Functional and Structural Insights into Novel Bacteriophage Defence Islands

PICTON, DAVID,MARK

Functional and Structural Insights into Novel Bacteriophage Defence Islands

David Mark Picton

BSc Hons, MPhil

A thesis submitted for the degree of Doctor of Philosophy

March 2021

Department of Biosciences

Durham University

Abstract

Bacteriophages are the most abundant organisms on the planet and are a major driving force in bacterial evolution. As obligate intracellular parasites, phages are reliant on their bacterial host for propagation, but bacteria have evolved means to prevent phage infections. Bacteriophage exclusion (BREX) is a novel phage-resistance system that confers resistance to a wide array of phages, functioning independently of restriction-modification, CRISPR-Cas and abortive infection mechanisms. BREX loci are present in ~10% of bacterial and archaeal genomes, including pathogenic strains such as non-typhoidal invasive *Salmonella enterica* and multidrug resistant *Escherichia fergusonii*.

Whilst investigating the mechanism of BREX in *E. fergusonii,* a putative endonuclease was discovered, clustered within the BREX locus. This enzyme, BrxU, was biochemically and structurally characterised, and shown to be a standalone phage defence system that targets modified phage genomes. It became clear that the BREX and BrxU phage defence systems were organised into a phage defence island, constituting a bacterial immune system capable of resisting multiple phage types.

Both systems detailed in this thesis represent novel antiphage mechanisms with potential for biotechnological application. The BrxU endonuclease structure has been solved to 2.12 Å and reveals insight into key protein domains implicated in type IV restriction enzymes. BrxU has been observed to utilise a range of nucleotide and metal cofactors and confers extensive protection to its bacterial host against phage infection.

Table of contents

List of Figures

List of Tables

List of Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Å	Angstrom
aa	amino acid(s)
abortive infection	abortive infection
ADP	adenosine diphosphate
AMP	adenosine monophosphate
AMP PnP	adenylyl-imidodiphosphate
Ap(R)	ampicillin (resistance)
APS	ammonium persulphate
ATP	adenosine triphosphate
base pairs	base pairs
BBSRC	Biotechnology and Biosciences Research Council
BREX	bacteriophage exclusion
c	complementary strand
Cm(R)	chloramphenicol (resistance)
CRISPR	clustered regularly interspaced short palindromic repeat
CTD	C-terminal domain
DNA	deoxyribonucleic acid
DMF	dimethyl formamide
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EOP	efficiency of plating
EtBr	ethidium bromide
EtOH	ethanol
FPLC	fast protein liquid chromatography
FW	forward
g	grams
glc	glycosyl

glu	glucose
HEPES	N-(2-hydroxylethyl)-piperazine-N′-(2-ethanesulfonic acid)
IMAC	Immobilised Metal Affinity Chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
L-ara	L-arabinose
LB(A)	Luria broth (agar)
LIC	ligation independent cloning
kb	kilobase pair(s)
kDa	kiloDalton(s)
Km(R)	kanamycin (resistance)
KO	knock out
mAU	milli-arbitrary unit
MES	2-(N-morpholino) ethanesulfonic acid
min	minutes
mM	millimolar
MME	monomethyl ether
MOI	multiplicity of infection
MQ	Milli Q water
MS	mass spectrometry
NWL	Northumbrian Water Limited
NTD	N-terminal domain
NTP	nucleotide triphosphate
OD600	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PDB	protein data bank
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming unit(s)
phage(s)	bacteriophage(s)

RACE	rapid amplification of cDNA ends
RBS	ribosome binding site
REase	restriction endonuclease
RM	restriction modification
RNA	ribonucleic acid
rpm	revolutions per minute
RV	reverse
S200i	superdex 200 Increase GL 5/150
SAD	single wavelength anomalous dispersion
SD	Shine-Dalgarno
SDM	site-directed mutagenesis
seq	sequencing
SENP	Sentrin protease
SDS	sodium dodecyl sulphate
StDv	standard deviation
SeM	selenomethionine
Sm(R)	streptomycin (resistance)
Sp(R)	spectinomycin (resistance)
SUMO	small ubiqutin-like modifier
TA	toxin-antitoxin
TAE	tris-acetate-EDTA
TEMED	N,N,N′,N′-tetramethylethylenediamine
Tc(R/S)	tetracycline (resistance/susceptibility)
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
(w)HTH	(winged) helix-turn-helix
WT	wild type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

List of Publications

Research performed within the duration of this study has contributed to two articles published in peer reviewed journals, which are shown at the end of this thesis.

1. Rodwell, Ella V., Nicolas Wenner, Caisey V. Pulford, Yueyi Cai, Arthur Bowers-Barnard, Alison Beckett, Jonathan Rigby, David M. Picton, Tim R. Blower, Nicholas A. Feasey, Jay C. D. Hinton, and Blanca M. Perez-Sepulveda. 2021. 'Isolation and characterisation of bacteriophages with activity against invasive non-typhoidal *Salmonella* causing bloodstream infection in Malawi', *Viruses*, 13: 478.

2. Cross, J. M., T. R. Blower, A. D. H. Kingdon, R. Pal, D. M. Picton, and J. W. Walton. 2020. 'Anticancer Ruthenium Complexes with HDAC Isoform Selectivity', Molecules, 25.

Declaration

The work presented in this thesis was carried out at Durham University between September 2016 and March 2021. The work is my own original research unless otherwise stated or indicated by citation. This work has not been submitted for any other qualification.

Statement of Copyright

Acknowledgements

Writing fifty thousand words detailing the intricate workings of bacteriophage resistance systems was relatively easy compared to finding the words to express my gratitude for the tutelage and friendship of Dr Tim Blower. It has been an honour to be Dr Blower's first PhD student and I am looking forward to continuing working together. I would like to thank Durham University and the BBSRC for funding this project and giving me an opportunity to chase a dream.

I am thankful for the support and friendship of fellow PhD students Ben Usher and Izaak Beck, who have been an infinite source of support and laughter since the beginning of my studies at Durham University. The fellowship in Digbeth will never be forgotten or spoken of ever again. More recently, it has been a pleasure to welcome Dr Abbie Kelly and Sam Went to the Blower lab, who have been constant sources of support and knowledge. I would like to thank Dr Vicki Linthwaite and Dr Tessa Young for their expertise and aid in experimental design. Vicki's troubleshooting and advice has aided me throughout my PhD, and Tessa was instrumental in the initial characterisation of BrxU. I would like to thank Dr Gary Sharples, Dr Paul Denny, Professor Ehmke Pohl, Professor Martin Cann and all of their staff and students for their support over the years. Professor Nigel Robinson, Dr Peter Chivers and Dr Karrera Djoko have been invaluable sources of knowledge in their respective fields and have helped the experimental procedures used in this study. Thank you to Ian Edwards for all his technical assistance. It has been a pleasure to work alongside Dr James Walton and his lab, who have made us biology outcasts feel so welcome. CG234 has been a wonderful environment to work in and I am excited to continue working here.

I am particularly indebted to Professor Ehmke Pohl, Dr Stefi Pohl and members of their lab for their help for all things structural. By extension of this, Dr Arnaud Basle of Newcastle University has been a fountain of crystallographic wisdom and without his support, structures would likely be absent from this report. I would like to extend my gratitude to DLS and the beamline scientists who have taken the time to help strategise X-ray data collection. As a lab, we are indebted to Dr Rick Morgan of NEB for his assistance with PacBio, and to Dr Darren Smith of Northumbria University.

Dr Rachel Parsons has been an inspiration throughout my time at university, as both a scientist and all-round legend. It is an honour to be a Team Awesome alumnus and I credit her for inspiring my academic career. I cannot thank Dr Rodney Johnson enough for all of his support and for showing me the Bermuda triangle. Being in the Rad van 300 miles away from shore is something I will never forget. I would like to thank the Bermuda Institute of Ocean Sciences and all of their staff for the opportunities presented to me.

I am forever grateful for crossing paths with Dr Lauren Drage in 2014. Her camaraderie and mentorship has been invaluable, and I hope one day we can revisit the Mile Castle, if she is allowed back in. It was in the lab of Dr Philip Aldridge in that we became friends, who has been unrelenting in his enthusiasm and support during my time in Newcastle. Check 1, 2 to my boy Carl MC who has been the greatest friend I could've asked for since he first armbarred me in 2012. A special thank you to Dr Richard Jones and Zain Kazi, who I will always be grateful to. I would like to apologise to anyone I have missed out during these acknowledgements. My hands hurt from typing and I have a deadline at midnight.

Finally, I would like to thank my family for the life I have been given. To my girlfriend, Sam, thank you for your love and patience. Without your support, none of this would be possible. To my sister, Beth, who has been my best friend since the day she realised I existed, Mugatu would be proud of you. To my dad, Dr Picton Sr., one day I'll be able to bench more than you and then you'll be in trouble. Until then, I'll be in your debt for every lesson you have given me. Mum, you have always been a beacon of love and support and I am eternally grateful to be your son. I'm sorry for moving so far away, but you kept hoovering too early on Saturdays.

Hoffwn ggflwyno'r gwaith hwn er cof am fy nhadau Mel a fy mamgu June. Bu'n ffordd hir ac yn anodd ond byddaf yn ddiolchgar iddynt am byth am ddisglerio I lawr arnai.

"Mesur ddwywaith, torri unwaith."

