', *Molecular Microbiology*, 44(4). Available at: https://doi.org/10.1046/j.1365-2958.2002.02921.x.

Johnson, A. and O'Donnell, M. (2005) 'Cellular DNA replicases: Components and dynamics at the replication fork', *Annual Review of Biochemistry*. Available at: https://doi.org/10.1146/annurev.biochem.73.011303.073859.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873). Available at: https://doi.org/10.1038/s41586-021-03819-2.

Jurėnas, D. and van Melderen, L. (2020) 'The Variety in the Common Theme of Translation Inhibition by Type II Toxin–Antitoxin Systems', *Frontiers in Genetics*. Available at: https://doi.org/10.3389/fgene.2020.00262.

Jurėnas, D. *et al.* (2022) 'Biology and evolution of bacterial toxin–antitoxin systems', *Nature Reviews Microbiology*. Available at: https://doi.org/10.1038/s41579-021-00661-1.

Jurėnas, D., van Melderen, L. and Garcia-Pino, A. (2019) 'Mechanism of regulation and neutralization of the AtaR–AtaT toxin–antitoxin system', *Nature Chemical Biology*, 15(3). Available at: https://doi.org/10.1038/s41589-018-0216-z.

Kabir, S. *et al.* (2020) 'Fluoroquinolone resistance and mutational profile of gyrA in pulmonary MDR tuberculosis patients', *BMC Pulmonary Medicine*, 20(1). Available at: https://doi.org/10.1186/s12890-020-1172-4.

Kabsch, W. (2010a) 'Integration, scaling, space-group assignment and post-refinement.', *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp. 133–44. Available at: https://doi.org/10.1107/S0907444909047374.

Kabsch, W. (2010b) 'XDS.', *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp. 125–32. Available at: https://doi.org/10.1107/S0907444909047337.

Kamada, K., Hanaoka, F. and Burley, S.K. (2003) 'Crystal structure of the MazE/MazF complex: Molecular bases of antidote-toxin recognition', *Molecular Cell*, 11(4). Available at: https://doi.org/10.1016/S1097-2765(03)00097-2.

Kampranis, S.C. and Maxwell, A. (1996) 'Conversion of DNA gyrase into a conventional type II topoisomerase', *Proceedings of the National Academy of Sciences of the United States of America*, 93(25). Available at: https://doi.org/10.1073/pnas.93.25.14416.

Kampranis, S.C., Bates, A.D. and Maxwell, A. (1999) 'A model for the mechanism of strand passage by DNA gyrase', *Proceedings of the National Academy of Sciences of the United States of America*, 96(15). Available at: https://doi.org/10.1073/pnas.96.15.8414.

Kamruzzaman, M. and Iredell, J. (2019) 'A ParDE-family toxin antitoxin system in major resistance plasmids of Enterobacteriaceae confers antibiotic and heat tolerance', *Scientific Reports*, 9(1), p. 9872. Available at: https://doi.org/10.1038/s41598-019-46318-1.

Kana, B.D. *et al.* (2008) 'The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro', *Molecular Microbiology*, 67(3). Available at: https://doi.org/10.1111/j.1365-2958.2007.06078.x.

Kana, B.D., Mizrahi, V. and Gordhan, B.G. (2010) 'Depletion of resuscitation-promoting factors has limited impact on the drug susceptibility of Mycobacterium tuberculosis', *Journal of Antimicrobial Chemotherapy*, 65(8). Available at: https://doi.org/10.1093/jac/dkq199.

Kang, D.D. *et al.* (2011) 'Profiling early lung immune responses in the mouse model of tuberculosis', *PLoS ONE*, 6(1). Available at: https://doi.org/10.1371/journal.pone.0016161.

Kaplan, G. *et al.* (2003) 'Mycobacterium tuberculosis Growth at the Cavity Surface: A Microenvironment with Failed Immunity', *Infection and Immunity*, 71(12). Available at: https://doi.org/10.1128/IAI.71.12.7099-7108.2003.

Kapopoulou, A., Lew, J.M. and Cole, S.T. (2011) 'The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes', *Tuberculosis*, 91(1). Available at: https://doi.org/10.1016/j.tube.2010.09.006.

Karkare, S., Yousafzai, F., Mitchenall, Lesley A, *et al.* (2012) 'The role of Ca^{2+} in the activity of Mycobacterium tuberculosis DNA gyrase.', *Nucleic acids research*, 40(19), pp. 9774–87. Available at: https://doi.org/10.1093/nar/gks704.

Kaspy, I. *et al.* (2013) 'HipA-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase', *Nature Communications*, 4. Available at: https://doi.org/10.1038/ncomms4001.

Kato, J.I., Suzuki, H. and Ikeda, H. (1992) 'Purification and characterization of DNA topoisomerase IV in Escherichia coli', *Journal of Biological Chemistry*, 267(36). Available at: https://doi.org/10.1016/s0021-9258(18)35660-6.

Kawano, M., Aravind, L. and Storz, G. (2007) 'An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin', *Molecular Microbiology*, 64(3), pp. 738–754. Available at: https://doi.org/10.1111/j.1365-2958.2007.05688.x.

Kedzierska, B. and Hayes, F. (2016) 'Emerging roles of toxin-antitoxin modules in bacterial pathogenesis', *Molecules*. Available at: https://doi.org/10.3390/molecules21060790.

Keren, I. *et al.* (2011) 'Characterization and transcriptome analysis of Mycobacterium tuberculosis persisters.', *mBio*, 2(3), pp. e00100-11. Available at: https://doi.org/10.1128/mBio.00100-11.

Khan, N. *et al.* (2017) 'T-cell exhaustion in tuberculosis: pitfalls and prospects', *Critical Reviews in Microbiology*. Available at: https://doi.org/10.1080/1040841X.2016.1185603.

Kikuchi, A. and Asai, K. (1984) 'Reverse gyrase - A topoisomerase which introduces positive superhelical turns into DNA', *Nature*, 309(5970). Available at: https://doi.org/10.1038/309677a0.

Kim, J.S. and Wood, T.K. (2016) 'Persistent persister misperceptions', *Frontiers in Microbiology*. Available at: https://doi.org/10.3389/fmicb.2016.02134.

King, D.E., Malone, R. and Lilley, S.H. (2000) 'New classification and update on the quinolone antibiotics', *American Family Physician*, 61(9).

Kirkegaard, K. and Wang, J.C. (1985) 'Bacterial DNA topoisomerase I can relax positively supercoiled DNA containing a single-stranded loop', *Journal of Molecular Biology*, 185(3). Available at: https://doi.org/10.1016/0022-2836(85)90075-0.

Knight, G.M. *et al.* (2019) 'Global burden of latent multidrug-resistant tuberculosis: trends and estimates based on mathematical modelling', *The Lancet Infectious Diseases*, 19(8). Available at: https://doi.org/10.1016/S1473-3099(19)30307-X.

Korch, S.B., Contreras, H. and Clark-Curtiss, J.E. (2009) 'Three Mycobacterium tuberculosis rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages', *Journal of Bacteriology*, 191(5). Available at: https://doi.org/10.1128/JB.01318-08.

Kreuzer, K.N. and Cozzarelli, N.R. (1979) 'Escherichia coli mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: Effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth', *Journal of Bacteriology*, 140(2). Available at: https://doi.org/10.1128/jb.140.2.424-435.1979.

Krissinel, E. and Henrick, K. (2007) 'Inference of Macromolecular Assemblies from Crystalline State', *Journal of Molecular Biology*, 372(3). Available at: https://doi.org/10.1016/j.jmb.2007.05.022.

Kristoffersen, P. *et al.* (2000) 'Bacterial toxin-antitoxin gene system as containment control in yeast cells', *Applied and Environmental Microbiology*, 66(12). Available at: https://doi.org/10.1128/AEM.66.12.5524-5526.2000.

Kumar, K. *et al.* (2010) 'Discovery of anti-TB agents that target the cell-division protein FtsZ', *Future Medicinal Chemistry*. Available at: https://doi.org/10.4155/fmc.10.220.

Kumar, K., McHugh, T.D. and Lipman, M. (2017) 'Fluoroquinolones for treating tuberculosis', *Clinical Pharmacist*, 9(5). Available at: https://doi.org/10.1211/CP.2017.20202555.

Kumar, P. *et al.* (2008) 'Crystal Structure of Mycobacterium tuberculosis YefM Antitoxin Reveals that it is Not an Intrinsically Unstructured Protein', *Journal of Molecular Biology*, 383(3). Available at: https://doi.org/10.1016/j.jmb.2008.08.067.

Lam, C. wan, Yeung, W. lan and Law, C. yiu (2017) 'Global developmental delay and intellectual disability associated with a de novo TOP2B mutation', *Clinica Chimica Acta*, 469. Available at: https://doi.org/10.1016/j.cca.2017.03.022.

Lambert, M.P. and Neuhaus, F.C. (1972) 'Mechanism of D-cycloserine action: alanine racemase from Escherichia coli W.', *Journal of Bacteriology*, 110(3). Available at: https://doi.org/10.1128/jb.110.3.978-987.1972.

Landau, M. *et al.* (2005) 'ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures', *Nucleic Acids Research*, 33(SUPPL. 2). Available at: https://doi.org/10.1093/nar/gki370.

Lanz, M.A. and Klostermeier, D. (2012) 'The GyrA-box determines the geometry of DNA bound to gyrase and couples DNA binding to the nucleotide cycle', *Nucleic Acids Research*, 40(21). Available at: https://doi.org/10.1093/nar/gks852.

Laponogov, I. *et al.* (2009) 'Structural insight into the quinolone-DNA cleavage complex of type IIA topoisomerases', *Nature Structural and Molecular Biology*, 16(6). Available at: https://doi.org/10.1038/nsmb.1604.

Lee, C.M. *et al.* (2019) 'Topoisomerase III acts at the replication fork to remove precatenanes', *Journal of Bacteriology*, 201(7). Available at: https://doi.org/10.1128/JB.00563-18.

Lee, J. *et al.* (2011) 'Mycobacterium tuberculosis induces an atypical cell death mode to escape from infected macrophages', *PLoS ONE*, 6(3). Available at: https://doi.org/10.1371/journal.pone.0018367.

Lee, J.H. and Berger, J.M. (2019) 'Cell cycle-dependent control and roles of DNA topoisomerase II', *Genes*. Available at: https://doi.org/10.3390/genes10110859.

Lee, S.H. (2016) 'Tuberculosis infection and latent tuberculosis', *Tuberculosis and Respiratory Diseases*. Available at: https://doi.org/10.4046/trd.2016.79.4.201.

Lefebvre, J.S. and Haynes, L. (2012) 'Aging of the CD4 t cell compartment', *Open Longevity Science*, 6. Available at: https://doi.org/10.2174/1876326X01206010083.

Leonardi, R. and Jackowski, S. (2007) 'Biosynthesis of Pantothenic Acid and Coenzyme A', *EcoSal Plus*, 2(2). Available at: https://doi.org/10.1128/ecosalplus.3.6.3.4.

Leplae, R. *et al.* (2011) 'Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families', *Nucleic Acids Research*, 39(13), pp. 5513–5525. Available at: https://doi.org/10.1093/nar/gkr131.

LeRoux, M. *et al.* (2020) 'Stress Can Induce Transcription of Toxin-Antitoxin Systems without Activating Toxin', *Molecular Cell*, 79(2). Available at: https://doi.org/10.1016/j.molcel.2020.05.028.

Leslie, A.G.W. and Powell, H.R. (2007) 'Processing diffraction data with mosflm', in. Springer, Dordrecht, pp. 41–51. Available at: https://doi.org/10.1007/978-1-4020-6316-9_4.

Levine, C., Hiasa, H. and Marians, K.J. (1998) 'DNA gyrase and topoisomerase IV: Biochemical activities, physiological roles during chromosome replication, and drug sensitivities', *Biochimica et Biophysica Acta - Gene Structure and Expression*. Available at: https://doi.org/10.1016/S0167-4781(98)00126-2.

Lewis, J.M. and Sloan, D.J. (2015) 'The role of delamanid in the treatment of drug-resistant tuberculosis', *Therapeutics and Clinical Risk Management*. Available at: https://doi.org/10.2147/TCRM.S71076.

Li, G.Y. *et al.* (2009) 'Inhibitory mechanism of Escherichia coli RelE-RelB toxin-antitoxin module involves a helix displacement near an mRNA interferase active site', *Journal of Biological Chemistry*, 284(21). Available at: https://doi.org/10.1074/jbc.M809656200.

Li, M. and Liu, Y. (2016) 'Topoisomerase I in Human Disease Pathogenesis and Treatments', *Genomics, Proteomics and Bioinformatics*. Available at: https://doi.org/10.1016/j.gpb.2016.02.004.

Li, X. *et al.* (2019) 'The plasmid-borne quinolone resistance protein QnrB, a novel DnaA-binding protein, increases the bacterial mutation rate by triggering DNA replication stress', *Molecular Microbiology*, 111(6). Available at: https://doi.org/10.1111/mmi.14235.

Li, Y. *et al.* (2010) 'Escherichia coli condensin MukB stimulates topoisomerase IV activity by a direct physical interaction', *Proceedings of the National Academy of Sciences of the United States of America*, 107(44). Available at: https://doi.org/10.1073/pnas.1008678107.

Liang, Y. *et al.* (2014) 'Structural and functional characterization of Escherichia coli toxin-antitoxin complex dinJ-yafQ', *Journal of Biological Chemistry*, 289(30). Available at: https://doi.org/10.1074/jbc.M114.559773.

Linka, R.M. *et al.* (2007) 'C-Terminal regions of topoisomerase IIα and IIβ determine isoform-specific functioning of the enzymes in vivo', *Nucleic Acids Research*, 35(11). Available at: https://doi.org/10.1093/nar/gkm102.

Lite, T.L. v. *et al.* (2020) 'Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library', *eLife*, 9. Available at: https://doi.org/10.7554/eLife.60924.

Liu, J. *et al.* (2018) 'Echinacoside, an Inestimable Natural Product in Treatment of Neurological and other Disorders', *Molecules*. Available at: https://doi.org/10.3390/molecules23051213.

Liu, L.F. and Wang, J.C. (1987) 'Supercoiling of the DNA template during transcription.', *Proceedings of the National Academy of Sciences of the United States of America*, 84(20). Available at: https://doi.org/10.1073/pnas.84.20.7024.

López-García, P. (1999) 'DNA supercoiling and temperature adaptation: A clue to early diversification of life?', *Journal of Molecular Evolution*, 49(4). Available at: https://doi.org/10.1007/PL00006567.

Lunge, A. *et al.* (2020) 'The unfoldase ClpC1 of mycobacterium tuberculosis regulates the expression of a distinct subset of proteins having intrinsically disordered termini', *Journal of Biological Chemistry*, 295(28). Available at: https://doi.org/10.1074/jbc.ra120.013456.

Madl, T. *et al.* (2006) 'Structural Basis for Nucleic Acid and Toxin Recognition of the Bacterial Antitoxin CcdA', *Journal of Molecular Biology*, 364(2), pp. 170–185. Available at: https://doi.org/10.1016/j.jmb.2006.08.082.

Maehigashi, T. *et al.* (2015) 'Molecular basis of ribosome recognition and mRNA hydrolysis by the E. coli YafQ toxin', *Nucleic Acids Research*, 43(16). Available at: https://doi.org/10.1093/nar/gkv791.

Magnuson, R.D. (2007) 'Hypothetical functions of toxin-antitoxin systems', *Journal of Bacteriology*. Available at: https://doi.org/10.1128/JB.00958-07.

Maltezou, H.C., Spyridis, P. and Kafetzis, D.A. (2000) 'Extra-pulmonary tuberculosis in children', *Archives of Disease in Childhood*, 83(4). Available at: https://doi.org/10.1136/adc.83.4.342.

Marianovsky, I. *et al.* (2001) 'The regulation of the Escherichia coli mazEF promoter involves an unusual alternating palindrome.', *The Journal of biological chemistry*, 276(8), pp. 5975–84. Available at: https://doi.org/10.1074/jbc.M008832200.

Massé, E. and Drolet, M. (1999) 'Relaxation of transcription-induced negative supercoiling is an essential function of Escherichia coli DNA topoisomerase I', *Journal of Biological Chemistry*, 274(23). Available at: https://doi.org/10.1074/jbc.274.23.16654.

Masuda, H. and Inouye, M. (2017) 'Toxins of prokaryotic toxin-antitoxin systems with sequence-specific endoribonuclease activity', *Toxins*. Available at: https://doi.org/10.3390/toxins9040140.

Masuda, H., Tan, Q., Awano, N., Wu, K.-P., *et al.* (2012) 'YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli', *Molecular Microbiology*, 84(5), pp. 979–989. Available at: https://doi.org/10.1111/j.1365-2958.2012.08068.x.

Mathema, B. *et al.* (2017) 'Drivers of Tuberculosis Transmission', *Journal of Infectious Diseases*, 216. Available at: https://doi.org/10.1093/infdis/jix354.

Mattison, K. *et al.* (2006) 'Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs', *Journal of Biological Chemistry*, 281(49). Available at: https://doi.org/10.1074/jbc.M605198200.

Mazal, H. and Haran, G. (2019) 'Single-molecule FRET methods to study the dynamics of proteins at work', *Current Opinion in Biomedical Engineering*. Available at: https://doi.org/10.1016/j.cobme.2019.08.007.

Mazurek, Ł. *et al.* (2021) 'Pentapeptide repeat protein QnrB1 requires ATP hydrolysis to rejuvenate poisoned gyrase complexes', *Nucleic acids research*, 49(3). Available at: https://doi.org/10.1093/nar/gkaa1266.

McCoy, A.J. and IUCr (2007) 'Solving structures of protein complexes by molecular replacement with *Phaser*', *Acta Crystallographica Section D Biological Crystallography*, 63(1), pp. 32–41. Available at: https://doi.org/10.1107/S0907444906045975.

McKenzie, J.L. *et al.* (2012) 'A vapbc toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria', *Journal of Bacteriology*, 194(9). Available at: https://doi.org/10.1128/JB.06790-11.

McKie, S.J., Neuman, K.C. and Maxwell, A. (2021) 'DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis', *BioEssays*. Available at: https://doi.org/10.1002/bies.202000286.

Mérens, A. *et al.* (2009) 'The pentapeptide repeat proteins MfpAMt and QnrB4 exhibit opposite effects on DNA gyrase catalytic reactions and on the Ternary gyrase-DNA-quinolone complex', *Journal of Bacteriology*, 191(5). Available at: https://doi.org/10.1128/JB.01205-08.

Miallau, L. *et al.* (2013) 'Comparative proteomics identifies the cell-associated lethality of M. tuberculosis RelBE-like toxin-antitoxin complexes', *Structure*, 21(4). Available at: https://doi.org/10.1016/j.str.2013.02.008.

Mierendorf, R.C. *et al.* (2003) 'Expression and Purification of Recombinant Proteins Using the pET System', in *Molecular Diagnosis of Infectious Diseases*. Available at: https://doi.org/10.1385/0-89603-485-2:257.

Millet, J.P. *et al.* (2013) 'Factors that influence current tuberculosis epidemiology', *European Spine Journal*. Available at: https://doi.org/10.1007/s00586-012-2334-8.

Mirambeau, G., Duguet, M. and Forterre, P. (1984) 'ATP-dependent DNA topoisomerase from the archaebacterium Sulfolobus acidocaldarius. Relaxation of supercoiled DNA at high temperature', *Journal of Molecular Biology*, 179(3). Available at: https://doi.org/10.1016/0022-2836(84)90080-9.

Mirdita, M. *et al.* (2022) 'ColabFold: making protein folding accessible to all', *Nature Methods*, 19(6). Available at: https://doi.org/10.1038/s41592-022-01488-1.

Mittal, E. *et al.* (2018) 'Mycobacterium tuberculosis type VII secretion system effectors differentially impact the ESCRT endomembrane damage response', *mBio*, 9(6). Available at: https://doi.org/10.1128/mBio.01765-18.

Mogues, T. *et al.* (2001) 'The relative importance of T cell subsets in immunity and immunopathology of airborne Mycobacterium tuberculosis infection in mice', *Journal of Experimental Medicine*, 193(3). Available at: https://doi.org/10.1084/jem.193.3.271.

Morrison, A., Higgins, N.P. and Cozzarelli, N.R. (1980) 'Interaction between DNA gyrase and its cleavage site on DNA.', *Journal of Biological Chemistry*, 255(5). Available at: https://doi.org/10.1016/s0021-9258(19)86016-7.

Mukamolova, G. v. *et al.* (2002) 'The rpf gene of Micrococcus luteus encodes an essential secreted growth factor', *Molecular Microbiology*, 46(3). Available at: https://doi.org/10.1046/j.1365-2958.2002.03183.x.

Murphy, H.N. *et al.* (2005) 'The OtsAB pathway is essential for trehalose biosynthesis in Mycobacterium tuberculosis', *Journal of Biological Chemistry*, 280(15). Available at: https://doi.org/10.1074/jbc.M414232200.

Muthuramalingam, M. *et al.* (2018) 'The toxin from a ParDE toxin-antitoxin system found in *Pseudomonas aeruginosa* offers protection to cells challenged with anti-gyrase antibiotics', *Molecular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/mmi.14165.

Muthuramalingam, M. *et al.* (2018) 'The toxin from a ParDE toxin-antitoxin system found in *Pseudomonas aeruginosa* offers protection to cells challenged with anti-gyrase antibiotics', *Molecular Microbiology* [Preprint]. Available at: https://doi.org/10.1111/mmi.14165.

Mutschler, H. *et al.* (2011) 'A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis', *PLoS Biology*, 9(3). Available at: https://doi.org/10.1371/journal.pbio.1001033.

Nadkarni, S., Mauri, C. and Ehrenstein, M.R. (2007) 'Anti-TNF-α therapy induces a distinct regulatory T cell population in patients with rheumatoid arthritis via TGF-β', *Journal of Experimental Medicine*, 204(1). Available at: https://doi.org/10.1084/jem.20061531.

Nagaraja, V. *et al.* (2017) 'DNA topoisomerase I and DNA gyrase as targets for TB therapy', *Drug Discovery Today*. Available at: https://doi.org/10.1016/j.drudis.2016.11.006.

Neubauer, C. *et al.* (2009) 'The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease RelE', *Cell*, 139(6). Available at: https://doi.org/10.1016/j.cell.2009.11.015.

Neuman, K.C. *et al.* (2009) 'Mechanisms of chiral discrimination by topoisomerase IV', *Proceedings of the National Academy of Sciences of the United States of America*, 106(17). Available at: https://doi.org/10.1073/pnas.0900574106.

Nguyen, T.N.A. *et al.* (2019) 'Molecular diagnosis of drug-resistant tuberculosis; A literature review', *Frontiers in Microbiology*, 10(MAR). Available at: https://doi.org/10.3389/fmicb.2019.00794.

Nichols, M.D. *et al.* (1999) 'Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11', *EMBO Journal*, 18(21). Available at: https://doi.org/10.1093/emboj/18.21.6177.

Nieto, C. *et al.* (2007) 'The yefM-yoeB toxin-antitoxin systems of Escherichia coli and Streptococcus pneumoniae: functional and structural correlation.', *Journal of bacteriology*, 189(4), pp. 1266–78. Available at: https://doi.org/10.1128/JB.01130-06.

Niklasson, M. *et al.* (2015) 'Robust and convenient analysis of protein thermal and chemical stability', *Protein Science*. Available at: https://doi.org/10.1002/pro.2809.

Nonejuie, P. *et al.* (2013) 'Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules', *Proceedings of the National Academy of Sciences of the United States of America*, 110(40). Available at: https://doi.org/10.1073/pnas.1311066110.

Noss, E.H. *et al.* (2001) ' Toll-Like Receptor 2-Dependent Inhibition of Macrophage Class II MHC Expression and Antigen Processing by 19-kDa Lipoprotein of Mycobacterium tuberculosis ', *The Journal of Immunology*, 167(2). Available at: https://doi.org/10.4049/jimmunol.167.2.910.

Nurse, P. *et al.* (2003) 'Topoisomerase III can serve as the cellular decatenase in Escherichia coli', *Journal of Biological Chemistry*, 278(10). Available at: https://doi.org/10.1074/jbc.M211211200.

O'Fágáin, C., Cummins, P.M. and O'Connor, B.F. (2011) 'Gel-filtration chromatography.', *Methods in molecular biology (Clifton, N.J.)*. Available at: https://doi.org/10.1007/978-1-60761-913-0_2.

Oberer, M. *et al.* (2007) 'The solution structure of ParD, the antidote of the ParDE toxin-antitoxin module, provides the structural basis for DNA and toxin binding', *Protein Science*, 16(8), pp. 1676–1688. Available at: https://doi.org/10.1110/ps.062680707.

Ogawa, T. *et al.* (2015) 'Direct observation of DNA overwinding by reverse gyrase', *Proceedings of the National Academy of Sciences of the United States of America*, 112(24). Available at: https://doi.org/10.1073/pnas.1422203112.

Ogura, T. and Hiraga, S. (1983) 'Mini-F plasmid genes that couple host cell division to plasmid proliferation.', *Proceedings of the National Academy of Sciences of the United States of America*, 80(15). Available at: https://doi.org/10.1073/pnas.80.15.4784.

Oren, A. and Garrity, G. (2020) 'List of new names and new combinations previously effectively, but not validly, published', *International Journal of Systematic and Evolutionary Microbiology*, 70(7). Available at: https://doi.org/10.1099/ijsem.0.004244.

Orphanides, G. and Maxwell, A. (1994) 'Evidence for a conformational change in the DNA gyrase-DNA complex from hydroxyl radical footprinting', *Nucleic Acids Research*, 22(9). Available at: https://doi.org/10.1093/nar/22.9.1567.

Overgaard, M. *et al.* (2008) 'Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity', *Molecular Microbiology*, 69(4), pp. 841–857. Available at: https://doi.org/10.1111/j.1365-2958.2008.06313.x.

Overgaard, M., Borch, J. and Gerdes, K. (2009) 'RelB and RelE of Escherichia coli Form a Tight Complex That Represses Transcription via the Ribbon–Helix–Helix Motif in RelB', *Journal of Molecular Biology*, 394(2), pp. 183–196. Available at: https://doi.org/10.1016/J.JMB.2009.09.006.

Page, R. and Peti, W. (2016) 'Toxin-antitoxin systems in bacterial growth arrest and persistence', *Nature Chemical Biology* [Preprint]. Available at: https://doi.org/10.1038/nchembio.2044.

Pakamwong, B. *et al.* (2022) 'Identification of Potent DNA Gyrase Inhibitors Active against Mycobacterium tuberculosis', *Journal of Chemical Information and Modeling*, 62(7). Available at: https://doi.org/10.1021/acs.jcim.1c01390.

Pandey, D.P. and Gerdes, K. (2005) 'Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes', *Nucleic Acids Research*, 33(3). Available at: https://doi.org/10.1093/nar/gki201.

Park, D.W. *et al.* (2020) 'Crystal structure of proteolyzed VapBC and DNA-bound VapBC from Salmonella enterica Typhimurium LT2 and VapC as a putative Ca2+-dependent ribonuclease', *FASEB Journal*, 34(2). Available at: https://doi.org/10.1096/fj.201901989R.

Park, J.Y. *et al.* (2020) 'Induced DNA bending by unique dimerization of HigA antitoxin', *IUCrJ*, 7. Available at: https://doi.org/10.1107/S2052252520006466.

Peltier, J. *et al.* (2020) 'Type I toxin-antitoxin systems contribute to the maintenance of mobile genetic elements in Clostridioides difficile', *Communications Biology*, 3(1). Available at: https://doi.org/10.1038/s42003-020-01448-5.

Perez-Cheeks, B.A. *et al.* (2012) 'A role for topoisomerase III in Escherichia coli chromosome segregation', *Molecular Microbiology*, 86(4). Available at: https://doi.org/10.1111/mmi.12039.

Petit, T. *et al.* (2004) 'Comparative value of tumour grade, hormonal receptors, Ki-67, HER-2 and topoisomerase II alpha status as predictive markers in breast cancer patients treated with neoadjuvant anthracycline-based chemotherapy', *European Journal of Cancer*, 40(2). Available at: https://doi.org/10.1016/S0959-8049(03)00675-0.

Petrella, S. *et al.* (2019) 'Overall Structures of Mycobacterium tuberculosis DNA Gyrase Reveal the Role of a Corynebacteriales GyrB-Specific Insert in ATPase Activity', *Structure* [Preprint]. Available at: https://doi.org/10.1016/J.STR.2019.01.004.

Pfyffer, G.E. (2015) ' Mycobacterium: General Characteristics, Laboratory Detection, and Staining Procedures ', in *Manual of Clinical Microbiology* . Available at: https://doi.org/10.1128/9781555817381.ch30.

Piddington, D.L. *et al.* (2001) 'Cu,Zn superoxide dismutase of Mycobacterium tuberculosis contributes to survival in activated macrophages that are generating an oxidative burst', *Infection and Immunity*, 69(8). Available at: https://doi.org/10.1128/IAI.69.8.4980-4987.2001.

Piddock, L.J.V., Walters, R.N. and Diver, J.M. (1990) 'Correlation of quinolone MIC and inhibition of DNA, RNA, and protein synthesis and induction of the SOS response in Escherichia coli',

Antimicrobial Agents and Chemotherapy, 34(12). Available at: https://doi.org/10.1128/AAC.34.12.2331.

Pierrat, O.A. and Maxwell, A. (2003) 'The Action of the Bacterial Toxin Microcin B17', *Journal of Biological Chemistry*, 278(37). Available at: https://doi.org/10.1074/jbc.m304516200.

Piton, J. *et al.* (2010) 'Structural insights into the quinolone resistance mechanism of Mycobacterium tuberculosis DNA gyrase', *PLoS ONE*, 5(8). Available at: https://doi.org/10.1371/journal.pone.0012245.

Pitts, S.L. *et al.* (2011) 'Use of divalent metal ions in the DNA cleavage reaction of topoisomerase IV', *Nucleic Acids Research*, 39(11), p. 4808. Available at: https://doi.org/10.1093/NAR/GKR018.

Pommier, Y. (2013) 'Drugging topoisomerases: Lessons and Challenges', *ACS Chemical Biology*. Available at: https://doi.org/10.1021/cb300648v.

Pommier, Y. *et al.* (2010) 'DNA topoisomerases and their poisoning by anticancer and antibacterial drugs', *Chemistry and Biology*. Available at: https://doi.org/10.1016/j.chembiol.2010.04.012.

Postow, L. *et al.* (2001) 'Topological challenges to DNA replication: Conformations at the fork', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15). Available at: https://doi.org/10.1073/pnas.111006998.

Pranger, A.D. *et al.* (2019) 'The Role of Fluoroquinolones in the Treatment of Tuberculosis in 2019.', *Drugs*, 79(2), pp. 161–171. Available at: https://doi.org/10.1007/s40265-018-1043-y.

Prosser, G.A. and de Carvalho, L.P.S. (2013) 'Kinetic mechanism and inhibition of Mycobacterium tuberculosis d-alanine: D-alanine ligase by the antibiotic d-cycloserine', *FEBS Journal*, 280(4). Available at: https://doi.org/10.1111/febs.12108.

Prysak, M.H. *et al.* (2009) 'Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage', *Molecular Microbiology*, 71(5). Available at: https://doi.org/10.1111/j.1365-2958.2008.06572.x.

Qian, H. *et al.* (2019) 'Toxin–antitoxin operon kacAT of Klebsiella pneumoniae is regulated by conditional cooperativity via a W-shaped KacA–KacT complex', *Nucleic Acids Research*, 47(14), pp. 7690–7702. Available at: https://doi.org/10.1093/nar/gkz563.

Qiao, Y. *et al.* (2021) 'Single-molecular förster resonance energy transfer measurement on structures and interactions of biomolecules', *Micromachines*. Available at: https://doi.org/10.3390/mi12050492.

Rachman, H., Strong, M., Schaible, U., *et al.* (2006) 'Mycobacterium tuberculosis gene expression profiling within the context of protein networks', *Microbes and Infection*, 8(3), pp. 747–757. Available at: https://doi.org/10.1016/j.micinf.2005.09.011.

Rachman, H., Strong, M., Ulrichs, T., *et al.* (2006) 'Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis.', *Infection and immunity*, 74(2), pp. 1233–42. Available at: https://doi.org/10.1128/IAI.74.2.1233-1242.2006.

Racko, D. *et al.* (2018) 'Chromatin loop extrusion and chromatin unknotting', *Polymers*, 10(10). Available at: https://doi.org/10.3390/polym10101126.

Rajan, R., Osterman, A. and Mondragón, A. (2016) 'Methanopyrus kandleri topoisomerase v contains three distinct AP lyase active sites in addition to the topoisomerase active site', *Nucleic Acids Research*, 44(7). Available at: https://doi.org/10.1093/nar/gkw122.

Ramage, H.R., Connolly, L.E. and Cox, J.S. (2009) 'Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution', *PLoS Genetics*. Edited by S.M. Rosenberg, 5(12), p. e1000767. Available at: https://doi.org/10.1371/journal.pgen.1000767.

Ramakrishnan, L. (2012) 'Revisiting the role of the granuloma in tuberculosis', *Nature Reviews Immunology*. Available at: https://doi.org/10.1038/nri3211.

Rasmussen, K.K. *et al.* (2018) 'Structural basis of the bacteriophage TP901-1 CI repressor dimerization and interaction with DNA', *FEBS Letters*, 592(10). Available at: https://doi.org/10.1002/1873-3468.13060.

Remot, A., Doz, E. and Winter, N. (2019) 'Neutrophils and close relatives in the hypoxic environment of the tuberculous granuloma: New avenues for host-directed therapies?', *Frontiers in Immunology*. Available at: https://doi.org/10.3389/fimmu.2019.00417.

Rempel, S. *et al.* (2020) 'A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds', *Nature*, 580(7803). Available at: https://doi.org/10.1038/s41586-020-2072-8.

Rengarajan, J., Bloom, B.R. and Rubin, E.J. (2005) 'Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), pp. 8327–32. Available at: https://doi.org/10.1073/pnas.0503272102.

Rice, P., Longden, L. and Bleasby, A. (2000) 'EMBOSS: The European Molecular Biology Open Software Suite', *Trends in Genetics*. Available at: https://doi.org/10.1016/S0168-9525(00)02024-2.

Roberts, R.C., Ström, A.R. and Helinski, D.R. (1994) 'The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss', *Journal of Molecular Biology*, 237(1). Available at: https://doi.org/10.1006/jmbi.1994.1207.

Robinson, J.M. (2008) 'Reactive oxygen species in phagocytic leukocytes', *Histochemistry and Cell Biology*. Available at: https://doi.org/10.1007/s00418-008-0461-4.

Robson, J. *et al.* (2009) 'The vapBC Operon from Mycobacterium smegmatis Is An Autoregulated Toxin-Antitoxin Module That Controls Growth via Inhibition of Translation', *Journal of Molecular Biology*, 390(3). Available at: https://doi.org/10.1016/j.jmb.2009.05.006.

Rocker, A. and Meinhart, A. (2016) 'Type II toxin: antitoxin systems. More than small selfish entities?', *Current genetics*, 62(2), pp. 287–90. Available at: https://doi.org/10.1007/s00294-015-0541-7.

Rodrigo, T. *et al.* (1997) 'Characteristics of tuberculosis patients who generate secondary cases', *International Journal of Tuberculosis and Lung Disease*, 1(4).