Chapter 1: Introduction

1.1 Introduction to Bacteriophage Biology

Bacteriophages (phages) are viruses that are natural predators of bacteria, and outnumber their prey by a factor of 10 (Stern and Sorek 2011a). Phages are the most ubiquitous set of organisms on the planet, and are the major driving force in the evolution of bacterial populations (Sulakvelidze 2011; Pingoud, Wilson, and Wende 2014). As obligate intracellular parasites, phages are reliant on their bacterial host for propagation, using the host machinery to replicate their genome and form progeny (Salmond and Fineran 2015). Phage genomes can be RNA based such as that of MS2, which was the first genome to ever be sequenced (Fiers et al. 1976). However, the majority of studies performed on phage resistance focus on dsDNA phages, therefore this study has excluded RNA phages, as host interactions with phage DNA form a substantial chapter of phage-resistance literature. Occasionally, phage encoded DNA can be recombined into the host genome, which may confer the host with an evolutionary advantage such as antibiotic resistance, or the acquisition of an exotoxin (Safa, Nair, and Kong 2010). However, the majority of phage interactions with the host result in the propagation of progeny phages, and the subsequent lysis of the host. Constant interactions between bacterial hosts and phages has led to the co-evolution of mechanisms for phage propagation, and bacterial phage-resistance mechanisms (Labrie, Samson, and Moineau 2010). The use of phages as a means to treat bacterial infections, known as “phage therapy”, has become a re-emerging field after being somewhat abandoned following the discovery of antibiotics in the late 1920s, which provided much cheaper, more efficacious options for treating bacterial infections. Phage-host interactions were poorly understood during this period, and even before phages could be harnessed for their bactericidal effects, researchers were already becoming aware of bacterial phage-resistance arising in secondary cultures (Summers 2001).

1.1.1 Phage therapy and the discovery of targeted antimicrobial agents

The majority of interest in phage biology following their discovery was to generate methods of utilising phages as antimicrobial agents. Parallels are often drawn between the effects of the simple chemical antibiotics and biological machines that phages represent due to their shared suppression of bacteria. Much of the interest in phage biology stems from the highly selective nature of phage-host interactions and the potential to selectively target a bacterial species for removal. Phage-based therapies are being developed and are currently used to treat bacterial infections that have not responded to antibiotic treatments.

Penicillin was discovered in 1928 by Sir Alexander Fleming, who noticed a clearance zone on an agar plate with a lawn of bacteria. The clearance zone formed a perimeter around a growth of *Penicillium notatum*, and Fleming named the active compound penicillin. The first antimicrobial agents were discovered in 1910 by Paul Ehrlich, with arsphenamine and its derivative neoarsphenamine being used to treat *Treponema pallidum*, causative agent of syphilis. Significant efforts were made to find new drugs with similar effects, but these were unsuccessful until Fleming's discovery. Gerhard Domagk observed the antibacterial activity of compound in an oil dye called sulfamidochrysoidine. This was developed to produce the closely related sulphanilamide (van Miert 1994). The desire to increase development of these 'magic bullet' compounds shifted much of the interest in phage therapy to researching new antibiotics (Rohwer and Segall 2015). Whilst interest in phage therapy waned, however, they remained a vital model system for our molecular understanding of the biological central dogma. A new era of phage biology has emerged, and with a greater understanding of phage interactions and processes, the biotechnological and medical potential is markedly increased.

1.1.2 Phages as a model for understanding DNA

Alfred Hershey and Martha Chase demonstrated the process of DNA injection by invading phages, showing that the genetic material entered the host whilst the remainder of the phage did not. Radiolabelled phage DNA was found within bacterial hosts after they had been sheared with a high-speed blender which effectively removed the adsorbed phages. It was this observation that showed that DNA is the genetic material responsible for encoding genomes (Hershey and Chase 1952). The first sequenced genome was of the phage ΦX174, and as more genomes were discovered, it was observed that significant transfer between genomes was occurring (Sanger et al. 1977). This led to the understanding of horizontal gene transfer (HGT) between organisms as a major driving force of genetic diversity (Hendrix et al. 2000).

Much of the phage research during the dawn of molecular biology was focused on a handful of coliphages such as λ and P1. Through the study of phages, restriction modification was discovered, as well as the usage of a 3-nucleotide sequence to code for each amino acid. With the development of new technologies and understanding of life sciences, much of the scientific community focused on other research areas such as cancer and human pathogens such as HIV (Rohwer and Segall 2015).

1.1.3 Classification of bacteriophages

There are four realms of viruses, separated by the nature of their genomes. The Monodnaviria are single stranded DNA viruses that have a replicative doubled stranded intermediate (NCBI 2020). The Riboviria encode RNA-dependent polymerases, such as reverse transcriptase and RNA-dependent RNA polymerase (Venkataraman, Prasad, and Selvarajan 2018). The Varidnaviria are double stranded DNA viruses which encode vertical jelly roll-type major capsid proteins (Koonin et al. 2019). The Duplodnaviria encompasses all double stranded DNA

viruses that encode the major capsid protein HK97 (Koonin et al. 2019). Also with the Duplodnaviria are the *Herpesvirales* that infect eukaryotes. It is within this realm the *Caudovirales* order exist, which contain viruses that solely infect prokaryotes

The *Caudovirales* (*caudo* – Latin 'tail') order consists of three families, classified mainly on the properties of their tails (Figure 1.1). The Podoviridae, such as T7 and P22, have a simple tail morphology which is non-contractile. The Myoviridae such as T4, have long contractile tails and the Siphoviridae, such as λ, have long, non-contractile tails. The Ackermannviridae, Autographiviridae, Chaseviridae, Demerecviridae, Drexlerviridae and Herelleviridae account for the remainder of the Caudovirales, but are significantly less common and much less studied (Virus Metadata Resource 2019). Caudovirales are non-enveloped viruses with a single linear double stranded DNA genome packaged within an icosahedral protein head. Their DNA genome is injected into the bacterial host though the tail, which is linked to the head via head-to-tail connector proteins. The *Caudovirales* account for 95% of all studied phages.

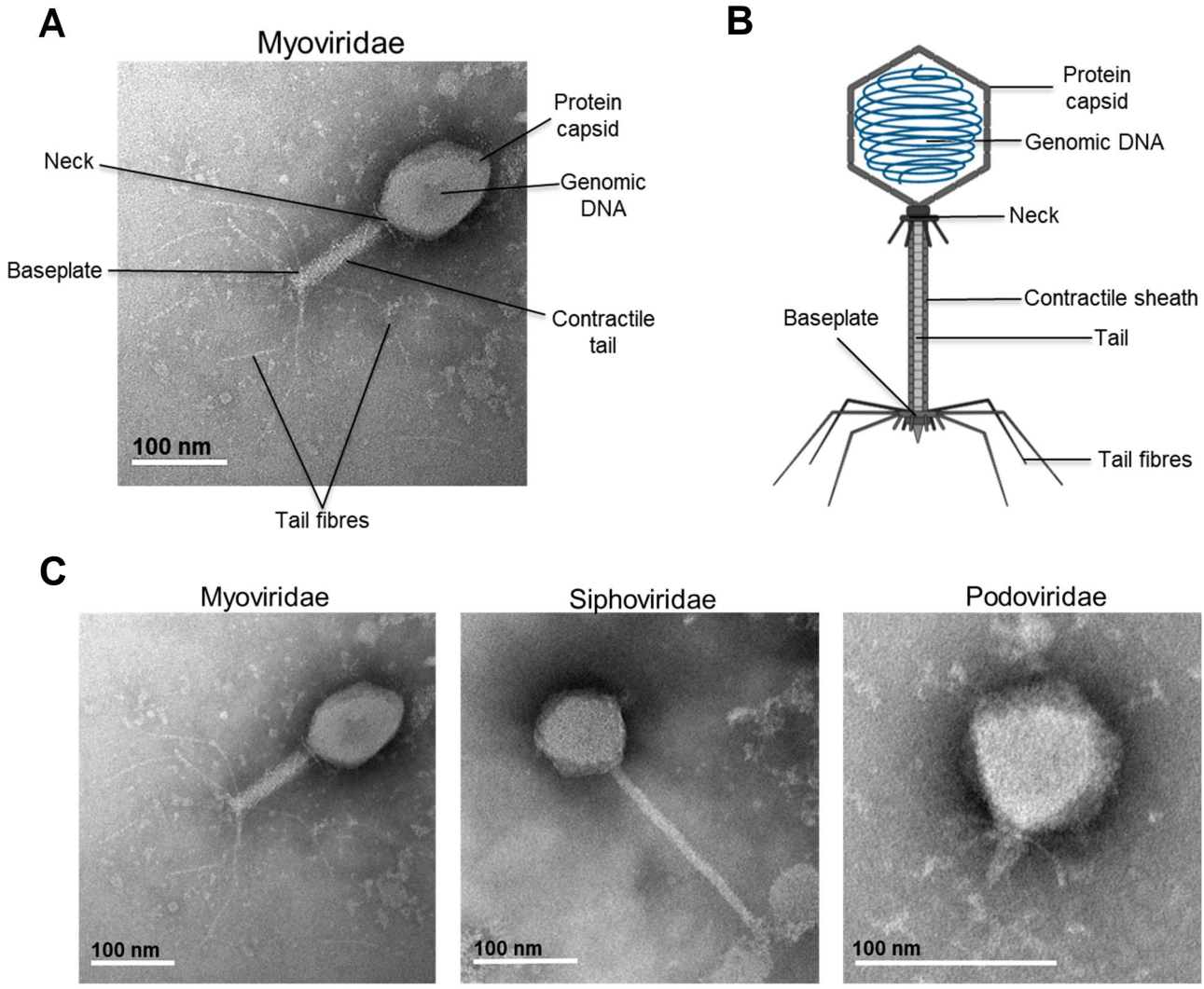


Figure 1.1: Structural features of tailed coliphages. *A) Negatively stained electron micrograph of a Myoviridae. B) Schematic labelling of typical Myoviridae phage. C) Negatively stained electron micrographs of the three most common families of Caudovirales that infect* Escherichia coli. *Images provided by Dr Tim Blower, from the Durham University undergraduate Microbiology Workshop.*

1.1.4 Phage adsorption and injection

The life cycle of a phage begins with the attachment of the tail fibres to the host receptor, known as adsorption. There are many modes of action for different phages, and the processes of the Myoviridae T4 and Siphoviridae λ will be detailed. The T4 phage binds OmpC and lipopolysaccharide (LPS) on the outer membrane of *E. coli*, resulting in an irreversible binding to the bacterial surface, facilitated by a rotation of the long tail fibres and a large conformational change in the baseplate (Figure 1.2). A contraction of the sheath occurs, piercing the bacterial surface allowing for the translocation of the viral genome into the host (Maghsoodi et al. 2019). Piercing the host's membrane is a three-step process. First, the needle of the tail tip is thrust through the outer membrane mechanically, followed by lysozyme driven degradation of the periplasmic peptidoglycan layers (Nakagawa, Arisaka, and Ishii 1985). The lysozyme domain of gp5 shares structural similarities to hen egg white lysozyme (Yap and Rossmann 2014). The last step involves the fusion of the tail tube with the cytoplasmic membrane and the dissociation of the tail tip from the main body of the tail tube (Maghsoodi et al. 2019). Phage λ does not have a contractile tail, and as such, cannot mechanically pierce the outer membrane. Instead, once bound to the maltose outer membrane protein, DNA translocates through the membrane pore directly into the host periplasm, and then is passed through the mannose permease complex into the cytoplasm (Erni 2006). Effectively, both phages must inject their genomes into the host in order to establish an infection.

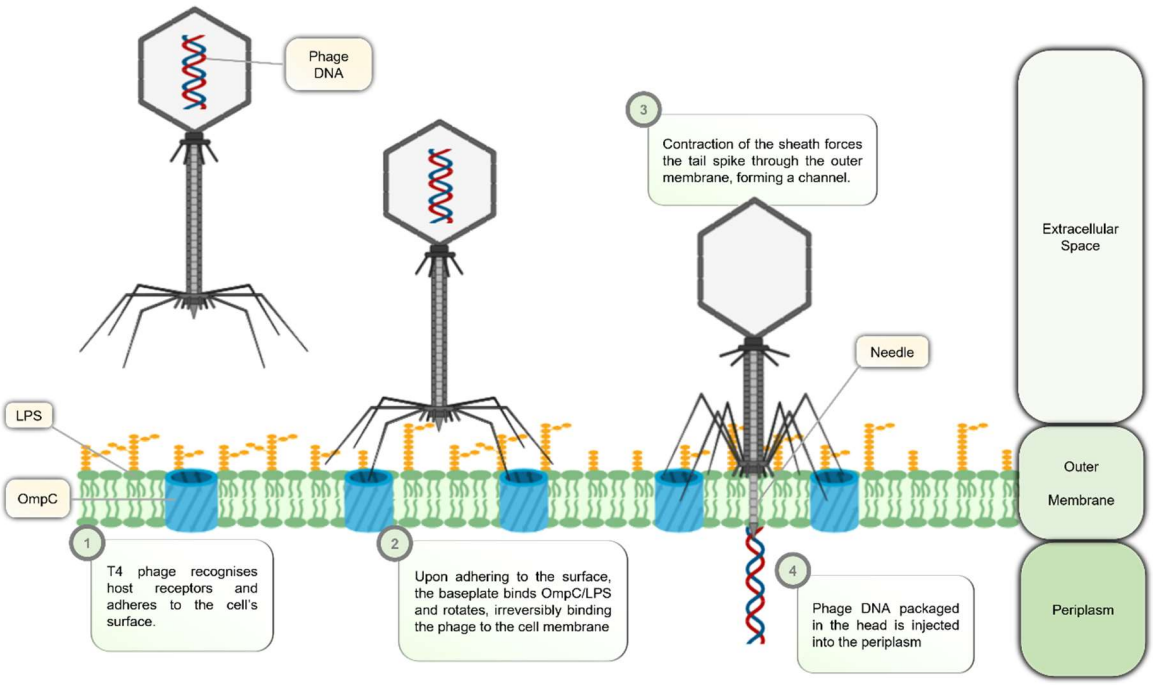


Figure 1.2: Adsorption of a bacteriophage. *Binding of T4 phage to the bacterial outer membrane via OmpC/LPS results in the rotation of the phage baseplate, irreversibly locking the membrane to the phage. The contraction of the sheath pierces the outer membrane, allowing the phage genome to translocate into the periplasm before entering the cytoplasm.*

1.1.5 Lytic cycle of T4

T4 is a lytic phage and does not integrate into the host genome. Replication proceeds in phases (Figure 1.3). In the early phase, the T4 genome has entered its bacterial host, host gene expression is arrested, and host DNA is degraded. T4 relies on reassigning host RNA polymerases (RNAP) to transcribe its DNA as it does not encode its own (Hinton 2010). The early phase genes are transcribed as their promoter sequences strongly compete for host σ70-RNAP complexes. The early phase of T4 involves the transcription of factors to hijack host processes, such as preventing transcription using deoxycytosine. This is facilitated by the production of T4 Alc, which terminates transcription of deoxycytosine containing DNA.

T4 overcomes this transcription arrest by incorporating modified cytosines that have been methylated into its genome, and further modifying these bases by glycosylation (Severinov et al. 1994). The anti-sigma factor AsiA is produced during early infection, preventing RNA polymerase from binding host promoters, and instead paving the way for transcription of middle phase genes. This is driven by the activity of T4 MotA (modified of transcription) which directs the RNAP-AsiA complex to middle phase gene promoters (Hinton 1991). This drives the production of components for genome replication.

Finally, late phase genes for the production of the structural proteins for the head and tail are transcribed, as well as key assembly proteins. Late phase transcription is activated by the production of T4 encoded sigma-55 (Miller et al. 2003). Once the biosynthesis of progeny components is complete, the procapsid is produced and matured. Gp16, gp17 and gp20 form the terminase complex which translocates the linear T4 dsDNA genome into the head. This mechanism requires hydrolysis of ATP in order to package unit-length DNA into a condensed structure within the phage head (Rao and Black 2005).

Once new phage virions are assembled, host lysis is induced through phage-encoded holins and lysozyme that disrupt and degrades the host's cell wall, leading to the destruction of the cell's integrity and the release of progeny phage. T4 can detect superinfection of the host by another phage, and through lysis inhibition it can delay the cycle, allowing T4 to accumulate a significantly number of progeny phage, resulting in a greater "burst size" upon lysis (Burch et al. 2011). Released progeny reinitiate the phage life cycle by adsorbing to another host cell.

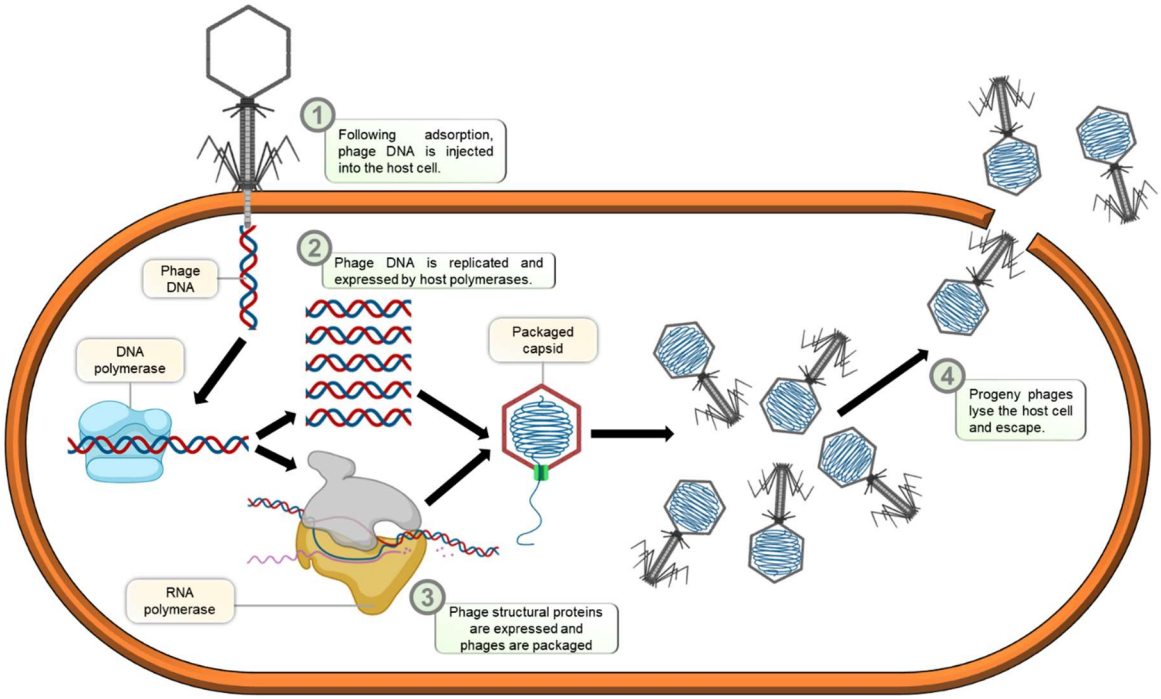


Figure 1.3: Lytic life cycle of T4. *Phage DNA is replicated and expressed by host machinery and progeny phages are packaged before being released.*