Rogerson, B.J. *et al.* (2006) 'Expression levels of Mycobacterium tuberculosis antigen-encoding genes versus production levels of antigen-specific T cells during stationary level lung infection in mice', *Immunology*, 118(2). Available at: https://doi.org/10.1111/j.1365-2567.2006.02355.x.

Rouse, D.A. *et al.* (1995) 'Characterization of the katG and inhA genes of isoniazid-resistant clinical isolates of Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 39(11). Available at: https://doi.org/10.1128/AAC.39.11.2472.

Roy, R., Hohng, S. and Ha, T. (2008) 'A practical guide to single-molecule FRET', *Nature Methods*. Available at: https://doi.org/10.1038/nmeth.1208.

Rozwarski, D.A. *et al.* (1998) 'Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis', *Science*, 279(5347). Available at: https://doi.org/10.1126/science.279.5347.98.

Rueden, C.T. *et al.* (2017) 'ImageJ2: ImageJ for the next generation of scientific image data', *BMC Bioinformatics*, 18(1). Available at: https://doi.org/10.1186/s12859-017-1934-z.

Ruiz-Echevarría, M.J., de la Cueva, G. and Díaz-Orejas, R. (1995) 'Translational coupling and limited degradation of a polycistronic messenger modulate differential gene expression in the parD stability system of plasmid R1', *MGG Molecular & General Genetics*, 248(5). Available at: https://doi.org/10.1007/BF02423456.

Rullas, J. *et al.* (2015) 'Combinations of β-lactam antibiotics currently in clinical trials are efficacious in a dhp-i-deficient mouse model of tuberculosis infection', *Antimicrobial Agents and Chemotherapy*, 59(8). Available at: https://doi.org/10.1128/AAC.01063-15.

Russell-Goldman, E. *et al.* (2008) 'A Mycobacterium tuberculosis Rpf double-knockout strain exhibits profound defects in reactivation from chronic tuberculosis and innate immunity phenotypes', *Infection and Immunity*, 76(9). Available at: https://doi.org/10.1128/IAI.01735-07.

Rustad, T.R. *et al.* (2009) 'Hypoxia: A window into Mycobacterium tuberculosis latency', *Cellular Microbiology*. Available at: https://doi.org/10.1111/j.1462-5822.2009.01325.x.

Sala, A., Bordes, P. and Genevaux, P. (2014) 'Multiple toxin-antitoxin systems in Mycobacterium tuberculosis', *Toxins*. Available at: https://doi.org/10.3390/toxins6031002.

Samson, J.E. *et al.* (2013) 'Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system', *Molecular Microbiology*, 87(4), pp. 756–768. Available at: https://doi.org/10.1111/mmi.12129.

Sarathy, J.P., Dartois, V. and Lee, E.J.D. (2012) 'The role of transport mechanisms in Mycobacterium Tuberculosis drug resistance and tolerance', *Pharmaceuticals*. Available at: https://doi.org/10.3390/ph5111210.

Sarathy, J.P., Gruber, G. and Dick, T. (2019) 'Re-understanding the mechanisms of action of the anti-mycobacterial drug bedaquiline', *Antibiotics*. Available at: https://doi.org/10.3390/antibiotics8040261.

Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) 'Genes required for mycobacterial growth defined by high density mutagenesis', *Molecular Microbiology*, 48(1), pp. 77–84. Available at: https://doi.org/10.1046/j.1365-2958.2003.03425.x.

Sauvage, E. and Terrak, M. (2016) 'Glycosyltransferases and transpeptidases/penicillin-binding proteins: Valuable targets for new antibacterials', *Antibiotics*. Available at: https://doi.org/10.3390/antibiotics5010012.

Schaaf, H.S. *et al.* (2009) 'Ethionamide cross-and co-resistance in children with isoniazid-resistant tuberculosis', *International Journal of Tuberculosis and Lung Disease*, 13(11).

Schmid, M.B. and Sawitzke, J.A. (1993) 'Multiple bacterial topoisomerases: Specialization or redundancy?', *BioEssays*. Available at: https://doi.org/10.1002/bies.950150703.

Schnappinger, D. *et al.* (2003) 'Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment.', *The Journal of experimental medicine*, 198(5), pp. 693–704. Available at: https://doi.org/10.1084/jem.20030846.

Schoeffler, A.J., May, A.P. and Berger, J.M. (2010) 'A domain insertion in Escherichia coli GyrB adopts a novel fold that plays a critical role in gyrase function', *Nucleic Acids Research*, 38(21). Available at: https://doi.org/10.1093/nar/gkq665.

Schuller, M. *et al.* (2021) 'Molecular basis for DarT ADP-ribosylation of a DNA base', *Nature*, 596(7873). Available at: https://doi.org/10.1038/s41586-021-03825-4.

Schumacher, M.A. *et al.* (2009) 'Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB', *Science*, 323(5912). Available at: https://doi.org/10.1126/science.1163806.

Schvartzman, J.B. and Stasiak, A. (2004) 'A topological view of the replicon', *EMBO Reports*. Available at: https://doi.org/10.1038/sj.embor.7400101.

Seok, K.K. *et al.* (2007) 'Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen Streptococcus pneumoniae', *Journal of Biological Chemistry*, 282(27). Available at: https://doi.org/10.1074/jbc.M701703200.

Shafiani, S. *et al.* (2010) 'Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis', *Journal of Experimental Medicine*, 207(7). Available at: https://doi.org/10.1084/jem.20091885.

Shah, S. and Heddle, J.G. (2014) 'Squaring up to DNA: pentapeptide repeat proteins and DNA mimicry', *Applied Microbiology and Biotechnology*. Available at: https://doi.org/10.1007/s00253-014-6151-3.

Shan Chang, D.P. and Guan, X.L. (2021) 'Metabolic versatility of mycobacterium tuberculosis during infection and dormancy', *Metabolites*, 11(2). Available at: https://doi.org/10.3390/metabo11020088.

Sharma, S. and Tyagi, J.S. (2016) 'Mycobacterium tuberculosis DevR/DosR dormancy regulator activation mechanism: Dispensability of phosphorylation, cooperativity and essentiality of α10 Helix', *PLoS ONE*, 11(8). Available at: https://doi.org/10.1371/journal.pone.0160723.

Shi, L., North, R. and Gennaro, M.L. (2004) 'Effect of Growth State on Transcription Levels of Genes Encoding Major Secreted Antigens of Mycobacterium tuberculosis in the Mouse Lung', *Infection and Immunity*, 72(4). Available at: https://doi.org/10.1128/IAI.72.4.2420-2424.2004.

Shiloh, M.U., Manzanillo, P. and Cox, J.S. (2008) 'Mycobacterium tuberculosis Senses Host-Derived Carbon Monoxide during Macrophage Infection', *Cell Host and Microbe*, 3(5). Available at: https://doi.org/10.1016/j.chom.2008.03.007.

Sievers, F. and Higgins, D.G. (2014) 'Clustal Omega', *Current Protocols in Bioinformatics*, 2014. Available at: https://doi.org/10.1002/0471250953.bi0313s48.

Simanshu, D.K. *et al.* (2013) 'Structural Basis of mRNA Recognition and Cleavage by Toxin MazF and Its Regulation by Antitoxin MazE in Bacillus subtilis', *Molecular Cell*, 52(3). Available at: https://doi.org/10.1016/j.molcel.2013.09.006.

Singh, R., Barry, C.E. and Boshoff, H.I.M. (2010) 'The three RelE homologs of Mycobacterium tuberculosis have individual, drug-specific effects on bacterial antibiotic tolerance', *Journal of Bacteriology*, 192(5). Available at: https://doi.org/10.1128/JB.01285-09.

Singletary, L.A. *et al.* (2009) 'An SOS-regulated type 2 toxin-antitoxin system.', *Journal of bacteriology*, 191(24), pp. 7456–65. Available at: https://doi.org/10.1128/JB.00963-09.

Sissi, C. and Palumbo, M. (2009) 'Effects of magnesium and related divalent metal ions in topoisomerase structure and function', *Nucleic Acids Research*, 37(3). Available at: https://doi.org/10.1093/nar/gkp024.

Sivaramakrishnan, S. and de Montellano, P.R.O. (2013) 'The DosS-DosT/DosR Mycobacterial Sensor System', *Biosensors*. Available at: https://doi.org/10.3390/bios3030259.

Skipper, H.E., Mitchell, J.H. and Bennett, L.L. (1950) 'Inhibition of Nucleic Acid Synthesis by Folic Acid Antagonists', *Cancer Research*, 10(8).

Slayden, R.A., Dawson, C.C. and Cummings, J.E. (2018) 'Toxin–antitoxin systems and regulatory mechanisms in Mycobacterium tuberculosis', *Pathogens and Disease*, 76(4). Available at: https://doi.org/10.1093/femspd/fty039.

Slesarev, A.I. *et al.* (1993) 'DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote', *Nature*, 364(6439). Available at: https://doi.org/10.1038/364735a0.

Smith, C.G. *et al.* (1956) 'Streptonivicin, a new antibiotic. I. Discovery and biologic studies', *Antibiotics & chemotherapy*, 6(2).

Smith, I. (2003) 'Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence.', *Clinical microbiology reviews*, 16(3), pp. 463–96. Available at: https://doi.org/10.1128/CMR.16.3.463-496.2003.

Snead, K.J., Moore, L.L. and Bourne, C.R. (2022) 'ParD Antitoxin Hotspot Alters a Disorder-to-Order Transition upon Binding to Its Cognate ParE Toxin, Lessening Its Interaction Affinity and Increasing Its Protease Degradation Kinetics', *Biochemistry*, 61(1). Available at: https://doi.org/10.1021/acs.biochem.1c00584.

Solano-Gutierrez, J.S., Pino, C. and Robledo, J. (2019) 'Toxin–antitoxin systems shows variability among *Mycobacterium tuberculosis* lineages', *FEMS Microbiology Letters*, 366(1). Available at: https://doi.org/10.1093/femsle/fny276.

Song, S. and Wood, T.K. (2020) 'Toxin/Antitoxin System Paradigms: Toxins Bound to Antitoxins Are Not Likely Activated by Preferential Antitoxin Degradation', *Advanced Biosystems*. Available at: https://doi.org/10.1002/adbi.201900290.

Songailiene, I. *et al.* (2020) 'HEPN-MNT Toxin-Antitoxin System: The HEPN Ribonuclease Is Neutralized by OligoAMPylation', *Molecular Cell*, 80(6). Available at: https://doi.org/10.1016/j.molcel.2020.11.034.

Soo, V.W.C. and Wood, T.K. (2013) 'Antitoxin MqsA Represses Curli Formation Through the Master Biofilm Regulator CsgD', *Scientific Reports*, 3(1), p. 3186. Available at: https://doi.org/10.1038/srep03186.

Soren, B.C. *et al.* (2020) 'Topoisomerase IB: a relaxing enzyme for stressed DNA', *Cancer Drug Resistance*. Available at: https://doi.org/10.20517/cdr.2019.106.

Sousa, E.H.S. *et al.* (2007) ' DosT and DevS are oxygen-switched kinases in Mycobacterium tuberculosis ', *Protein Science*, 16(8). Available at: https://doi.org/10.1110/ps.072897707.

Souza, C.J.F. and Garcia-Rojas, E.E. (2015) 'Effects of salt and protein concentrations on the association and dissociation of ovalbumin-pectin complexes', *Food Hydrocolloids*, 47. Available at: https://doi.org/10.1016/j.foodhyd.2015.01.010.

Speer, S.L. *et al.* (2021) 'The intracellular environment affects protein-protein interactions', *Proceedings of the National Academy of Sciences of the United States of America*, 118(11). Available at: https://doi.org/10.1073/pnas.2019918118.

Stalker, D.J. and Jungbluth, G.L. (2003) 'Clinical Pharmacokinetics of Linezolid, a Novel Oxazolidinone Antibacterial', *Clinical Pharmacokinetics*. Available at: https://doi.org/10.2165/00003088-200342130-00004.

Sterckx, Y.G.J. *et al.* (2015) 'An efficient method for the purification of proteins from four distinct toxin-antitoxin modules', *Protein Expression and Purification*, 108. Available at: https://doi.org/10.1016/j.pep.2015.01.001.

Sterckx, Y.G.J. *et al.* (2016) 'A unique hetero-hexadecameric architecture displayed by the Escherichia coli O157 PaaA2-ParE2 antitoxin-toxin complex', *Journal of Molecular Biology*, 428(8). Available at: https://doi.org/10.1016/j.jmb.2016.03.007.

Stewart, G.R. *et al.* (2005) 'Mycobacterial mutants with defective control of phagosomal acidification', *PLoS Pathogens*, 1(3). Available at: https://doi.org/10.1371/journal.ppat.0010033.

Stieber, D., Gabant, P. and Szpirer, C.Y. (2008) 'The art of selective killing: Plasmid toxin/antitoxin systems and their technological applications', *BioTechniques*, 45(3). Available at: https://doi.org/10.2144/000112955.

Stinear, T.P. *et al.* (2008) 'Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis', *Genome Research*, 18(5). Available at: https://doi.org/10.1101/gr.075069.107.

Stivers, J.T., Harris, T.K. and Mildvan, A.S. (1997) 'Vaccinia DNA topoisomerase I: Evidence supporting a free rotation mechanism for DNA supercoil relaxation', *Biochemistry*, 36(17). Available at: https://doi.org/10.1021/bi962880t.

Stone, M.D. *et al.* (2003) 'Chirality sensing by Escherichia coli topoisomerase IV and the mechanism of type II topoisomerases', *Proceedings of the National Academy of Sciences of the United States of America*, 100(15). Available at: https://doi.org/10.1073/pnas.1133178100.

Suarez, J. *et al.* (2009) 'An oxyferrous heme/protein-based radical intermediate is catalytically competent in the catalase reaction of Mycobacterium tuberculosis catalase-peroxidase (KatG)', *Journal of Biological Chemistry*, 284(11). Available at: https://doi.org/10.1074/jbc.M808106200.

Sugimoto-Shirasu, K. *et al.* (2002) 'DNA topoisomerase VI is essential for endoreduplication in Arabidopsis', *Current Biology*, 12(20). Available at: https://doi.org/10.1016/S0960-9822(02)01198-3.

Suski, C. and Marians, K.J. (2008) 'Resolution of Converging Replication Forks by RecQ and Topoisomerase III', *Molecular Cell*, 30(6). Available at: https://doi.org/10.1016/j.molcel.2008.04.020.

Takagi, H. *et al.* (2005) 'Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects', *Nature Structural and Molecular Biology*, 12(4). Available at: https://doi.org/10.1038/nsmb911.

Takahashi, H., Hayakawa, I. and Akimoto, T. (2003) 'The history of the development and changes of quinolone antibacterial agents', *Yakushigaku zasshi. The Journal of Japanese history of pharmacy*, 38(2).

Takahashi, T.S. *et al.* (2020) 'Expanding the type IIB DNA topoisomerase family: Identification of new topoisomerase and topoisomerase-like proteins in mobile genetic elements', *NAR Genomics and Bioinformatics*, 2(1). Available at: https://doi.org/10.1093/nargab/lqz021.

Takayama, K., Wang, C. and Besra, G.S. (2005) 'Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis', *Clinical Microbiology Reviews*. Available at: https://doi.org/10.1128/CMR.18.1.81-101.2005.

Tam, J.E. and Kline, B.C. (1989) 'The F plasmid ccd autorepressor is a complex of CcdA and CcdB proteins', *MGG Molecular & General Genetics*, 219(1–2). Available at: https://doi.org/10.1007/BF00261153.

Tan, K. *et al.* (2015) 'Structural basis for suppression of hypernegative DNA supercoiling by E. coli topoisomerase I', *Nucleic Acids Research*, 43(22). Available at: https://doi.org/10.1093/nar/gkv1073.

Tan, K. *et al.* (2016) 'Insights from the Structure of Mycobacterium tuberculosis Topoisomerase i with a Novel Protein Fold', *Journal of Molecular Biology*, 428(1). Available at: https://doi.org/10.1016/j.jmb.2015.11.024.

Tan, Q., Awano, N. and Inouye, M. (2011) 'YeeV is an Escherichia coli toxin that inhibits cell division by targeting the cytoskeleton proteins, FtsZ and MreB.', *Molecular microbiology*, 79(1), pp. 109–18. Available at: https://doi.org/10.1111/j.1365-2958.2010.07433.x.

Taneja, B. *et al.* (2006) 'Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases', *EMBO Journal*, 25(2). Available at: https://doi.org/10.1038/sj.emboj.7600922.

Tang, S. *et al.* (2015) 'Efficacy, safety and tolerability of linezolid for the treatment of XDR-TB: a study in China', *European Respiratory Journal*, 45(1), pp. 161–170. Available at: https://doi.org/10.1183/09031936.00035114.

Tao, J. *et al.* (2013) 'Mycobacterium fluoroquinolone resistance protein B, a novel small GTPase, is involved in the regulation of DNA gyrase and drug resistance', *Nucleic Acids Research*, 41(4). Available at: https://doi.org/10.1093/nar/gks1351.

Texier, P. *et al.* (2021) 'ClpXP-mediated Degradation of the TAC Antitoxin is Neutralized by the SecB-like Chaperone in Mycobacterium tuberculosis', *Journal of Molecular Biology*, 433(5). Available at: https://doi.org/10.1016/j.jmb.2021.166815.

Thanna, S. and Sucheck, S.J. (2016) 'Targeting the trehalose utilization pathways of Mycobacterium tuberculosis', *MedChemComm*. Available at: https://doi.org/10.1039/c5md00376h.

Thisted, T. and Gerdes, K. (1992) 'Mechanism of post-segregational killing by the hok/sok system of plasmid R1: Sok antisense RNA regulates hok gene expression indirectly through the overlapping mok gene', *Journal of Molecular Biology*, 223(1), pp. 41–54. Available at: https://doi.org/10.1016/0022-2836(92)90714-U.

Thompson, A.M. *et al.* (2017) 'Antitubercular Nitroimidazoles Revisited: Synthesis and Activity of the Authentic 3-Nitro Isomer of Pretomanid', *ACS Medicinal Chemistry Letters*, 8(12). Available at: https://doi.org/10.1021/acsmedchemlett.7b00356.

Timmins, G.S. *et al.* (2004) 'Nitric oxide generated from isoniazid activation by KatG: Source of nitric oxide and activity against Mycobacterium tuberculosis', *Antimicrobial Agents and Chemotherapy*, 48(8). Available at: https://doi.org/10.1128/AAC.48.8.3006-3009.2004.

Tolia, N.H. and Joshua-Tor, L. (2006) 'Strategies for protein coexpression in Escherichia coli', *Nature Methods*, 3(1). Available at: https://doi.org/10.1038/nmeth0106-55.

Torres, N.M.C. *et al.* (2019) 'Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis', *PLoS ONE*, 14(12). Available at: https://doi.org/10.1371/journal.pone.0226507.

Torrey, H.L. *et al.* (2016) 'High Persister Mutants in Mycobacterium tuberculosis.', *PloS one*, 11(5), p. e0155127. Available at: https://doi.org/10.1371/journal.pone.0155127.

Tran, J.H. and Jacoby, G.A. (2002) 'Mechanism of plasmid-mediated quinolone resistance', *Proceedings of the National Academy of Sciences of the United States of America*, 99(8). Available at: https://doi.org/10.1073/pnas.082092899.

Trentini, D.B. *et al.* (2016) 'Arginine phosphorylation marks proteins for degradation by a Clp protease', *Nature*, 539(7627). Available at: https://doi.org/10.1038/nature20122.

Tretter, E.M. and Berger, J.M. (2012) 'Mechanisms for defining supercoiling set point of DNA gyrase orthologs: II. The shape of the GyrA subunit C-terminal domain (CTD) is not a sole determinant for controlling supercoiling efficiency', *Journal of Biological Chemistry*, 287(22). Available at: https://doi.org/10.1074/jbc.M112.345736.

Trovatti, E. *et al.* (2008) 'Peptides based on CcdB protein as novel inhibitors of bacterial topoisomerases', *Bioorganic and Medicinal Chemistry Letters*, 18(23). Available at: https://doi.org/10.1016/j.bmcl.2008.10.008.

Tufariello, J.A.M. *et al.* (2006) 'Deletion of the Mycobacterium tuberculosis resuscitation-promoting factor Rv1009 gene results in delayed reactivation from chronic tuberculosis', *Infection and Immunity*, 74(5). Available at: https://doi.org/10.1128/IAI.74.5.2985-2995.2006.

Turnbull, K.J. and Gerdes, K. (2017) 'HicA toxin of Escherichia coli derepresses hicAB transcription to selectively produce HicB antitoxin', *Molecular Microbiology*, 104(5). Available at: https://doi.org/10.1111/mmi.13662.

Vagin, A.A. *et al.* (2004) '*REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp. 2184–2195. Available at: https://doi.org/10.1107/S0907444904023510.

van Crevel, R., Ottenhoff, T.H.M. and van der Meer, J.W.M. (2002) 'Innate immunity to Mycobacterium tuberculosis.', *Clinical microbiology reviews*, 15(2), pp. 294–309. Available at: https://doi.org/10.1128/CMR.15.2.294-309.2002.

van Melderen, L. and de Bast, M.S. (2009) 'Bacterial toxin-Antitoxin systems: More than selfish entities?', *PLoS Genetics*. Available at: https://doi.org/10.1371/journal.pgen.1000437.

Vandal, O.H. *et al.* (2008) 'A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis', *Nature Medicine*, 14(8). Available at: https://doi.org/10.1038/nm.1795.

vanden Broeck, A. *et al.* (2019) 'Cryo-EM structure of the complete E. coli DNA gyrase nucleoprotein complex', *Nature Communications*, 10(1), p. 4935. Available at: https://doi.org/10.1038/s41467-019-12914-y.

Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1). Available at: https://doi.org/10.1093/nar/gkab1061.

Vergne, I., Chua, J. and Deretic, V. (2003) 'Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca 2+/calmodulin-PI3K hVPS34 cascade', *Journal of Experimental Medicine*, 198(4). Available at: https://doi.org/10.1084/jem.20030527.

Vergne, I., Gilleron, M. and Nigou, J. (2014) 'Manipulation of the endocytic pathway and phagocyte functions by Mycobacterium tuberculosis lipoarabinomannan', *Frontiers in Cellular and Infection Microbiology*, 4(DEC). Available at: https://doi.org/10.3389/fcimb.2014.00187.

Vincent, A.T. *et al.* (2018) 'The mycobacterial cell envelope: A relict from the past or the result of recent evolution?', *Frontiers in Microbiology*. Available at: https://doi.org/10.3389/fmicb.2018.02341.

Vogel, J. *et al.* (2004) 'The small RNA istR inhibits synthesis of an SOS-induced toxic peptide', *Current Biology*, 14(24). Available at: https://doi.org/10.1016/j.cub.2004.12.003.

Wallgren, A. (1948) 'The time-table of tuberculosis', *Tubercle*. Available at: https://doi.org/10.1016/S0041-3879(48)80033-4.

Wang, J.C. (1971) 'Interaction between DNA and an Escherichia coli protein ω', *Journal of Molecular Biology*, 55(3). Available at: https://doi.org/10.1016/0022-2836(71)90334-2.

Wang, J.C. (1998) 'Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine', *Quarterly Reviews of Biophysics*. Available at: https://doi.org/10.1017/S0033583598003424.

Wang, X. *et al.* (2012) 'A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS', *Nature Chemical Biology*, 8(10), pp. 855–861. Available at: https://doi.org/10.1038/nchembio.1062.

Warner, D.F. and Mizrahi, V. (2006) 'Tuberculosis chemotherapy: The influence of bacillary stress and damage response pathways on drug efficacy', *Clinical Microbiology Reviews*. Available at: https://doi.org/10.1128/CMR.00060-05.

Waterhouse, A.M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9). Available at: https://doi.org/10.1093/bioinformatics/btp033.

Wayne, L.G. (1994) 'Dormancy ofMycobacterium tuberculosis and latency of disease', *European Journal of Clinical Microbiology & Infectious Diseases*, 13(11), pp. 908–914. Available at: https://doi.org/10.1007/BF02111491.

Weel-Sneve, R. *et al.* (2013) 'Single Transmembrane Peptide DinQ Modulates Membrane-Dependent Activities', *PLoS Genetics*, 9(2). Available at: https://doi.org/10.1371/journal.pgen.1003260.

Weidlich, D. and Klostermeier, D. (2020) 'Functional interactions between gyrase subunits are optimized in a species-specific manner', *Journal of Biological Chemistry*, 295(8). Available at: https://doi.org/10.1074/jbc.RA119.010245.

Wen, W. *et al.* (2018) 'Autoregulation and Virulence Control by the Toxin-Antitoxin System SavRS in Staphylococcus aureus.', *Infection and immunity*, 86(5). Available at: https://doi.org/10.1128/IAI.00032-18.

Wen, Y., Behiels, E. and Devreese, B. (2014) 'Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity', *Pathogens and Disease*, 70(3), pp. 240–249. Available at: https://doi.org/10.1111/2049-632X.12145.

Wendorff, T.J. and Berger, J.M. (2018) 'Topoisomerase VI senses and exploits both DNA crossings and bends to facilitate strand passage', *eLife*, 7. Available at: https://doi.org/10.7554/eLife.31724.

WHO. (2004). Global Tuberculosis Report 2004. [Online]. World Health Organisation. Last Updated:October 2004. Available at: https://apps.who.int/iris/bitstream/handle/10665/42889/9241562641.pdf?sequence=2 [Accessed 01 November 2022].

WHO. (2005). Global Tuberculosis Report 2005. [Online]. World Health Organisation. Last Updated: October 2005. Available at: https://apps.who.int/iris/bitstream/handle/10665/137094/9789241564809_eng.pdf?sequence=1 [Accessed 01 November 2022].

WHO. (2014). Global Tuberculosis Report 2014. [Online]. World Health Organisation. Last Updated: October 2014. Available at: https://apps.who.int/iris/bitstream/handle/10665/144569/9241562919_eng.pdf?sequence=1 [Accessed 01 November 2022].

WHO. (2016). Global Tuberculosis Report 2016. [Online]. World Health Organisation. Last Updated: October 2016. Available at: https://apps.who.int/iris/bitstream/handle/10665/250441/9789241565394-eng.pdf?sequence=1 [Accessed 01 November 2022].

WHO. (2017). Global Tuberculosis Report 2017. [Online]. World Health Organisation. Last Updated: October 2017. Available at: https://apps.who.int/iris/handle/10665/259366 [Accessed 01 November 2022].

WHO. (2021). Global Tuberculosis Report 2021. [Online]. World Health Organisation. Last Updated: October 2021. Available at: https://www.who.int/publications/i/item/9789240037021 [Accessed 01 November 2022].

WHO. (2022). Global Tuberculosis Report 2022. [Online]. World Health Organisation. Last Updated: October 2022. Available at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 [Accessed 01 November 2022].

WHO. (2022b). Rapid communication: key changes to the treatment of drug-resistant tuberculosis. [Online]. World Health Organisation. Last Updated: May 2022. Available at: https://www.who.int/publications/i/item/WHO-UCN-TB-2022-2 [Accessed 01 November 2022].

Wilmaerts, D. *et al.* (2018) 'The persistence-inducing toxin hokb forms dynamic pores that cause ATP leakage', *mBio*, 9(4). Available at: https://doi.org/10.1128/mBio.00744-18.

Winn, M.D. *et al.* (2011) 'Overview of the CCP4 suite and current developments.', *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4), pp. 235–42. Available at: https://doi.org/10.1107/S0907444910045749.

Winter, G. and IUCr (2010) '*xia2* : an expert system for macromolecular crystallography data reduction', *Journal of Applied Crystallography*, 43(1), pp. 186–190. Available at: https://doi.org/10.1107/S0021889809045701.

Winther, K.S. *et al.* (2013) 'VapC20 of mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA', *Nature Communications*, 4. Available at: https://doi.org/10.1038/ncomms3796.

Wohlkonig, A. *et al.* (2010) 'Structural basis of quinolone inhibition of type IIA topoisomerases and target-mediated resistance', *Nature Structural and Molecular Biology*, 17(9). Available at: https://doi.org/10.1038/nsmb.1892.

Woldringh, C.L., Jensen, P.R. and Westerhoff, H. v. (1995) 'Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces?', *FEMS Microbiology Letters*, 131(3). Available at: https://doi.org/10.1111/j.1574-6968.1995.tb07782.x.

Wolf, A.J. *et al.* (2007) 'Mycobacterium tuberculosis Infects Dendritic Cells with High Frequency and Impairs Their Function In Vivo', *The Journal of Immunology*, 179(4). Available at: https://doi.org/10.4049/jimmunol.179.4.2509.

Wolf, A.J. *et al.* (2008) 'Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs', *Journal of Experimental Medicine*, 205(1). Available at: https://doi.org/10.1084/jem.20071367.

Wood, T.K., Knabel, S.J. and Kwan, B.W. (2013) 'Bacterial persister cell formation and dormancy', *Applied and Environmental Microbiology*. Available at: https://doi.org/10.1128/AEM.02636-13.

Worby, C. a *et al.* (2010) 'The Fic domain : A new paradigm for adenylylation', *Molecular cell*, 34(1).

Xavier Gomis-Rüth, F. *et al.* (1998) 'The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator', *EMBO Journal*, 17(24). Available at: https://doi.org/10.1093/emboj/17.24.7404.

Xavier, A.S. and Lakshmanan, M. (2014) 'Delamanid: A new armor in combating drug-resistant tuberculosis', *Journal of Pharmacology and Pharmacotherapeutics*, 5(3). Available at: https://doi.org/10.4103/0976-500X.136121.

Xie, Y. *et al.* (2018) 'TADB 2.0: An updated database of bacterial type II toxin-antitoxin loci', *Nucleic Acids Research*, 46(D1), pp. D749–D753. Available at: https://doi.org/10.1093/nar/gkx1033.

Xu, J. *et al.* (2018) 'Identification of three type II toxin-antitoxin systems in Streptococcus suis serotype 2', *Toxins*, 10(11). Available at: https://doi.org/10.3390/toxins10110467.

Xue, L. *et al.* (2020) 'Distinct oligomeric structures of the YoeB-YefM complex provide insights into the conditional cooperativity of type II toxin-antitoxin system', *Nucleic Acids Research*, 48(18). Available at: https://doi.org/10.1093/nar/gkaa706.

Yamaguchi, Y. and Inouye, M. (2011) 'Regulation of growth and death in Escherichia coli by toxin-antitoxin systems', *Nature Reviews Microbiology*. Available at: https://doi.org/10.1038/nrmicro2651.

Yamaguchi, Y. and Inouye, M. (2016) 'Toxin-Antitoxin Systems in Bacteria and Archaea', in *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*. Available at: https://doi.org/10.1002/9781119004813.ch8.

Yang, J. *et al.* (2016) 'Structural insight into the E. coli HigBA complex', *Biochemical and Biophysical Research Communications*, 478(4). Available at: https://doi.org/10.1016/j.bbrc.2016.08.131.

Yang, Q. *et al.* (2012) 'Physical and functional interactions between 3-methyladenine DNA glycosylase and topoisomerase i in mycobacteria', *Biochemistry (Moscow)*, 77(4). Available at: https://doi.org/10.1134/S0006297912040098.

Yang, W. (2011) 'Nucleases: Diversity of structure, function and mechanism', *Quarterly Reviews of Biophysics*, 44(1). Available at: https://doi.org/10.1017/S0033583510000181.

Yano, T. *et al.* (2011) 'Reduction of Clofazimine by Mycobacterial Type 2 NADH:Quinone Oxidoreductase', *Journal of Biological Chemistry*, 286(12). Available at: https://doi.org/10.1074/jbc.m110.200501.

Yu, X. *et al.* (2020) 'Characterization of a toxin-antitoxin system in Mycobacterium tuberculosis suggests neutralization by phosphorylation as the antitoxicity mechanism', *Communications Biology*, 3(1), p. 216. Available at: https://doi.org/10.1038/s42003-020-0941-1.

Yuan, J. *et al.* (2010) 'Vibrio cholerae ParE2 poisons DNA gyrase via a mechanism distinct from other gyrase inhibitors', *Journal of Biological Chemistry*, 285(51), pp. 40397–40408. Available at: https://doi.org/10.1074/jbc.M110.138776.

Zainuddin, Z.F. and Dale, J.W. (1990) 'Does Mycobacterium tuberculosis have plasmids?', *Tubercle*, 71(1). Available at: https://doi.org/10.1016/0041-3879(90)90060-L.

Zaw, M.T., Emran, N.A. and Lin, Z. (2018) 'Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis', *Journal of Infection and Public Health*. Available at: https://doi.org/10.1016/j.jiph.2018.04.005.

Zechiedrich, E.L., Khodursky, A.B. and Cozzarelli, N.R. (1997) 'Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli', *Genes and Development*, 11(19). Available at: https://doi.org/10.1101/gad.11.19.2580.

Zhai, W. *et al.* (2019) 'The immune escape mechanisms of Mycobacterium Tuberculosis', *International Journal of Molecular Sciences*. Available at: https://doi.org/10.3390/ijms20020340.

Zhang, D. *et al.* (2012) 'Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics', *Biology Direct*, 7. Available at: https://doi.org/10.1186/1745-6150-7-18.

Zhang, S.-P. *et al.* (2020) 'Type II toxin–antitoxin system in bacteria: activation, function, and mode of action', *Biophysics Reports*, 6(2–3). Available at: https://doi.org/10.1007/s41048-020-00109-8.

Zhao, R. *et al.* (2019) 'Structure and allosteric coupling of type Ⅱ antitoxin CopASO', *Biochemical and Biophysical Research Communications*, 514(4). Available at: https://doi.org/10.1016/j.bbrc.2019.05.049.

Zheng, J. *et al.* (2013) 'Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in mycobacterium tuberculosis', *Journal of Biological Chemistry*, 288(32). Available at: https://doi.org/10.1074/jbc.M113.475798.

Zhou, J. *et al.* (2021) 'Insights into the neutralization and dna binding of toxin–antitoxin system pareso-copaso by structure-function studies', *Microorganisms*, 9(12). Available at: https://doi.org/10.3390/microorganisms9122506.

Zhu, L. *et al.* (2008) 'The mRNA interferases, MazF-mt3 and MazF-mt7 from Mycobacterium tuberculosis target unique pentad sequences in single-stranded RNA', *Molecular Microbiology*, 69(3). Available at: https://doi.org/10.1111/j.1365-2958.2008.06284.x.

Ziemski, M. *et al.* (2021) 'Genome-wide interaction screen for Mycobacterium tuberculosis ClpCP protease reveals toxin–antitoxin systems as a major substrate class', *FEBS Journal*, 288(1). Available at: https://doi.org/10.1111/febs.15335.

Zimmer, A.J. *et al.* (2022) 'Tuberculosis in times of COVID-19', *Journal of epidemiology and community health*, 76(3). Available at: https://doi.org/10.1136/jech-2021-217529.

Supplementary Figures

Figure S1. Structures of ParDE system complexes in the PDB

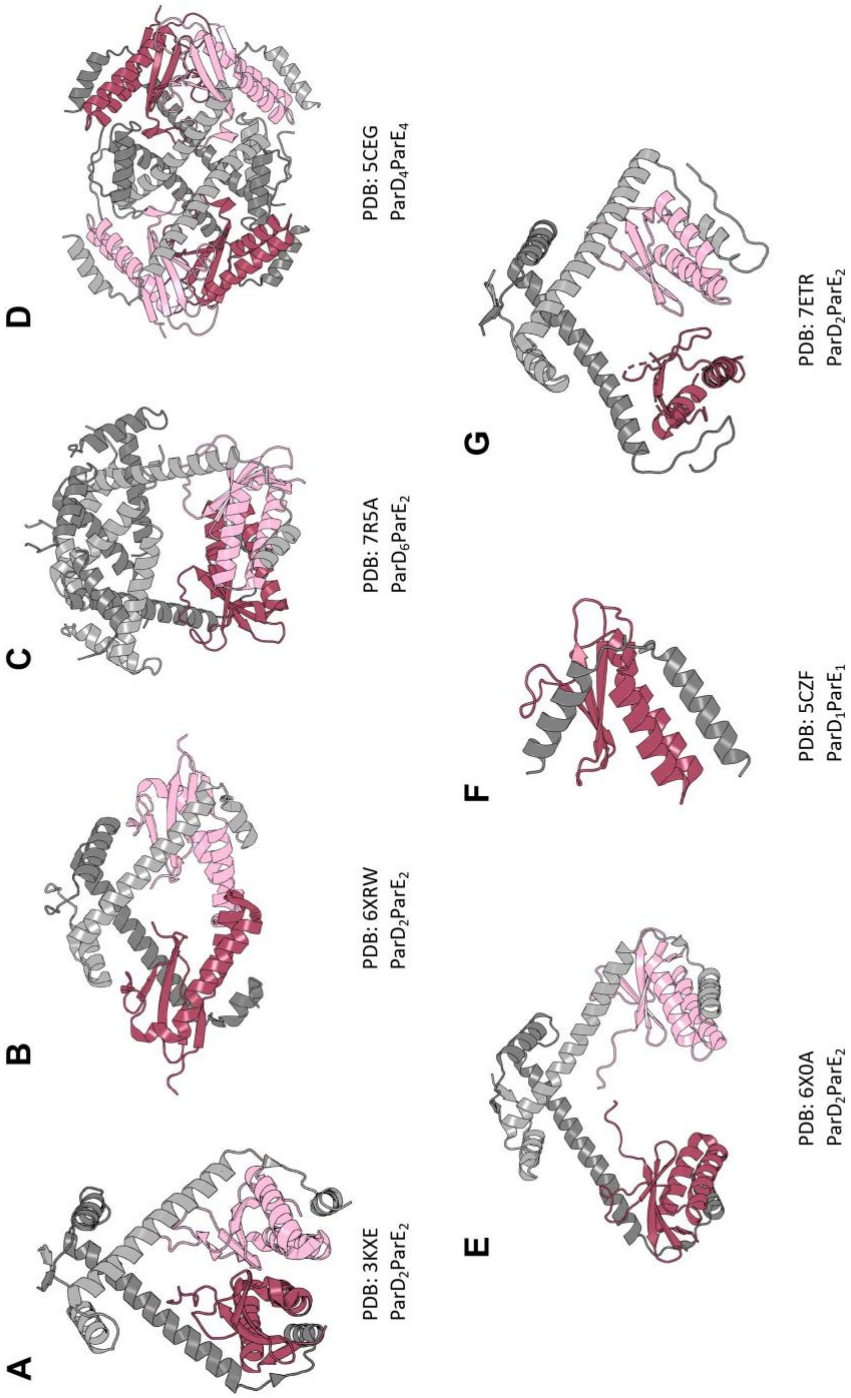
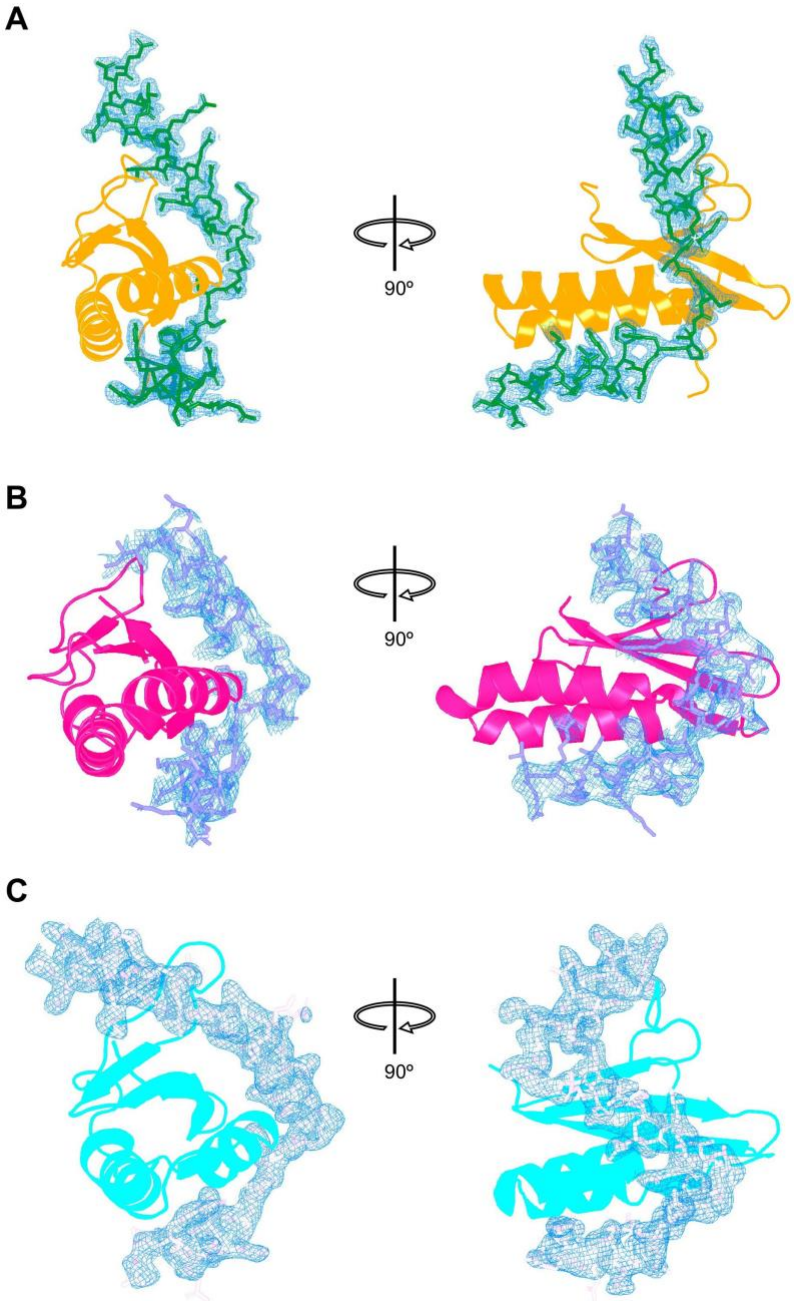


Figure legend overleaf

Figure S1. Structures of ParDE system complexes in the PDB

ParD Antitoxin chains are coloured in pairs as light and dark grey. ParE toxins are coloured light pink and raspberry for clarity. **A**) *C. crescentus* heterotetrameric $ParD_2ParE_2$ complex (PDB: 3KXE); **B**) *P. aeruginosa* heterotetrameric $ParD_2ParE_2$ complex (PDB: 6XRW); **C**) *V. cholerae* heterohexameric $ParD_4ParE_2$ complex (PDB: 7R5A); **D**) *M. opportunistium* heterooctameric $ParD_4ParE_4$ complex (PDB: 5CEG); **E**) *M. opportunistium* heterotetrameric $ParD_2ParE_2$ complex (PDB: 6X0A); **F**) *E. coli* dimeric $ParD_1ParE_1$ complex (PDB: 5CZF); **G**) *S. oneidensis* heterotetrameric $ParD_2ParE_2$ complex (PDB: 7ETR). Each structure demonstrates the ParD-ParE interaction at the highly conserved ParE toxin interface. All structures above, with the exemption of that shown in F, contain the ParD CopG dimerisation domain and a similar V-shaped structure. Notable differences are seen in A, whereby the ParE toxins form an interface; in C, whereby an additional $ParD_2$ CopG domain is inserted in-between the dimers bound to cognate toxins; in D, whereby two tetramers have become entangled; and in F, whereby the ParD chain only exists as the toxin-binding region.

Figure S2. 2Fo-Fc representation of antitoxin chains at the toxin-antitoxin interface



A) ParDE1 conserved interface showing the ParD1 chain (residues 40 – 80, forest green) with 2Fo-Fc electron density map. ParE1 is rendered bright orange; **B**) ParDE2 conserved interface showing the ParD2 chain (residues 36 – 71, slate blue) with 2Fo-Fc electron density map. ParE2 is rendered magenta; **C**) RelBE1 conserved interface showing the RelB chain (residues 43 – 88, light pink) with 2Fo-Fc electron density map. RelE1 is rendered cyan. Electron density maps are contoured at 2σ.

Supplementary Files

Supplementary file S1 – p1B-6His-TEV-GyrB (pTRB312) ORF

```
atg aaa tct tct cac cat cac cat cac cat gaa aac ctg tac ttc caa tcc aat gca gct
 M   K   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   S   N   A   A
gcc cag aaa aag aag gcc caa gac gaa tac ggc gct gcg tct atc acc att ctc gaa ggg
 A   Q   K   K   K   A   Q   D   E   Y   G   A   A   S   I   T   I   L   E   G
ctg gag gcc gtc cgc aaa cgt ccc ggc atg tac att ggc tcg acc ggt gag cgc ggt tta
 L   E   A   V   R   K   R   P   G   M   Y   I   G   S   T   G   E   R   G   L
cac cat ctc att tgg gag gtg gtc gac aac gcg gtc gac gag gcg atg gcc ggt tat gca
 H   H   L   I   W   E   V   V   D   N   A   V   D   E   A   M   A   G   Y   A
acc aca gtg aac gta gtg ctg ctt gag gat ggc ggt gtc gag gtc gcc gac gac ggc cgc
 T   T   V   N   V   V   L   L   E   D   G   G   V   E   V   A   D   D   G   R
ggc att ccg gtc gcc acc cac gcc tcc ggc ata ccg acc gtc gac gtg gtg atg aca caa
 G   I   P   V   A   T   H   A   S   G   I   P   T   V   D   V   V   M   T   Q
cta cat gcc ggc ggc aag ttc gac tcg gac gcg tat gcg ata tct ggt ggt ctg cac ggc
 L   H   A   G   G   K   F   D   S   D   A   Y   A   I   S   G   G   L   H   G
gtc ggc gtg tcg gtg gtt aac gcg cta tcc acc cgg ctc gaa gtc gag atc aag cgc gac
 V   G   V   S   V   V   N   A   L   S   T   R   L   E   V   E   I   K   R   D
ggg tac gag tgg tct cag gtt tat gag aag tcg gaa ccc ctg ggc ctc aag caa ggg gcg
 G   Y   E   W   S   Q   V   Y   E   K   S   E   P   L   G   L   K   Q   G   A
ccg acc aag aag acg ggg tca acg gtg cgg ttc tgg gcc gac ccc gct gtt ttc gaa acc
 P   T   K   K   T   G   S   T   V   R   F   W   A   D   P   A   V   F   E   T
acg gaa tac gac ttc gaa acc gtc gcc cgc cgg ctg caa gag atg gcg ttc ctc aac aag
 T   E   Y   D   F   E   T   V   A   R   R   L   Q   E   M   A   F   L   N   K
ggg ctg acc atc aac ctg acc gac gag agg gtg acc caa gac gag gtc gtc gac gaa gtg
 G   L   T   I   N   L   T   D   E   R   V   T   Q   D   E   V   V   D   E   V
gtc agc gac gtc gcc gag gcg ccg aag tcg gca agt gaa cgc gca gcc gaa tcc act gca
 V   S   D   V   A   E   A   P   K   S   A   S   E   R   A   A   E   S   T   A
ccg cac aaa gtt aag agc cgc acc ttt cac tat ccg ggt ggc ctg gtg gac ttc gtg aaa
 P   H   K   V   K   S   R   T   F   H   Y   P   G   G   L   V   D   F   V   K
cac atc aac cgc acc aag aac gcg att cat agc agc atc gtg gac ttt tcc ggc aag ggc
 H   I   N   R   T   K   N   A   I   H   S   S   I   V   D   F   S   G   K   G
acc ggg cac gag gtg gag atc gcg atg caa tgg aac gcc ggg tat tcg gag tcg gtg cac
 T   G   H   E   V   E   I   A   M   Q   W   N   A   G   Y   S   E   S   V   H
acc ttc gcc aac acc atc aac acc cac gag ggc ggc acc cac gaa gag ggc ttc cgc agc
 T   F   A   N   T   I   N   T   H   E   G   G   T   H   E   E   G   F   R   S
gcg ctg acg tcg gtg gtg aac aag tac gcc aag gac cgc aag cta ctg aag gac aag gac
 A   L   T   S   V   V   N   K   Y   A   K   D   R   K   L   L   K   D   K   D
ccc aac ctc acc ggt gac gat atc cgg gaa ggc ctg gcc gct gtg atc tcg gtg aag gtc
 P   N   L   T   G   D   D   I   R   E   G   L   A   A   V   I   S   V   K   V
agc gaa ccg cag ttc gag ggc cag acc aag acc aag ttg ggc aac acc gag gtc aaa tcg
 S   E   P   Q   F   E   G   Q   T   K   T   K   L   G   N   T   E   V   K   S
ttt gtg cag aag gtc tgt aac gaa cag ctg acc cac tgg ttt gaa gcc aac ccc acc gac
 F   V   Q   K   V   C   N   E   Q   L   T   H   W   F   E   A   N   P   T   D
gcg aaa gtc gtt gtg aac aag gct gtg tcc tcg gcg caa gcc cgt atc gcg gca cgt aag
 A   K   V   V   V   N   K   A   V   S   S   A   Q   A   R   I   A   A   R   K
gca cga gag ttg gtg cgg cgt aag agc gcc acc gac atc ggt gga ttg ccc ggc aag ctg
 A   R   E   L   V   R   R   K   S   A   T   D   I   G   G   L   P   G   K   L
gcc gat tgc cgt tcc acg gat ccg cgc aag tcc gaa ctg tat gtc gta gaa ggt gac tcg
 A   D   C   R   S   T   D   P   R   K   S   E   L   Y   V   V   E   G   D   S
gcc ggc ggt tct gca aaa agc ggt cgc gat tcg atg ttc cag gcg ata ctt ccg ctg cgc
 A   G   G   S   A   K   S   G   R   D   S   M   F   Q   A   I   L   P   L   R
ggc aag atc atc aat gtg gag aaa gcg cgc atc gac cgg gtg cta aag aac acc gaa gtt
 G   K   I   I   N   V   E   K   A   R   I   D   R   V   L   K   N   T   E   V
cag gcg atc atc acg gcg ctg ggc acc ggg atc cac gac gag ttc gat atc ggc aag ctg
 Q   A   I   I   T   A   L   G   T   G   I   H   D   E   F   D   I   G   K   L
cgc tac cac aag atc gtg ctg atg gcc gac gcc gat gtt gac ggc caa cat att tcc acg
 R   Y   H   K   I   V   L   M   A   D   A   D   V   D   G   Q   H   I   S   T
ctg ttg ttg acg ttg ttg ttc cgg ttc atg cgg ccg ctc atc gag aac ggg cat gtg ttt
 L   L   L   T   L   L   F   R   F   M   R   P   L   I   E   N   G   H   V   F
ttg gca caa ccg ccg ctg tac aaa ctc aag tgg cag cgc agt gac ccg gaa ttc gca tac
 L   A   Q   P   P   L   Y   K   L   K   W   Q   R   S   D   P   E   F   A   Y
tcc gac cgc gag cgc gac ggt ctg ctg gag gcg ggg ctg aag gcc ggg aag aag atc aac
 S   D   R   E   R   D   G   L   L   E   A   G   L   K   A   G   K   K   I   N
aag gaa gac ggc att cag cgg tac aag ggt cta ggt gaa atg gac gct aag gag ttg tgg
 K   E   D   G   I   Q   R   Y   K   G   L   G   E   M   D   A   K   E   L   W
gag acc acc atg gat ccc tcg gtt cgt gtg ttg cgt caa gtg acg ctg gac gac gcc gcc
 E   T   T   M   D   P   S   V   R   V   L   R   Q   V   T   L   D   D   A   A
gcc gcc gac gag ttg ttc tcc atc ctg atg ggc gag gac gtc gac gcg cgg cgc agc ttt
 A   A   D   E   L   F   S   I   L   M   G   E   D   V   D   A   R   R   S   F
atc acc cgc aac gcc aag gat gtt cgg ttc ctg gat gtc taa
 I   T   R   N   A   K   D   V   R   F   L   D   V   -
```

Supplementary file S2 – p1B-6His-TEV-GyrA (pTRB696) ORF

```
atg aaa tct tct cac cat cac cat cac cat gaa aac ctg tac ttc caa tcc aat gca aca
 M   K   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   S   N   A   T
```

```
gac acg acg ttg ccg cct gac gac tcg ctc gac cgg atc gaa ccg gtt gac atc gag cag
 D   T   T   L   P   P   D   D   S   L   D   R   I   E   P   V   D   I   E   Q
gag atg cag cgc agc tac atc gac tat gcg atg agc gtg atc gtc ggc cgc gcg ctg ccg
 E   M   Q   R   S   Y   I   D   Y   A   M   S   V   I   V   G   R   A   L   P
gag gtg cgc gac ggg ctc aag ccc gtg cat cgc cgg gtg ctc tat gca atg ttc gat tcc
 E   V   R   D   G   L   K   P   V   H   R   R   V   L   Y   A   M   F   D   S
ggc ttc cgc ccg gac cgc agc cac gcc aag tcg gcc cgg tcg gtt gcc gag acc atg ggc
 G   F   R   P   D   R   S   H   A   K   S   A   R   S   V   A   E   T   M   G
aac tac cac ccg cac ggc gac gcg tcg atc tac gac agc ctg gtg cgc atg gcc cag ccc
 N   Y   H   P   H   G   D   A   S   I   Y   D   S   L   V   R   M   A   Q   P
tgg tcg ctg cgc tac ccg ctg gtg gac ggc cag ggc aac ttc ggc tcg cca ggc aat gac
 W   S   L   R   Y   P   L   V   D   G   Q   G   N   F   G   S   P   G   N   D
cca ccg gcg gcg atg agg tac acc gaa gcc cgg ctg acc ccg ttg gcg atg gag atg ctg
 P   P   A   A   M   R   Y   T   E   A   R   L   T   P   L   A   M   E   M   L
agg gaa atc gac gag gag aca gtc gat ttc atc cct aac tac gac ggc cgg gtg caa gag
 R   E   I   D   E   E   T   V   D   F   I   P   N   Y   D   G   R   V   Q   E
ccg acg gtg cta ccc agc cgg ttc ccc aac ctg ctg gcc aac ggg tca ggc ggc atc gcg
 P   T   V   L   P   S   R   F   P   N   L   L   A   N   G   S   G   G   I   A
gtc ggc atg gca acc aat atc ccg ccg cac aac ctg cgt gag ctg gcc gac gcg gtg ttc
 V   G   M   A   T   N   I   P   P   H   N   L   R   E   L   A   D   A   V   F
tgg gcg ctg gag aat cac gac gcc gac gaa gag gag acc ctg gcc gcg gtc atg ggg cgg
 W   A   L   E   N   H   D   A   D   E   E   E   T   L   A   A   V   M   G   R
gtt aaa ggc ccg gac ttc ccg acc gcc gga ctg atc gtc gga tcc cag ggc acc gct gat
 V   K   G   P   D   F   P   T   A   G   L   I   V   G   S   Q   G   T   A   D
gcc tac aaa act ggc cgc ggc tcc att cga atg cgc gga gtt gtt gag gta gaa gag gat
 A   Y   K   T   G   R   G   S   I   R   M   R   G   V   V   E   V   E   E   D
tcc cgc ggt cgt acc tcg ctg gtg atc acc gag ttg ccg tat cag gtc aac cac gac aac
 S   R   G   R   T   S   L   V   I   T   E   L   P   Y   Q   V   N   H   D   N
ttc atc act tcg atc gcc gaa cag gtc cga gac ggc aag ctg gcc ggc att tcc aac att
 F   I   T   S   I   A   E   Q   V   R   D   G   K   L   A   G   I   S   N   I
gag gac cag tct agc gat cgg gtc ggt tta cgc atc gtc atc gag atc aag cgc gat gcg
 E   D   Q   S   S   D   R   V   G   L   R   I   V   I   E   I   K   R   D   A
gtg gcc aag gtg gtg atc aat aac ctt tac aag cac acc cag ctg cag acc agc ttt ggc
 V   A   K   V   V   I   N   N   L   Y   K   H   T   Q   L   Q   T   S   F   G
gcc aac atg cta gcg atc gtc gac ggg gtg ccg cgc acg ctg cgg ctg gac cag ctg atc
 A   N   M   L   A   I   V   D   G   V   P   R   T   L   R   L   D   Q   L   I
cgc tat tac gtt gac cac caa ctc gac gtc att gtg cgg cgc acc acc tac cgg ctg cgc
 R   Y   Y   V   D   H   Q   L   D   V   I   V   R   R   T   T   Y   R   L   R
aag gca aac gag cga gcc cac att ctg cgc ggc ctg gtt aaa gcg ctc gac gcg ctg gac
 K   A   N   E   R   A   H   I   L   R   G   L   V   K   A   L   D   A   L   D
gag gtc att gca ctg atc cgg gcg tcg gag acc gtc gat atc gcc cgg gcc gga ctg atc
 E   V   I   A   L   I   R   A   S   E   T   V   D   I   A   R   A   G   L   I
gag ctg ctc gac atc gac gag atc cag gcc cag gca atc ctg gac atg cag ttg cgg cgc
 E   L   L   D   I   D   E   I   Q   A   Q   A   I   L   D   M   Q   L   R   R
ctg gcc gca ctg gaa cgc cag cgc atc atc gac gac ctg gcc aaa atc gag gcc gag atc
 L   A   A   L   E   R   Q   R   I   I   D   D   L   A   K   I   E   A   E   I
gcc gat ctg gaa gac atc ctg gca aaa ccc gag cgg cag cgt ggg atc gtg cgc gac gaa
 A   D   L   E   D   I   L   A   K   P   E   R   Q   R   G   I   V   R   D   E
ctc gcc gaa atc gtg gac agg cac ggc gac gac cgg cgt acc cgg atc atc gcg gcc gac
 L   A   E   I   V   D   R   H   G   D   D   R   R   T   R   I   I   A   A   D
gga gac gtc agc gac gag gat ttg atc gcc cgc gag gac gtc gtt gtc act atc acc gaa
 G   D   V   S   D   E   D   L   I   A   R   E   D   V   V   V   T   I   T   E
acg gga tac gcc aag cgc acc aag acc gat ctg tat cgc agc cag aaa cgc ggc ggc aag
 T   G   Y   A   K   R   T   K   T   D   L   Y   R   S   Q   K   R   G   G   K
ggc gtg cag ggt gcg ggg ttg aag cag gac gac atc gtc gcg cac ttc ttc gtg tgc tcc
 G   V   Q   G   A   G   L   K   Q   D   D   I   V   A   H   F   F   V   C   S
acc cac gat ttg atc ctg ttc ttc acc acc cag gga cgg gtt tat cgg gcc aag gcc tac
 T   H   D   L   I   L   F   F   T   T   Q   G   R   V   Y   R   A   K   A   Y
gac ttg ccc gag gcc tcc cgg acg gcg cgc ggg cag cac gtg gcc aac ctg tta gcc ttc
 D   L   P   E   A   S   R   T   A   R   G   Q   H   V   A   N   L   L   A   F
cag ccc gag gaa cgc atc gcc cag gtc atc cag att cgc ggc tac acc gac gcc ccg tac
 Q   P   E   E   R   I   A   Q   V   I   Q   I   R   G   Y   T   D   A   P   Y
ctg gtg ctg gcc act cgc aac ggg ctg gtg aaa aag tcc aag ctg acc gac ttc gac tcc
 L   V   L   A   T   R   N   G   L   V   K   K   S   K   L   T   D   F   D   S
aat cgc tcg ggc gga atc gtg gcg gtc aac ctg cgc gac aac gac gag ctg gtc ggt gcg
 N   R   S   G   G   I   V   A   V   N   L   R   D   N   D   E   L   V   G   A
gtg ctg tgt tcg gcc ggc gac gac ctg ctg ctg gtc tcg gcc aac ggg cag tcc atc agg
 V   L   C   S   A   G   D   D   L   L   L   V   S   A   N   G   Q   S   I   R
ttc tcg gcg acc gac gag gcg ctg cgg cca atg ggt cgt gcc acc tcg ggt gtg cag ggc
 F   S   A   T   D   E   A   L   R   P   M   G   R   A   T   S   G   V   Q   G
atg cgg ttc aat atc gac gac cgg ctg ctg tcg ctg aac gtc gtg cgt gaa ggc acc tat
 M   R   F   N   I   D   D   R   L   L   S   L   N   V   V   R   E   G   T   Y
ctg ctg gtg gcg acg tca ggg ggc tat gcg aaa cgt acc gcg atc gag gaa tac ccg gta
 L   L   V   A   T   S   G   G   Y   A   K   R   T   A   I   E   E   Y   P   V
cag ggc cgc ggc ggt aaa ggt gtg ctg acg gtc atg tac gac cgc cgg cgc ggc agg ttg
 Q   G   R   G   G   K   G   V   L   T   V   M   Y   D   R   R   R   G   R   L
gtt ggg gcg ttg att gtc gac gac gac agc gag ctg tat gcc gtc act tcc ggc ggt ggc
 V   G   A   L   I   V   D   D   D   S   E   L   Y   A   V   T   S   G   G   G
gtg atc cgc acc gcg gca cgc cag gtt cgc aag gcg gga cgg cag acc aag ggt gtt cgg
 V   I   R   T   A   A   R   Q   V   R   K   A   G   R   Q   T   K   G   V   R
ttg atg aat ctg ggc gag ggc gac aca ctg ttg gcc atc gcg cgc aac gcc gaa gaa agt
 L   M   N   L   G   E   G   D   T   L   L   A   I   A   R   N   A   E   E   S
```

```
ggc gac gat aat gcc gtg gac gcc aac ggc gca gac cag acg ggc aat taa
 G   D   D   N   A   V   D   A   N   G   A   D   Q   T   G   N   -
```

Supplementary file S3 – pET28-MHL-6His-TEV-GyrBA (pTRB697) ORF

```
atg gga tct tcc cat cat cat cac cac cat gaa aac ctg tac ttt caa agt aac gct gct
 M   G   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   S   N   A   A
gcc cag aaa aag aag gcc caa gac gaa tac ggc gct gcg tct atc acc att ctc gaa ggg
 A   Q   K   K   K   A   Q   D   E   Y   G   A   A   S   I   T   I   L   E   G
ctg gag gcc gtc cgc aaa cgt ccc ggc atg tac att ggc tcg acc ggt gag cgc ggt tta
 L   E   A   V   R   K   R   P   G   M   Y   I   G   S   T   G   E   R   G   L
cac cat ctc att tgg gag gtg gtc gac aac gcg gtc gac gag gcg atg gcc ggt tat gca
 H   H   L   I   W   E   V   V   D   N   A   V   D   E   A   M   A   G   Y   A
acc aca gtg aac gta gtg ctg ctt gag gat ggc ggt gtc gag gtc gcc gac gac ggc cgc
 T   T   V   N   V   V   L   L   E   D   G   G   V   E   V   A   D   D   G   R
ggc att ccg gtc gcc acc cac gcc tcc ggc ata ccg acc gtc gac gtg gtg atg aca caa
 G   I   P   V   A   T   H   A   S   G   I   P   T   V   D   V   V   M   T   Q
cta cat gcc ggc ggc aag ttc gac tcg gac gcg tat gcg ata tct ggt ggt ctg cac ggc
 L   H   A   G   G   K   F   D   S   D   A   Y   A   I   S   G   G   L   H   G
gtc ggc gtg tcg gtg gtt aac gcg cta tcc acc cgg ctc gaa gtc gag atc aag cgc gac
 V   G   V   S   V   V   N   A   L   S   T   R   L   E   V   E   I   K   R   D
ggg tac gag tgg tct cag gtt tat gag aag tcg gaa ccc ctg ggc ctc aag caa ggg gcg
 G   Y   E   W   S   Q   V   Y   E   K   S   E   P   L   G   L   K   Q   G   A
ccg acc aag aag acg ggg tca acg gtg cgg ttc tgg gcc gac ccc gct gtt ttc gaa acc
 P   T   K   K   T   G   S   T   V   R   F   W   A   D   P   A   V   F   E   T
acg gaa tac gac ttc gaa acc gtc gcc cgc cgg ctg caa gag atg gcg ttc ctc aac aag
 T   E   Y   D   F   E   T   V   A   R   R   L   Q   E   M   A   F   L   N   K
ggg ctg acc atc aac ctg acc gac gag agg gtg acc caa gac gag gtc gtc gac gaa gtg
 G   L   T   I   N   L   T   D   E   R   V   T   Q   D   E   V   V   D   E   V
gtc agc gac gtc gcc gag gcg ccg aag tcg gca agt gaa cgc gca gcc gaa tcc act gca
 V   S   D   V   A   E   A   P   K   S   A   S   E   R   A   A   E   S   T   A
ccg cac aaa gtt aag agc cgc acc ttt cac tat ccg ggt ggc ctg gtg gac ttc gtg aaa
 P   H   K   V   K   S   R   T   F   H   Y   P   G   G   L   V   D   F   V   K
cac atc aac cgc acc aag aac gcg att cat agc agc atc gtg gac ttt tcc ggc aag ggc
 H   I   N   R   T   K   N   A   I   H   S   S   I   V   D   F   S   G   K   G
acc ggg cac gag gtg gag atc gcg atg caa tgg aac gcc ggg tat tcg gag tcg gtg cac
 T   G   H   E   V   E   I   A   M   Q   W   N   A   G   Y   S   E   S   V   H
acc ttc gcc aac acc atc aac acc cac gag ggc ggc acc cac gaa gag ggc ttc cgc agc
 T   F   A   N   T   I   N   T   H   E   G   G   T   H   E   E   G   F   R   S
gcg ctg acg tcg gtg gtg aac aag tac gcc aag gac cgc aag cta ctg aag gac aag gac
 A   L   T   S   V   V   N   K   Y   A   K   D   R   K   L   L   K   D   K   D
ccc aac ctc acc ggt gac gat atc cgg gaa ggc ctg gcc gct gtg atc tcg gtg aag gtc
 P   N   L   T   G   D   D   I   R   E   G   L   A   A   V   I   S   V   K   V
agc gaa ccg cag ttc gag ggc cag acc aag acc aag ttg ggc aac acc gag gtc aaa tcg
 S   E   P   Q   F   E   G   Q   T   K   T   K   L   G   N   T   E   V   K   S
ttt gtg cag aag gtc tgt aac gaa cag ctg acc cac tgg ttt gaa gcc aac ccc acc gac
 F   V   Q   K   V   C   N   E   Q   L   T   H   W   F   E   A   N   P   T   D
gcg aaa gtc gtt gtg aac aag gct gtg tcc tcg gcg caa gcc cgt atc gcg gca cgt aag
 A   K   V   V   V   N   K   A   V   S   S   A   Q   A   R   I   A   A   R   K
gca cga gag ttg gtg cgg cgt aag agc gcc acc gac atc ggt gga ttg ccc ggc aag ctg
 A   R   E   L   V   R   R   K   S   A   T   D   I   G   G   L   P   G   K   L
gcc gat tgc cgt tcc acg gat ccg cgc aag tcc gaa ctg tat gtc gta gaa ggt gac tcg
 A   D   C   R   S   T   D   P   R   K   S   E   L   Y   V   V   E   G   D   S
gcc ggc ggt tct gca aaa agc ggt cgc gat tcg atg ttc cag gcg ata ctt ccg ctg cgc
 A   G   G   S   A   K   S   G   R   D   S   M   F   Q   A   I   L   P   L   R
ggc aag atc atc aat gtg gag aaa gcg cgc atc gac cgg gtg cta aag aac acc gaa gtt
 G   K   I   I   N   V   E   K   A   R   I   D   R   V   L   K   N   T   E   V
cag gcg atc atc acg gcg ctg ggc acc ggg atc cac gac gag ttc gat atc ggc aag ctg
 Q   A   I   I   T   A   L   G   T   G   I   H   D   E   F   D   I   G   K   L
cgc tac cac aag atc gtg ctg atg gcc gac gcc gat gtt gac ggc caa cat att tcc acg
 R   Y   H   K   I   V   L   M   A   D   A   D   V   D   G   Q   H   I   S   T
ctg ttg ttg acg ttg ttg ttc cgg ttc atg cgg ccg ctc atc gag aac ggg cat gtg ttt
 L   L   L   T   L   L   F   R   F   M   R   P   L   I   E   N   G   H   V   F
ttg gca caa ccg ccg ctg tac aaa ctc aag tgg cag cgc agt gac ccg gaa ttc gca tac
 L   A   Q   P   P   L   Y   K   L   K   W   Q   R   S   D   P   E   F   A   Y
tcc gac cgc gag cgc gac ggt ctg ctg gag gcg ggg ctg aag gcc ggg aag aag atc aac
 S   D   R   E   R   D   G   L   L   E   A   G   L   K   A   G   K   K   I   N
aag gaa gac ggc att cag cgg tac aag ggt cta ggt gaa atg gac gct aag gag ttg tgg
 K   E   D   G   I   Q   R   Y   K   G   L   G   E   M   D   A   K   E   L   W
gag acc acc atg gat ccc tcg gtt cgt gtg ttg cgt caa gtg acg ctg gac gac gcc gcc
 E   T   T   M   D   P   S   V   R   V   L   R   Q   V   T   L   D   D   A   A
gcc gcc gac gag ttg ttc tcc atc ctg atg ggc gag gac gtc gac gcg cgg cgc agc ttt
 A   A   D   E   L   F   S   I   L   M   G   E   D   V   D   A   R   R   S   F
atc acc cgc aac gcc aag gat gtt cgg ttc ctg gat gtc aca gac acg acg ttg ccg cct
 I   T   R   N   A   K   D   V   R   F   L   D   V   T   D   T   T   L   P   P
gac gac tcg ctc gac cgg atc gaa ccg gtt gac atc gag cag gag atg cag cgc agc tac
 D   D   S   L   D   R   I   E   P   V   D   I   E   Q   E   M   Q   R   S   Y
atc gac tat gcg atg agc gtg atc gtc ggc cgc gcg ctg ccg gag gtg cgc gac ggg ctc
 I   D   Y   A   M   S   V   I   V   G   R   A   L   P   E   V   R   D   G   L
```

```
aag ccc gtg cat cgc cgg gtg ctc tat gca atg ttc gat tcc ggc ttc cgc ccg gac cgc
 K   P   V   H   R   R   V   L   Y   A   M   F   D   S   G   F   R   P   D   R
agc cac gcc aag tcg gcc cgg tcg gtt gcc gag acc atg ggc aac tac cac ccg cac ggc
 S   H   A   K   S   A   R   S   V   A   E   T   M   G   N   Y   H   P   H   G
gac gcg tcg atc tac gac agc ctg gtg cgc atg gcc cag ccc tgg tcg ctg cgc tac ccg
 D   A   S   I   Y   D   S   L   V   R   M   A   Q   P   W   S   L   R   Y   P
ctg gtg gac ggc cag ggc aac ttc ggc tcg cca ggc aat gac cca ccg gcg gcg atg agg
 L   V   D   G   Q   G   N   F   G   S   P   G   N   D   P   P   A   A   M   R
tac acc gaa gcc cgg ctg acc ccg ttg gcg atg gag atg ctg agg gaa atc gac gag gag
 Y   T   E   A   R   L   T   P   L   A   M   E   M   L   R   E   I   D   E   E
aca gtc gat ttc atc cct aac tac gac ggc cgg gtg caa gag ccg acg gtg cta ccc agc
 T   V   D   F   I   P   N   Y   D   G   R   V   Q   E   P   T   V   L   P   S
cgg ttc ccc aac ctg ctg gcc aac ggg tca ggc ggc atc gcg gtc ggc atg gca acc aat
 R   F   P   N   L   L   A   N   G   S   G   G   I   A   V   G   M   A   T   N
atc ccg ccg cac aac ctg cgt gag ctg gcc gac gcg gtg ttc tgg gcg ctg gag aat cac
 I   P   P   H   N   L   R   E   L   A   D   A   V   F   W   A   L   E   N   H
gac gcc gac gaa gag gag acc ctg gcc gcg gtc atg ggg cgg gtt aaa ggc ccg gac ttc
 D   A   D   E   E   E   T   L   A   A   V   M   G   R   V   K   G   P   D   F
ccg acc gcc gga ctg atc gtc gga tcc cag ggc acc gct gat gcc tac aaa act ggc cgc
 P   T   A   G   L   I   V   G   S   Q   G   T   A   D   A   Y   K   T   G   R
ggc tcc att cga atg cgc gga gtt gtt gag gta gaa gag gat tcc cgc ggt cgt acc tcg
 G   S   I   R   M   R   G   V   V   E   V   E   E   D   S   R   G   R   T   S
ctg gtg atc acc gag ttg ccg tat cag gtc aac cac gac aac ttc atc act tcg atc gcc
 L   V   I   T   E   L   P   Y   Q   V   N   H   D   N   F   I   T   S   I   A
gaa cag gtc cga gac ggc aag ctg gcc ggc att tcc aac att gag gac cag tct agc gat
 E   Q   V   R   D   G   K   L   A   G   I   S   N   I   E   D   Q   S   S   D
cgg gtc ggt tta cgc atc gtc atc gag atc aag cgc gat gcg gtg gcc aag gtg gtg atc
 R   V   G   L   R   I   V   I   E   I   K   R   D   A   V   A   K   V   V   I
aat aac ctt tac aag cac acc cag ctg cag acc agc ttt ggc gcc aac atg cta gcg atc
 N   N   L   Y   K   H   T   Q   L   Q   T   S   F   G   A   N   M   L   A   I
gtc gac ggg gtg ccg cgc acg ctg cgg ctg gac cag ctg atc cgc tat tac gtt gac cac
 V   D   G   V   P   R   T   L   R   L   D   Q   L   I   R   Y   Y   V   D   H
caa ctc gac gtc att gtg cgg cgc acc acc tac cgg ctg cgc aag gca aac gag cga gcc
 Q   L   D   V   I   V   R   R   T   T   Y   R   L   R   K   A   N   E   R   A
cac att ctg cgc ggc ctg gtt aaa gcg ctc gac gcg ctg gac gag gtc att gca ctg atc
 H   I   L   R   G   L   V   K   A   L   D   A   L   D   E   V   I   A   L   I
cgg gcg tcg gag acc gtc gat atc gcc cgg gcc gga ctg atc gag ctg ctc gac atc gac
 R   A   S   E   T   V   D   I   A   R   A   G   L   I   E   L   L   D   I   D
gag atc cag gcc cag gca atc ctg gac atg cag ttg cgg cgc ctg gcc gca ctg gaa cgc
 E   I   Q   A   Q   A   I   L   D   M   Q   L   R   R   L   A   A   L   E   R
cag cgc atc atc gac gac ctg gcc aaa atc gag gcc gag atc gcc gat ctg gaa gac atc
 Q   R   I   I   D   D   L   A   K   I   E   A   E   I   A   D   L   E   D   I
ctg gca aaa ccc gag cgg cag cgt ggg atc gtg cgc gac gaa ctc gcc gaa atc gtg gac
 L   A   K   P   E   R   Q   R   G   I   V   R   D   E   L   A   E   I   V   D
agg cac ggc gac gac cgg cgt acc cgg atc atc gcg gcc gac gga gac gtc agc gac gag
 R   H   G   D   D   R   R   T   R   I   I   A   A   D   G   D   V   S   D   E
gat ttg atc gcc cgc gag gac gtc gtt gtc act atc acc gaa acg gga tac gcc aag cgc
 D   L   I   A   R   E   D   V   V   V   T   I   T   E   T   G   Y   A   K   R
acc aag acc gat ctg tat cgc agc cag aaa cgc ggc ggc aag ggc gtg cag ggt gcg ggg
 T   K   T   D   L   Y   R   S   Q   K   R   G   G   K   G   V   Q   G   A   G
ttg aag cag gac gac atc gtc gcg cac ttc ttc gtg tgc tcc acc cac gat ttg atc ctg
 L   K   Q   D   D   I   V   A   H   F   F   V   C   S   T   H   D   L   I   L
ttc ttc acc acc cag gga cgg gtt tat cgg gcc aag gcc tac gac ttg ccc gag gcc tcc
 F   F   T   T   Q   G   R   V   Y   R   A   K   A   Y   D   L   P   E   A   S
cgg acg gcg cgc ggg cag cac gtg gcc aac ctg tta gcc ttc cag ccc gag gaa cgc atc
 R   T   A   R   G   Q   H   V   A   N   L   L   A   F   Q   P   E   E   R   I
gcc cag gtc atc cag att cgc ggc tac acc gac gcc ccg tac ctg gtg ctg gcc act cgc
 A   Q   V   I   Q   I   R   G   Y   T   D   A   P   Y   L   V   L   A   T   R
aac ggg ctg gtg aaa aag tcc aag ctg acc gac tcc aat cgc tcg ggc gga atc
 N   G   L   V   K   K   S   K   L   T   D   F   D   S   N   R   S   G   G   I
gtg gcg gtc aac ctg cgc gac aac gac gag ctg gtc ggt gcg gtg ctg tgt tcg gcc ggc
 V   A   V   N   L   R   D   N   D   E   L   V   G   A   V   L   C   S   A   G
gac gac ctg ctg ctg gtc tcg gcc aac ggg cag tcc atc agg ttc tcg gcg acc gac gag
 D   D   L   L   L   V   S   A   N   G   Q   S   I   R   F   S   A   T   D   E
gcg ctg cgg cca atg ggt cgt gcc acc tcg ggt gtg cag ggc atg cgg ttc aat atc gac
 A   L   R   P   M   G   R   A   T   S   G   V   Q   G   M   R   F   N   I   D
gac cgg ctg ctg tcg ctg aac gtc gtg cgt gaa ggc acc tat ctg ctg gtg gcg acg tca
 D   R   L   L   S   L   N   V   V   R   E   G   T   Y   L   L   V   A   T   S
ggg ggc tat gcg aaa cgt acc gcg atc gag gaa tac ccg gta cag ggc cgc ggc ggt aaa
 G   G   Y   A   K   R   T   A   I   E   E   Y   P   V   Q   G   R   G   G   K
ggt gtg ctg acg gtc atg tac gac cgc cgg cgc ggc agg ttg gtt ggg gcg ttg att gtc
 G   V   L   T   V   M   Y   D   R   R   R   G   R   L   V   G   A   L   I   V
gac gac gac agc gag ctg tat gcc gtc act tcc ggc ggt ggc gtg atc cgc acc gcg gca
 D   D   D   S   E   L   Y   A   V   T   S   G   G   G   V   I   R   T   A   A
cgc cag gtt cgc aag gcg gga cgg cag acc aag ggt gtt cgg ttg atg aat ctg ggc gag
 R   Q   V   R   K   A   G   R   Q   T   K   G   V   R   L   M   N   L   G   E
ggc gac aca ctg ttg gcc atc gcg cgc aac gcc gaa gaa agt ggc gac gat aat gcc gtg
 G   D   T   L   L   A   I   A   R   N   A   E   E   S   G   D   D   N   A   V
gac gcc aac ggc gca gac cag acg ggc aat taa
 D   A   N   G   A   D   Q   T   G   N   *
```