1.1.6 Lysogenic life cycle of phage λ

Phage λ exhibits both lytic and lysogenic activity which is largely modulated by cI (Figure 1.4). The repressor cI is the sole product of the lysogenic cycle λ, preventing transcription from the P_L and P_R regions (Lewis et al. 2011). The first mRNAs produced are cro and N, transcribed in opposite directions. N binds terminator regions allowing for polycistronic transcription by allowing RNA polymerase to continue transcribing past terminators, expressing the *cII* and *cIII* genes, which drive bacteriostasis favouring lysogeny (Rajamanickam and Hayes 2018). The presence of cI prevents expression of lytic genes, maintaining the lysogenic cycle as the prophage genome is replicated with every replication of the host chromosome. Under cellular stresses such as UV light, DNA SOS repair induces the cleavage of cI due to its similar structure to the LexA transcription repressor, allowing for lytic genes to be expressed (Little 1984).

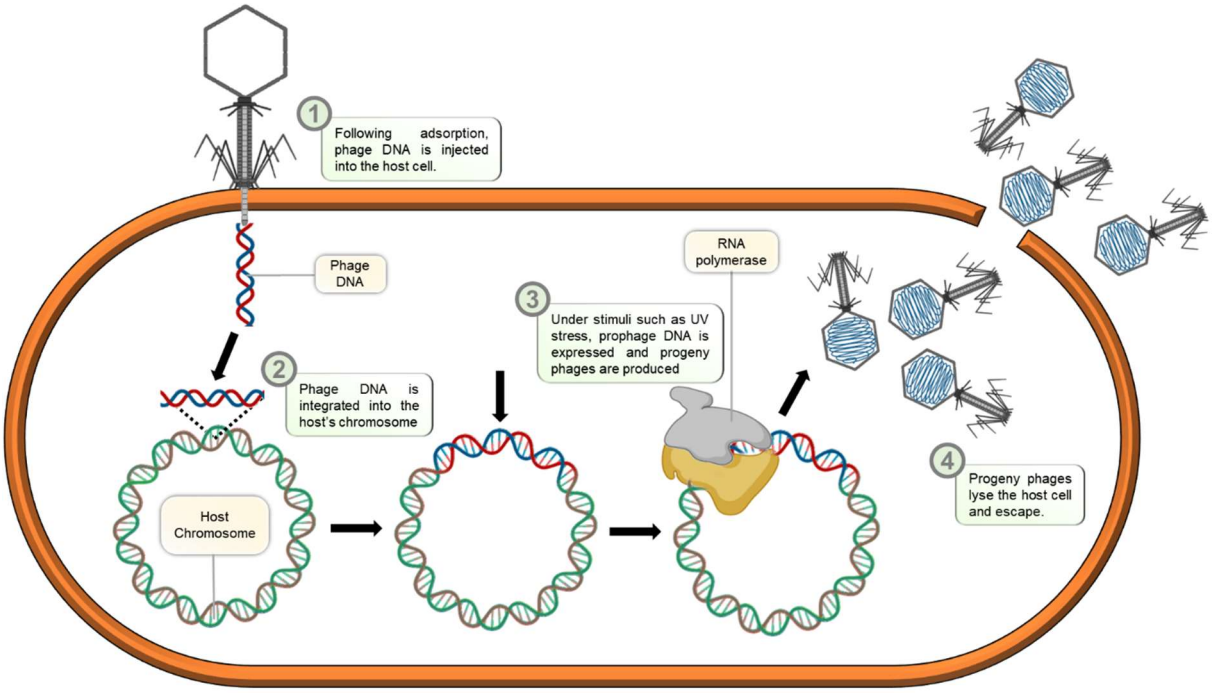


Figure 1.4: Lysogenic cycle of λ phage. *Phage DNA is introduced into the host genome. Expression of the prophage DNA leads to the production of progeny phages, resulting in the release of progeny phage.*

1.1.7 Prevention of superinfection

Phage λ encodes the Rex two-component system, consisting of a cytoplasmic phage recognition unit RexA, and an ion channel forming protein RexB. Phage λ undergoes lysogenic cycles within the *E. coli* host and the Rex system is expressed. RexA detects a DNA-protein complex induced by a new phage infection, inducing the activity of the membrane pore produced by RexB. This leads to a loss of membrane potential due to the movement of monovalent cations, leading to cell death (Snyder 1995). As a result, the newly invading phage cannot produce progeny and establish a wider phage community. T5 phage uses the outer membrane protein FhuA, an iron uptake protein, as its receptor and covers it with the lipoprotein Llp. This prevents other phages such as T1 from binding and establishing a superinfection (Pedruzzi, Rosenbusch, and Locher 1998).

1.1.8 Environmental bacteriophage interactions with their bacterial hosts

It is estimated that there are 10^{31} phages within the biosphere, making them ten times more abundant than their bacterial hosts (Whitman, Coleman, and Wiebe 1998). Parsons et al. estimate that there are a total of 10^{30} virioplankton in the oceans, accounting for 94% of the total number of nucleic-acid-containing particles and organisms (Parsons et al. 2012; Suttle 2007). As a result, phages are a major driving force for shaping the microbial ecology of oceans, turning over roughly 20% of the marine microbial biomass every day. This releases dissolved organic carbon back into the marine environment, affecting biogeochemical cycling of nutrients (Suttle 2007). Phages are also highly abundant in soil, with estimates of up to 10^{10} per gram in rich soils. The infection of key biogeochemical cycling prokaryotes means that the activity of soil phages is a nutrient cycling factor with the soil environment (Braga et al. 2020). The human intestines contain a diverse community of microorganisms, with estimates of total cell counts exceeding 10^{14} cells, with an average of 500 different species contributing to the microbiota (Thursby and Juge 2017). Despite the abundance of food sources for microorganisms within the human intestines, there are significant pressures to maintain microbial communities. Phages are highly abundant within the intestine and directly shape the microbiome, and as a result, the intestine represents another niche is which bacteria rely on anti-phage mechanisms to ensure their survival (Sausset et al. 2020).

1.2 Gene Regulation in Bacteria

Control of bacterial gene expression is largely achieved by regulation of transcription and translation. Bacterial genes are commonly clustered together in polycistronic operons and are transcribed under the control of a single upstream promoter. Studies of the lac operon of *E. coli* laid the foundations of gene expression research and detailed expression dynamics such as inducers and repressors. Sequence specific DNA-binding proteins bind operator sequences in close proximity to the promoter sequence, preventing RNAP from binding. The λ cI repressor functions this way by repressing transcription from the *pR* promoter (Rojo 1999). Bacteria mRNA transcripts are much less stable than eukaryote transcripts, and therefore the coupling transcription and translation is stringently coupled (Rauhut and Klug 1999).

1.2.1 WYL domain-containing proteins and transcriptional regulation

The WYL domain contains a Sm-like SH3 beta-barrel fold domain, and proteins containing this domain belong to the WYL-like superfamily. The domain is named after the three conserved residues WYL, which has subsequently been shown to only be present in a small subset of these proteins (Makarova et al. 2014). WYL containing proteins are predicted to have a ligand-binding domain that recognises negatively changed elements such as DNA or nucleotides. These proteins have been implicated in the detection of modified nucleotides and their derivatives. A WYL domain protein sl7009 has been shown to negatively regulate the I-D CRISPR-Cas system in cyanobatcteria by repressing transcription. This protein also contains a helix-turn-helix motif for DNA binding (Hein et al. 2013). It is speculated that these proteins represent a viral DNA sensing system that activates anti-phage effector proteins.