Supplementary file S4 – pET-Duet1-6His-TEV-GyrBA^{56} (pTRB642) ORF

```
atg aaa tct tct cac cat cac cat cac cat gaa aac ctg tac ttc caa tcc aat gct gcc
 M   K   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   S   N   A   A
cag aaa aag aag gcc caa gac gaa tac ggc gct gcg tct atc acc att ctc gaa ggg ctg
 Q   K   K   K   A   Q   D   E   Y   G   A   A   S   I   T   I   L   E   G   L
gag gcc gtc cgc aaa cgt ccc ggc atg tac att ggc tcg acc ggt gag cgc ggt tta cac
 E   A   V   R   K   R   P   G   M   Y   I   G   S   T   G   E   R   G   L   H
cat ctc att tgg gag gtg gtc gac aac gcg gtc gac gag gcg atg gcc ggt tat gca acc
 H   L   I   W   E   V   V   D   N   A   V   D   E   A   M   A   G   Y   A   T
aca gtg aac gta gtg ctg ctt gag gat ggc ggt gtc gag gtc gcc gac gac ggc cgc ggc
 T   V   N   V   V   L   L   E   D   G   G   V   E   V   A   D   D   G   R   G
att ccg gtc gcc acc cac gcc tcc ggc ata ccg acc gtc gac gtg gtg atg aca caa cta
 I   P   V   A   T   H   A   S   G   I   P   T   V   D   V   V   M   T   Q   L
cat gcc ggc ggc aag ttc gac tcg gac gcg tat gcg ata tct ggt ggt ctg cac ggc gtc
 H   A   G   G   K   F   D   S   D   A   Y   A   I   S   G   G   L   H   G   V
ggc gtg tcg gtg gtt aac gcg cta tcc acc cgg ctc gaa gtc gag atc aag cgc gac ggg
 G   V   S   V   V   N   A   L   S   T   R   L   E   V   E   I   K   R   D   G
tac gag tgg tct cag gtt tat gag aag tcg gaa ccc ctg ggc ctc aag caa ggg gcg ccg
 Y   E   W   S   Q   V   Y   E   K   S   E   P   L   G   L   K   Q   G   A   P
acc aag aag acg ggg tca acg gtg cgg ttc tgg gcc gac ccc gct gtt ttc gaa acc acg
 T   K   K   T   G   S   T   V   R   F   W   A   D   P   A   V   F   E   T   T
gaa tac gac ttc gaa acc gtc gcc cgc cgg ctg caa gag atg gcg ttc ctc aac aag ggg
 E   Y   D   F   E   T   V   A   R   R   L   Q   E   M   A   F   L   N   K   G
ctg acc atc aac ctg acc gac gag agg gtg acc caa gac gag gtc gtc gac gaa gtg gtc
 L   T   I   N   L   T   D   E   R   V   T   Q   D   E   V   V   D   E   V   V
agc gac gtc gcc gag gcg ccg aag tcg gca agt gaa cgc gca gcc gaa tcc act gca ccg
 S   D   V   A   E   A   P   K   S   A   S   E   R   A   A   E   S   T   A   P
cac aaa gtt aag agc cgc acc ttt cac tat ccg ggt ggc ctg gtg gac ttc gtg aaa cac
 H   K   V   K   S   R   T   F   H   Y   P   G   G   L   V   D   F   V   K   H
atc aac cgc acc aag aac gcg att cat agc agc atc gtg gac ttt tcc ggc aag ggc acc
 I   N   R   T   K   N   A   I   H   S   S   I   V   D   F   S   G   K   G   T
ggg cac gag gtg gag atc gcg atg caa tgg aac gcc ggg tat tcg gag tcg gtg cac acc
 G   H   E   V   E   I   A   M   Q   W   N   A   G   Y   S   E   S   V   H   T
ttc gcc aac acc atc aac acc cac gag ggc ggc acc cac gaa gag ggc ttc cgc agc gcg
 F   A   N   T   I   N   T   H   E   G   G   T   H   E   E   G   F   R   S   A
ctg acg tcg gtg gtg aac aag tac gcc aag gac cgc aag cta ctg aag gac aag gac ccc
 L   T   S   V   V   N   K   Y   A   K   D   R   K   L   L   K   D   K   D   P
aac ctc acc ggt gac gat atc cgg gaa ggc ctg gcc gct gtg atc tcg gtg aag gtc agc
 N   L   T   G   D   D   I   R   E   G   L   A   A   V   I   S   V   K   V   S
gaa ccg cag ttc gag ggc cag acc aag acc aag ttg ggc aac acc gag gtc aaa tcg ttt
 E   P   Q   F   E   G   Q   T   K   T   K   L   G   N   T   E   V   K   S   F
gtg cag aag gtc tgt aac gaa cag ctg acc cac tgg ttt gaa gcc aac ccc acc gac gcg
 V   Q   K   V   C   N   E   Q   L   T   H   W   F   E   A   N   P   T   D   A
aaa gtc gtt gtg aac aag gct gtg tcc tcg gcg caa gcc cgt atc gcg gca cgt aag gca
 K   V   V   V   N   K   A   V   S   S   A   Q   A   R   I   A   A   R   K   A
cga gag ttg gtg cgg cgt aag agc gcc acc gac atc ggt gga ttg ccc ggc aag ctg gcc
 R   E   L   V   R   R   K   S   A   T   D   I   G   G   L   P   G   K   L   A
gat tgc cgt tcc acg gat ccg cgc aag tcc gaa ctg tat gtc gta gaa ggt gac tcg gcc
 D   C   R   S   T   D   P   R   K   S   E   L   Y   V   V   E   G   D   S   A
ggc ggt tct gca aaa agc ggt cgc gat tcg atg ttc cag gcg ata ctt ccg ctg cgc ggc
 G   G   S   A   K   S   G   R   D   S   M   F   Q   A   I   L   P   L   R   G
aag atc atc aat gtg gag aaa gcg cgc atc gac cgg gtg cta aag aac acc gaa gtt cag
 K   I   I   N   V   E   K   A   R   I   D   R   V   L   K   N   T   E   V   Q
gcg atc atc acg gcg ctg ggc acc ggg atc cac gac gag ttc gat atc ggc aag ctg cgc
 A   I   I   T   A   L   G   T   G   I   H   D   E   F   D   I   G   K   L   R
tac cac aag atc gtg ctg atg gcc gac gcc gat gtt gac ggc caa cat att tcc acg ctg
 Y   H   K   I   V   L   M   A   D   A   D   V   D   G   Q   H   I   S   T   L
ttg ttg acg ttg ttg ttc cgg ttc atg cgg ccg ctc atc gag aac ggg cat gtg ttt ttg
 L   L   T   L   L   F   R   F   M   R   P   L   I   E   N   G   H   V   F   L
gca caa ccg ccg ctg tac aaa ctc aag tgg cag cgc agt gac ccg gaa ttc gca tac tcc
 A   Q   P   P   L   Y   K   L   K   W   Q   R   S   D   P   E   F   A   Y   S
gac cgc gag cgc gac ggt ctg ctg gag gcg ggg ctg aag gcc ggg aag aag atc aac aag
 D   R   E   R   D   G   L   L   E   A   G   L   K   A   G   K   K   I   N   K
gaa gac ggc att cag cgg tac aag ggt cta ggt gaa atg gac gct aag gag ttg tgg gag
 E   D   G   I   Q   R   Y   K   G   L   G   E   M   D   A   K   E   L   W   E
acc acc atg gat ccc tcg gtt cgt gtg ttg cgt caa gtg acg ctg gac gac gcc gcc gcc
 T   T   M   D   P   S   V   R   V   L   R   Q   V   T   L   D   D   A   A   A
gcc gac gag ttg ttc tcc atc ctg atg ggc gag gac gtc gac gcg cgg cgc agc ttt atc
 A   D   E   L   F   S   I   L   M   G   E   D   V   D   A   R   R   S   F   I
acc cgc aac gcc aag gat gtt cgg ttc ctg gat gtc aca gac acg acg ttg ccg cct gac
 T   R   N   A   K   D   V   R   F   L   D   V   T   D   T   T   L   P   P   D
```

```
gac tcg ctc gac cgg atc gaa ccg gtt gac atc gag cag gag atg cag cgc agc tac atc
 D   S   L   D   R   I   E   P   V   D   I   E   Q   E   M   Q   R   S   Y   I
gac tat gcg atg agc gtg atc gtc ggc cgc gcg ctg ccg gag gtg cgc gac ggg ctc aag
 D   Y   A   M   S   V   I   V   G   R   A   L   P   E   V   R   D   G   L   K
ccc gtg cat cgc cgg gtg ctc tat gca atg ttc gat tcc ggc ttc cgc ccg gac cgc agc
 P   V   H   R   R   V   L   Y   A   M   F   D   S   G   F   R   P   D   R   S
cac gcc aag tcg gcc cgg tcg gtt gcc gag acc atg ggc aac tac cac ccg cac ggc gac
 H   A   K   S   A   R   S   V   A   E   T   M   G   N   Y   H   P   H   G   D
gcg tcg atc tac gac agc ctg gtg cgc atg gcc cag ccc tgg tcg ctg cgc tac ccg ctg
 A   S   I   Y   D   S   L   V   R   M   A   Q   P   W   S   L   R   Y   P   L
gtg gac ggc cag ggc aac ttc ggc tcg cca ggc aat gac cca ccg gcg gcg atg agg tac
 V   D   G   Q   G   N   F   G   S   P   G   N   D   P   P   A   A   M   R   Y
acc gaa gcc cgg ctg acc ccg ttg gcg atg gag atg ctg agg gaa atc gac gag gag aca
 T   E   A   R   L   T   P   L   A   M   E   M   L   R   E   I   D   E   E   T
gtc gat ttc atc cct aac tac gac ggc cgg gtg caa gag ccg acg gtg cta ccc agc cgg
 V   D   F   I   P   N   Y   D   G   R   V   Q   E   P   T   V   L   P   S   R
ttc ccc aac ctg ctg gcc aac ggg tca ggc ggc atc gcg gtc ggc atg gca acc aat atc
 F   P   N   L   L   A   N   G   S   G   G   I   A   V   G   M   A   T   N   I
ccg ccg cac aac ctg cgt gag ctg gcc gac gcg gtg ttc tgg gcg ctg gag aat cac gac
 P   P   H   N   L   R   E   L   A   D   A   V   F   W   A   L   E   N   H   D
gcc gac gaa gag gag acc ctg gcc gcg gtc atg ggg cgg gtt aaa ggc ccg gac ttc ccg
 A   D   E   E   E   T   L   A   A   V   M   G   R   V   K   G   P   D   F   P
acc gcc gga ctg atc gtc gga tcc cag ggc acc gct gat gcc tac aaa act ggc cgc ggc
 T   A   G   L   I   V   G   S   Q   G   T   A   D   A   Y   K   T   G   R   G
tcc att cga atg cgc gga gtt gtt gag gta gaa gag gat tcc cgc ggt cgt acc tcg ctg
 S   I   R   M   R   G   V   V   E   V   E   E   D   S   R   G   R   T   S   L
gtg atc acc gag ttg ccg tat cag gtc aac cac gac aac ttc atc act tcg atc gcc gaa
 V   I   T   E   L   P   Y   Q   V   N   H   D   N   F   I   T   S   I   A   E
cag gtc cga gac ggc aag ctg gcc ggc att tcc aac att gag gac cag tct agc gat cgg
 Q   V   R   D   G   K   L   A   G   I   S   N   I   E   D   Q   S   S   D   R
gtc ggt tta cgc atc gtc atc gag atc aag cgc gat gcg gtg gcc aag gtg gtg atc aat
 V   G   L   R   I   V   I   E   I   K   R   D   A   V   A   K   V   V   I   N
aac ctt tac aag cac acc cag ctg cag acc agc ttt ggc gcc aac atg cta gcg atc gtc
 N   L   Y   K   H   T   Q   L   Q   T   S   F   G   A   N   M   L   A   I   V
gac ggg gtg ccg cgc acg ctg cgg ctg gac cag ctg atc cgc tat tac gtt gac cac caa
 D   G   V   P   R   T   L   R   L   D   Q   L   I   R   Y   Y   V   D   H   Q
ctc gac gtc att gtg cgg cgc acc acc tac cgg ctg cgc aag gca aac gag cga gcc cac
 L   D   V   I   V   R   R   T   T   Y   R   L   R   K   A   N   E   R   A   H
att ctg cgc ggc ctg gtt aaa gcg ctc gac gcg ctg gac gag gtc att gca ctg atc cgg
 I   L   R   G   L   V   K   A   L   D   A   L   D   E   V   I   A   L   I   R
gcg tcg gag acc gtc gat atc gcc cgg gcc gga ctg atc gag ctg ctc gac atc gac gag
 A   S   E   T   V   D   I   A   R   A   G   L   I   E   L   L   D   I   D   E
atc cag gcc cag gca atc ctg gac atg cag ttg cgg cgc ctg gcc gca ctg gaa cgc cag
 I   Q   A   Q   A   I   L   D   M   Q   L   R   R   L   A   A   L   E   R   Q
cgc atc atc gac gac ctg gcc aaa atc gag gcc gag atc gcc gat ctg gaa gac atc ctg
 R   I   I   D   D   L   A   K   I   E   A   E   I   A   D   L   E   D   I   L
gca aaa ccc gag cgg cag cgt ggg atc gtg cgc gac gaa ctc gcc gaa atc gtg gac agg
 A   K   P   E   R   Q   R   G   I   V   R   D   E   L   A   E   I   V   D   R
cac ggc gac gac cgg cgt acc cgg atc atc gcg taa
 H   G   D   D   R   R   T   R   I   I   A   *
```

Supplementary file S5 – pET-Duet1-6His-TEV-GyrB^{28}A^{56} (pTRB643) ORF

```
atg aaa tct tct cac cat cac cat cac cat cac gaa aac ctg tac ttc caa tcc aat ttg gtg
 M   K   S   S   H   H   H   H   H   H   E   N   L   Y   F   Q   S   N   L   V
cgg cgt aag agc gcc acc gac atc ggt gga ttg ccc ggc aag ctg gcc gat tgc cgt tcc
 R   R   K   S   A   T   D   I   G   G   L   P   G   K   L   A   D   C   R   S
acg gat ccg cgc aag tcc gaa ctg tat gtc gta gaa ggt gac tcg gcc ggc ggt tct gca
 T   D   P   R   K   S   E   L   Y   V   V   E   G   D   S   A   G   G   S   A
aaa agc ggt cgc gat tcg atg ttc cag gcg ata ctt ccg ctg cgc ggc aag atc atc aat
 K   S   G   R   D   S   M   F   Q   A   I   L   P   L   R   G   K   I   I   N
gtg gag aaa gcg cgc atc gac cgg gtg cta aag aac acc gaa gtt cag gcg atc atc acg
 V   E   K   A   R   I   D   R   V   L   K   N   T   E   V   Q   A   I   I   T
gcg ctg ggc acc ggg atc cac gac gag ttc gat atc ggc aag ctg cgc tac cac aag atc
 A   L   G   T   G   I   H   D   E   F   D   I   G   K   L   R   Y   H   K   I
gtg ctg atg gcc gac gcc gat gtt gac ggc caa cat att tcc acg ctg ttg ttg acg ttg
 V   L   M   A   D   A   D   V   D   G   Q   H   I   S   T   L   L   L   T   L
ttg ttc cgg ttc atg cgg ccg ctc atc gag aac ggg cat gtg ttt ttg gca caa ccg ccg
 L   F   R   F   M   R   P   L   I   E   N   G   H   V   F   L   A   Q   P   P
ctg tac aaa ctc aag tgg cag cgc agt gac ccg gaa ttc gca tac tcc gac cgc gag cgc
 L   Y   K   L   K   W   Q   R   S   D   P   E   F   A   Y   S   D   R   E   R
gac ggt ctg ctg gag gcg ggg ctg aag gcc ggg aag aag atc aac aag gaa gac ggc att
 D   G   L   L   E   A   G   L   K   A   G   K   K   I   N   K   E   D   G   I
cag cgg tac aag ggt cta ggt gaa atg gac gct aag gag ttg tgg gag acc acc atg gat
 Q   R   Y   K   G   L   G   E   M   D   A   K   E   L   W   E   T   T   M   D
ccc tcg gtt cgt gtg ttg cgt caa gtg acg ctg gac gac gcc gcc gcc gcc gac gag ttg
 P   S   V   R   V   L   R   Q   V   T   L   D   D   A   A   A   A   D   E   L
ttc tcc atc ctg atg ggc gag gac gtc gac gcg cgg cgc agc ttt atc acc cgc aac gcc
 F   S   I   L   M   G   E   D   V   D   A   R   R   S   F   I   T   R   N   A
```

```
aag gat gtt cgg ttc ctg gat gtc aca gac acg acg ttg ccg cct gac gac tcg ctc gac
 K   D   V   R   F   L   D   V   T   D   T   T   L   P   P   D   D   S   L   D
cgg atc gaa ccg gtt gac atc gag cag gag atg cag cgc agc tac atc gac tat gcg atg
 R   I   E   P   V   D   I   E   Q   E   M   Q   R   S   Y   I   D   Y   A   M
agc gtg atc gtc ggc cgc gcg ctg ccg gag gtg cgc gac ggg ctc aag ccc gtg cat cgc
 S   V   I   V   G   R   A   L   P   E   V   R   D   G   L   K   P   V   H   R
cgg gtg ctc tat gca atg ttc gat tcc ggc ttc cgc ccg gac cgc agc cac gcc aag tcg
 R   V   L   Y   A   M   F   D   S   G   F   R   P   D   R   S   H   A   K   S
gcc cgg tcg gtt gcc gag acc atg ggc aac tac cac ccg cac ggc gac gcg tcg atc tac
 A   R   S   V   A   E   T   M   G   N   Y   H   P   H   G   D   A   S   I   Y
gac agc ctg gtg cgc atg gcc cag ccc tgg tcg ctg cgc tac ccg ctg gtg gac ggc cag
 D   S   L   V   R   M   A   Q   P   W   S   L   R   Y   P   L   V   D   G   Q
ggc aac ttc ggc tcg cca ggc aat gac cca ccg gcg gcg atg agg tac acc gaa gcc cgg
 G   N   F   G   S   P   G   N   D   P   P   A   A   M   R   Y   T   E   A   R
ctg acc ccg ttg gcg atg gag atg ctg agg gaa atc gac gag gag aca gtc gat ttc atc
 L   T   P   L   A   M   E   M   L   R   E   I   D   E   E   T   V   D   F   I
cct aac tac gac ggc cgg gtg caa gag ccg acg gtg cta ccc agc cgg ttc ccc aac ctg
 P   N   Y   D   G   R   V   Q   E   P   T   V   L   P   S   R   F   P   N   L
ctg gcc aac ggg tca ggc ggc atc gcg gtc ggc atg gca acc aat atc ccg ccg cac aac
 L   A   N   G   S   G   G   I   A   V   G   M   A   T   N   I   P   P   H   N
ctg cgt gag ctg gcc gac gcg gtg ttc tgg gcg ctg gag aat cac gac gcc gac gaa gag
 L   R   E   L   A   D   A   V   F   W   A   L   E   N   H   D   A   D   E   E
gag acc ctg gcc gcg gtc atg ggg cgg gtt aaa ggc ccg gac ttc ccg acc gcc gga ctg
 E   T   L   A   A   V   M   G   R   V   K   G   P   D   F   P   T   A   G   L
atc gtc gga tcc cag ggc acc gct gat gcc tac aaa act ggc cgc ggc tcc att cga atg
 I   V   G   S   Q   G   T   A   D   A   Y   K   T   G   R   G   S   I   R   M
cgc gga gtt gtt gag gta gaa gag gat tcc cgc ggt cgt acc tcg ctg gtg atc acc gag
 R   G   V   V   E   V   E   E   D   S   R   G   R   T   S   L   V   I   T   E
ttg ccg tat cag gtc aac cac gac aac ttc atc act tcg atc gcc gaa cag gtc cga gac
 L   P   Y   Q   V   N   H   D   N   F   I   T   S   I   A   E   Q   V   R   D
ggc aag ctg gcc ggc att tcc aac att gag gac cag tct agc gat cgg gtc ggt tta cgc
 G   K   L   A   G   I   S   N   I   E   D   Q   S   S   D   R   V   G   L   R
atc gtc atc gag atc aag cgc gat gcg gtg gcc aag gtg gtg atc aat aac ctt tac aag
 I   V   I   E   I   K   R   D   A   V   A   K   V   V   I   N   N   L   Y   K
cac acc cag ctg cag acc agc ttt ggc gcc aac atg cta gcg atc gtc gac ggg gtg ccg
 H   T   Q   L   Q   T   S   F   G   A   N   M   L   A   I   V   D   G   V   P
cgc acg ctg cgg ctg gac cag ctg atc cgc tat tac gtt gac cac caa ctc gac gtc att
 R   T   L   R   L   D   Q   L   I   R   Y   Y   V   D   H   Q   L   D   V   I
gtg cgg cgc acc acc tac cgg ctg cgc aag gca aac gag cga gcc cac att ctg cgc ggc
 V   R   R   T   T   Y   R   L   R   K   A   N   E   R   A   H   I   L   R   G
ctg gtt aaa gcg ctc gac gcg ctg gac gag gtc att gca ctg atc cgg gcg tcg gag acc
 L   V   K   A   L   D   A   L   D   E   V   I   A   L   I   R   A   S   E   T
gtc gat atc gcc cgg gcc gga ctg atc gag ctg ctc gac atc gac gag atc cag gcc cag
 V   D   I   A   R   A   G   L   I   E   L   L   D   I   D   E   I   Q   A   Q
gca atc ctg gac atg cag ttg cgg cgc ctg gcc gca ctg gaa cgc cag cgc atc atc gac
 A   I   L   D   M   Q   L   R   R   L   A   A   L   E   R   Q   R   I   I   D
gac ctg gcc aaa atc gag gcc gag atc gcc gat ctg gaa gac atc ctg gca aaa ccc gag
 D   L   A   K   I   E   A   E   I   A   D   L   E   D   I   L   A   K   P   E
cgg cag cgt ggg atc gtg cgc gac gaa ctc gcc gaa atc gtg gac agg cac ggc gac gac
 R   Q   R   G   I   V   R   D   E   L   A   E   I   V   D   R   H   G   D   D
cgg cgt acc cgg atc atc gcg taa
 R   R   T   R   I   I   A   *
```

Supplementary file S6 – pET-Duet1-6His-SUMO-ParD1-ParE1 (pTRB569) ORFs

6His-SUMO-ParE1 in frame (first ORF, bold text)

```
atg ggc agt ggc cat cat cat cat cat cat ccg tca gga gtc aag act gag
 M   G   S   G   H   H   H   H   H   H   P   S   G   V   K   T   E
aac aac gat cat att aat ttg aag gtg gcg ggg cag gat ggt tct gtg gtg cag ttt aag
 N   N   D   H   I   N   L   K   V   A   G   Q   D   G   S   V   V   Q   F   K
att aag agg cat aca cca ctt agt aaa cta atg aaa gcc tat tgt gaa cga cag gga ttg
 I   K   R   H   T   P   L   S   K   L   M   K   A   Y   C   E   R   Q   G   L
tca atg agg cag atc aga ttc cga ttt gac ggg caa cca atc aat gaa aca gac aca cct
 S   M   R   Q   I   R   F   R   F   D   G   Q   P   I   N   E   T   D   T   P
gca cag ttg gaa atg gag gat gaa gat aca att gat gtg ttt caa cag cag acg gga ggt
 A   Q   L   E   M   E   D   E   D   T   I   D   V   F   Q   Q   Q   T   G   G
agt agc cga tac ctt ctc tcg cct gcc gcg cag gca cat ctg gaa gag atc tgg gac tgc
 S   S   R   Y   L   L   S   P   A   A   Q   A   H   L   E   E   I   W   D   C
acc tat gac cgt tgg ggt gtc gat cag gcc gag cag tac ctg cgc gaa ctt caa cac gct
 T   Y   D   R   W   G   V   D   Q   A   E   Q   Y   L   R   E   L   Q   H   A
atc gac cgt gct gcg gca aac ccg cgg atc gga cga gcg tgc gac gag att cgc ccc ggc
 I   D   R   A   A   A   N   P   R   I   G   R   A   C   D   E   I   R   P   G
tat cgc aag ctc tcg gcc ggg tca cac acg ttg ttc tat cgg gtg act ggc gaa ggc acc
 Y   R   K   L   S   A   G   S   H   T   L   F   Y   R   V   T   G   E   G   T
atc gac gtc gtg cga gtc ctg cac caa cgg atg gac gtc gac cgg aac ctc tga aag ctt
 I   D   V   V   R   V   L   H   Q   R   M   D   V   D   R   N   L   *   K   L
gcg gcc gca taa tgc tta agt cga aca gaa agt aat cgt att gta cac ggc cgc ata atc
 A   A   A   -   C   L   S   R   T   E   S   N   R   I   V   H   G   R   I   I
gaa att aat acg act cac tat agg gga att gtg agc gga taa caa ttc ccc atc tta gta
 E   I   N   T   T   H   Y   R   G   I   V   S   G   -   Q   F   P   I   L   V
tat tag tta agt ata aga agg aga tat aca tat ggg taa gaa cac gtc ctt cgt cct cga
 Y   -   L   S   I   R   R   R   Y   T   Y   G   -   E   H   V   L   R   P   R
cga gca cta cag cgc ctt cat cga cgg cga gat cgc cgc ggg ccg cta ccg gtc ggc cag
 R   A   L   Q   R   L   H   R   R   R   D   R   R   G   P   L   P   V   G   Q
tga agt cat ccg ctc cgc gtt gcg act gct cga gga ccg tga aac cca gct gcg cgc gct
 -   S   H   P   L   R   V   A   T   A   R   G   P   -   N   P   A   A   R   A
tcg tga ggc tct cga ggc cgg cga acg cag cgg cag ctc gac acc gtt cga ctt cga cgg
 S   -   G   S   R   G   R   R   T   Q   R   Q   L   D   T   V   R   L   R   R
att cct cgg tcg taa gcg ggc tga cgc ctc gcg tgg ccg gtg agg
 I   P   R   S   -   A   G   -   R   L   A   W   P   V   R
```

ParD1 in frame (second ORF, bold text)

```
cca tgg gca gtg gcc atc atc atc atc atc atc cgt cag gag tca aga ctg aga
 D   I   P   W   A   V   A   I   I   I   I   I   I   R   Q   E   S   R   L   R
aca acg atc ata tta att tga agg tgg cgg ggc agg atg gtt ctg tgg tgc agt tta aga
 T   T   I   I   L   I   -   R   W   R   G   R   M   V   L   W   C   S   L   R
tta aga ggc ata cac cac tta gta aac taa tga aag cct att gtg aac gac agg gat tgt
 L   R   G   I   H   H   L   V   N   -   -   K   P   I   V   N   D   R   D   C
caa tga ggc aga tca gat tcc gat ttg acg ggc aac caa tca atg aaa cag aca cac ctg
 Q   -   G   R   S   D   S   D   L   T   G   N   Q   S   M   K   Q   T   H   L
cac agt tgg aaa tgg agg atg aag ata caa ttg atg tgt ttc aac agc aga cgg gag gta
 H   S   W   K   W   R   M   K   I   Q   L   M   C   F   N   S   R   R   E   V
gta gcc gat acc ttc tct cgc ctg ccg cgc agg cac atc tgg aag aga tct ggg act gca
 V   A   D   T   F   S   R   L   P   R   R   H   I   W   K   R   S   G   T   A
cct atg acc gtt ggg gtg tcg atc agg ccg agc agt acc tgc gcg aac ttc aac acg cta
 P   M   T   V   G   V   S   I   R   P   S   S   T   C   A   N   F   N   T   L
tcg acc gtg ctg cgg caa acc cgc gga tcg gac gag cgt gcg acg aga ttc gcc ccg gct
 S   T   V   L   R   Q   T   R   G   S   D   E   R   A   T   R   F   A   P   A
atc gca agc tct cgg ccg ggt cac aca cgt tgt tct atc ggg tga ctg gcg aag gca cca
 I   A   S   S   R   P   G   H   T   R   C   S   I   G   -   L   A   K   A   P
tcg acg tcg tgc gag tcc tgc acc aac gga tgg acg tcg acc gga acc tct gaa agc ttg
 S   T   S   C   E   S   C   T   N   G   W   T   S   T   G   T   S   E   S   L
cgg ccg cat aat gct taa gtc gaa cag aaa gta atc gta ttg tac acg gcc gca taa tcg
```

```
 R   P   H   N   A   -   V   E   Q   K   V   I   V   L   Y   T   A   A   -   S
aaa tta ata cga ctc act ata ggg gaa ttg tga gcg gat aac aat tcc cca tct tag tat
 K   L   I   R   L   T   I   G   E   L   -   A   D   N   N   S   P   S   -   Y
att agt taa gta taa gaa gga gat ata cat atg ggt aag aac acg tcc ttc gtc ctc gac
 I   S   -   V   -   E   G   D   I   H   M   G   K   N   T   S   F   V   L   D
gag cac tac agc gcc ttc atc gac ggc gag atc gcc gcg ggc cgc tac cgg tcg gcc agt
 E   H   Y   S   A   F   I   D   G   E   I   A   A   G   R   Y   R   S   A   S
gaa gtc atc cgc tcc gcg ttg cga ctg ctc gag gac cgt gaa acc cag ctg cgc gcg ctt
 E   V   I   R   S   A   L   R   L   L   E   D   R   E   T   Q   L   R   A   L
cgt gag gct ctc gag gcc ggc gaa cgc agc ggc agc tcg aca ccg ttc gac ttc gac gga
 R   E   A   L   E   A   G   E   R   S   G   S   S   T   P   F   D   F   D   G
ttc ctc ggt cgt aag cgg gct gac gcc tcg cgt ggc cgg tga
 F   L   G   R   K   R   A   D   A   S   R   G   R   -
```

Supplementary file S7 – pET-Duet1-6His-SUMO-ParD2-ParE2 (pTRB570) ORFs

6His-SUMO-ParE2 in frame (first ORF, bold text)

```
atg ggc agt ggc cat cat cat cat cat cat cat ccg tca gga gtc aag act gag
 M   G   S   G   H   H   H   H   H   H   P   S   G   V   K   T   E
aac aac gat cat att aat ttg aag gtg gcg ggg cag gat ggt tct gtg gtg cag ttt aag
 N   N   D   H   I   N   L   K   V   A   G   Q   D   G   S   V   V   Q   F   K
att aag agg cat aca cca ctt agt aaa cta atg aaa gcc tat tgt gaa cga cag gga ttg
 I   K   R   H   T   P   L   S   K   L   M   K   A   Y   C   E   R   Q   G   L
tca atg agg cag atc aga ttc cga ttt gac ggg caa cca atc aat gaa aca gac aca cct
 S   M   R   Q   I   R   F   R   F   D   G   Q   P   I   N   E   T   D   T   P
gca cag ttg gaa atg gag gat gaa gat aca att gat gtg ttt caa cag cag acg gga ggt
 A   Q   L   E   M   E   D   E   D   T   I   D   V   F   Q   Q   Q   T   G   G
acg cgc agg ctg cgc gtc cat aac ggg gtt gaa gac gac cta ttc gag gcg ttt tcc tac
 T   R   R   L   R   V   H   N   G   V   E   D   D   L   F   E   A   F   S   Y
tac gcg gac gcg gcg cca gat cag atc gat cgg ctt tac aac ttg ttt gtc gat gcc gtg
 Y   A   D   A   A   P   D   Q   I   D   R   L   Y   N   L   F   V   D   A   V
acg aag cgg att ccg cag gct ccg aac gcg ttt gcg ccg tta ttc aag cac tat cga cac
 T   K   R   I   P   Q   A   P   N   A   F   A   P   L   F   K   H   Y   R   H
atc tac ctc cgg ccc ttc agg tac tac gtt gcc tat cgg acg acc gac gag gct atc gac
 I   Y   L   R   P   F   R   Y   Y   V   A   Y   R   T   T   D   E   A   I   D
ata ctg gct gtt cgg cac gga atg gag aac ccg aac gcg gtc gag gct gag atc tct ggc
 I   L   A   V   R   H   G   M   E   N   P   N   A   V   E   A   E   I   S   G
cgc acc ttc gag tga aag ctt gcg gcc gca taa tgc tta agt cga aca gaa agt aat cgt
 R   T   F   E   -   K   L   A   A   A   -   C   L   S   R   T   E   S   N   R
att gta cac ggc cgc ata atc gaa att aat acg act cac tat agg gga att gtg agc gga
 I   V   H   G   R   I   I   E   I   N   T   T   H   Y   R   G   I   V   S   G
taa caa ttc ccc atc tta gta tat tag tta agt ata aga agg aga tat aca tat ggt ggt
 -   Q   F   P   I   L   V   Y   -   L   S   I   R   R   R   Y   T   Y   G   G
caa ccg ggc att gct ggc gag cgt cga cgc act gtc gcg tga tga gca gat tga gct cgt
 Q   P   G   I   A   G   E   R   R   R   T   V   A   -   -   A   D   -   A   R
cga gca cat caa cgg aaa cct agc cga ggg cat gca tat cag cga ggc caa cca ggc gct
 R   A   H   Q   R   K   P   S   R   G   H   A   Y   Q   R   G   Q   P   G   A
cat cga agc gcg ggc caa tga cac cga cga tgc tca ttg gtc cac cat tga tga ctt cga
 H   R   S   A   G   Q   -   H   R   R   C   S   L   V   H   H   -   -   L   R
caa gcg gat ccg cgc ccg gct cgg atg agg tac cct cga gtc tgg taa aga aac cgc tgc
 Q   A   D   P   R   P   A   R   M   R   Y   P   R   V   W   -   R   N   R   C
tgc gaa att tga acg cca gca cat gga ctc gtc tac tag cgc agc tta att
 C   E   I   -   T   P   A   H   G   L   V   Y   -   R   S   L   I
```

ParD2 in frame (second ORF, bold text)

```
cca tgg gca gtg gcc atc atc atc atc atc atc cgt cag gag tca aga ctg aga
 P   W   A   V   A   I   I   I   I   I   I   R   Q   E   S   R   L   R
aca acg atc ata tta att tga agg tgg cgg ggc agg atg gtt ctg tgg tgc agt tta aga
 T   T   I   I   L   I   -   R   W   R   G   R   M   V   L   W   C   S   L   R
tta aga ggc ata cac cac tta gta aac taa tga aag cct att gtg aac gac agg gat tgt
 L   R   G   I   H   H   L   V   N   -   -   K   P   I   V   N   D   R   D   C
caa tga ggc aga tca gat tcc gat ttg acg ggc aac caa tca atg aaa cag aca cac ctg
 Q   -   G   R   S   D   S   D   L   T   G   N   Q   S   M   K   Q   T   H   L
cac agt tgg aaa tgg agg atg aag ata caa ttg atg tgt ttc aac agc aga cgg gag gta
 H   S   W   K   W   R   M   K   I   Q   L   M   C   F   N   S   R   R   E   V
cgc gca ggc tgc gcg tcc ata acg ggg ttg aag acg acc tat tcg agg cgt ttt cct act
 R   A   G   C   A   S   I   T   G   L   K   T   T   Y   S   R   R   F   P   T
acg cgg acg cgg cgc cag atc aga tcg atc ggc ttt aca act tgt ttg tcg atg ccg tga
 T   R   T   R   R   Q   I   R   S   I   G   F   T   T   C   L   S   M   P   -
cga agc gga ttc cgc agg ctc cga acg cgt ttg cgc cgt tat tca agc act atc gac aca
 R   S   G   F   R   R   L   R   T   R   L   R   R   Y   S   S   T   I   D   T
tct acc tcc ggc cct tca ggt act acg ttg cct atc gga cga ccg acg agg cta tcg aca
```

```
 S   T   S   G   P   S   G   T   T   L   P   I   G   R   P   T   R   L   S   T
tac tgg ctg ttc ggc acg gaa tgg aga acc cga acg cgg tcg agg ctg aga tct ctg gcc
 Y   W   L   F   G   T   E   W   R   T   R   T   R   S   R   L   R   S   L   A
gca cct tcg agt gaa agc ttg cgg ccg cat aat gct taa gtc gaa cag aaa gta atc gta
 A   P   S   S   E   S   L   R   P   H   N   A   -   V   E   Q   K   V   I   V
ttg tac acg gcc gca taa tcg aaa tta ata cga ctc act ata ggg gaa ttg tga gcg gat
 L   Y   T   A   A   -   S   K   L   I   R   L   T   I   G   E   L   -   A   D
aac aat tcc cca tct tag tat att agt taa gta taa gaa gga gat ata cat atg gtg gtc
 N   N   S   P   S   -   Y   I   S   -   V   -   E   G   D   I   H   M   V   V
aac cgg gca ttg ctg gcg agc gtc gac gca ctg tcg cgt gat gag cag att gag ctc gtc
 N   R   A   L   L   A   S   V   D   A   L   S   R   D   E   Q   I   E   L   V
gag cac atc aac gga aac cta gcc gag ggc atg cat atc agc gag gcc aac cag gcg ctc
 E   H   I   N   G   N   L   A   E   G   M   H   I   S   E   A   N   Q   A   L
atc gaa gcg cgg gcc aat gac acc gac gat gct cat tgg tcc acc att gat gac ttc gac
 I   E   A   R   A   N   D   T   D   D   A   H   W   S   T   I   D   D   F   D
aag cgg atc cgc gcc cgg ctc gga tga ggt acc ctc gag tct ggt aaa gaa acc gct gct
 K   R   I   R   A   R   L   G   -   G   T   L   E   S   G   K   E   T   A   A
gcg aaa ttt gaa cgc cag cac atg gac tcg tct act agc gca gct taa
 A   K   F   E   R   Q   H   M   D   S   S   T   S   A   A   -
```

Supplementary file S8 – pET-Duet1-6His-SUMO-RelE1-RelB1 (pTRB638) ORFs

6His-SUMO-RelE1 in frame (first ORF, bold text)

```
atg agt ggc cat cat cat cat cat cat ccg tca gga gtc aag act gag aac aac gat cat
 M   S   G   H   H   H   H   H   H   P   S   G   V   K   T   E   N   N   D   H
att aat ttg aag gtg gcg ggg cag gat ggt tct gtg gtg cag ttt aag att aag agg cat
 I   N   L   K   V   A   G   Q   D   G   S   V   V   Q   F   K   I   K   R   H
aca cca ctt agt aaa cta atg aaa gcc tat tgt gaa cga cag gga ttg tca atg agg cag
 T   P   L   S   K   L   M   K   A   Y   C   E   R   Q   G   L   S   M   R   Q
atc aga ttc cga ttt gac ggg caa cca atc aat gaa aca gac aca cct gca cag ttg gaa
 I   R   F   R   F   D   G   Q   P   I   N   E   T   D   T   P   A   Q   L   E
atg gag gat gaa gat aca att gat gtg ttt caa cag cag acg gga ggc agc gac gac cat
 M   E   D   E   D   T   I   D   V   F   Q   Q   Q   T   G   G   S   D   D   H
ccc tac cac gtg gcg atc acc gcg aca gcg gca cgc gac ctg caa cgc tta ccc gaa aag
 P   Y   H   V   A   I   T   A   T   A   A   R   D   L   Q   R   L   P   E   K
atc gcc gcc gca tgt gtc gag ttt gtt ttc gga ccg ctg ctt aac aac ccg cat agg ttg
 I   A   A   A   C   V   E   F   V   F   G   P   L   L   N   N   P   H   R   L
ggc aag ccg ctg cgc aat gac ctt gaa ggc ctc cac tca gcc cgc cgc ggt gat tac cgc
 G   K   P   L   R   N   D   L   E   G   L   H   S   A   R   R   G   D   Y   R
gtc gtc tac gcc atc gac gac ggc cac cac cga gtc gag atc atc cac atc gct cgt cgc
 V   V   Y   A   I   D   D   G   H   H   R   V   E   I   I   H   I   A   R   R
agt gcc agc tac cga atg aac ccg tgc cgg cca cgt taa aag ctt gcg gcc gca taa tgc
 S   A   S   Y   R   M   N   P   C   R   P   R   -   K   L   A   A   A   -   C
tta agt cga aca gaa agt aat cgt att gta cac ggc cgc ata atc gaa att aat acg act
 L   S   R   T   E   S   N   R   I   V   H   G   R   I   I   E   I   N   T   T
cac tat agg gga att gtg agc gga taa caa ttc ccc atc tta gta tat tag tta agt ata
 H   Y   R   G   I   V   S   G   -   Q   F   P   I   L   V   Y   -   L   S   I
aga agg aga tat aca tat ggc tgt tgt ccc act ggg cga agt ccg caa tcg cct ctc tga
 R   R   R   Y   T   Y   G   C   C   P   T   G   R   S   P   Q   S   P   L   -
gta cgt cgc cga agt tga gct gac aca cga gcg cat cac gat aac ccg gca cgg tca tcc
 V   R   R   R   S   -   A   D   T   R   A   H   H   D   N   P   A   R   S   S
ggc ggc ggt att gat ctc ggc cga tga cct ggc gtc cat cga gga aac gct gga ggt gct
 G   G   G   I   D   L   G   R   -   P   G   V   H   R   G   N   A   G   G   A
acg cac ccc tgg cgc cag cga ggc cat tcg tga agg cct cgc cga tgt tgc cgc agg gcg
 T   H   P   W   R   Q   R   G   H   S   -   R   P   R   R   C   C   R   R   A
ctt cgt gag caa cga cga gat ccg caa ccg tta cac cgc gcg gtg agg
 L   R   E   Q   R   R   D   P   Q   P   L   H   R   A   V   R
```

RelB1 in frame (second ORF, bold text)

```
atga gtg gcc atc atc atc atc atc atc cgt cag gag tca aga ctg aga aca acg atc ata
  -   V   A   I   I   I   I   I   I   R   Q   E   S   R   L   R   T   T   I   I
tta att tga agg tgg cgg ggc agg atg gtt ctg tgg tgc agt tta aga tta aga ggc ata
 L   I   -   R   W   R   G   R   M   V   L   W   C   S   L   R   L   R   G   I
cac cac tta gta aac taa tga aag cct att gtg aac gac agg gat tgt caa tga ggc aga
 H   H   L   V   N   -   -   K   P   I   V   N   D   R   D   C   Q   -   G   R
tca gat tcc gat ttg acg ggc aac caa tca atg aaa cag aca cac ctg cac agt tgg aaa
 S   D   S   D   L   T   G   N   Q   S   M   K   Q   T   H   L   H   S   W   K
tgg agg atg aag ata caa ttg atg tgt ttc aac agc aga cgg gag gca gcg acg acc atc
 W   R   M   K   I   Q   L   M   C   F   N   S   R   R   E   A   A   T   T   I
cct acc acg tgg cga tca ccg cga cag cgg cac gcg acc tgc aac gct tac ccg aaa aga
 P   T   T   W   R   S   P   R   Q   R   H   A   T   C   N   A   Y   P   K   R
tcg ccg ccg cat gtg tcg agt ttg ttt tcg gac cgc tgc tta aca acc cgc ata ggt tgg
 S   P   P   H   V   S   S   L   F   S   D   R   C   L   T   T   R   I   G   W
gca agc cgc tgc gca atg acc ttg aag gcc tcc act cag ccc gcc gcg gtg att acc gcg
 A   S   R   C   A   M   T   L   K   A   S   T   Q   P   A   A   V   I   T   A
tcg tct acg cca tcg acg acg gcc acc acc gag tcg aga tca tcc aca tcg ctc gtc gca
```

```
 S   S   T   P   S   T   T   A   T   T   E   S   R   S   S   T   S   L   V   A
gtg cca gct acc gaa tga acc cgt gcc ggc cac gtt aaa agc ttg cgg ccg cat aat gct
 V   P   A   T   E   -   T   R   A   G   H   V   K   S   L   R   P   H   N   A
taa gtc gaa cag aaa gta atc gta ttg tac acg gcc gca taa tcg aaa tta ata cga ctc
 -   V   E   Q   K   V   I   V   L   Y   T   A   A   -   S   K   L   I   R   L
act ata ggg gaa ttg tga gcg gat aac aat tcc cca tct tag tat att agt taa gta taa
 T   I   G   E   L   -   A   D   N   N   S   P   S   -   Y   I   S   -   V   -
gaa gga gat ata cat atg gct gtt gtc cca ctg ggc gaa gtc cgc aat cgc ctc tct gag
 E   G   D   I   H   M   A   V   V   P   L   G   E   V   R   N   R   L   S   E
tac gtc gcc gaa gtt gag ctg aca cac gag cgc atc acg ata acc cgg cac ggt cat ccg
 Y   V   A   E   V   E   L   T   H   E   R   I   T   I   T   R   H   G   H   P
gcg gcg gta ttg atc tcg gcc gat gac ctg gcg tcc atc gag gaa acg ctg gag gtg cta
 A   A   V   L   I   S   A   D   D   L   A   S   I   E   E   T   L   E   V   L
cgc acc cct ggc gcc agc gag gcc att cgt gaa ggc ctc gcc gat gtt gcc gca ggg cgc
 R   T   P   G   A   S   E   A   I   R   E   G   L   A   D   V   A   A   G   R
ttc gtg agc aac gac gag atc cgc aac cgt tac acc gcg cgg tga
 F   V   S   N   D   E   I   R   N   R   Y   T   A   R   -
```

Supplementary file S9 - RelE protein sequences

```
>JW1555
MAYFLDFDERALKEWRKLGSTVREQLKKKLVEVLESPRIEANKLRGMPDCYKIKLRSSGYRLVYQVIDEKVVVFVISVGKRERSEVYSEAVKRIL
>TTHC013
MGYRIEFDPRAEKELEKLDREVARRILRFLRERVATLEDPRSLGEPLRGPELGRFWKYRVGDYRLICHIQDREATVLVLRVGHARDVYR
>TBFG_11272
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>MRA_1255
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>BCG_1306c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Mb1278c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Rv1246c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>TBFG_12882
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>MRA_2891
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>BCG_2888
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Mb2891
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Rv2866
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Rv3358
VRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>Mb3393
MRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>Lxx22677
MSWDVQFAPAAIRGLDRLPPRVVAAVVEFVTVTLPGNPYRMSKPLQGDLEGYYSARRGDYRVLFSLDEDRRVLLVGRIAHRADVYRPR
>Francci3_2679
MTGAAGPYRLEITGPAARALAGRIPEKVATAVHEFITTTLLENPHRLGKRLLYPPYAGTWSARRGMYRVLYEIDEENRIVLVTAVEHRADAYRGR
>Noca_2680
MSAPDEGGTGYEVVFTRGARRALEWDLPAAVAMAAFEFIRGPLREAPRRVGKPLLEPLTPLWSARRGEYRILYRILDRRLVIAVVTIAHRRDAYGRRE
>TDE0735
MKVVLTETFKKQLKKLDATISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGDYRILCRICDDKLIITVIEIGHRSTVYR
>Suden_1966
MSYKLLIDDKVIKDLKQIDKLWQKKIIEVIKTKLVENPHLGKPLVGNLSPYYRLRVFDYRVIYEINDDEVVVIVIKIGHRKDIYK
>TDE1978
MKVILTETFKKQLKKLDAAISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGNYRILCRIHDDRLIITVIEIGHRSTVYR
>Paes_2094
MVWTIEFAATAEKELSKLDKSAAKRILKFLKERVATDPRSSGKALRGDHAGLWRYRIGDYRVICEFRDQTVSVLVVRIGHRKEVYR
>Mhun_1873
MSKFTLLISKGAERDLAQLPKFARRSLEVAFAELELLENPREKLGPLKGRAKGLYSLRTGEYRAILEIFDNKLMLLVIEAGHRKTIYRKYQS
>WD0126
MKYDIVYSKNFTERDFLNLPKTIRSRITKAINERFTTDPIKFGKPLRGRLKGYRRIRVSDYRIIYTVNIAKHKVFVATAGHRDTIYEKAT
>FN0211
MKYDVEYSKTAMNTIKKMDSSTSKLIRTWIEKNLINTENPRIKGKALTGDLKGLWRYRIGDYRILAEIQDDKIVILILDIGHRSKIYL-p
>WD0122
MGLERYKVKSLKSVVEKDLPNLPKIIRLRVQKAIKERLTVDPINFGEPLHHNLKGCRRLRVGDYRVIYRVNQLDHIVTITEIGHRDDIYKK
>WD0600
MGIIYQIGYLEGVDTEDLPSLPKTIRLRVQKAIEKRLTIIPDKVGEALSHKWVGYFRLRVGDYRVIYLIDNSEHMVKIAAIGHRKEIYKRSPE
>Cpha266_1376
MVWKIEFASSAEKELARLDKSAARRIVKYLRERVAIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEVYR
>RF_0898
MFKVIWENKARAELAELPYPLKILDKVESYLAQNPIVLGKPLKGEYKNLYRYRFGNYRIIYSVSIEKSTVTVIKIGHRANIY
>asl4561
MIYQIEITTRAAKQLKKLSEDIKLKIEEKIQELSNNPRSNDVVKLEGEEDTYRIRVGNYRILYEIKDDLLIVKVVKISHRRDVYRRK
>Clim_1338
MVWKIEFAASAEKELARRDKSAARRIVKYLRERVEIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEIYR
>Gura_2099
MTYRIELTKTAERDLLAVPKPVLKRLDACILGLADDPLPPGVKKLKNSDGLYRVRVSDYRIIYRIEQEILTVLVVKIGHRREVYR
>Gmet_A3569
MVWKVEIDPAARRELKKLDPQISGRVLKFLFERVARLDDPRSIGEALKGSRFGDFWKYRVGDYRIITSIEDEALVILVVRVGNRREVYER
>Rru_A3196
MTWKIEFDPSALRELDKLDPQIAARVLRFLRDRVAVLENPRSLGEALKGPRLGAFWKYRVSDYRIIAHIEDDTLRILVLRIGNRREVYR
>WD0404
MKTSGNKTYTIKFLKNVIEKEIPALPAKIKLMVQEAIKKRLTVDPFNLGKPLCHSFRGQYRIRVSNYRIIYIINHSERKVLITAVGYRKNIYKHRRLHN
```

>Plav_1617
MAWTIDYTDDAINQLRKLDKQSARRIVEYMDERIAGAANPRSFGKPLSGPLGQFWRYRVGDFRIICDIEDEVLRVLVVRIGDRKDVYRKAKR
>SCATT_39270
MSEYRTVFRPEAQAELRKVPRDMALRILAKLTELESDPLGFNTTALVSQPDRRRLRVGDYRVIYTIDNGELVVWVVHVGHRSTVYGA
>XfasM23_1261
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>PD1184
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>WD0269
MKTSGNKRYTIKYLKHVLKRNLPSLPEAIKPKIKDAIREYLATDPIGNGVLLRNRLKGHRRIRVDDYRVVYRVNTAERKVTIVSIGHRDNIYKQAILDLLKH
>FN0497
MKKYEVKFSEAAIKELKKLDKPTATMIKLWVIQNLENTINPRQHGKSLTANYSGKWRYRVGNYRLLAEIYDDEILILIFKVAHRSIVYKK
>RF_1272
MEMNYKICYLEEVTNKHIPMLSSNAKTLIKCAIEERLMFDPIAFGKPLRYSLKGHRRFRISDYRIIYRIEQETSTVIIIAIKHRKEIYQEFI
>Pcar_1777
MKMSGGRVAFSVIYHPDVKGRDIPKINGDVRVRIKKAIETRLMVAPQEYGEPLRKTLKGYWKLRVGDYRIVFKIDGDEILILGICHRKGVYPLMESRQ
>LPC_2245
MMTPGSKLYQIEYIEDVVKNDMPSLSTSAKKLIKKAIEERLMADPIGFGKPLRYSLKGHRRLRVSDYRIVYRIEAETNTVVIIAIKHRKEVYDDF
>MGAS10270_Spy1927
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>M28_Spy1840
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>SSU98_0616
MAYKLVLSDDALKQLKKMDRHVGMMLAKDLKKRLDGLENPRQFGKALVGDYKGLWRYRVGNYRVICDIIDNKMVILALEIGHRKEIYKK
>Xfasm12_1292
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCDIQDSVLRVLVVRVGHRGEIYR

>pTC-F14_p08
MAWRIEFDDKAKKDLAALDKSVAKRITAFLRERVAHLDDPRSIGEALKGSKLGDFWKYRVGDWRIIASIEDEALRILVVRIGNRREVYRK
>Sbal_4390
MAWTIDYSERALKSLKKMDKQNAKRILDFLEQRIAILDDPRTSGKPLKGDLGIFWRYRVGDYRVLCEIQDSKLVILTALIGHRKEIYE
>Npun_R1755
MPSVEMSERYSLRIAKTAEKDLLDLQAKLYKQVVSKILSLQGNSRPQDCKALKGYEGGYRVDQGEYRILYTIDEESKLIDVFRVGKRNDGEVYKNL
>RSal33209_0100
MTYRIAYTPRVIKDLKKLDKQTVRRVKDFFDRLNRDNPRSLGKALVGQDFWRYRIGDHRILVAIQDDVLTVLVVKVAHRREVYKER
>MA1694
MTYQVVLSPDFEKETKIFFKKDPVLYGRFKKTVNSILENPECGKPLRNVLKGLRRVHIGHFVLIYEIDNTNETITFLKFSPHDKAYK
>Noc_0433
MYAIVVHRRAARYLRKLPQDQQVKIKHVLAQMKNGPLGLSGIKSMVGDWAGYRRIRVGNVRIIFWIDELKNVVYVDHIGPRGDVYKDKT
>RSc3279
MNAIHWTAWAARQLRKLDRQHQRVLVEAVGQLEAMPHCRQVRALREHRYGYRLRVGDYRVLSDWDDGIRIVDIQEVSKRDERTYRH
>Ppha_0985
MYKIIFTKEAQKALLRLPGSTAVQVCQKLEQLAADSYAPIANAKKLQNRSGYRLRIGDWRVIYEIQNDKLVVLVLKIAQRGEVYR
>asl2100
MSDRYTLRIARTAEKDLLDLQPKQFKQVVSKILSLQGTPRPQDCKALKGYEGGYRVDQGEYRILYTIDDETQLVDVFRVGKRNDDEVYQNL
>BH07080
MAWTIRYEKKALSFLKKCDKKEARRIVDFLDQHVAPLEDVRVIGKPLKGQFSGLWRYRVGDYRILCELYDKELVVLVLAVGHRKNIYKG
>TK0965
MSYELILSGKSEKALKKAPPEDRKRIVSALFKLKENPWAMQYKKLRGYPFYRVRVGDWRIIYTVDDEARIVYVVRLGKREGVYDSL
>SSU05_0886
MKTCYKLVPTSRFIKQLKKLDKFTQKQITNYLSSHVTDNPRQYGKALTANRSGQWRYRIGNYRVIVNIEDDKLIVTSIKMRNGSI
>Ppha_1272
MAYSVGYKKSVQCDLRQLSRLDAQRIYDQIEQELVKNPKSNPLLKGRFVGLRKYRVGDYRVIYCVLDEEVIILRIAHRRDVYKRDI
>Mhun_0373
MEPYGISMAPSALKNYKKFPEKLREKIKSEALHIARNPYIYEELSEPLKGIRSYHFTFNSTQYRIAYQINESSREIEILLVKTRENFYDKLFRTFR
>ECA0674
MAHIVWTGKAVKDLRKLPANDQKAIQANVNSLGDYPATKSKPLDITKLTDRGSEYRLRVGNYRVLFEIQKGEPVIIEIQRVLRRTSTTY
>HY04AAS1_1546
MWKVVITKQAKKDLVNIYRAGLKSKFEKLVEEIKKDPYTSQCEKLIGDLEGYYSCRINRKHRLVYEIDDEQKIVKIVSVWNHY
>AM1_2740
MEFHIELTELALEMIGAIKDRREQQGIIHRIQKLKREPLQQGKPLTGDLKGLYSVRAVGQRYRVVYQVKSEKIIVVVVGVGRRKEGDKKDVYTLLKKLLERPDG
>Francci3_1442
MTRLAAEPRPSGARALTGQPTGILRIRVGEYRVIYQVDHTRVLVTIVHVAHRREVYRHL
>Memar_0664
MIWRLILMPVAERVLNNIPDPDAGRIKEELYALADEPYPRFHVKKLKGHQNSPLYSLRVGQYRIILVIEDNVMVITVIEIGNRSKIYRKY
>Exig_1264
MTTYTVEFERGAQKSLKKMDPQQSRIIMSWIKKNLVGTDDPRRHGKGLISNRSGEWRYRIGDYRLIADIQDDKVLILILEIGHRRRDIYK
>RF_1286
MQYKLSFSKTALKNLLKISTNKRKAILEKLEQLKLNPYKENNNIKKLIGYDGYRLRIGDYRVIYRINKGKLEILVINVDVRG
>AAur_0707
MSYAVQVAPAAVRQLRKIPPEARRRIQAAIEILAETPRPPGAKKLSGSSGDWRVRTGDYRIIYEIRDAQLIVLVVAMGHRRDIYQH
>FMG_1568
MRVIYSEKSLKSLKKLDKPIQKMIIHYMEKVGQLEEPRARGKALSANLRGFWRYRVSNYRIICEIDDDKLIICVVEVDHRKNIYKC
>PHS013
MTYRVKIHKQVVKALQSLPKAHYRRFLEFRDILEYEPVPREKFDVIKLEGTGDLDLYRARLGDYRVIYSVNWKDKVIKILKLKPRGRAYK
>A1I_07635
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>RBE_1375
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>Suden_0812
MVYNIQYDPKALKQLKKLDKSIALLILDGIEEFASNPVLTKIKKLKTPFDGAYRLRICDYRVVFYQEDNLMLISKIAHRKDVYI
>Nmul_A0326
MTYRISFNPAAQRDLDKMDPQARQRLLKYLNGRIVLLEDARCLGEPLLASQFFGYWRYRAGDYRITCDIQDEELHVLIVKGGNRRVCGVG
>spr1103
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>DNO_0273
MIHSFADKDTYDFFHGTICKRFESFASVARRKLTMLDNAATLEFLRSPPANRLESLKGDRAGQYSIRINDRYRICFRWDSGAHDVEIVDYH
>SPH_1337
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SPD_1081
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SP_1223
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>Atu0674

MIWTIEYHTLVQKEMRKINPEVRRRIRSFLHERLAALDDPRQIGATLQGSELGNFWRYRVGDYRIICDIQDQKLVVLVVEIGHRREIYR
>MA0376
MYRIIYSPAAKRDLKRLPADVQDRVHDALEEIADDPYAHVKKLKTPYNSPIFAYRVGKYRVIMSIHDFELIILVLEVGDRKNIYRKF
>SPCG_1079
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQINVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>CKO_pCKO2p07161
MDKQNARRIVDFMSLRIAVAADPRQSGKPLKGELGEFWRYRVGDYRVLCEIRDDELVILAATIGHRREVYD
>AF2342
MFRVVVHRKATQELKRLKKAHLKKFGVLLETLKTDPIPWKRFDVKKIEGEENTYRIRIGDFRVIYFLDKPTKTVHILKVERRGKVYD
>USA300HOU_2446
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAOUHSC_02756
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAS2348
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SACOL2464
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>MW2380
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>Rpal_3615
MIQSFRGKFARAILHDRKAPKGLPADLLGTARRKLVQLNAAAALADLAIPPGNRLEALRGDLQGLHSIRINDQWRIVFRWKDTGPEDVEIVDYH
>SAHV_2440
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SA2245
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAV2456
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>CJA_2979
MTRSLAWTDAAWSDYLYWQGQDRKTLKRINQLIRETLREPFQGIGKPEPLKESLAGFWSRRIDDTHRLVYAVDDNHLTIIACRYHYER

>pc0990
MKYELFVNPRVEKALSKIDKHMALKIRNNIRSLAANPRPLGVKKIKGNDNAYRIRVGDYRIIYEIYDSKILILIVNVGHRKEVYE
>SAHV_2391
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SA2195
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAV2407
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAR2497
MVRLNITFSPQAFEDYKYFQQNDKKMVKKINELIKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>USA300HOU_2387
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAOUHSC_02691
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAS2298
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SACOL2404
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>MW2329
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAB2287c
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>TK0791
MSFENRILISKRALKELKNVPESQRDIIKDRISKLAFFPLVKLDVQKLKGYDNVYRLRVGEYRVIFEYNKEERIVMILKVGKRGNVYS
>MM_2605
MSFDVKLHPDAVKFLVSLNPETKERLKSGIKNLEMDPFKSRPHADIKKLKGTKKRNDLYRLRIGDYRMIYSVEENTIFILEIIPRERGYDWL
>SPG_1114
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVXQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>MAE_18810
MNYRVIIPKPIQKQLNNLPKQQRERLITAIRLLTDTPRPSGVKKLKGYDETYRIRIGDYRIIYKIQDQEMLIIILSSIHRKDAY

Supplementary file S10 - ParE protein sequences

>Mb2166c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE
>TBFG_11987
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MRA_1968
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>BCG_1998c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Mb1994c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MT2008
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Rv1959c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Plav_0711
MSAYVFSPAARADLEEIWNYSVEQWGAERAETYILQIRDACEALAGGKKKGRAIDGIRPGYLKLAAGSHFLFFRIGNGGVIDVVRILHQRMDVSAHLNEP
>Mext_1517
MSVRLSPRARADLSRIWDDSAERWGADQADCYIRLLAGGFDRLAEDPARGLGADEIRKGYFRLSVGSHILFYRLGAEGGIEVIRILHGRMDFKRHF
>Nmul_B2801
MKRYILSPAAKTDITNIRKYTIQKWGKSQADKYTLELRERMRWLADNPMLGRARDEVKEGYRSFKEGSHIIFYRVVESTIEIIGIPHQNMDIEQNLEEEILLFTN
IADGGFEGD
>Glov_2585
MPSFTLTNMAKADLKEIAKFTQNRWGREQRDLYLQMLDVSFRQLAVNPLKGKDCSDIRIGYRKLLAGSHVIFYRQTLTDTIEIVRVLHGHMDIETRLSEP
>VCA0359
MKPFNLTVAAKADLRDIALFTQRRWGKEQRNVYLKQFDDSFWLLAENPDIGKSCDEIREGYRKFPQGSHVIFYQQTGSQQIRVIRILHKSMDVNPIFGA
>VV1_2525
MRPFQLTNKAKSDLRDIALFTSRRWGREQRNIYLKQFDDSFWLLAENPDIGKACDEIRDGYRKFPQGSHVIFYRQIGSQNIEIIRILHKSMDVNPIFGA
>Mpop_5138
MKVRLSAAARRDLSGIWTYSAKRWDEAQADRYVRLFADSFDGLGRGIVKGRKADDVRPGYFKLAVGSHLIFYRMGAADVIEVVRILHQRMDIDRHL
>NE2114