1.2.2 Bacterial RNA polymerase and sigma factors

Dissimilar to eukaryotes which employ several RNAPs, bacteria utilises only one, comprised of two α subunits, two β subunits and a single ω subunit, arranged as α2ββ′ω (Rauhut and Klug 1999). The RNAP complex binds a σ-factor to form the RNAP holoenzyme (Figure 1.5). The dimerisation of the α subunits allows for the binding of the other subunits via the N-terminals, whilst regulatory elements interact with the C-terminal domains (Liu et al. 1996). The β and β' subunits comprise the catalytic core of RNA synthesis and bind double stranded nucleic acid molecules. The beta subunits are recruited via the ω subunit, which also maintains the structural integrity of the enzyme (Mathew and Chatterji 2006; Zhou et al. 2013)

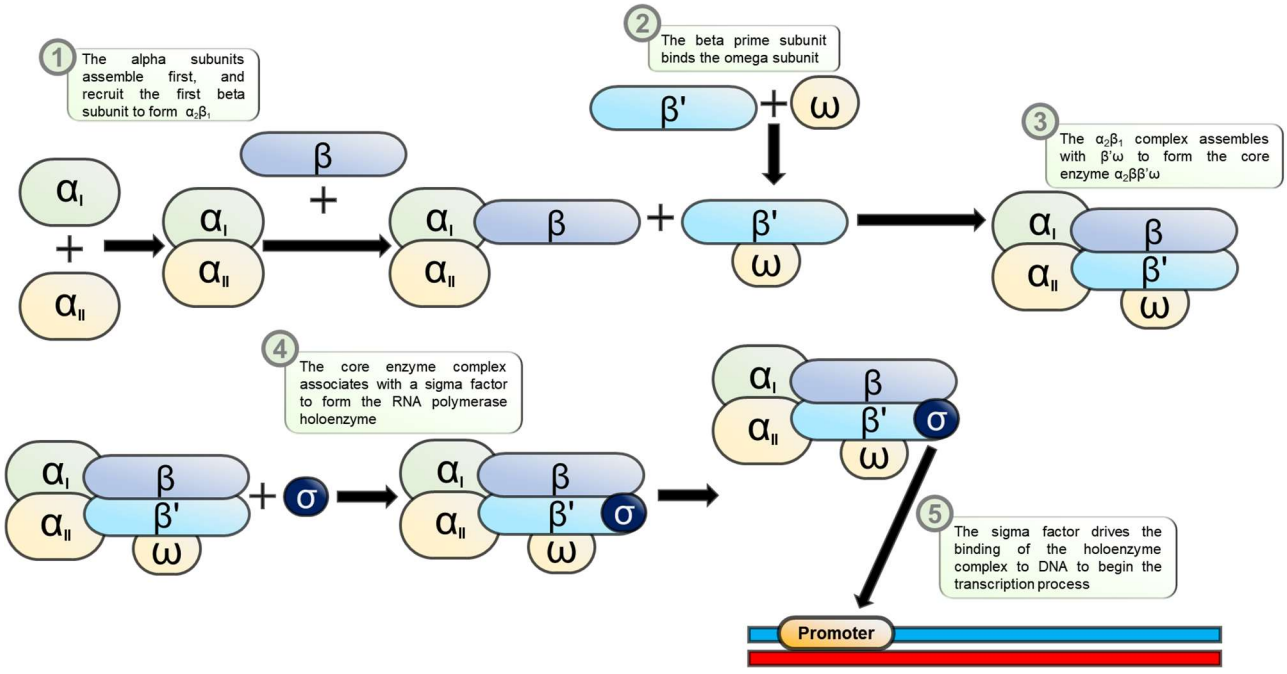


Figure 1.5: Recruitment of RNA polymerase. *RNA-polymerase subunits form the RNA polymerase holoenzyme capable of transcribing DNA. Alpha subunits organise first and beta and omega subunits to form an active complex when bound to a sigma factor.*

1.2.3 Sigma factors

σ factors direct RNAP to promoter sequences at positions upstream of the transcription start sites (TSSs). The σ^{70} family contains the housekeeping sigma factors for expression during normal growth conditions in *E. coli*, such as RpoD in exponential phase, and directs RNAP to conserved hexanucleotide sequences at positions -35 and -10 relative to the TSS. Bacteria employ many other sigma factors to drive a specific gene expression profile (Bervoets and Charlier 2019). Under stress conditions such as starvation RpoS is utilised driving a change in 10% of *E. coli* genes (Tripathi, Zhang, and Lin 2014). Similarly, RpoF drives flagellar synthesis, RpoN is produced in response to nitrogen starvation, and RpoH initiates the expression of heat shock response proteins such as DNA repair enzymes and chaperones (Narberhaus and Balsiger 2003; Kazmierczak, Wiedmann, and Boor 2005). Alternative σ factors are encoded by phages to drive specific gene expression from their own genomes. Once transcription has been initiated, σ factors are displaced.

1.2.4 The *lac* operon as a regulatory model and its experimental use

Much of the current understanding of gene expression is resultant of studies on the lactose regulatory system in *E. coli.* The *lac* operon consists of the effector genes *lacZ, lacY* and *lacA.* Lactose is hydrolysed by the product of *lacZ,* β-galactosidase, producing galactose and glucose (Figure 1.6). LacY is a galactoside transporter which imports lactose, and LacA catalyses acetylation of thiogalactosides (Lewis 2005). In the absence of allolactose in the growth media, the LacI repressor binds the operator sequence with high specificity, preventing RNAP from transcribing the downstream effector genes (Agnes Ullmann 2009).

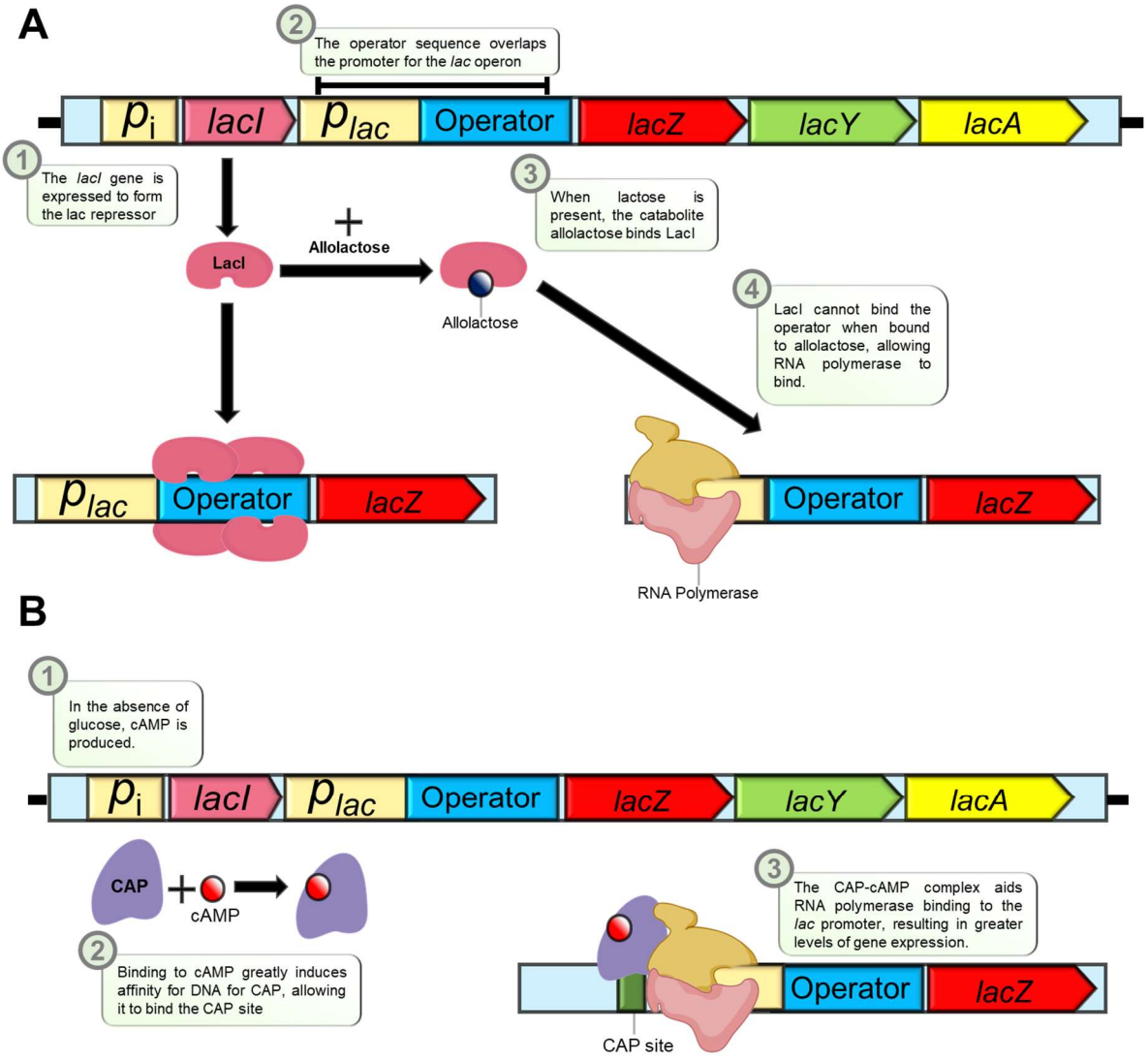


Figure 1.6: Regulation of the lac operon. *A) Binding of allolactose displaces the LacI complex from the operator sequence, allowing RNA polymerase to bind the lac promoter. B) Production of cAMP activates CAP which aids RNA polymerase binding to the lac promoter.*

1.2.5 Helix-turn-helix motifs

Helix-turn-helix (HTH) motifs bind the major groove of DNA via hydrogen bonds and Van der Waals interactions. Monomers consists of α-helices separated by a short region of amino acids. HTH motifs were originally discovered due to similarities being observed between several regulatory components encoded by *E. coli* and phage λ. This led to the observation that Cro, CAP and the λ repressor all contain a 20-25 amino acid sequence that recognises DNA (Matthews et al. 1982; Anderson et al. 1981; McKay and Steitz 1981). Each monomer contains at least 2 α-helices, with one helix contributing to DNA recognition, supported by the other helix/ helices (Matthews et al. 1982). Winged HTH domains consists of a 3-helix bundle, and a 3 or 4-stranded beta sheet that forms the 'wing'. The α-helical regions contain the DNA recognition helix, and the wing stabilise the protein-DNA interaction by making separate contacts such as at the minor groove (Gajiwala and Burley 2000).