MKHYLLSPEAKTDITNIRQYTTQQWGKTQADKYILRLRERMRWLADNPMLGRARDEIKEGYRSFSEGDHVIFYRMAGSAIEVIGIPHQNMDIEQNLSSGNLLLPD
IADYEPEDG
>Jann_0024
MKELVIAPRADADLIGIYDYTSDTWGLDQADKYLDELHGRISGLVTGATVSRSAEEVGPGLRRALAGQHVVFFREDAETVKVVRVLHQRMDLGR
>CC_0873
MKPYRLSRRAKADLDDIWTYSEQRWGVEQAADYARELQATIEMIAEHPGMGQPDENLRAGYRRCASGSHVVFYRVGVRVEIIRVLHQSMNARAHLG
>NE0261
MGSFILRQKAMDDLLSIGRYTRKEWGKTQQIRYLTQLDRAFHELADKPGLGRACDDIREGYFKYGVGKHVIFYRHTGKDQIEIIRILHGRMDIEQHL
>Dde_2911
MAKVRITAAAKAHLLDIWAYTEAAWGASQADVYLKDIGAAFDRLAHTPHLGKARPEIHNDYRSLPVKIHVIFYTVTNSATNSVTFVNIIGVLHAKMDVAGRI
>gsl4116
MAATPRPPGGAQRNNYLAKLDASFQLLAQEPQRGRACDELRPGYRKYHVGRHLIFYRESEAGLEIIRVLHERMDIDSHLEDG
>HCH_10015
MTDNTFKLRPKAEADLVSIYQFSLREWGADKAEAYIREINEAFLTLVNNKTLGSDRSYVRPSLRAYYVGSHVVFYKPTVYGVAVIRVLHQSMDYVRHL
>GFO_2358
MSNDKYRISQKALEDLDKIWIYTFKKWSKVQADRYYNLIIEEIHFIADNFLTGKSAEQTRKNYRVTKIKSHLIFYRKDENQTVEIIRILHERMDIKRRL
>Xfasm12_1561
MAKSYRLTPLAEADLEEIWFYTFRHWSIEQADSYHRSLVAVFEGLAAGTKQGCPSVLPDFNKYLCGSHVVYFMDYADHLDIIRILHQRQDTERYL
>Fjoh_4525
MAKYHLTNKAVEDLAVIWNYTFDEWSENQADKYYLLLLDSCQEAAENPSLGKKYDIVTEKLLGYKSNEHILFYQIISENEIEIVRILHGKMDLKSKF
>Dshi_3750
MPENSRWVIRPAAEKDLSDIWHYGADNWGIEQADHYSDSLFALFDLLADFPEMARERSEFTPPVRIHPSGAHLVIYRVEGQGVAIIRILHVHRNLTAYLLDG
>Npun_AF028
MATILIKPLAQADLLDIWNFIASDSFDKADQLLKKIDSQLKMLASNPGMGRKRDSLTPNLRSFPVGNYLIFYRTINQGIEVIRVIHGARDIQSLFEKDDE
>Cphamn1_1161
MVSYRLTNKAVEDLAGIWSYTVDTWSENQADQYFQMLLDSFQDIATGRVIGKHYDGILQSLQGKKAGKHIIFYRKVDSDVVEIVRILHEQMDLKNRIRGNTGNS
>Sbal195_4670
MPLYHLTQEAQSDLIEIRRYTVQQWGDVQSKKYLSELEQTIQLLATNPRLGRLKPDVGSNVMSFPHASHVIYYVTNEPQLIVFGVLHKRMVPFNHLGERDMIS
>MAE_15460
MNRYIIAPSANQDLNKIADYFLVVNIEAGEKLFRKFYKKCQQLVQFPKLGRSYSHIKPSLRGLPLDGYIIFYRVINETVEILRIVNGRQDLDALFSEIK
>PMI0551
MYKLSKLAEEDIYQIARYTIQQFGVTQAKKYHNDLKQTFELLAKAPWIGRECNWVCNGMRRFEFKKHSIYYLPKNDTLFITRLIHHSIDVDFVDFPK

>plu2251
MADDDIFHIARYTIQQFGQNQAKRYHDELKRTFELLAISPWIGRECDWICAGMRRFEFKKHSIYYMLQDDDIFISRVIHQSMDVDILDFPE
>SO_A0079
MYKLSNLAAEDFERIFEYTLLNFGVKQADDYTVSMHNALLAITEQPLIGHECLEIAKELRRHNHHKHAIFYKKQPDGIYILRILHQQMEPLRHFYPDTD
>Mmar10_1629
MTLRVRLTPRGVADIKAIARYTTQTWGHSQCQSYLRSLDDRFRWLAEHPASGRPRDDVAESSRSFRHNAHIVFYRTRDDGIDVLAVVHSAMDIDGI
>Atu1510
MNNYRLSTQAENEILDIFLYGIERFGLNQARLYKDGMESCFQLLGNNPRMGRSATIVGEGIRRHEHGSHVIFYETDGSGVLILTIVHGRSIRRLKL
>Clim_2032
MPKIIIRPLVVEDLAEIWSYIAEDSSNRADSFVDFIDGKFHELAQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDDQLNPQQ
>Mpop_5431
MGSYQLTGRADADILDVFLYGLTEFGMAQAERYQGILEHTFQLLADNPRMGRAAEVIGAGVRRHECGAHVILYEEATDGVLILAVVHSRSVRRLSL
>Rru_A0699
MPITLSDLALKDLEEIRHYTVRQWGREQWLRYYRGLVSTFERIEQSPESGRSRDLFLPGLRSISYGKHSVFYAPIAAAGGAIVILRIVHQKRHLPALTYYEDIE
>YPTS_4253
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YpAngola_B0004
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPDSF_4012
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPA_CD0085
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>pYV0026
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>NE1525
MAEYRLSPAAQRDLDGIFNYTFQQWGAAQAVRYIDILEAACTELVETSSQGQDCSYIRPGYRRRHVERHITTE
>Shewmr4_0332
MNNKRYKLSRLAQIHLQQIKDYTLQHFSESQWHKYKESLISGLQMLANNPGLGRSCNDIYPNGFFFPIGKHTAYFTKEADFILIVALLGQPQLPQNHLK
>Clim_1299
MPTIRIRPLVFDDLAEIWSYIAEDSPNRAVEFIDSIDGKFHELSQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDEQLNPKQ
>RPC_0159
MARKFRLSEQAEADLNSIWRYIASADVRAANGQIDRLTEAMDLLAEAPLAGRARPEFGAQLRSFFLRGYTIFYVVTPGGIEIARVLHERMDIDPADFS
>gll1910
MSRYILNVLAAQDLDEIADRFVAASVEAGERFFEEFDRKCLQLVTFPNSGRSYGTLRTGLRGLPLGEYIIFYRVLDDGIEILRVVSGRRDLPTLFEESGS
>PBPRC0063
MSKYVLSTGAQTDLIDIRQYTLDKWGQVQWKGYFAELKSTMILLAENPAIGAKTLDLGDNYFRFPLKHHVIYYIQQPTQIVIVAVLGKSMSPQKHFEIIN
>SO_A0088
MAVTYHLTPDAQSDLIGIHRFTLAQWGATQSKTYLSGLKQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLTHLAERETI
>Sfri_3471
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRVNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_1990
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_0191
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sala_2662
MTAFRVSAIASRRLDEIFVYSLDTWGQEQAETYIRELFACFDRIARRELLWRAIPAEFSVDGYYCRHEHHYVYWRLLADGDVGIVTILHERMHQMDRFREDDGA
>Shewmr7_3694
MNNKRYKLSRLAQTHLLKIKDYTLQHFSESQWHKYKEGLISGLQMLANNPGLGRSCNDIYPKGFYFPIGKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>VIBHAR_p08247
MTRYKLSPAAQTDLIDIRRYTLENWGGAQWTNYFGELKQSMALLASNELIGIDMPELGTGYCRFPLKHHVIYYIRKPDHIVIAAVLGRNMSPAKHFQRQL
>CC_2756
MGRVIRTRPVSGDLDRVFRDVCENNGVKVASAQLNRIESVFHRLSAFPRLGRDRSDLRPGLRTFSVKPWQVLYRLNGEDVVILRILDGRMNLAAQLGKKT
>SbBS512_E1151
MRIIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>YEP0057
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNLELGAECNWICSDMRRFQYKKHGIYYITLSNDILISRVLHQSIDIDAQDFPEYEQ
>Sputcn32_0428
MNNKRYKLSRLSQTHLQQIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>ECs0462
MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Z0510

MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Sputw3181_0282
MKNKRYKLSKLAQTHLLKIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>plu4785
MYKLTDQAAEDFAGIYDYTFLQFGETQADHYTEALEAFFDTLAEMPHIGREYPSVPGVMLVEFHRHTVFYTIRDTDILIASILHQQMNHPRYFQRSLSAI
>EcSMS35_0443
MYKLSGKAVEDFRGIYDYTLGKFGEEQADRYTDSLGIFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>CC_2984
MWIMSYRLSRKAEQDLIDIYVAGVGLFGVAQAERYQDTLEAAFGAIAAFPHIGRERPELRPPVRVHPCKSHIILYVLDERGALIVRVRHAGEDWVGEAGG
>Plav_1047
MLFGAAQAETYHARLEQAFDFLSANPRAARERLEITPPVRCHPLGVHIIIYLIEENDDVLILRVRHSREDWDASPI
>Npun_F4900
MSNICRFTVTASRDIEAIIDYIADNSSFNAAESLLSKINNKCERLAKFPGMGRRRDELAPNLRSFPVNDYLIFYRSIEEGVEILRVVSGYQDLEGLFLGQDED
>Rsph17025_1825
MAIRVQEAASVRLDEIYRYTRDRWGEAQAESYIKGLFAAFEQIETRGVVSRPVPAEFGAEGYFFRYERHFVYWRRLGNGDIGIVTILHERMHQIARLKDDFG
>msr1301
MGFRLSLAAEEDIIGIAEQGVRLFGAVQARQYHDELFAIFDLIAAGPRMARERLELSPPMRIHPFKAHLVVYRIEADGDVFIVRVRHGHEDWANEGTR
>SDY_2258
MPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNVGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>MAE_01370
MNQYIISTEAIRDMEQILDYLANTNINAGEKFLEEFSKKCRYLSQFPLMGRSYREIRPYLRGLPMKNYIIFYRLTEQGLEIMPFIKGERDLEAFFFENP
>RBE_1165
MMSKILPYKLTRIAESDLLQIKYYSEEVWGQDITKKYLTQIHNKCIEICNFPHIRINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKNMDITKIYSKKLT
>Nwi_3092
MSFRLSLAAEDDIIGIAEEGVRLFGPAQARQYHDDLFAILDLIAANPRMARERHELSPPMRIHPFKAHLVIYRIEEDGVLIVRVRHGHEDWVSEGAR
>PA1S_21500
MAKYRISHDAQADIIDILRFTHNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNIGKVVRPRHFVFFRVATDQVLEVVRVLHD
SMDLDHHLPQR
>gsr3505
MAQVLKTRQAERDIEDIWFYIALEDLQAADRWLEGMSAQAQLVASQPRMGRVRPELGTEIRSFAAGRYVLFYRPLPDGIELVRVLHGARDLDALFGGDL
>PMIP32
MSVLFTKKAKEHIRTIRRYSLTHWGKNVAETYINCLRITIINIIEKQPSIGIDRSDDLFTGIRSFPVESHIIYYREVENGIEVLAVLHQSQDPYIHIKRD
>Ent638_4215
MKEIELTPKAEEDLEAIWDNSFRQFGVVQADAYIGRIAAVFDVLAMHDIGTHRAELGENICSLPVEQHMVYFLSSHSVVTIIRILSQSQDTMRHEPWR
>PSHAa2625
MRKFKLSNDAKEDLRRIYQYGCKEYSQQQADTYFYAFFTLFEKLAANPYPYQSVAHIRKGYRRAVCGSDSIYYRVYETDVEIMAVLGGQDIEEWL
>PA14_60050
MAKYRISHDAQADIVDILRFTHNHFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCHSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
AMDLDQHLPQR
>EcSMS35_0978
MRTIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPEEIIVIRILNQNQYASRHLHWS
>XF2070
MAKSYRLTPLAEADLEEIWFYTFRHWSIGQADSYHRSLVAVFEGLAAGTKLRPPFCSAGLQ
>PSPA7_4525
MAKYRISHDAQADIIDILRFTQNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
SMELERHLARNRLR
>Mpop_4792
MRRAVFLASARADLVQILEDITRASGSLATGQIFVRQLRAQCHRLAALPGTLGRARPDLLPDIRSTPYRGYIIFFRYLGDTFEVVNILHSRRDIDDFFATDATS
>MAE_00450
MSKIIFSPSARLDLKQVNSYLAGKNPQAARILKEKIQQACKKLAKFPNLGRRRDELIPRLRSFPVEDYLIFYFPLENGIEIARVVSGYRDLDAMFDVD
>A1I_01455
MMSKILPYKLTRIAESDLLQIKYYSEEIWGQDITKKYLTQIHNKCIEICNFPHIGINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKTWI
>VC0395_0937
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0385
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0311
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>CJA_3710
MKIEWTTKALMDLARLYDFLAPVNQAVAARLVQQLTQAPNKLLANPRIGEKLDEFLPREVRRLLIGTYEIRYEIGNNTIYLLRIWHTREAR
>RPD_0573
MAHNRRISPKARVDLDAIWSFIAADSEKAADAMIEQITAAFAMLTDNPNAGRLRPEISAAVRSFPVKRYLIFYVAEASGIKIIRVLHGRQDRSRQDLAP
>Rsph17029_3497
MAIRVQETASVRLDEIYRYTRDRWGTAQAETYITGLFAAFEQIDTRGVMSRPVPAEFGVEGYSFRYERHFVYWRRLSNADIGIVTILHERMHQMDRFKQDFG
>PSEEN3278
MVRISYHARTDIVDILRYTEVKFGAAARGRYQDLLQAAFRAIACEPGRVGSAARDELSPGLRSLHLFFCRLEVTSQRVVRPRHIVFYRAAAGEVIEIVRILHDVM
EVSSHLEHLHQ
>STM2954.1n
MVKLTPKASEDLENIWHYGWQHFGEIQADRYINHLSEIFSIMSANNIGTPRPELGEYIYALPFKRHIIYFIQSVTEVIVIRILSQNQDAGKHVNWL
>Saro_2087
MRLEQSRLAKDDLESILEYGLLYFGTERVLAYLDHIESRFQQLLAYPRSGRIEQDLPGTVYSTSCEAHRIYYEPTDDAVVIVRVLQKSMDVSRWIG
>PD0960
MTGYILTAAAETDLRSIIRYTRKQWGDAQMRRYIATLEQDMASLAAGRGVFRNMSVLFPALRMGRCEHHYVFCLPREGAPALIVAIFHERMDLMTRLADRLK
>RB10182
MSRYLLSHSANANLDEIAGDASNAVAILEALHNTFQVLANHPGVGTLREDLLPGIRVFSPPRPANNYVIFFYPISGGIEVAAVIHGSRDWISMFTDGFRPKKS
>XF2032
MTGYVLTEAAESDLRGIVRYTRKQWGNAQVRHYIATLERGIASLAEGRGAFNDMSSLFPALRMGRYEHHYVFCLPREEAPALIVAIFHERMDLMTRLADRLK
>Pfl01_3062
MPQYRISNAARVDIVDILRLSQTQFGDQARQRYQALILAALQALADTPYRIGSHERGEVAPGLRSYHLSYSRQQAKHPHGTVKSPRHVVFYRVANDEVIEVVRLL
HDAMDVQLHLADD
>PA1S_21315
MELKWTNKGLSDLTRLYEFLAAVNRSAAARTVQQLTSAPTKLLANPRIGERLEEFDPRDVRRILVGHYEMRYEIAGSTIYLLRLWHTREDR
>Mmar10_0699
MKIEWTARASGDLARLFDFLEPVAPDAAVRVVQEIARAPNRLLYHPRIGEKLDVYEPREVRRIIIGHYEMRYEIAAGTLFILRVWHSREDRDFGPEA
>Mpop_0593
MGEVVRRPRARRDLLDLWDYIANDDEMAADGFLDRIEGALAMLSDNPKAGRHRPELVTTLRSFPVESIVLFYFPMQTGIELVRVLSGYRDIGRDDFDA
>Rru_A3239
MPRLVILPAARLDLIEIGDFIALDNPERAASFVAEIEARMIQAADRPASFPTRDELHEGLRSARHGRYLIFFIEDGDEVRVVRVLHGARDFQRIMG
>plu0251
MSVRFTNKAREHIRAIKLYSMRRWGANVAEAYSTSLRVTMTEILDRHPSPGRDRSEDLYFGVLSFPVESHIIYYREVPTGIEVLAVLHQTQDPHNHLYPMDFKMH
RDGKGANPREHRERCDRGERVQPTKRQLER

>pRL120082
MAPRVLILPTALDNYRLAVRETARKWSTEQAKAYSRLLRAGFEGIPEAYARVRIKKDERVGNSLFRLYKIEHHYAVYIVVDDSTFVIAAVLHERMDIPAQLRTIE
RLTDREYAALMGHPRPKS
>SMc00694
MSRELVFTPAALADLEETFWFVAADNPRRARSYVAEIEQACRNLCETPLMGRGRPDLRPNLFIFPLWRRVLIAYELPDNRVDILRVFSGGQDYEAIMSGE
>Dole_2119
MKIVWSPLAVERASEIADYISQDKPAAATNWINVLFSKVDQLRANPEIGRIVPEINDRQFRELIYGNYRIIYHIGAKQISILTVRHGRQILPTDEIKA
>ECs2281
MLPVLWLESADTDLDDITSYIARFDIDAAERLWQRLRGCVLPLSEHPYLYPPSDRVPGLREIVAHPNYIILYRVTTSSVEVVNVIHARRQFP
>Amet_0747
MYKILRTDKAEDQLRDIIFYIADDSGDVEVALKYLDKIEAAINRLQEFPKSGSIPRYSILKKRGYRVVTVEKHLVFYRIDEEEETVIIYAIVDGRREYRNLI
>plu4594
MPYHVMLTKNAEADLEDIYDYIVENDSSEKADYVLDQLLKTADSLANFPEKGNYPKELQALGIRDFRQTFFKPYRVIYQITGKQVVIFVIADGRRDMQTLLTHRL
LSVSA
>azo3077
MTRIELAPEVADDFEWILDHLDHYRANDPAARIREIIDAIGVLASNPLIGRPTDNDNRELVIGRRSHGYLALYRYVAGIDTVFVLAVRSQREAGFER
>Mpop_5181
MAHDLVFHRLARADLFEIYDYIEERSGAARAGGFLDRIEAACRGLTEFPGKGTPRDDVPGGLRTWALERRVLIAYRMTPGHIEILRVLYAGRDFRADEIPH
>Gura_1318
MKIKWTNEALEQLIEIEEFISKDSPERTAVFVDQLIEHAEDSLPDNPRMGRTVLEIANPDIRELIFRKYRIVYRLTALSGLVCASRLLSPIRGVRM
>RSc3224
MRLAITPLAEQDLESIADYIAQDNPARAVTFVRDLREQCQRLVMNPPGYRLRPELGDDIRSCAYGRYVIFFVAAPDEVIVIRILHGARDLPAVFHADEP
>RPE_4765
MKCVISSTAEADLLAIHAYLSERNPAAADRIITRFFHRFDELCEFPFLGPDRSELRASLRGLRVDGFVAFYIIEPNQIIVVRVLDGRMDIEGKFAE
>RPC_2216
MGAVRYTRKARQDLIDIWRHIAAESPATADQCLDRIEARCKQLAAFPEIGRERRDIAPDARMLVVERWIAIYRVVEQGVQIVRIVDGARDLSRLALPQK
>NJAUSS_0644
MRLTKQASDDLTAIYRFIAIELQSPLTADKNIHLFEQSIKSLSTFPERCPILEGFESEGIVIRKLIVKNYVIFYRFVGEVVTVLRVLHGTSNIDALLRDIAEDND
VN
>Swol_1971
MKYKILRTDKAEEQLREIIFYIADDSGDIDIALGYLDKIETAINRLQEFPESGSVPRYSILKKQGYRVVIVERHLVFYKINEADKLVIIYAIVDGRREYRNLI
>Spro_0056
MKIKWLRKAASNFDDAYDWFYQESPQVATQFAREVFRLVNLLATNPALGRAGRVMGTREIVMKTFPYLIPYRIKNNEIHILRVFHMRRHPPKKSW
>RF_0787
MVKCDFTNSAKRDIEDIADYSLKNWGRQPTLKYLDEIYRKTLDLSINPNIGVLRSDIYPNLLSFPIRKQLDLLHNVANKEEFEGDTSLRTAAYT
>msr9191
MIPLRNGAKRRPSATYLHKHLQTLSETPALWRKLPGNLAIPADLKLDAYFSHHGRHYVFFRKLSGDRIGVISILHDRMDVPVRLAEDLQALQSRSEDR

>Amet_0378
MHKLRINPMATEDLIEIRDYIMKELENPTAAVNVVRKIIESYEQLKEFPMLGVDLSTKVNVQTDFRYLVSGSYIVFYQADNESVSIYRILYARRDYLKILFPNEV
DID
>Veis_4341
MLPIFWLETADTDLADITQYIGLRDINAAERMWHRLRNCVLPLSEHPYLYRISERVPGLREIVAHPNYLVLYRVTATRIEVVNVVHTRREFPQHSNDGP
>Mext_4629
MAAEVVWSRLARTDLLDIYVSIGSHDPAAAERLYDRIEERARQLALQPRMGPRRPEIRPSTRILVETPFLILYETIPDTDDAPVREVEIVRVLDGRRDVSVVLHA
GENSAEW
>Ppha_0144
MERLHAFLYEINPDAAARAARAILEGAGFLESMPDIGRPMDDDTGRREWFISFSAGAFVLRYMWNKNDTIVIVRVWHSKEKRT
>RHE_CH02732
MTTTYRLTRTTDAVLSGIDEYANLHFGEAQADAYLLDWDRIFILLSRVPAMGDECDALGAGLRRLHIAFYREIPNGILVIDIIGADRLAEGHLQSNRRSGPTAPH
AP
>PSHAa1041
MAKFIAFDKPSAADKWVNDIFDRTDLLGTQPELGREVPELLGSRYRELIFGSYRIIYKVEHEIKILTLRNSRQLLSLSDIEQ
>Rv2142c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE

Supplementary file S11 - RelE/ParE protein sequences

>Rv2142c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE
>Mb2166c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE
>Rv3358
VRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>Mb3393
MRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>TBFG_11987
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MRA_1968
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>BCG_1998c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Mb1994c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MT2008
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Rv1959c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Plav_0711
MSAYVFSPAARADLEEIWNYSVEQWGAERAETYILQIRDACEALAGGKKKGRAIDGIRPGYLKLAAGSHFLFFRIGNGGVIDVVRILHQRMDVSAHLNEP
>Mext_1517
MSVRLSPRARADLSRIWDDSAERWGADQADCYIRLLAGGFDRLAEDPARGLGADEIRKGYFRLSVGSHILFYRLGAEGGIEVIRILHGRMDFKRHF
>Nmul_B2801
MKRYILSPAAKTDITNIRKYTIQKWGKSQADKYTLELRERMRWLADNPMLGRARDEVKEGYRSFKEGSHIIFYRVVESTIEIIGIPHQNMDIEQNLEEEILLFTN
IADGGFEGD
>Glov_2585
MPSFTLTNMAKADLKEIAKFTQNRWGREQRDLYLQMLDVSFRQLAVNPLKGKDCSDIRIGYRKLLAGSHVIFYRQTLTDTIEIVRVLHGHMDIETRLSEP
>VCA0359
MKPFNLTVAAKADLRDIALFTQRRWGKEQRNVYLKQFDDSFWLLAENPDIGKSCDEIREGYRKFPQGSHVIFYQQTGSQQIRVIRILHKSMDVNPIFGA
>VV1_2525
MRPFQLTNKAKSDLRDIALFTSRRWGREQRNIYLKQFDDSFWLLAENPDIGKACDEIRDGYRKFPQGSHVIFYRQIGSQNIEIIRILHKSMDVNPIFGA
>Mpop_5138
MKVRLSAAARRDLSGIWTYSAKRWDEAQADRYVRLFADSFDGLGRGIVKGRKADDVRPGYFKLAVGSHLIFYRMGAADVIEVVRILHQRMDIDRHL

>NE2114
MKHYLLSPEAKTDITNIRQYTTQQWGKTQADKYILRLRERMRWLADNPMLGRARDEIKEGYRSFSEGDHVIFYRMAGSAIEVIGIPHQNMDIEQNLSSGNLLLPD
IADYEPEDG
>Jann_0024
MKELVIAPRADADLIGIYDYTSDTWGLDQADKYLDELHGRISGLVTGATVSRSAEEVGPGLRRALAGQHVVFFREDAETVKVVRVLHQRMDLGR
>CC_0873
MKPYRLSRRAKADLDDIWTYSEQRWGVEQAADYARELQATIEMIAEHPGMGQPDENLRAGYRRCASGSHVVFYRVGVRVEIIRVLHQSMNARAHLG
>NE0261
MGSFILRQKAMDDLLSIGRYTRKEWGKTQQIRYLTQLDRAFHELADKPGLGRACDDIREGYFKYGVGKHVIFYRHTGKDQIEIIRILHGRMDIEQHL
>Dde_2911
MAKVRITAAAKAHLLDIWAYTEAAWGASQADVYLKDIGAAFDRLAHTPHLGKARPEIHNDYRSLPVKIHVIFYTVTNSATNSVTFVNIIGVLHAKMDVAGRI
>gsl4116
MAATPRPPGGAQRNNYLAKLDASFQLLAQEPQRGRACDELRPGYRKYHVGRHLIFYRESEAGLEIIRVLHERMDIDSHLEDG
>HCH_10015
MTDNTFKLRPKAEADLVSIYQFSLREWGADKAEAYIREINEAFLTLVNNKTLGSDRSYVRPSLRAYYVGSHVVFYKPTVYGVAVIRVLHQSMDYVRHL
>GFO_2358
MSNDKYRISQKALEDLDKIWIYTFKKWSKVQADRYYNLIIEEIHFIADNFLTGKSAEQTRKNYRVTKIKSHLIFYRKDENQTVEIIRILHERMDIKRRL
>Xfasm12_1561
MAKSYRLTPLAEADLEEIWFYTFRHWSIEQADSYHRSLVAVFEGLAAGTKQGCPSVLPDFNKYLCGSHVVYFMDYADHLDIIRILHQRQDTERYL
>Fjoh_4525
MAKYHLTNKAVEDLAVIWNYTFDEWSENQADKYYLLLLDSCQEAAENPSLGKKYDIVTEKLLGYKSNEHILFYQIISENEIEIVRILHGKMDLKSKF
>Dshi_3750
MPENSRWVIRPAAEKDLSDIWHYGADNWGIEQADHYSDSLFALFDLLADFPEMARERSEFTPPVRIHPSGAHLVIYRVEGQGVAIIRILHVHRNLTAYLLDG
>Npun_AF028
MATILIKPLAQADLLDIWNFIASDSFDKADQLLKKIDSQLKMLASNPGMGRKRDSLTPNLRSFPVGNYLIFYRTINQGIEVIRVIHGARDIQSLFEKDDE
>Cphamn1_1161
MVSYRLTNKAVEDLAGIWSYTVDTWSENQADQYFQMLLDSFQDIATGRVIGKHYDGILQSLQGKKAGKHIIFYRKVDSDVVEIVRILHEQMDLKNRIRGNTGNS
>Sbal195_4670
MPLYHLTQEAQSDLIEIRRYTVQQWGDVQSKKYLSELEQTIQLLATNPRLGRLKPDVGSNVMSFPHASHVIYYVTNEPQLIVFGVLHKRMVPFNHLGERDMIS
>MAE_15460
MNRYIIAPSANQDLNKIADYFLVVNIEAGEKLFRKFYKKCQQLVQFPKLGRSYSHIKPSLRGLPLDGYIIFYRVINETVEILRIVNGRQDLDALFSEIK
>PMI0551
MYKLSKLAEEDIYQIARYTIQQFGVTQAKKYHNDLKQTFELLAKAPWIGRECNWVCNGMRRFEFKKHSIYYLPKNDTLFITRLIHHSIDVDFVDFPK
>plu2251
MADDDIFHIARYTIQQFGQNQAKRYHDELKRTFELLAISPWIGRECDWICAGMRRFEFKKHSIYYMLQDDDIFISRVIHQSMDVDILDFPE
>SO_A0079
MYKLSNLAAEDFERIFEYTLLNFGVKQADDYTVSMHNALLAITEQPLIGHECLEIAKELRRHNHHKHAIFYKKQPDGIYILRILHQQMEPLRHFYPDTD
>Mmar10_1629
MTLRVRLTPRGVADIKAIARYTTQTWGHSQCQSYLRSLDDRFRWLAEHPASGRPRDDVAESSRSFRHNAHIVFYRTRDDGIDVLAVVHSAMDIDGI
>Atu1510
MNNYRLSTQAENEILDIFLYGIERFGLNQARLYKDGMESCFQLLGNNPRMGRSATIVGEGIRRHEHGSHVIFYETDGSGVLILTIVHGRSIRRLKL
>Clim_2032
MPKIIIRPLVVEDLAEIWSYIAEDSSNRADSFVDFIDGKFHELAQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDDQLNPQQ
>Mpop_5431
MGSYQLTGRADADILDVFLYGLTEFGMAQAERYQGILEHTFQLLADNPRMGRAAEVIGAGVRRHECGAHVILYEEATDGVLILAVVHSRSVRRLSL
>Rru_A0699
MPITLSDLALKDLEEIRHYTVRQWGREQWLRYYRGLVSTFERIEQSPESGRSRDLFLPGLRSISYGKHSVFYAPIAAAGGAIVILRIVHQKRHLPALTYYEDIE
>YPTS_4253
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YpAngola_B0004
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPDSF_4012
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPA_CD0085
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>pYV0026
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>NE1525
MAEYRLSPAAQRDLDGIFNYTFQQWGAAQAVRYIDILEAACTELVETSSQGQDCSYIRPGYRRRHVERHITTE
>Shewmr4_0332
MNNKRYKLSRLAQIHLQQIKDYTLQHFSESQWHKYKESLISGLQMLANNPGLGRSCNDIYPNGFFFPIGKHTAYFTKEADFILIVALLGQPQLPQNHLK
>Clim_1299
MPTIRIRPLVFDDLAEIWSYIAEDSPNRAVEFIDSIDGKFHELSQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDEQLNPKQ
>RPC_0159
MARKFRLSEQAEADLNSIWRYIASADVRAANGQIDRLTEAMDLLAEAPLAGRARPEFGAQLRSFFLRGYTIFYVVTPGGIEIARVLHERMDIDPADFS
>gll1910
MSRYILNVLAAQDLDEIADRFVAASVEAGERFFEEFDRKCLQLVTFPNSGRSYGTLRTGLRGLPLGEYIIFYRVLDDGIEILRVVSGRRDLPTLFEESGS
>PBPRC0063
MSKYVLSTGAQTDLIDIRQYTLDKWGQVQWKGYFAELKSTMILLAENPAIGAKTLDLGDNYFRFPLKHHVIYYIQQPTQIVIVAVLGKSMSPQKHFEIIN
>SO_A0088
MAVTYHLTPDAQSDLIGIHRFTLAQWGATQSKTYLSGLKQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLTHLAERETI
>Sfri_3471
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRVNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_1990
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_0191
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sala_2662
MTAFRVSAIASRRLDEIFVYSLDTWGQEQAETYIRELFACFDRIARRELLWRAIPAEFSVDGYYCRHEHHYVYWRLLADGDVGIVTILHERMHQMDRFREDDGA
>Shewmr7_3694
MNNKRYKLSRLAQTHLLKIKDYTLQHFSESQWHKYKEGLISGLQMLANNPGLGRSCNDIYPKGFYFPIGKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>VIBHAR_p08247
MTRYKLSPAAQTDLIDIRRYTLENWGGAQWTNYFGELKQSMALLASNELIGIDMPELGTGYCRFPLKHHVIYYIRKPDHIVIAAVLGRNMSPAKHFQRQL
>CC_2756
MGRVIRTRPVSGDLDRVFRDVCENNGVKVASAQLNRIESVFHRLSAFPRLGRDRSDLRPGLRTFSVKPWQVLYRLNGEDVVILRILDGRMNLAAQLGKKT
>SbBS512_E1151
MRIIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>YEP0057
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNLELGAECNWICSDMRRFQYKKHGIYYITLSNDILISRVLHQSIDIDAQDFPEYEQ
>Sputcn32_0428
MNNKRYKLSRLSQTHLQQIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>ECs0462
MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Z0510

MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Sputw3181_0282
MKNKRYKLSKLAQTHLLKIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>plu4785
MYKLTDQAAEDFAGIYDYTFLQFGETQADHYTEALEAFFDTLAEMPHIGREYPSVPGVMLVEFHRHTVFYTIRDTDILIASILHQQMNHPRYFQRSLSAI
>EcSMS35_0443
MYKLSGKAVEDFRGIYDYTLGKFGEEQADRYTDSLGIFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>CC_2984
MWIMSYRLSRKAEQDLIDIYVAGVGLFGVAQAERYQDTLEAAFGAIAAFPHIGRERPELRPPVRVHPCKSHIILYVLDERGALIVRVRHAGEDWVGEAGG
>Plav_1047
MLFGAAQAETYHARLEQAFDFLSANPRAARERLEITPPVRCHPLGVHIIIYLIEENDDVLILRVRHSREDWDASPI
>Npun_F4900
MSNICRFTVTASRDIEAIIDYIADNSSFNAAESLLSKINNKCERLAKFPGMGRRRDELAPNLRSFPVNDYLIFYRSIEEGVEILRVVSGYQDLEGLFLGQDED
>Rsph17025_1825
MAIRVQEAASVRLDEIYRYTRDRWGEAQAESYIKGLFAAFEQIETRGVVSRPVPAEFGAEGYFFRYERHFVYWRRLGNGDIGIVTILHERMHQIARLKDDFG
>msr1301
MGFRLSLAAEEDIIGIAEQGVRLFGAVQARQYHDELFAIFDLIAAGPRMARERLELSPPMRIHPFKAHLVVYRIEADGDVFIVRVRHGHEDWANEGTR
>SDY_2258
MPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNVGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>MAE_01370
MNQYIISTEAIRDMEQILDYLANTNINAGEKFLEEFSKKCRYLSQFPLMGRSYREIRPYLRGLPMKNYIIFYRLTEQGLEIMPFIKGERDLEAFFFENP
>RBE_1165
MMSKILPYKLTRIAESDLLQIKYYSEEVWGQDITKKYLTQIHNKCIEICNFPHIRINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKNMDITKIYSKKLT
>Nwi_3092
MSFRLSLAAEDDIIGIAEEGVRLFGPAQARQYHDDLFAILDLIAANPRMARERHELSPPMRIHPFKAHLVIYRIEEDGVLIVRVRHGHEDWVSEGAR
>PA1S_21500
MAKYRISHDAQADIIDILRFTHNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNIGKVVRPRHFVFFRVATDQVLEVVRVLHD
SMDLDHHLPQR
>gsr3505
MAQVLKTRQAERDIEDIWFYIALEDLQAADRWLEGMSAQAQLVASQPRMGRVRPELGTEIRSFAAGRYVLFYRPLPDGIELVRVLHGARDLDALFGGDL
>PMIP32
MSVLFTKKAKEHIRTIRRYSLTHWGKNVAETYINCLRITIINIIEKQPSIGIDRSDDLFTGIRSFPVESHIIYYREVENGIEVLAVLHQSQDPYIHIKRD
>Ent638_4215
MKEIELTPKAEEDLEAIWDNSFRQFGVVQADAYIGRIAAVFDVLAMHDIGTHRAELGENICSLPVEQHMVYFLSSHSVVTIIRILSQSQDTMRHEPWR
>PSHAa2625
MRKFKLSNDAKEDLRRIYQYGCKEYSQQQADTYFYAFFTLFEKLAANPYPYQSVAHIRKGYRRAVCGSDSIYYRVYETDVEIMAVLGGQDIEEWL
>PA14_60050
MAKYRISHDAQADIVDILRFTHNHFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCHSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
AMDLDQHLPQR
>EcSMS35_0978
MRTIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPEEIIVIRILNQNQYASRHLHWS
>XF2070
MAKSYRLTPLAEADLEEIWFYTFRHWSIGQADSYHRSLVAVFEGLAAGTKLRPPFCSAGLQ

>PSPA7_4525
MAKYRISHDAQADIIDILRFTQNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
SMELERHLARNRLR
>Mpop_4792
MRRAVFLASARADLVQILEDITRASGSLATGQIFVRQLRAQCHRLAALPGTLGRARPDLLPDIRSTPYRGYIIFFRYLGDTFEVVNILHSRRDIDDFFATDATS
>MAE_00450
MSKIIFSPSARLDLKQVNSYLAGKNPQAARILKEKIQQACKKLAKFPNLGRRRDELIPRLRSFPVEDYLIFYFPLENGIEIARVVSGYRDLDAMFDVD
>A1I_01455
MMSKILPYKLTRIAESDLLQIKYYSEEIWGQDITKKYLTQIHNKCIEICNFPHIGINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKTWI
>VC0395_0937
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0385
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0311
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>CJA_3710
MKIEWTTKALMDLARLYDFLAPVNQAVAARLVQQLTQAPNKLLANPRIGEKLDEFLPREVRRLLIGTYEIRYEIGNNTIYLLRIWHTREAR
>RPD_0573
MAHNRRISPKARVDLDAIWSFIAADSEKAADAMIEQITAAFAMLTDNPNAGRLRPEISAAVRSFPVKRYLIFYVAEASGIKIIRVLHGRQDRSRQDLAP
>Rsph17029_3497
MAIRVQETASVRLDEIYRYTRDRWGTAQAETYITGLFAAFEQIDTRGVMSRPVPAEFGVEGYSFRYERHFVYWRRLSNADIGIVTILHERMHQMDRFKQDFG
>PSEEN3278
MVRISYHARTDIVDILRYTEVKFGAAARGRYQDLLQAAFRAIACEPGRVGSAARDELSPGLRSLHLFFCRLEVTSQRVVRPRHIVFYRAAAGEVIEIVRILHDVM
EVSSHLEHLHQ
>STM2954.1n
MVKLTPKASEDLENIWHYGWQHFGEIQADRYINHLSEIFSIMSANNIGTPRPELGEYIYALPFKRHIIYFIQSVTEVIVIRILSQNQDAGKHVNWL
>Saro_2087
MRLEQSRLAKDDLESILEYGLLYFGTERVLAYLDHIESRFQQLLAYPRSGRIEQDLPGTVYSTSCEAHRIYYEPTDDAVVIVRVLQKSMDVSRWIG
>PD0960
MTGYILTAAAETDLRSIIRYTRKQWGDAQMRRYIATLEQDMASLAAGRGVFRNMSVLFPALRMGRCEHHYVFCLPREGAPALIVAIFHERMDLMTRLADRLK
>RB10182
MSRYLLSHSANANLDEIAGDASNAVAILEALHNTFQVLANHPGVGTLREDLLPGIRVFSPPRPANNYVIFFYPISGGIEVAAVIHGSRDWISMFTDGFRPKKS
>XF2032
MTGYVLTEAAESDLRGIVRYTRKQWGNAQVRHYIATLERGIASLAEGRGAFNDMSSLFPALRMGRYEHHYVFCLPREEAPALIVAIFHERMDLMTRLADRLK
>Pfl01_3062
MPQYRISNAARVDIVDILRLSQTQFGDQARQRYQALILAALQALADTPYRIGSHERGEVAPGLRSYHLSYSRQQAKHPHGTVKSPRHVVFYRVANDEVIEVVRLL
HDAMDVQLHLADD
>PA1S_21315
MELKWTNKGLSDLTRLYEFLAAVNRSAAARTVQQLTSAPTKLLANPRIGERLEEFDPRDVRRILVGHYEMRYEIAGSTIYLLRLWHTREDR
>Mmar10_0699
MKIEWTARASGDLARLFDFLEPVAPDAAVRVVQEIARAPNRLLYHPRIGEKLDVYEPREVRRIIIGHYEMRYEIAAGTLFILRVWHSREDRDFGPEA
>Mpop_0593
MGEVVRRPRARRDLLDLWDYIANDDEMAADGFLDRIEGALAMLSDNPKAGRHRPELVTTLRSFPVESIVLFYFPMQTGIELVRVLSGYRDIGRDDFDA
>Rru_A3239
MPRLVILPAARLDLIEIGDFIALDNPERAASFVAEIEARMIQAADRPASFPTRDELHEGLRSARHGRYLIFFIEDGDEVRVVRVLHGARDFQRIMG
>plu0251

MSVRFTNKAREHIRAIKLYSMRRWGANVAEAYSTSLRVTMTEILDRHPSPGRDRSEDLYFGVLSFPVESHIIYYREVPTGIEVLAVLHQTQDPHNHLYPMDFKMH
RDGKGANPREHRERCDRGERVQPTKRQLER
>pRL120082
MAPRVLILPTALDNYRLAVRETARKWSTEQAKAYSRLLRAGFEGIPEAYARVRIKKDERVGNSLFRLYKIEHHYAVYIVVDDSTFVIAAVLHERMDIPAQLRTIE
RLTDREYAALMGHPRPKS
>SMc00694
MSRELVFTPAALADLEETFWFVAADNPRRARSYVAEIEQACRNLCETPLMGRGRPDLRPNLFIFPLWRRVLIAYELPDNRVDILRVFSGGQDYEAIMSGE
>Dole_2119
MKIVWSPLAVERASEIADYISQDKPAAATNWINVLFSKVDQLRANPEIGRIVPEINDRQFRELIYGNYRIIYHIGAKQISILTVRHGRQILPTDEIKA
>ECs2281
MLPVLWLESADTDLDDITSYIARFDIDAAERLWQRLRGCVLPLSEHPYLYPPSDRVPGLREIVAHPNYIILYRVTTSSVEVVNVIHARRQFP
>Amet_0747
MYKILRTDKAEDQLRDIIFYIADDSGDVEVALKYLDKIEAAINRLQEFPKSGSIPRYSILKKRGYRVVTVEKHLVFYRIDEEEETVIIYAIVDGRREYRNLI
>plu4594
MPYHVMLTKNAEADLEDIYDYIVENDSSEKADYVLDQLLKTADSLANFPEKGNYPKELQALGIRDFRQTFFKPYRVIYQITGKQVVIFVIADGRRDMQTLLTHRL
LSVSA
>azo3077
MTRIELAPEVADDFEWILDHLDHYRANDPAARIREIIDAIGVLASNPLIGRPTDNDNRELVIGRRSHGYLALYRYVAGIDTVFVLAVRSQREAGFER
>Mpop_5181
MAHDLVFHRLARADLFEIYDYIEERSGAARAGGFLDRIEAACRGLTEFPGKGTPRDDVPGGLRTWALERRVLIAYRMTPGHIEILRVLYAGRDFRADEIPH
>Gura_1318
MKIKWTNEALEQLIEIEEFISKDSPERTAVFVDQLIEHAEDSLPDNPRMGRTVLEIANPDIRELIFRKYRIVYRLTALSGLVCASRLLSPIRGVRM
>RSc3224
MRLAITPLAEQDLESIADYIAQDNPARAVTFVRDLREQCQRLVMNPPGYRLRPELGDDIRSCAYGRYVIFFVAAPDEVIVIRILHGARDLPAVFHADEP
>RPE_4765
MKCVISSTAEADLLAIHAYLSERNPAAADRIITRFFHRFDELCEFPFLGPDRSELRASLRGLRVDGFVAFYIIEPNQIIVVRVLDGRMDIEGKFAE
>RPC_2216
MGAVRYTRKARQDLIDIWRHIAAESPATADQCLDRIEARCKQLAAFPEIGRERRDIAPDARMLVVERWIAIYRVVEQGVQIVRIVDGARDLSRLALPQK
>NJAUSS_0644
MRLTKQASDDLTAIYRFIAIELQSPLTADKNIHLFEQSIKSLSTFPERCPILEGFESEGIVIRKLIVKNYVIFYRFVGEVVTVLRVLHGTSNIDALLRDIAEDND
VN
>Swol_1971
MKYKILRTDKAEEQLREIIFYIADDSGDIDIALGYLDKIETAINRLQEFPESGSVPRYSILKKQGYRVVIVERHLVFYKINEADKLVIIYAIVDGRREYRNLI
>Spro_0056
MKIKWLRKAASNFDDAYDWFYQESPQVATQFAREVFRLVNLLATNPALGRAGRVMGTREIVMKTFPYLIPYRIKNNEIHILRVFHMRRHPPKKSW
>RF_0787
MVKCDFTNSAKRDIEDIADYSLKNWGRQPTLKYLDEIYRKTLDLSINPNIGVLRSDIYPNLLSFPIRKQLDLLHNVANKEEFEGDTSLRTAAYT
>msr9191
MIPLRNGAKRRPSATYLHKHLQTLSETPALWRKLPGNLAIPADLKLDAYFSHHGRHYVFFRKLSGDRIGVISILHDRMDVPVRLAEDLQALQSRSEDR
>Amet_0378
MHKLRINPMATEDLIEIRDYIMKELENPTAAVNVVRKIIESYEQLKEFPMLGVDLSTKVNVQTDFRYLVSGSYIVFYQADNESVSIYRILYARRDYLKILFPNEV
DID
>Veis_4341
MLPIFWLETADTDLADITQYIGLRDINAAERMWHRLRNCVLPLSEHPYLYRISERVPGLREIVAHPNYLVLYRVTATRIEVVNVVHTRREFPQHSNDGP
>Mext_4629
MAAEVVWSRLARTDLLDIYVSIGSHDPAAAERLYDRIEERARQLALQPRMGPRRPEIRPSTRILVETPFLILYETIPDTDDAPVREVEIVRVLDGRRDVSVVLHA
GENSAEW

>Ppha_0144
MERLHAFLYEINPDAAARAARAILEGAGFLESMPDIGRPMDDDTGRREWFISFSAGAFVLRYMWNKNDTIVIVRVWHSKEKRT
>RHE_CH02732
MTTTYRLTRTTDAVLSGIDEYANLHFGEAQADAYLLDWDRIFILLSRVPAMGDECDALGAGLRRLHIAFYREIPNGILVIDIIGADRLAEGHLQSNRRSGPTAPH
AP
>PSHAa1041
MAKFIAFDKPSAADKWVNDIFDRTDLLGTQPELGREVPELLGSRYRELIFGSYRIIYKVEHEIKILTLRNSRQLLSLSDIEQ
>TBFG_11272
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>MRA_1255
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>BCG_1306c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Mb1278c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Rv1246c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>TBFG_12882
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>MRA_2891
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>BCG_2888
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Mb2891
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Rv2866
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Lxx22677
MSWDVQFAPAAIRGLDRLPPRVVAAVVEFVTVTLPGNPYRMSKPLQGDLEGYYSARRGDYRVLFSLDEDRRVLLVGRIAHRADVYRPR
>Francci3_2679
MTGAAGPYRLEITGPAARALAGRIPEKVATAVHEFITTTLLENPHRLGKRLLYPPYAGTWSARRGMYRVLYEIDEENRIVLVTAVEHRADAYRGR
>Noca_2680
MSAPDEGGTGYEVVFTRGARRALEWDLPAAVAMAAFEFIRGPLREAPRRVGKPLLEPLTPLWSARRGEYRILYRILDRRLVIAVVTIAHRRDAYGRRE
>TDE0735
MKVVLTETFKKQLKKLDATISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGDYRILCRICDDKLIITVIEIGHRSTVYR
>Suden_1966
MSYKLLIDDKVIKDLKQIDKLWQKKIIEVIKTKLVENPHLGKPLVGNLSPYYRLRVFDYRVIYEINDDEVVVIVIKIGHRKDIYK
>TDE1978
MKVILTETFKKQLKKLDAAISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGNYRILCRIHDDRLIITVIEIGHRSTVYR
>Paes_2094
MVWTIEFAATAEKELSKLDKSAAKRILKFLKERVATDPRSSGKALRGDHAGLWRYRIGDYRVICEFRDQTVSVLVVRIGHRKEVYR
>Mhun_1873
MSKFTLLISKGAERDLAQLPKFARRSLEVAFAELELLENPREKLGPLKGRAKGLYSLRTGEYRAILEIFDNKLMLLVIEAGHRKTIYRKYQS
>WD0126
MKYDIVYSKNFTERDFLNLPKTIRSRITKAINERFTTDPIKFGKPLRGRLKGYRRIRVSDYRIIYTVNIAKHKVFVATAGHRDTIYEKAT
>FN0211

MKYDVEYSKTAMNTIKKMDSSTSKLIRTWIEKNLINTENPRIKGKALTGDLKGLWRYRIGDYRILAEIQDDKIVILILDIGHRSKIYL
>WD0122
MGLERYKVKSLKSVVEKDLPNLPKIIRLRVQKAIKERLTVDPINFGEPLHHNLKGCRRLRVGDYRVIYRVNQLDHIVTITEIGHRDDIYKK
>WD0600
MGIIYQIGYLEGVDTEDLPSLPKTIRLRVQKAIEKRLTIIPDKVGEALSHKWVGYFRLRVGDYRVIYLIDNSEHMVKIAAIGHRKEIYKRSPE
>Cpha266_1376
MVWKIEFASSAEKELARLDKSAARRIVKYLRERVAIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEVYR
>RF_0898
MFKVIWENKARAELAELPYPLKILDKVESYLAQNPIVLGKPLKGEYKNLYRYRFGNYRIIYSVSIEKSTVTVIKIGHRANIY
>asl4561
MIYQIEITTRAAKQLKKLSEDIKLKIEEKIQELSNNPRSNDVVKLEGEEDTYRIRVGNYRILYEIKDDLLIVKVVKISHRRDVYRRK
>Clim_1338
MVWKIEFAASAEKELARRDKSAARRIVKYLRERVEIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEIYR
>Gura_2099
MTYRIELTKTAERDLLAVPKPVLKRLDACILGLADDPLPPGVKKLKNSDGLYRVRVSDYRIIYRIEQEILTVLVVKIGHRREVYR
>Gmet_A3569
MVWKVEIDPAARRELKKLDPQISGRVLKFLFERVARLDDPRSIGEALKGSRFGDFWKYRVGDYRIITSIEDEALVILVVRVGNRREVYER
>Rru_A3196
MTWKIEFDPSALRELDKLDPQIAARVLRFLRDRVAVLENPRSLGEALKGPRLGAFWKYRVSDYRIIAHIEDDTLRILVLRIGNRREVYR
>WD0404
MKTSGNKTYTIKFLKNVIEKEIPALPAKIKLMVQEAIKKRLTVDPFNLGKPLCHSFRGQYRIRVSNYRIIYIINHSERKVLITAVGYRKNIYKHRRLHN
>Plav_1617
MAWTIDYTDDAINQLRKLDKQSARRIVEYMDERIAGAANPRSFGKPLSGPLGQFWRYRVGDFRIICDIEDEVLRVLVVRIGDRKDVYRKAKR
>SCATT_39270
MSEYRTVFRPEAQAELRKVPRDMALRILAKLTELESDPLGFNTTALVSQPDRRRLRVGDYRVIYTIDNGELVVWVVHVGHRSTVYGA
>XfasM23_1261
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>PD1184
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>WD0269
MKTSGNKRYTIKYLKHVLKRNLPSLPEAIKPKIKDAIREYLATDPIGNGVLLRNRLKGHRRIRVDDYRVVYRVNTAERKVTIVSIGHRDNIYKQAILDLLKH
>FN0497
MKKYEVKFSEAAIKELKKLDKPTATMIKLWVIQNLENTINPRQHGKSLTANYSGKWRYRVGNYRLLAEIYDDEILILIFKVAHRSIVYKK
>RF_1272
MEMNYKICYLEEVTNKHIPMLSSNAKTLIKCAIEERLMFDPIAFGKPLRYSLKGHRRFRISDYRIIYRIEQETSTVIIIAIKHRKEIYQEFI
>Pcar_1777
MKMSGGRVAFSVIYHPDVKGRDIPKINGDVRVRIKKAIETRLMVAPQEYGEPLRKTLKGYWKLRVGDYRIVFKIDGDEILILGICHRKGVYPLMESRQ
>LPC_2245
MMTPGSKLYQIEYIEDVVKNDMPSLSTSAKKLIKKAIEERLMADPIGFGKPLRYSLKGHRRLRVSDYRIVYRIEAETNTVVIIAIKHRKEVYDDF
>MGAS10270_Spy1927
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>M28_Spy1840
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>SSU98_0616
MAYKLVLSDDALKQLKKMDRHVGMMLAKDLKKRLDGLENPRQFGKALVGDYKGLWRYRVGNYRVICDIIDNKMVILALEIGHRKEIYKK
>Xfasm12_1292
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCDIQDSVLRVLVVRVGHRGEIYR
>pTC-F14_p08
MAWRIEFDDKAKKDLAALDKSVAKRITAFLRERVAHLDDPRSIGEALKGSKLGDFWKYRVGDWRIIASIEDEALRILVVRIGNRREVYRK
>Sbal_4390
MAWTIDYSERALKSLKKMDKQNAKRILDFLEQRIAILDDPRTSGKPLKGDLGIFWRYRVGDYRVLCEIQDSKLVILTALIGHRKEIYE
>Npun_R1755
MPSVEMSERYSLRIAKTAEKDLLDLQAKLYKQVVSKILSLQGNSRPQDCKALKGYEGGYRVDQGEYRILYTIDEESKLIDVFRVGKRNDGEVYKNL
>RSal33209_0100
MTYRIAYTPRVIKDLKKLDKQTVRRVKDFFDRLNRDNPRSLGKALVGQDFWRYRIGDHRILVAIQDDVLTVLVVKVAHRREVYKER
>MA1694
MTYQVVLSPDFEKETKIFFKKDPVLYGRFKKTVNSILENPECGKPLRNVLKGLRRVHIGHFVLIYEIDNTNETITFLKFSPHDKAYK
>Noc_0433
MYAIVVHRRAARYLRKLPQDQQVKIKHVLAQMKNGPLGLSGIKSMVGDWAGYRRIRVGNVRIIFWIDELKNVVYVDHIGPRGDVYKDKT
>RSc3279
MNAIHWTAWAARQLRKLDRQHQRVLVEAVGQLEAMPHCRQVRALREHRYGYRLRVGDYRVLSDWDDGIRIVDIQEVSKRDERTYRH
>Ppha_0985
MYKIIFTKEAQKALLRLPGSTAVQVCQKLEQLAADSYAPIANAKKLQNRSGYRLRIGDWRVIYEIQNDKLVVLVLKIAQRGEVYR
>asl2100
MSDRYTLRIARTAEKDLLDLQPKQFKQVVSKILSLQGTPRPQDCKALKGYEGGYRVDQGEYRILYTIDDETQLVDVFRVGKRNDDEVYQNL
>BH07080
MAWTIRYEKKALSFLKKCDKKEARRIVDFLDQHVAPLEDVRVIGKPLKGQFSGLWRYRVGDYRILCELYDKELVVLVLAVGHRKNIYKG
>TK0965
MSYELILSGKSEKALKKAPPEDRKRIVSALFKLKENPWAMQYKKLRGYPFYRVRVGDWRIIYTVDDEARIVYVVRLGKREGVYDSL
>SSU05_0886
MKTCYKLVPTSRFIKQLKKLDKFTQKQITNYLSSHVTDNPRQYGKALTANRSGQWRYRIGNYRVIVNIEDDKLIVTSIKMRNGSI
>Ppha_1272
MAYSVGYKKSVQCDLRQLSRLDAQRIYDQIEQELVKNPKSNPLLKGRFVGLRKYRVGDYRVIYCVLDEEVIILRIAHRRDVYKRDI
>Mhun_0373
MEPYGISMAPSALKNYKKFPEKLREKIKSEALHIARNPYIYEELSEPLKGIRSYHFTFNSTQYRIAYQINESSREIEILLVKTRENFYDKLFRTFR
>ECA0674
MAHIVWTGKAVKDLRKLPANDQKAIQANVNSLGDYPATKSKPLDITKLTDRGSEYRLRVGNYRVLFEIQKGEPVIIEIQRVLRRTSTTY
>HY04AAS1_1546
MWKVVITKQAKKDLVNIYRAGLKSKFEKLVEEIKKDPYTSQCEKLIGDLEGYYSCRINRKHRLVYEIDDEQKIVKIVSVWNHY
>AM1_2740
MEFHIELTELALEMIGAIKDRREQQGIIHRIQKLKREPLQQGKPLTGDLKGLYSVRAVGQRYRVVYQVKSEKIIVVVVGVGRRKEGDKKDVYTLLKKLLERPDG
>Francci3_1442
MTRLAAEPRPSGARALTGQPTGILRIRVGEYRVIYQVDHTRVLVTIVHVAHRREVYRHL
>Memar_0664
MIWRLILMPVAERVLNNIPDPDAGRIKEELYALADEPYPRFHVKKLKGHQNSPLYSLRVGQYRIILVIEDNVMVITVIEIGNRSKIYRKY
>Exig_1264
MTTYTVEFERGAQKSLKKMDPQQSRIIMSWIKKNLVGTDDPRRHGKGLISNRSGEWRYRIGDYRLIADIQDDKVLILILEIGHRRDIYK
>RF_1286
MQYKLSFSKTALKNLLKISTNKRKAILEKLEQLKLNPYKENNNIKKLIGYDGYRLRIGDYRVIYRINKGKLEILVINVDVRG
>AAur_0707
MSYAVQVAPAAVRQLRKIPPEARRRIQAAIEILAETPRPPGAKKLSGSSGDWRVRTGDYRIIYEIRDAQLIVLVVAMGHRRDIYQH
>FMG_1568
MRVIYSEKSLKSLKKLDKPIQKMIIHYMEKVGQLEEPRARGKALSANLRGFWRYRVSNYRIICEIDDDKLIICVVEVDHRKNIYKC
>PHS013

MTYRVKIHKQVVKALQSLPKAHYRRFLEFRDILEYEPVPREKFDVIKLEGTGDLDLYRARLGDYRVIYSVNWKDKVIKILKLKPRGRAYK
>A1I_07635
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>RBE_1375
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>Suden_0812
MVYNIQYDPKALKQLKKLDKSIALLILDGIEEFASNPVLTKIKKLKTPFDGAYRLRICDYRVVFYQEDNLMLISKIAHRKDVYI
>Nmul_A0326
MTYRISFNPAAQRDLDKMDPQARQRLLKYLNGRIVLLEDARCLGEPLLASQFFGYWRYRAGDYRITCDIQDEELHVLIVKGGNRRVCGVG
>spr1103
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>DNO_0273
MIHSFADKDTYDFFHGTICKRFESFASVARRKLTMLDNAATLEFLRSPPANRLESLKGDRAGQYSIRINDRYRICFRWDSGAHDVEIVDYH
>SPH_1337
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SPD_1081
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SP_1223
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>Atu0674
MIWTIEYHTLVQKEMRKINPEVRRRIRSFLHERLAALDDPRQIGATLQGSELGNFWRYRVGDYRIICDIQDQKLVVLVVEIGHRREIYR
>MA0376
MYRIIYSPAAKRDLKRLPADVQDRVHDALEEIADDPYAHVKKLKTPYNSPIFAYRVGKYRVIMSIHDFELIILVLEVGDRKNIYRKF
>SPCG_1079
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQINVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>CKO_pCKO2p07161
MDKQNARRIVDFMSLRIAVAADPRQSGKPLKGELGEFWRYRVGDYRVLCEIRDDELVILAATIGHRREVYD
>AF2342
MFRVVVHRKATQELKRLKKAHLKKFGVLLETLKTDPIPWKRFDVKKIEGEENTYRIRIGDFRVIYFLDKPTKTVHILKVERRGKVYD
>USA300HOU_2446
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAOUHSC_02756
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAS2348
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SACOL2464
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>MW2380
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>Rpal_3615
MIQSFRGKFARAILHDRKAPKGLPADLLGTARRKLVQLNAAAALADLAIPPGNRLEALRGDLQGLHSIRINDQWRIVFRWKDTGPEDVEIVDYH
>SAHV_2440
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SA2245
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAV2456
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>CJA_2979
MTRSLAWTDAAWSDYLYWQGQDRKTLKRINQLIRETLREPFQGIGKPEPLKESLAGFWSRRIDDTHRLVYAVDDNHLTIIACRYHYER
>pc0990
MKYELFVNPRVEKALSKIDKHMALKIRNNIRSLAANPRPLGVKKIKGNDNAYRIRVGDYRIIYEIYDSKILILIVNVGHRKEVYE
>SAHV_2391
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY

>SA2195
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAV2407
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAR2497
MVRLNITFSPQAFEDYKYFQQNDKKMVKKINELIKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>USA300HOU_2387
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAOUHSC_02691
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAS2298
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SACOL2404
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>MW2329
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAB2287c
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>TK0791
MSFENRILISKRALKELKNVPESQRDIIKDRISKLAFFPLVKLDVQKLKGYDNVYRLRVGEYRVIFEYNKEERIVMILKVGKRGNVYS
>MM_2605
MSFDVKLHPDAVKFLVSLNPETKERLKSGIKNLEMDPFKSRPHADIKKLKGTKKRNDLYRLRIGDYRMIYSVEENTIFILEIIPRERGYDWL
>SPG_1114
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVXQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>MAE_18810
MNYRVIIPKPIQKQLNNLPKQQRERLITAIRLLTDTPRPSGVKKLKGYDETYRIRIGDYRIIYKIQDQEMLIIILSSIHRKDAY

Supplementary file S12 - RNase T1/RelE/ParE protein sequences

>RNaseT1
MMYSKLLTLTTLLLPTALALPSLVERACDYTCGSNCYSSSDVSTAQAAGYQLHEDGETVGSNSYPHKYNNYEGFDFSVSSPYYEWPILSSGDVYSGGSPGADRVV
FNENNQLAGVITHTG
ASGNNFVECT
>Rv2142c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE
>Mb2166c
MTRRLRVHNGVEDDLFEAFSYYADAAPDQIDRLYNLFVDAVTKRIPQAPNAFAPLFKHYRHIYLRPFRYYVAYRTTDEAIDILAVRHGMENPNAVEAEISGRTFE
>Rv3358
VRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>Mb3393

MRSVNFDPDAWEDFLFWLAADRKTARRITRLIGEIQRDPFSGIGKPEPLQGELSGYWSRRIDDEHRLVYRAGDDEVTMLKARYHY
>TBFG_11987
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MRA_1968
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>BCG_1998c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Mb1994c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>MT2008
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Rv1959c
MSSRYLLSPAAQAHLEEIWDCTYDRWGVDQAEQYLRELQHAIDRAAANPRIGRACDEIRPGYRKLSAGSHTLFYRVTGEGTIDVVRVLHQRMDVDRNL
>Plav_0711
MSAYVFSPAARADLEEIWNYSVEQWGAERAETYILQIRDACEALAGGKKKGRAIDGIRPGYLKLAAGSHFLFFRIGNGGVIDVVRILHQRMDVSAHLNEP
>Mext_1517
MSVRLSPRARADLSRIWDDSAERWGADQADCYIRLLAGGFDRLAEDPARGLGADEIRKGYFRLSVGSHILFYRLGAEGGIEVIRILHGRMDFKRHF
>Nmul_B2801
MKRYILSPAAKTDITNIRKYTIQKWGKSQADKYTLELRERMRWLADNPMLGRARDEVKEGYRSFKEGSHIIFYRVVESTIEIIGIPHQNMDIEQNLEEEILLFTN
IADGGFEGD
>Glov_2585
MPSFTLTNMAKADLKEIAKFTQNRWGREQRDLYLQMLDVSFRQLAVNPLKGKDCSDIRIGYRKLLAGSHVIFYRQTLTDTIEIVRVLHGHMDIETRLSEP
>VCA0359
MKPFNLTVAAKADLRDIALFTQRRWGKEQRNVYLKQFDDSFWLLAENPDIGKSCDEIREGYRKFPQGSHVIFYQQTGSQQIRVIRILHKSMDVNPIFGA
>VV1_2525
MRPFQLTNKAKSDLRDIALFTSRRWGREQRNIYLKQFDDSFWLLAENPDIGKACDEIRDGYRKFPQGSHVIFYRQIGSQNIEIIRILHKSMDVNPIFGA
>Mpop_5138
MKVRLSAAARRDLSGIWTYSAKRWDEAQADRYVRLFADSFDGLGRGIVKGRKADDVRPGYFKLAVGSHLIFYRMGAADVIEVVRILHQRMDIDRHL
>NE2114
MKHYLLSPEAKTDITNIRQYTTQQWGKTQADKYILRLRERMRWLADNPMLGRARDEIKEGYRSFSEGDHVIFYRMAGSAIEVIGIPHQNMDIEQNLSSGNLLLPD
IADYEPEDG
>Jann_0024
MKELVIAPRADADLIGIYDYTSDTWGLDQADKYLDELHGRISGLVTGATVSRSAEEVGPGLRRALAGQHVVFFREDAETVKVVRVLHQRMDLGR
>CC_0873
MKPYRLSRRAKADLDDIWTYSEQRWGVEQAADYARELQATIEMIAEHPGMGQPDENLRAGYRRCASGSHVVFYRVGVRVEIIRVLHQSMNARAHLG
>NE0261
MGSFILRQKAMDDLLSIGRYTRKEWGKTQQIRYLTQLDRAFHELADKPGLGRACDDIREGYFKYGVGKHVIFYRHTGKDQIEIIRILHGRMDIEQHL
>Dde_2911
MAKVRITAAAKAHLLDIWAYTEAAWGASQADVYLKDIGAAFDRLAHTPHLGKARPEIHNDYRSLPVKIHVIFYTVTNSATNSVTFVNIIGVLHAKMDVAGRI
>gsl4116
MAATPRPPGGAQRNNYLAKLDASFQLLAQEPQRGRACDELRPGYRKYHVGRHLIFYRESEAGLEIIRVLHERMDIDSHLEDG
>HCH_10015
MTDNTFKLRPKAEADLVSIYQFSLREWGADKAEAYIREINEAFLTLVNNKTLGSDRSYVRPSLRAYYVGSHVVFYKPTVYGVAVIRVLHQSMDYVRHL
>GFO_2358
MSNDKYRISQKALEDLDKIWIYTFKKWSKVQADRYYNLIIEEIHFIADNFLTGKSAEQTRKNYRVTKIKSHLIFYRKDENQTVEIIRILHERMDIKRRL
>Xfasm12_1561
MAKSYRLTPLAEADLEEIWFYTFRHWSIEQADSYHRSLVAVFEGLAAGTKQGCPSVLPDFNKYLCGSHVVYFMDYADHLDIIRILHQRQDTERYL
>Fjoh_4525
MAKYHLTNKAVEDLAVIWNYTFDEWSENQADKYYLLLLDSCQEAAENPSLGKKYDIVTEKLLGYKSNEHILFYQIISENEIEIVRILHGKMDLKSKF
>Dshi_3750
MPENSRWVIRPAAEKDLSDIWHYGADNWGIEQADHYSDSLFALFDLLADFPEMARERSEFTPPVRIHPSGAHLVIYRVEGQGVAIIRILHVHRNLTAYLLDG
>Npun_AF028
MATILIKPLAQADLLDIWNFIASDSFDKADQLLKKIDSQLKMLASNPGMGRKRDSLTPNLRSFPVGNYLIFYRTINQGIEVIRVIHGARDIQSLFEKDDE
>Cphamn1_1161
MVSYRLTNKAVEDLAGIWSYTVDTWSENQADQYFQMLLDSFQDIATGRVIGKHYDGILQSLQGKKAGKHIIFYRKVDSDVVEIVRILHEQMDLKNRIRGNTGNS

>Sbal195_4670
MPLYHLTQEAQSDLIEIRRYTVQQWGDVQSKKYLSELEQTIQLLATNPRLGRLKPDVGSNVMSFPHASHVIYYVTNEPQLIVFGVLHKRMVPFNHLGERDMIS
>MAE_15460
MNRYIIAPSANQDLNKIADYFLVVNIEAGEKLFRKFYKKCQQLVQFPKLGRSYSHIKPSLRGLPLDGYIIFYRVINETVEILRIVNGRQDLDALFSEIK
>PMI0551
MYKLSKLAEEDIYQIARYTIQQFGVTQAKKYHNDLKQTFELLAKAPWIGRECNWVCNGMRRFEFKKHSIYYLPKNDTLFITRLIHHSIDVDFVDFPK
>plu2251
MADDDIFHIARYTIQQFGQNQAKRYHDELKRTFELLAISPWIGRECDWICAGMRRFEFKKHSIYYMLQDDDIFISRVIHQSMDVDILDFPE
>SO_A0079
MYKLSNLAAEDFERIFEYTLLNFGVKQADDYTVSMHNALLAITEQPLIGHECLEIAKELRRHNHHKHAIFYKKQPDGIYILRILHQQMEPLRHFYPDTD
>Mmar10_1629
MTLRVRLTPRGVADIKAIARYTTQTWGHSQCQSYLRSLDDRFRWLAEHPASGRPRDDVAESSRSFRHNAHIVFYRTRDDGIDVLAVVHSAMDIDGI
>Atu1510
MNNYRLSTQAENEILDIFLYGIERFGLNQARLYKDGMESCFQLLGNNPRMGRSATIVGEGIRRHEHGSHVIFYETDGSGVLILTIVHGRSIRRLKL
>Clim_2032
MPKIIIRPLVVEDLAEIWSYIAEDSSNRADSFVDFIDGKFHELAQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDDQLNPQQ
>Mpop_5431
MGSYQLTGRADADILDVFLYGLTEFGMAQAERYQGILEHTFQLLADNPRMGRAAEVIGAGVRRHECGAHVILYEEATDGVLILAVVHSRSVRRLSL
>Rru_A0699
MPITLSDLALKDLEEIRHYTVRQWGREQWLRYYRGLVSTFERIEQSPESGRSRDLFLPGLRSISYGKHSVFYAPIAAAGGAIVILRIVHQKRHLPALTYYEDIE
>YPTS_4253
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YpAngola_B0004
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPDSF_4012
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>YPA_CD0085
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>pYV0026
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNPELGAECNWICSDIRRFQYKKHGIYYITLSNDILISRVLHQSIDIDVQDFPEHE
>NE1525
MAEYRLSPAAQRDLDGIFNYTFQQWGAAQAVRYIDILEAACTELVETSSQGQDCSYIRPGYRRRHVERHITTE
>Shewmr4_0332
MNNKRYKLSRLAQIHLQQIKDYTLQHFSESQWHKYKESLISGLQMLANNPGLGRSCNDIYPNGFFFPIGKHTAYFTKEADFILIVALLGQPQLPQNHLK
>Clim_1299
MPTIRIRPLVFDDLAEIWSYIAEDSPNRAVEFIDSIDGKFHELSQSPRIGRSRNELLPGLCSFPVGRYIIFYLIIPDGIEVVRVLHASRDIDEQLNPKQ
>RPC_0159
MARKFRLSEQAEADLNSIWRYIASADVRAANGQIDRLTEAMDLLAEAPLAGRARPEFGAQLRSFFLRGYTIFYVVTPGGIEIARVLHERMDIDPADFS