1.3 Overview of Phage Defence Systems

Phage defence systems are usually driven through a specific mode of action, such as the recognition of a specific DNA motif which is targeted by an effector protein. As a result, a single phage defence system provides limited resistance. This is due to a number of reasons, such as the small temporal margins during a phage infection in which effectors must be active, but also due to phage encoded counter-resistance mechanisms. Therefore, bacteria employ a range of phage resistance systems, forming a multi-pronged defence.

1.3.1 Individual DNA sensing phage defence systems

Prokaryotes employ a wide range of anti-phage systems to prevent infections. Detailed examination of selected systems is provided in section 1.4, whereas this section provides an overview including restriction modification in which foreign DNA is degraded, abortive infection systems that prevents the spread of phage progeny by arresting bacterial cell growth, and BREX which functions independently of previously characterised systems.

1.3.2 Restriction Modification

Restriction modification (RM), primarily a two-component system, relies on the recognition of foreign DNA that can be distinguished from host DNA. The majority of RM systems employ a methyltransferase that modifies host DNA to ensure the cognate restriction endonuclease can no longer target it (Vasu and Nagaraja 2013). The endonuclease targets specific DNA sequences, generating double stranded breaks producing blunt or overhanging ends. The methyltransferase targets the same DNA sequence, adding a methyl group to the N^6 amino group of adenine, or the C^5 carbon or N^4 amino group of cytosine (Cheng and Roberts 2001).

1.3.3 Abortive infection

Abortive infection is an altruistic mechanism that prevents the spread of a phage infection by inducing cellular suicide, preventing phage progeny from spreading to neighbouring cells. Abortive infections covers a wide array of strategies and can be used to denote systems that upon detecting a phage infection, induce cell death before mature phages can be replicated (Lopatina, Tal, and Sorek 2020). These systems utilise a phage sensing element, and an element that induces cell death. As the activity of abortive infections systems results in the death of the host, it is seen as a second line defence, to be utilised when other systems such as RM have failed to prevent an established phage infection.

1.3.4 CRISPR-Cas

CRISPR-Cas systems comprise two components. The *cas* genes encode effector proteins such as the Cas9 nuclease that degrades target DNA. The CRISPR array is made of sections of direct repeats of DNA, separated by small spacer regions which have originated from previous encounters with foreign DNA, which can be introduced during phage infections. The CRISPR arrays form libraries of previously encountered phage DNA, allowing the host to identify phages in subsequent infections. This provides the host with an adaptive immune response (Stanley and Maxwell 2018).

1.3.5 Cyclic-oligonucleotide-based antiphage signalling systems

Cyclic-oligonucleotide-based antiphage signalling systems (CBASS) comprise another form of two-component system, employing a cyclic oligonucleotide cyclase that produces cGAMP upon sensing viral DNA, and a phospholipase that targets the host's inner membrane inducing cell death (Cohen et al. 2019). CBASS induces cell death and therefore falls under the category of abortive infection systems with a distinct detection and effector profile.

1.3.6 Defence island system associated with restriction-modification (DISARM)

DISARM represents a phage defence system that restricts incoming phage DNA, and is modulated by a DNA methyltransferase. Class I DISARM systems employ an adenine-methylase, a helicase domain containing protein, a DUF1988 domain protein, an SNF2-like helicase and a phospholipase-D domain protein. Class II systems vary by encoding a cytosine methyltransferase instead of an adenine methyltransferase and do not include a SNF2-like helicase (Ofir et al. 2018). DISARM is widespread in bacterial and archaeal genomes, and differs from classical RM systems, as it utilises multiple components to form a restriction module.

1.3.7 Phosphorothioate (PT) sensing phage defence

The Dnd system introduces sulphur atoms into the DNA backbone, protecting DNA from cleavage by RM systems. The Ssp system employs SspE which identifies PT modifications and can introduce single stranded DNA breaks to prevent phage DNA from replicating (Xiong et al. 2020).

1.3.8 Prevention of adsorption to cell surface

Biofilm formation can provide an indirect system of phage defence by allowing cells within the biofilm's interior to divide, whilst cells on periphery are susceptible to infection. The production of a dense extracellular matrix consisting of amyloid fibres prevents phages from disseminating to the colony's core (Vidakovic et al. 2018). Gram-negative bacteria such as *Vibrio cholerae* can produce outer membrane vesicles (OMVs) which contain phage receptors, effectively acting as a decoy landing site for phages. OMVs pinch of the host cell and form spherical structures that mimic the phage's bacterial target, ultimately reducing the infectively of the phage (Reyes-Robles et al. 2018).

1.3.9 Bacteriophage Exclusion

The Bacteriophage Exclusion (BREX) system shares similarities with other phage defence systems in that it utilises a methyltransferase to distinguish host DNA from foreign DNA, however it forms a distinctly different defence system by employing a variety of uncharacterised anti-phage components (Goldfarb et al. 2015).

1.4 Mechanisms of Phage Defence Systems

As BREX shares similarities with CRISPR-Cas and RM defence systems in its interactions with DNA, this review will focus on and compare these systems.

1.4.1 Initial characterisation of BREX

The simultaneous adaptation of the bacterial host through evolving phage-resistance mechanisms, alongside the rapid variation of phage genomes, has led to this field of research being regularly described as a militaristic "arms race" (Seed 2015; Stern and Sorek 2011b; Goldfarb et al. 2015; Nechaev and Severinov 2008). BREX systems have recently been described by Goldfarb et al. to confer resistance to an array of both lytic and temperate phages. These observations were made when BREX-negative *Bacillus subtilis* was transformed with the full BREX system isolated from *Bacillus cereus* (Goldfarb et al. 2015). Goldfarb et al. originally screened ~1,500 bacterial and archaeal strains for the presence of the phage growth limitation (Pgl) system, consisting of the essential alkaline phosphatase (*pglZ)*, as well *pglWXY*. This system has been shown to confer resistance in *Streptomyces coelicolor* to temperate phages (Sumby and Smith 2002). Upon screening for PglZ, Goldfarb et al. established that *pglZ* was encoded by ~10% of the screened bacterial and archaeal genomes, often within a gene cassette containing five other genes. BREX is commonly organised into two operons of *brxABCL* and *pglXZ.* BrxA is an RNA-binding anti-termination protein, BrxB is a protein with unknown function, BrxC is an ATP-binding protein, and BrxL is a Lon-like protease. PglX has been shown to be a methyltransferase that targets the host genome at specific motifs (Figure 1.7), and PglZ is a phosphatase (Barrangou and van der Oost 2015; Goldfarb et al. 2015). These six genes make up the core BREX system, denoted the BREX 1 or consensus system. Multiple similar systems exist however, often containing gene inversions, duplications, or additional genes with separate functions. Six distinct systems were identified and clustered based on their

gene content and organisation, as well as an additional minority group that did not subscribe to either of the other six groups and showed no homology to each other. A striking difference between BREX-encoding and Pgl-encoding strains, was that Pgl-encoding bacteria were susceptible to phage infection during the first infection cycle but not the second, however BREX-encoding strains were resistant during the first infection cycle.

1.4.2 Activity of BREX methyltransferases on phage DNA

Gordeeva et al. showed that progeny phages isolated from BREX-positive *E. coli* contain the same methylation pattern on the non-palindromic sequence GGTAA**G** as host genomic DNA (Gordeeva et al. 2019). The induction of λ prophages from BREX-positive cells produces resistant progeny phage, showing that the epigenetic modification of the phage genome confers protection from BREX. PacBio sequences of λ phage genomes showed that the complementary motif CTTACC was not methylated. Mutational analysis of the BREX operon showed that *brxA* was non-essential for methylation of GGTAA**G** motifs or for conferring resistance to phages. Deletion of *brxL* abolished resistance to phages, however methyltransferase activity of PglX was still observed. All other gene deletions resulted in a lack of methylation and abolished phage resistance (Gordeeva et al. 2019). REBASE predicts that PglX is a multifunctional restriction endonuclease/ methyltransferase fusion, however no endonuclease activity has been observed (Hui et al. 2019).