>gll1910
MSRYILNVLAAQDLDEIADRFVAASVEAGERFFEEFDRKCLQLVTFPNSGRSYGTLRTGLRGLPLGEYIIFYRVLDDGIEILRVVSGRRDLPTLFEESGS
>PBPRC0063
MSKYVLSTGAQTDLIDIRQYTLDKWGQVQWKGYFAELKSTMILLAENPAIGAKTLDLGDNYFRFPLKHHVIYYIQQPTQIVIVAVLGKSMSPQKHFEIIN
>SO_A0088
MAVTYHLTPDAQSDLIGIHRFTLAQWGATQSKTYLSGLKQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLTHLAERETI
>Sfri_3471
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRVNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_1990
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sputcn32_0191
MPVTYHLTPDAQSDLIGIHRFTLAQWGTTQSKTYLSGLRQTIQLLAETPTLGKNRPEVRMNVFSFPYSSHVIYYIQHEHQFVVFGILHKSMVPLAHLAEREII
>Sala_2662
MTAFRVSAIASRRLDEIFVYSLDTWGQEQAETYIRELFACFDRIARRELLWRAIPAEFSVDGYYCRHEHHYVYWRLLADGDVGIVTILHERMHQMDRFREDDGA
>Shewmr7_3694
MNNKRYKLSRLAQTHLLKIKDYTLQHFSESQWHKYKEGLISGLQMLANNPGLGRSCNDIYPKGFYFPIGKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>VIBHAR_p08247
MTRYKLSPAAQTDLIDIRRYTLENWGGAQWTNYFGELKQSMALLASNELIGIDMPELGTGYCRFPLKHHVIYYIRKPDHIVIAAVLGRNMSPAKHFQRQL
>CC_2756
MGRVIRTRPVSGDLDRVFRDVCENNGVKVASAQLNRIESVFHRLSAFPRLGRDRSDLRPGLRTFSVKPWQVLYRLNGEDVVILRILDGRMNLAAQLGKKT
>SbBS512_E1151
MRIIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>YEP0057
MYKLSELADEDIYNIASYTIRHFGVTQAKLYHENLAKVFELLAKNLELGAECNWICSDMRRFQYKKHGIYYITLSNDILISRVLHQSIDIDAQDFPEYEQ
>Sputcn32_0428
MNNKRYKLSRLSQTHLQQIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>ECs0462
MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Z0510
MYKLSGKAVEDFRGIYDYTLGKFGDEQADRYTDSLGTFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>Sputw3181_0282
MKNKRYKLSKLAQTHLLKIKDYTLQHFSESQWHKYKESLISGLQMLADNPGLGRSCEDIYPNGFYFPIAKHTAYFTKEDDFILIVALLGQPQLPQNHLK
>plu4785
MYKLTDQAAEDFAGIYDYTFLQFGETQADHYTEALEAFFDTLAEMPHIGREYPSVPGVMLVEFHRHTVFYTIRDTDILIASILHQQMNHPRYFQRSLSAI
>EcSMS35_0443
MYKLSGKAVEDFRGIYDYTLGKFGEEQADRYTDSLGIFLDTLSQMPEIGQDYDAIPEVKKIAFRFHTVYYVIRVDDILIARILHQLMEPRRHW
>CC_2984
MWIMSYRLSRKAEQDLIDIYVAGVGLFGVAQAERYQDTLEAAFGAIAAFPHIGRERPELRPPVRVHPCKSHIILYVLDERGALIVRVRHAGEDWVGEAGG
>Plav_1047
MLFGAAQAETYHARLEQAFDFLSANPRAARERLEITPPVRCHPLGVHIIIYLIEENDDVLILRVRHSREDWDASPI
>Npun_F4900
MSNICRFTVTASRDIEAIIDYIADNSSFNAAESLLSKINNKCERLAKFPGMGRRRDELAPNLRSFPVNDYLIFYRSIEEGVEILRVVSGYQDLEGLFLGQDED
>Rsph17025_1825
MAIRVQEAASVRLDEIYRYTRDRWGEAQAESYIKGLFAAFEQIETRGVVSRPVPAEFGAEGYFFRYERHFVYWRRLGNGDIGIVTILHERMHQIARLKDDFG
>msr1301
MGFRLSLAAEEDIIGIAEQGVRLFGAVQARQYHDELFAIFDLIAAGPRMARERLELSPPMRIHPFKAHLVVYRIEADGDVFIVRVRHGHEDWANEGTR
>SDY_2258
MPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNVGTPRPELGEGIFVLPFERHVIYFLQSPGEIIVIRILNQNQDATRHLHWS
>MAE_01370
MNQYIISTEAIRDMEQILDYLANTNINAGEKFLEEFSKKCRYLSQFPLMGRSYREIRPYLRGLPMKNYIIFYRLTEQGLEIMPFIKGERDLEAFFFENP
>RBE_1165
MMSKILPYKLTRIAESDLLQIKYYSEEVWGQDITKKYLTQIHNKCIEICNFPHIRINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKNMDITKIYSKKLT
>Nwi_3092
MSFRLSLAAEDDIIGIAEEGVRLFGPAQARQYHDDLFAILDLIAANPRMARERHELSPPMRIHPFKAHLVIYRIEEDGVLIVRVRHGHEDWVSEGAR
>PA1S_21500
MAKYRISHDAQADIIDILRFTHNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNIGKVVRPRHFVFFRVATDQVLEVVRVLHD
SMDLDHHLPQR
>gsr3505
MAQVLKTRQAERDIEDIWFYIALEDLQAADRWLEGMSAQAQLVASQPRMGRVRPELGTEIRSFAAGRYVLFYRPLPDGIELVRVLHGARDLDALFGGDL
>PMIP32
MSVLFTKKAKEHIRTIRRYSLTHWGKNVAETYINCLRITIINIIEKQPSIGIDRSDDLFTGIRSFPVESHIIYYREVENGIEVLAVLHQSQDPYIHIKRD
>Ent638_4215
MKEIELTPKAEEDLEAIWDNSFRQFGVVQADAYIGRIAAVFDVLAMHDIGTHRAELGENICSLPVEQHMVYFLSSHSVVTIIRILSQSQDTMRHEPWR
>PSHAa2625
MRKFKLSNDAKEDLRRIYQYGCKEYSQQQADTYFYAFFTLFEKLAANPYPYQSVAHIRKGYRRAVCGSDSIYYRVYETDVEIMAVLGGQDIEEWL
>PA14_60050
MAKYRISHDAQADIVDILRFTHNHFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCHSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
AMDLDQHLPQR
>EcSMS35_0978
MRTIKLMPKANEDLEGIWYYSYHHFGEPQADRYVEHLSDVLQILSNNNIGTPRPELGEGIFVLPFERHVIYFLQSPEEIIVIRILNQNQYASRHLHWS
>XF2070
MAKSYRLTPLAEADLEEIWFYTFRHWSIGQADSYHRSLVAVFEGLAAGTKLRPPFCSAGLQ
>PSPA7_4525
MAKYRISHDAQADIIDILRFTQNRFGDAARRRYQALIGAALEAVATDPQQVGSISREELGAGLRSIHLVYCRSMPNVGKVVRPRHFVFYRVATDQVLEVVRVLHD
SMELERHLARNRLR
>Mpop_4792
MRRAVFLASARADLVQILEDITRASGSLATGQIFVRQLRAQCHRLAALPGTLGRARPDLLPDIRSTPYRGYIIFFRYLGDTFEVVNILHSRRDIDDFFATDATS
>MAE_00450
MSKIIFSPSARLDLKQVNSYLAGKNPQAARILKEKIQQACKKLAKFPNLGRRRDELIPRLRSFPVEDYLIFYFPLENGIEIARVVSGYRDLDAMFDVD
>A1I_01455
MMSKILPYKLTRIAESDLLQIKYYSEEIWGQDITKKYLTQIHNKCIEICNFPHIGINRPEIFESVKSIAINSHMILYALVDNKIEILRVLHKTWI
>VC0395_0937
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0385
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>VCA0311
MSVYLNMQNKQYKLSQLAQEHLLKIKHYTIENFAEAQWQKYKSTLLSGFQTLADNPGLGKSCEDIYQNGFYFPVGKHMAYYTKEANFILIVAVLGQSQLPQKHLK
QSRFVS
>CJA_3710
MKIEWTTKALMDLARLYDFLAPVNQAVAARLVQQLTQAPNKLLANPRIGEKLDEFLPREVRRLLIGTYEIRYEIGNNTIYLLRIWHTREAR
>RPD_0573
MAHNRRISPKARVDLDAIWSFIAADSEKAADAMIEQITAAFAMLTDNPNAGRLRPEISAAVRSFPVKRYLIFYVAEASGIKIIRVLHGRQDRSRQDLAP

>Rsph17029_3497
MAIRVQETASVRLDEIYRYTRDRWGTAQAETYITGLFAAFEQIDTRGVMSRPVPAEFGVEGYSFRYERHFVYWRRLSNADIGIVTILHERMHQMDRFKQDFG
>PSEEN3278
MVRISYHARTDIVDILRYTEVKFGAAARGRYQDLLQAAFRAIACEPGRVGSAARDELSPGLRSLHLFFCRLEVTSQRVVRPRHIVFYRAAAGEVIEIVRILHDVM
EVSSHLEHLHQ
>STM2954.1n
MVKLTPKASEDLENIWHYGWQHFGEIQADRYINHLSEIFSIMSANNIGTPRPELGEYIYALPFKRHIIYFIQSVTEVIVIRILSQNQDAGKHVNWL
>Saro_2087
MRLEQSRLAKDDLESILEYGLLYFGTERVLAYLDHIESRFQQLLAYPRSGRIEQDLPGTVYSTSCEAHRIYYEPTDDAVVIVRVLQKSMDVSRWIG
>PD0960
MTGYILTAAAETDLRSIIRYTRKQWGDAQMRRYIATLEQDMASLAAGRGVFRNMSVLFPALRMGRCEHHYVFCLPREGAPALIVAIFHERMDLMTRLADRLK
>RB10182
MSRYLLSHSANANLDEIAGDASNAVAILEALHNTFQVLANHPGVGTLREDLLPGIRVFSPPRPANNYVIFFYPISGGIEVAAVIHGSRDWISMFTDGFRPKKS
>XF2032
MTGYVLTEAAESDLRGIVRYTRKQWGNAQVRHYIATLERGIASLAEGRGAFNDMSSLFPALRMGRYEHHYVFCLPREEAPALIVAIFHERMDLMTRLADRLK
>Pfl01_3062
MPQYRISNAARVDIVDILRLSQTQFGDQARQRYQALILAALQALADTPYRIGSHERGEVAPGLRSYHLSYSRQQAKHPHGTVKSPRHVVFYRVANDEVIEVVRLL
HDAMDVQLHLADD
>PA1S_21315
MELKWTNKGLSDLTRLYEFLAAVNRSAAARTVQQLTSAPTKLLANPRIGERLEEFDPRDVRRILVGHYEMRYEIAGSTIYLLRLWHTREDR
>Mmar10_0699
MKIEWTARASGDLARLFDFLEPVAPDAAVRVVQEIARAPNRLLYHPRIGEKLDVYEPREVRRIIIGHYEMRYEIAAGTLFILRVWHSREDRDFGPEA
>Mpop_0593
MGEVVRRPRARRDLLDLWDYIANDDEMAADGFLDRIEGALAMLSDNPKAGRHRPELVTTLRSFPVESIVLFYFPMQTGIELVRVLSGYRDIGRDDFDA
>Rru_A3239
MPRLVILPAARLDLIEIGDFIALDNPERAASFVAEIEARMIQAADRPASFPTRDELHEGLRSARHGRYLIFFIEDGDEVRVVRVLHGARDFQRIMG
>plu0251
MSVRFTNKAREHIRAIKLYSMRRWGANVAEAYSTSLRVTMTEILDRHPSPGRDRSEDLYFGVLSFPVESHIIYYREVPTGIEVLAVLHQTQDPHNHLYPMDFKMH
RDGKGANPREHRERCDRGERVQPTKRQLER
>pRL120082
MAPRVLILPTALDNYRLAVRETARKWSTEQAKAYSRLLRAGFEGIPEAYARVRIKKDERVGNSLFRLYKIEHHYAVYIVVDDSTFVIAAVLHERMDIPAQLRTIE
RLTDREYAALMGHPRPKS
>SMc00694
MSRELVFTPAALADLEETFWFVAADNPRRARSYVAEIEQACRNLCETPLMGRGRPDLRPNLFIFPLWRRVLIAYELPDNRVDILRVFSGGQDYEAIMSGE
>Dole_2119
MKIVWSPLAVERASEIADYISQDKPAAATNWINVLFSKVDQLRANPEIGRIVPEINDRQFRELIYGNYRIIYHIGAKQISILTVRHGRQILPTDEIKA
>ECs2281
MLPVLWLESADTDLDDITSYIARFDIDAAERLWQRLRGCVLPLSEHPYLYPPSDRVPGLREIVAHPNYIILYRVTTSSVEVVNVIHARRQFP
>Amet_0747
MYKILRTDKAEDQLRDIIFYIADDSGDVEVALKYLDKIEAAINRLQEFPKSGSIPRYSILKKRGYRVVTVEKHLVFYRIDEEEETVIIYAIVDGRREYRNLI
>plu4594
MPYHVMLTKNAEADLEDIYDYIVENDSSEKADYVLDQLLKTADSLANFPEKGNYPKELQALGIRDFRQTFFKPYRVIYQITGKQVVIFVIADGRRDMQTLLTHRL
LSVSA
>azo3077
MTRIELAPEVADDFEWILDHLDHYRANDPAARIREIIDAIGVLASNPLIGRPTDNDNRELVIGRRSHGYLALYRYVAGIDTVFVLAVRSQREAGFER
>Mpop_5181
MAHDLVFHRLARADLFEIYDYIEERSGAARAGGFLDRIEAACRGLTEFPGKGTPRDDVPGGLRTWALERRVLIAYRMTPGHIEILRVLYAGRDFRADEIPH
>Gura_1318
MKIKWTNEALEQLIEIEEFISKDSPERTAVFVDQLIEHAEDSLPDNPRMGRTVLEIANPDIRELIFRKYRIVYRLTALSGLVCASRLLSPIRGVRM
>RSc3224
MRLAITPLAEQDLESIADYIAQDNPARAVTFVRDLREQCQRLVMNPPGYRLRPELGDDIRSCAYGRYVIFFVAAPDEVIVIRILHGARDLPAVFHADEP
>RPE_4765
MKCVISSTAEADLLAIHAYLSERNPAAADRIITRFFHRFDELCEFPFLGPDRSELRASLRGLRVDGFVAFYIIEPNQIIVVRVLDGRMDIEGKFAE
>RPC_2216
MGAVRYTRKARQDLIDIWRHIAAESPATADQCLDRIEARCKQLAAFPEIGRERRDIAPDARMLVVERWIAIYRVVEQGVQIVRIVDGARDLSRLALPQK
>NJAUSS_0644
MRLTKQASDDLTAIYRFIAIELQSPLTADKNIHLFEQSIKSLSTFPERCPILEGFESEGIVIRKLIVKNYVIFYRFVGEVVTVLRVLHGTSNIDALLRDIAEDND
VN
>Swol_1971
MKYKILRTDKAEEQLREIIFYIADDSGDIDIALGYLDKIETAINRLQEFPESGSVPRYSILKKQGYRVVIVERHLVFYKINEADKLVIIYAIVDGRREYRNLI
>Spro_0056
MKIKWLRKAASNFDDAYDWFYQESPQVATQFAREVFRLVNLLATNPALGRAGRVMGTREIVMKTFPYLIPYRIKNNEIHILRVFHMRRHPPKKSW
>RF_0787
MVKCDFTNSAKRDIEDIADYSLKNWGRQPTLKYLDEIYRKTLDLSINPNIGVLRSDIYPNLLSFPIRKQLDLLHNVANKEEFEGDTSLRTAAYT
>msr9191
MIPLRNGAKRRPSATYLHKHLQTLSETPALWRKLPGNLAIPADLKLDAYFSHHGRHYVFFRKLSGDRIGVISILHDRMDVPVRLAEDLQALQSRSEDR
>Amet_0378
MHKLRINPMATEDLIEIRDYIMKELENPTAAVNVVRKIIESYEQLKEFPMLGVDLSTKVNVQTDFRYLVSGSYIVFYQADNESVSIYRILYARRDYLKILFPNEV
DID
>Veis_4341
MLPIFWLETADTDLADITQYIGLRDINAAERMWHRLRNCVLPLSEHPYLYRISERVPGLREIVAHPNYLVLYRVTATRIEVVNVVHTRREFPQHSNDGP
>Mext_4629
MAAEVVWSRLARTDLLDIYVSIGSHDPAAAERLYDRIEERARQLALQPRMGPRRPEIRPSTRILVETPFLILYETIPDTDDAPVREVEIVRVLDGRRDVSVVLHA
GENSAEW
>Ppha_0144
MERLHAFLYEINPDAAARAARAILEGAGFLESMPDIGRPMDDDTGRREWFISFSAGAFVLRYMWNKNDTIVIVRVWHSKEKRT
>RHE_CH02732
MTTTYRLTRTTDAVLSGIDEYANLHFGEAQADAYLLDWDRIFILLSRVPAMGDECDALGAGLRRLHIAFYREIPNGILVIDIIGADRLAEGHLQSNRRSGPTAPH
AP
>PSHAa1041
MAKFIAFDKPSAADKWVNDIFDRTDLLGTQPELGREVPELLGSRYRELIFGSYRIIYKVEHEIKILTLRNSRQLLSLSDIEQ
>TBFG_11272
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>MRA_1255
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>BCG_1306c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Mb1278c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>Rv1246c
MSDDHPYHVAITATAARDLQRLPEKIAAACVEFVFGPLLNNPHRLGKPLRNDLEGLHSARRGDYRVVYAIDDGHHRVEIIHIARRSASYRMNPCRPR
>TBFG_12882
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>MRA_2891

MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>BCG_2888
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Mb2891
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Rv2866
MPYTVRFTTTARRDLHKLPPRILAAVVEFAFGDLSREPLRVGKPLRRELAGTFSARRGTYRLLYRIDDEHTTVVILRVDHRADIYRR
>Lxx22677
MSWDVQFAPAAIRGLDRLPPRVVAAVVEFVTVTLPGNPYRMSKPLQGDLEGYYSARRGDYRVLFSLDEDRRVLLVGRIAHRADVYRPR
>Francci3_2679
MTGAAGPYRLEITGPAARALAGRIPEKVATAVHEFITTTLLENPHRLGKRLLYPPYAGTWSARRGMYRVLYEIDEENRIVLVTAVEHRADAYRGR
>Noca_2680
MSAPDEGGTGYEVVFTRGARRALEWDLPAAVAMAAFEFIRGPLREAPRRVGKPLLEPLTPLWSARRGEYRILYRILDRRLVIAVVTIAHRRDAYGRRE
>TDE0735
MKVVLTETFKKQLKKLDATISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGDYRILCRICDDKLIITVIEIGHRSTVYR
>Suden_1966
MSYKLLIDDKVIKDLKQIDKLWQKKIIEVIKTKLVENPHLGKPLVGNLSPYYRLRVFDYRVIYEINDDEVVVIVIKIGHRKDIYK
>TDE1978
MKVILTETFKKQLKKLDAAISKRVLDYLEQIELLDNPRSRGKALTSNLSGLWRYRVGNYRILCRIHDDRLIITVIEIGHRSTVYR
>Paes_2094
MVWTIEFAATAEKELSKLDKSAAKRILKFLKERVATDPRSSGKALRGDHAGLWRYRIGDYRVICEFRDQTVSVLVVRIGHRKEVYR
>Mhun_1873
MSKFTLLISKGAERDLAQLPKFARRSLEVAFAELELLENPREKLGPLKGRAKGLYSLRTGEYRAILEIFDNKLMLLVIEAGHRKTIYRKYQS
>WD0126
MKYDIVYSKNFTERDFLNLPKTIRSRITKAINERFTTDPIKFGKPLRGRLKGYRRIRVSDYRIIYTVNIAKHKVFVATAGHRDTIYEKAT
>FN0211
MKYDVEYSKTAMNTIKKMDSSTSKLIRTWIEKNLINTENPRIKGKALTGDLKGLWRYRIGDYRILAEIQDDKIVILILDIGHRSKIYL
>WD0122
MGLERYKVKSLKSVVEKDLPNLPKIIRLRVQKAIKERLTVDPINFGEPLHHNLKGCRRLRVGDYRVIYRVNQLDHIVTITEIGHRDDIYKK
>WD0600
MGIIYQIGYLEGVDTEDLPSLPKTIRLRVQKAIEKRLTIIPDKVGEALSHKWVGYFRLRVGDYRVIYLIDNSEHMVKIAAIGHRKEIYKRSPE
>Cpha266_1376
MVWKIEFASSAEKELARLDKSAARRIVKYLRERVAIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEVYR
>RF_0898
MFKVIWENKARAELAELPYPLKILDKVESYLAQNPIVLGKPLKGEYKNLYRYRFGNYRIIYSVSIEKSTVTVIKIGHRANIY
>asl4561
MIYQIEITTRAAKQLKKLSEDIKLKIEEKIQELSNNPRSNDVVKLEGEEDTYRIRVGNYRILYEIKDDLLIVKVVKISHRRDVYRRK
>Clim_1338
MVWKIEFAASAEKELARRDKSAARRIVKYLRERVEIDPRASGKSLRGDHAGLWRYRIGDYRVICEILDEKVSVLVVRVGHRKEIYR
>Gura_2099
MTYRIELTKTAERDLLAVPKPVLKRLDACILGLADDPLPPGVKKLKNSDGLYRVRVSDYRIIYRIEQEILTVLVVKIGHRREVYR
>Gmet_A3569
MVWKVEIDPAARRELKKLDPQISGRVLKFLFERVARLDDPRSIGEALKGSRFGDFWKYRVGDYRIITSIEDEALVILVVRVGNRREVYER
>Rru_A3196
MTWKIEFDPSALRELDKLDPQIAARVLRFLRDRVAVLENPRSLGEALKGPRLGAFWKYRVSDYRIIAHIEDDTLRILVLRIGNRREVYR
>WD0404
MKTSGNKTYTIKFLKNVIEKEIPALPAKIKLMVQEAIKKRLTVDPFNLGKPLCHSFRGQYRIRVSNYRIIYIINHSERKVLITAVGYRKNIYKHRRLHN
>Plav_1617
MAWTIDYTDDAINQLRKLDKQSARRIVEYMDERIAGAANPRSFGKPLSGPLGQFWRYRVGDFRIICDIEDEVLRVLVVRIGDRKDVYRKAKR
>SCATT_39270
MSEYRTVFRPEAQAELRKVPRDMALRILAKLTELESDPLGFNTTALVSQPDRRRLRVGDYRVIYTIDNGELVVWVVHVGHRSTVYGA
>XfasM23_1261
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>PD1184
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCAIQDSVLRVLVVRVGHRGEIYR
>WD0269
MKTSGNKRYTIKYLKHVLKRNLPSLPEAIKPKIKDAIREYLATDPIGNGVLLRNRLKGHRRIRVDDYRVVYRVNTAERKVTIVSIGHRDNIYKQAILDLLKH
>FN0497
MKKYEVKFSEAAIKELKKLDKPTATMIKLWVIQNLENTINPRQHGKSLTANYSGKWRYRVGNYRLLAEIYDDEILILIFKVAHRSIVYKK
>RF_1272
MEMNYKICYLEEVTNKHIPMLSSNAKTLIKCAIEERLMFDPIAFGKPLRYSLKGHRRFRISDYRIIYRIEQETSTVIIIAIKHRKEIYQEFI
>Pcar_1777
MKMSGGRVAFSVIYHPDVKGRDIPKINGDVRVRIKKAIETRLMVAPQEYGEPLRKTLKGYWKLRVGDYRIVFKIDGDEILILGICHRKGVYPLMESRQ
>LPC_2245
MMTPGSKLYQIEYIEDVVKNDMPSLSTSAKKLIKKAIEERLMADPIGFGKPLRYSLKGHRRLRVSDYRIVYRIEAETNTVVIIAIKHRKEVYDDF
>MGAS10270_Spy1927
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>M28_Spy1840
MRSWKMTYKLVVSDEVKKQLKKMDKHVGLMLAKDMKKRLDGLNNPRQFGKALTGQYKGLWRYRVGNYRVICDIVDNEMIILALEVGHRKEIYKR
>SSU98_0616
MAYKLVLSDDALKQLKKMDRHVGMMLAKDLKKRLDGLENPRQFGKALVGDYKGLWRYRVGNYRVICDIIDNKMVILALEIGHRKEIYKK
>Xfasm12_1292
MAWTIDYTDTAKQQLRKLDKHMARRIVDFMDERIAGLENPRSSGKALTGPLGGFWRYRVGDFRVVCDIQDSVLRVLVVRVGHRGEIYR
>pTC-F14_p08
MAWRIEFDDKAKKDLAALDKSVAKRITAFLRERVAHLDDPRSIGEALKGSKLGDFWKYRVGDWRIIASIEDEALRILVVRIGNRREVYRK
>Sbal_4390
MAWTIDYSERALKSLKKMDKQNAKRILDFLEQRIAILDDPRTSGKPLKGDLGIFWRYRVGDYRVLCEIQDSKLVILTALIGHRKEIYE
>Npun_R1755
MPSVEMSERYSLRIAKTAEKDLLDLQAKLYKQVVSKILSLQGNSRPQDCKALKGYEGGYRVDQGEYRILYTIDEESKLIDVFRVGKRNDGEVYKNL
>RSal33209_0100
MTYRIAYTPRVIKDLKKLDKQTVRRVKDFFDRLNRDNPRSLGKALVGQDFWRYRIGDHRILVAIQDDVLTVLVVKVAHRREVYKER
>MA1694
MTYQVVLSPDFEKETKIFFKKDPVLYGRFKKTVNSILENPECGKPLRNVLKGLRRVHIGHFVLIYEIDNTNETITFLKFSPHDKAYK
>Noc_0433
MYAIVVHRRAARYLRKLPQDQQVKIKHVLAQMKNGPLGLSGIKSMVGDWAGYRRIRVGNVRIIFWIDELKNVVYVDHIGPRGDVYKDKT
>RSc3279
MNAIHWTAWAARQLRKLDRQHQRVLVEAVGQLEAMPHCRQVRALREHRYGYRLRVGDYRVLSDWDDGIRIVDIQEVSKRDERTYRH
>Ppha_0985
MYKIIFTKEAQKALLRLPGSTAVQVCQKLEQLAADSYAPIANAKKLQNRSGYRLRIGDWRVIYEIQNDKLVVLVLKIAQRGEVYR
>asl2100
MSDRYTLRIARTAEKDLLDLQPKQFKQVVSKILSLQGTPRPQDCKALKGYEGGYRVDQGEYRILYTIDDETQLVDVFRVGKRNDDEVYQNL
>BH07080
MAWTIRYEKKALSFLKKCDKKEARRIVDFLDQHVAPLEDVRVIGKPLKGQFSGLWRYRVGDYRILCELYDKELVVLVLAVGHRKNIYKG
>TK0965

MSYELILSGKSEKALKKAPPEDRKRIVSALFKLKENPWAMQYKKLRGYPFYRVRVGDWRIIYTVDDEARIVYVVRLGKREGVYDSL
>SSU05_0886
MKTCYKLVPTSRFIKQLKKLDKFTQKQITNYLSSHVTDNPRQYGKALTANRSGQWRYRIGNYRVIVNIEDDKLIVTSIKMRNGSI
>Ppha_1272
MAYSVGYKKSVQCDLRQLSRLDAQRIYDQIEQELVKNPKSNPLLKGRFVGLRKYRVGDYRVIYCVLDEEVIILRIAHRRDVYKRDI
>Mhun_0373
MEPYGISMAPSALKNYKKFPEKLREKIKSEALHIARNPYIYEELSEPLKGIRSYHFTFNSTQYRIAYQINESSREIEILLVKTRENFYDKLFRTFR
>ECA0674
MAHIVWTGKAVKDLRKLPANDQKAIQANVNSLGDYPATKSKPLDITKLTDRGSEYRLRVGNYRVLFEIQKGEPVIIEIQRVLRRTSTTY
>HY04AAS1_1546
MWKVVITKQAKKDLVNIYRAGLKSKFEKLVEEIKKDPYTSQCEKLIGDLEGYYSCRINRKHRLVYEIDDEQKIVKIVSVWNHY
>AM1_2740
MEFHIELTELALEMIGAIKDRREQQGIIHRIQKLKREPLQQGKPLTGDLKGLYSVRAVGQRYRVVYQVKSEKIIVVVVGVGRRKEGDKKDVYTLLKKLLERPDG
>Francci3_1442
MTRLAAEPRPSGARALTGQPTGILRIRVGEYRVIYQVDHTRVLVTIVHVAHRREVYRHL
>Memar_0664
MIWRLILMPVAERVLNNIPDPDAGRIKEELYALADEPYPRFHVKKLKGHQNSPLYSLRVGQYRIILVIEDNVMVITVIEIGNRSKIYRKY
>Exig_1264
MTTYTVEFERGAQKSLKKMDPQQSRIIMSWIKKNLVGTDDPRRHGKGLISNRSGEWRYRIGDYRLIADIQDDKVLILILEIGHRRDIYK
>RF_1286
MQYKLSFSKTALKNLLKISTNKRKAILEKLEQLKLNPYKENNNIKKLIGYDGYRLRIGDYRVIYRINKGKLEILVINVDVRG
>AAur_0707
MSYAVQVAPAAVRQLRKIPPEARRRIQAAIEILAETPRPPGAKKLSGSSGDWRVRTGDYRIIYEIRDAQLIVLVVAMGHRRDIYQH
>FMG_1568
MRVIYSEKSLKSLKKLDKPIQKMIIHYMEKVGQLEEPRARGKALSANLRGFWRYRVSNYRIICEIDDDKLIICVVEVDHRKNIYKC
>PHS013
MTYRVKIHKQVVKALQSLPKAHYRRFLEFRDILEYEPVPREKFDVIKLEGTGDLDLYRARLGDYRVIYSVNWKDKVIKILKLKPRGRAYK
>A1I_07635
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>RBE_1375
MEYTLCFSKTSLKNLVKIFANKRKVILEKLEQLRLDPYKTNNNIKKLIGYDAYRLRVGDYRVIYKINQGRLEILVINIDVRGEVYK
>Suden_0812
MVYNIQYDPKALKQLKKLDKSIALLILDGIEEFASNPVLTKIKKLKTPFDGAYRLRICDYRVVFYQEDNLMLISKIAHRKDVYI
>Nmul_A0326
MTYRISFNPAAQRDLDKMDPQARQRLLKYLNGRIVLLEDARCLGEPLLASQFFGYWRYRAGDYRITCDIQDEELHVLIVKGGNRRVCGVG
>spr1103
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>DNO_0273
MIHSFADKDTYDFFHGTICKRFESFASVARRKLTMLDNAATLEFLRSPPANRLESLKGDRAGQYSIRINDRYRICFRWDSGAHDVEIVDYH
>SPH_1337
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SPD_1081
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>SP_1223
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>Atu0674
MIWTIEYHTLVQKEMRKINPEVRRRIRSFLHERLAALDDPRQIGATLQGSELGNFWRYRVGDYRIICDIQDQKLVVLVVEIGHRREIYR
>MA0376
MYRIIYSPAAKRDLKRLPADVQDRVHDALEEIADDPYAHVKKLKTPYNSPIFAYRVGKYRVIMSIHDFELIILVLEVGDRKNIYRKF
>SPCG_1079
MNNLYKLVPTRRFIKQLKKLDRYTQKLITNYLQINVLEDPRRHGKALVGNRVGQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>CKO_pCKO2p07161
MDKQNARRIVDFMSLRIAVAADPRQSGKPLKGELGEFWRYRVGDYRVLCEIRDDELVILAATIGHRREVYD
>AF2342
MFRVVVHRKATQELKRLKKAHLKKFGVLLETLKTDPIPWKRFDVKKIEGEENTYRIRIGDFRVIYFLDKPTKTVHILKVERRGKVYD
>USA300HOU_2446
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAOUHSC_02756
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAS2348
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SACOL2464
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>MW2380
MSNYTVKIKNSAKSDLKKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>Rpal_3615
MIQSFRGKFARAILHDRKAPKGLPADLLGTARRKLVQLNAAAALADLAIPPGNRLEALRGDLQGLHSIRINDQWRIVFRWKDTGPEDVEIVDYH

>SAHV_2440
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SA2245
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>SAV2456
MSNYTVKIKNSAKSDLRKIKHSYLKKSFLEIVETLKNDPYKITQSFEKLEPKYLERYSRRINHQHRVVYTVDDRNKEVLILSAWSHYD
>CJA_2979
MTRSLAWTDAAWSDYLYWQGQDRKTLKRINQLIRETLREPFQGIGKPEPLKESLAGFWSRRIDDTHRLVYAVDDNHLTIIACRYHYER
>pc0990
MKYELFVNPRVEKALSKIDKHMALKIRNNIRSLAANPRPLGVKKIKGNDNAYRIRVGDYRIIYEIYDSKILILIVNVGHRKEVYE
>SAHV_2391
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SA2195
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAV2407
MARLNITFSPQAFEDYKYFQQNNKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAR2497
MVRLNITFSPQAFEDYKYFQQNDKKMVKKINELIKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>USA300HOU_2387
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAOUHSC_02691
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAS2298
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SACOL2404
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY

>MW2329
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALEGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>SAB2287c
MARLNITFSPQAFEDYKYFQQNDKKMVKKINELLKSIDRNGALKGIGKPEKLKSNLTGYYSRRINHEHRLVYTVDDNHIKIASCKYHY
>TK0791
MSFENRILISKRALKELKNVPESQRDIIKDRISKLAFFPLVKLDVQKLKGYDNVYRLRVGEYRVIFEYNKEERIVMILKVGKRGNVYS
>MM_2605
MSFDVKLHPDAVKFLVSLNPETKERLKSGIKNLEMDPFKSRPHADIKKLKGTKKRNDLYRLRIGDYRMIYSVEENTIFILEIIPRERGYDWL
>SPG_1114
MYKLVPTRRFIKQLKKLDRYTQKLITNYLQTNVLEDPRRHGKALVGNRVXQWRYRIGNYRVIVQIVDDELVVATLEVGHRRDIY
>MAE_18810
MNYRVIIPKPIQKQLNNLPKQQRERLITAIRLLTDTPRPSGVKKLKGYDETYRIRIGDYRIIYKIQDQEMLIIILSSIHRKDAY