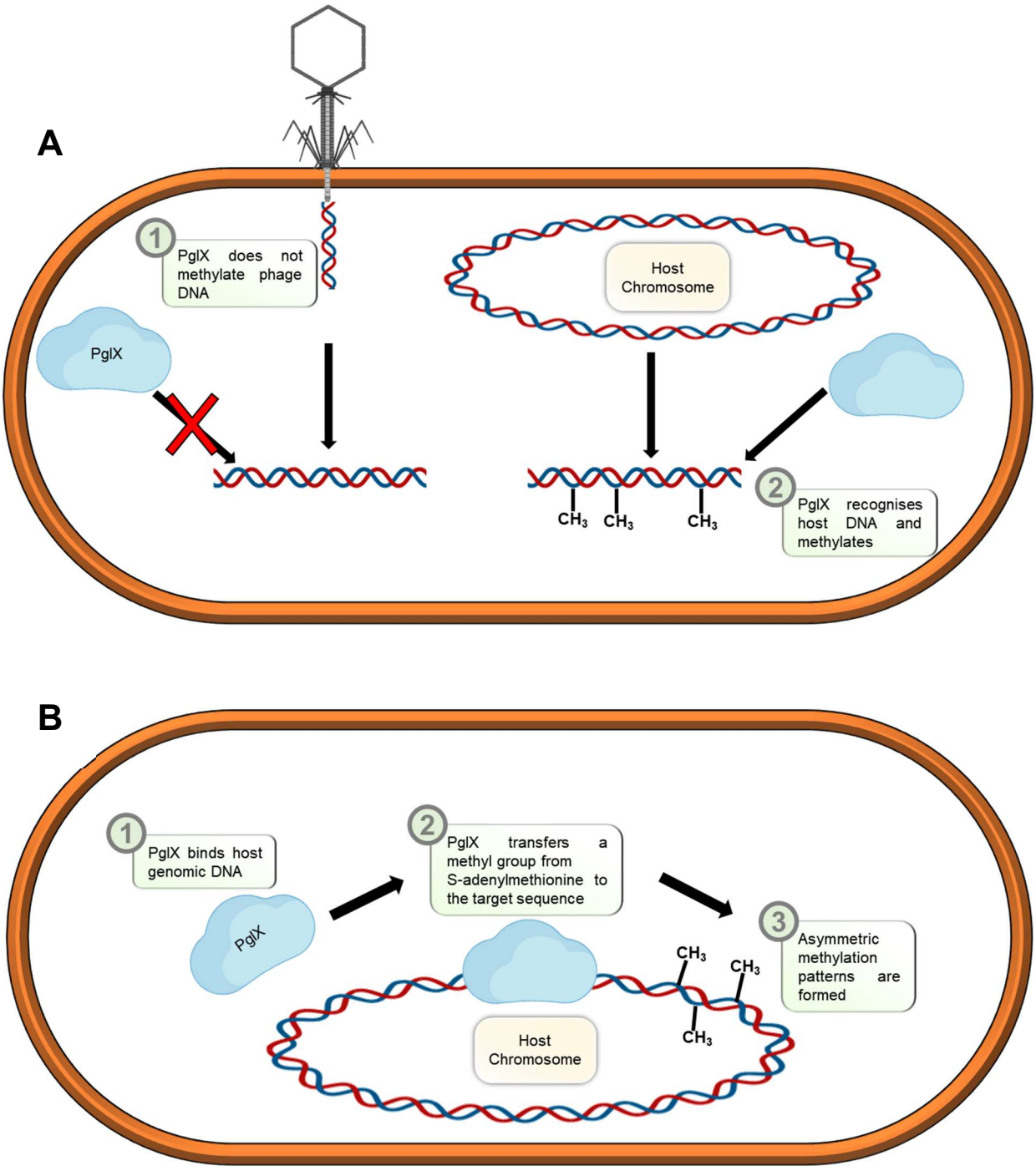


Figure 1.7: Activity of PglX on target DNA. *A) Incoming phage DNA remains unmethylated and remains susceptible to BREX effectors. B) Asymmetric, distinct methylation patterns are formed on host DNA rendering it protected from BREX effector proteins.*

1.4.3 Identifying distinctions between BREX and other resistance systems

BREX has been shown to operate in a distinct manner from other previously described resistance mechanisms. BREX has not been found to prevent the initial stages of infection such as phage adherence to the host surface or injection of DNA, and BREX does not mediate cleavage of phage DNA by restriction endonucleases. Structural modification or physical masking of surface receptors can lead to inhibition of phage adsorption, however this is not mediated by BREX (Coffey and Ross 2002). Phage receptors can be blocked by the production of host factors that competitively bind the surface receptor. Adsorption assays show that Φ3T adhere to BREX-positive *B. subtilis* as effectively as it can adhere to WT. Phage DNA is detected within cell lysates indicating that DNA injection following adsorption is successful and not inhibited by BREX (Goldfarb et al. 2015).

Another commonly employed phage resistance mechanism is the altruistic system of abortive infection (Abi), where an infected bacterium mediates its own death in order to prevent phages from spreading throughout the population (Chopin, Chopin, and Bidnenko 2005). BREX does not mediate this, as BREX-positive cultures do not exhibit arrested growth or culture decline, which is observed when abortive infection occurs (Goldfarb et al. 2015; Fineran et al. 2009). Therefore, abortive infection has been ruled out as a possible mechanism as to how BREX confers phage resistance.

1.4.4 Variations of BREX systems

Goldfarb et al. used a bioinformatic approach to identify 6 key types of BREX systems, clustered based on the components that they encode (Figure 1.8). The type I system detailed in 2015 is most commonly observed accounting for 55% of systems, however there are a further 5 classifications (Goldfarb et al. 2015). Type II accounts for 15% of BREX systems and

constitutes the phase growth limitation (Pgl) system, encoding a kinase PglW, an ATP-binding protein BrxD, a helicase BrxH, as well as PglX and PglZ. The ATPase BrxC is replaced by PglY, which shares the P-loop motif of BrxC and DUFs 2791 and 499. BREX type III encodes proteins of unknown function *brxF, brxA, brxC, pglXI, pglZ* and *brxHII.*

The type III systems have greater homology to type I systems, sharing the predicted RNA-binding protein *brxA* and *pglX.* However, type I pglX incorporates the methylase domain pfam13659 and type 3 *pglXI* pfam01555. The predicted helicase *pglHII* is present and brxL is absent. Type IV encodes only 5% of systems and comprise of only *brx*C/*pgl*Y, *PglZ, BrxL* and a phosphoadenynyl-sulphate (PAPS) reductase domain containing protein, brxP. There is no *pglX* encoded within these systems. The pfam10507 domain within *brxP* is implicated in the DND resistance system in which the DNA backbone is modified with sulphur additions. Type V also account for 5% of systems within this study and contains *brxA, brxB, brxC/pglY, pglX, pglZ* and *brxHII*, with no *brxL* present. It encodes the same helicase as type III, the same *pglX* as type I, as well as a duplication of *pglY/brxC* separated by *brxB*. Finally, type VI systems include *brxE*, a protein of unknown function, in addition to *brxA, brxB, brxC/pglY, pglX, pglZ, brxD* and *brxHI.* 7% of the systems found have not been classified into a specific group (Goldfarb et al. 2015).

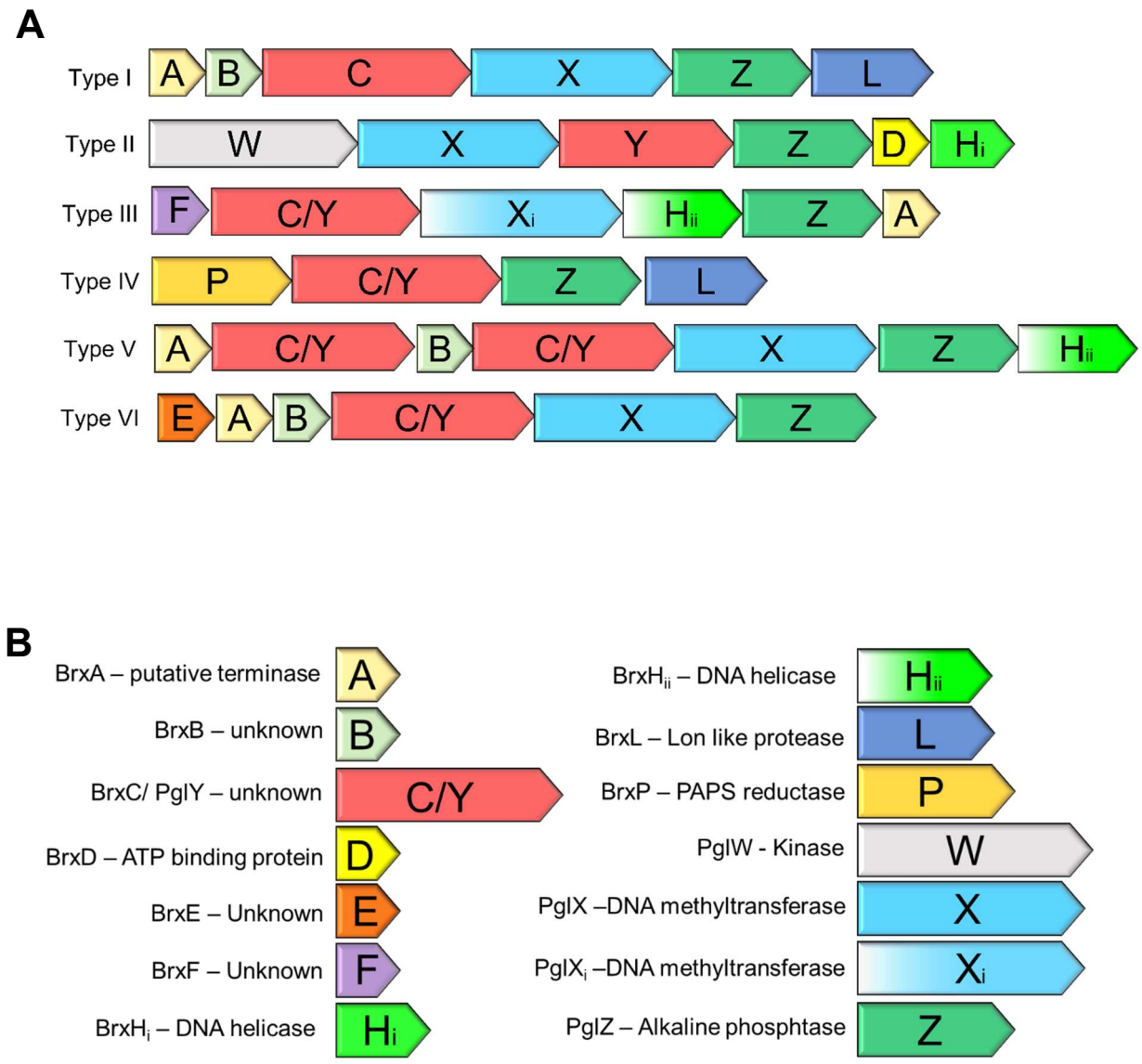


Figure 1.8: Genetic components of BREX. *A) Operonic representation of the 6 core BREX types. B) Annotation of individual BREX components with predicted function*

1.4.5 BREX protein analogues

Between the 6 classes of BREX systems, certain proteins are substituted for others which replace their function. BrxC which is found within the type I system, is replaced by another ATPase PglY in type II. Both proteins contain a nucleotide binding P-loop characteristic of kinases, helicases, and motor proteins (Thomsen and Berger 2008). However, only 4% of their sequences are shared which is primarily the result of the P-loop, indicating that they are significantly different proteins. However, they share a DUF499 and are predicted to have a shared role in BREX defence.

This variability is also observed for the predicted helicases BrxHI and BrxHII. Whilst BrxHI contains the COG1201 Lhr-like helicase domain, BrxHII contains a COG0553 DNA/RNA helicase domain. PglX and PglXI also incorporate different adenine-specific methyltransferase domains but are predicted to have the same function to their respective systems. However, not all systems encode a PglX or BrxH variant, whereas all systems encode a PglZ or BrxC/PglY component (Goldfarb et al. 2015).

1.4.6 The role of methyltransferases in phage resistance mechanisms

Bacteria employ restriction modification (R-M) systems in order to cleave injected phage DNA to prevent propagation and is often viewed as a primitive bacterial immune system. R-M systems are ubiquitous, and have almost infallible specificity to their targets (Vasu and Nagaraja 2013). R-M systems are often encoded by highly mobile elements such as plasmids or transposons, contributing to their ubiquity (Kobayashi 2001). R-M systems utilise methyltransferases to distinguish between viral and host DNA motifs, ensuring that host DNA is not cleaved by the endonuclease (Bickle 2004). Phage genomes that contain fewer restriction sites have a selective advantage as they are less prone to enzymatic cleavage (Samson et al.

2013). However, whilst BREX does encode a methyltransferase, phage DNA has been shown to not be degraded by BREX encoded proteins. *PglX* encodes for an adenine specific methyltransferase that methylates the non-palindromic TAGGAG motif at the fifth position.

Comparisons have been drawn between R-M and toxin antitoxin (TA) systems, where an antitoxin sequesters a toxin inhibiting it from causing cell death. The removal of the antitoxin causes the toxin to mediate its effect, and the removal of a methyltransferase from an R-M system can lead to cleavage of host DNA, both resulting in growth arrest or cell death (Stern and Sorek 2011a; Hayes 2003). The deletion of *pglX* in BREX-positive *B. subtilis* however does not result in cell toxicity, suggesting that PglX activity does not prevent a restriction endonuclease from inadvertently cleaving the host genome (Goldfarb et al. 2015). The specific structure and function of *pglX* has yet to be solved however, and therefore assessing the effect of the deletion of *pglX* is speculative.

Southern blot analysis has led to the hypothesis that BREX is not an R-M system, but rather mediates resistance by preventing the propagation of progeny phages by inhibiting DNA replication (Goldfarb et al. 2015). This has WT *B. subtilis* was shown to have an 81-fold increase in phage DNA 30 minutes post-infection (PI) with Φ3T when compared to 10 minutes PI, however BREX-positive *B. subtilis* exhibited no increase in phage DNA levels between 30 minutes and 10 minutes PI (Goldfarb et al. 2015). This evidence suggests that the BREX system mediates its action largely by preventing replication of phage DNA, preventing the propagation of any progeny phages. The mechanism in which this is achieved remains ambiguous.

1.4.7 Potential applications of BREX

Phage resistance mechanisms have been previously harnessed for powerful biotechnological functions, such as the use of restriction endonucleases for cloning, and CRISPR-Cas for genome editing (Sander and Joung 2014; Salmond and Fineran 2015). The potential for BREX to be utilised in such a manner is promising, as although its mechanisms are currently unclear, it represents a distinct variation from any of the previously described antiviral systems. One of the key potential uses of BREX proteins is the ability of PglX to perform sequence specific hemi-methylation. This could act as a molecular switch, in which the induction of PglX drives site-specific DNA methylation, driving different gene expression profiles. Bioengineering bacterial genomes to contain multiple target motifs would allow for the activation/repression of multiple genes under the control of one mode of action. The study of methyltransferase mechanics is key to understanding the changes of methylation patterns in human disease. in Distinct methylation patterns are observed in cancer patients, and these abnormalities can be screened for to identify tumours. There are a number of genes found to contain mutations involved in the oxidation of methylcytosine in cancer patients (Pfeifer et al. 2014). A recurrence of interest has emerged in the use of phages as a bactericidal treatment due to the stalled progress of antibiotic discovery (Bragg et al. 2014). In order to fully utilise phages for this purpose, it is important that the resistance mechanisms are understood, in order to prevent an analogous situation to that of the current antibiotic resistance crisis.

1.4.8 Phage encoded BREX counter-resistance mechanisms

Selective pressure from bacterial phage-resistance systems reciprocally drives phage evolution, and counter-resistance mechanisms have arisen allowing phages to evade bacterial antiviral systems. These counter-resistance mechanisms include the modification or reduction of restriction sites, encoding a molecular mimic to hijack the host abortive infection response and

initiating the stimulation of methyltransferases to methylate phage DNA to evade R-M systems (Kruger et al. 1988; Samson et al. 2013; Blower, Evans, et al. 2012).

The similarities shared between R-M systems and BREX regarding the use of a methyltransferase to distinguish between self and foreign DNA may give insight into potential counter-resistance mechanisms against BREX. A phage that could stimulate PglX to increase its activity, therefore driving the methylation of its own DNA before another element of BREX sequesters it, would give the phage a selective advantage. However, despite the similarities of the two systems, another hypothetical BREX resistance mechanism would be the modification of the TAGGAG motif of the bacterial chromosome that is methylated by PglX. This could be achieved either by incorporating unorthodox nucleotides or reducing the number of motifs present. This mechanism is employed by T-even coliphages by utilising hydroxymethyl cytosine instead of cytosine (Figures 1.9 and 1.10), effectively modifying the recognition motif (Kruger and Bickle 1983). BREX needs to be effectively characterised before an investigation into how other counter-resistance mechanisms might operate (Gordeeva et al. 2019).

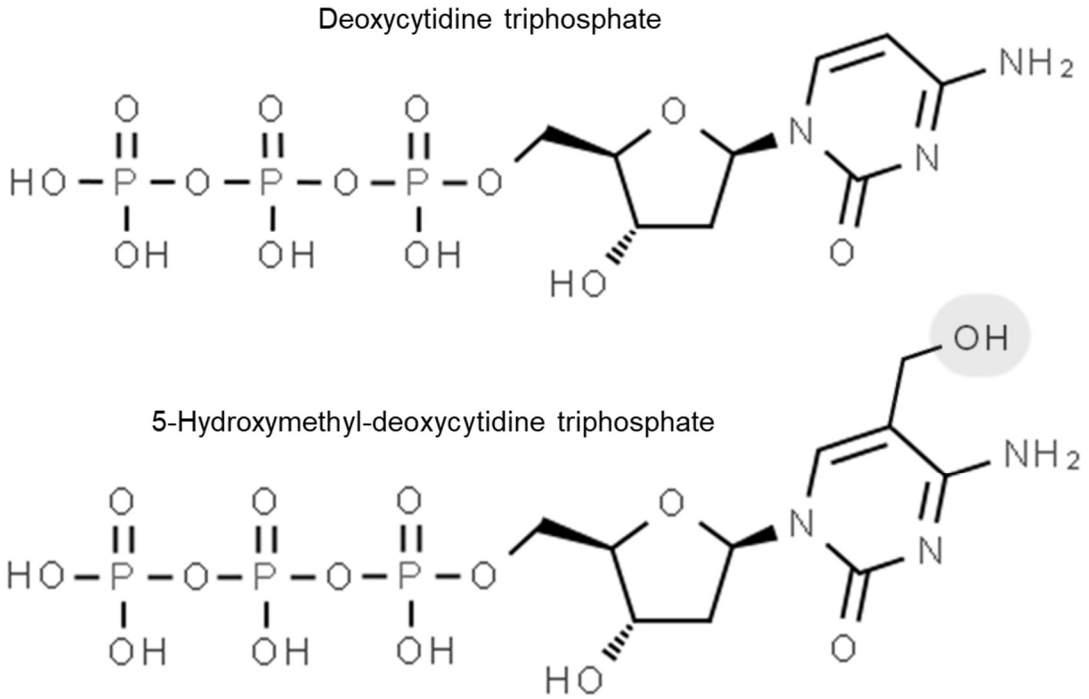


Figure 1.9: Structural formula of dCTP and hm-dCTP. *hm-dCTP (5hmC) is modified at the C5 carbon of the cytidine base.*

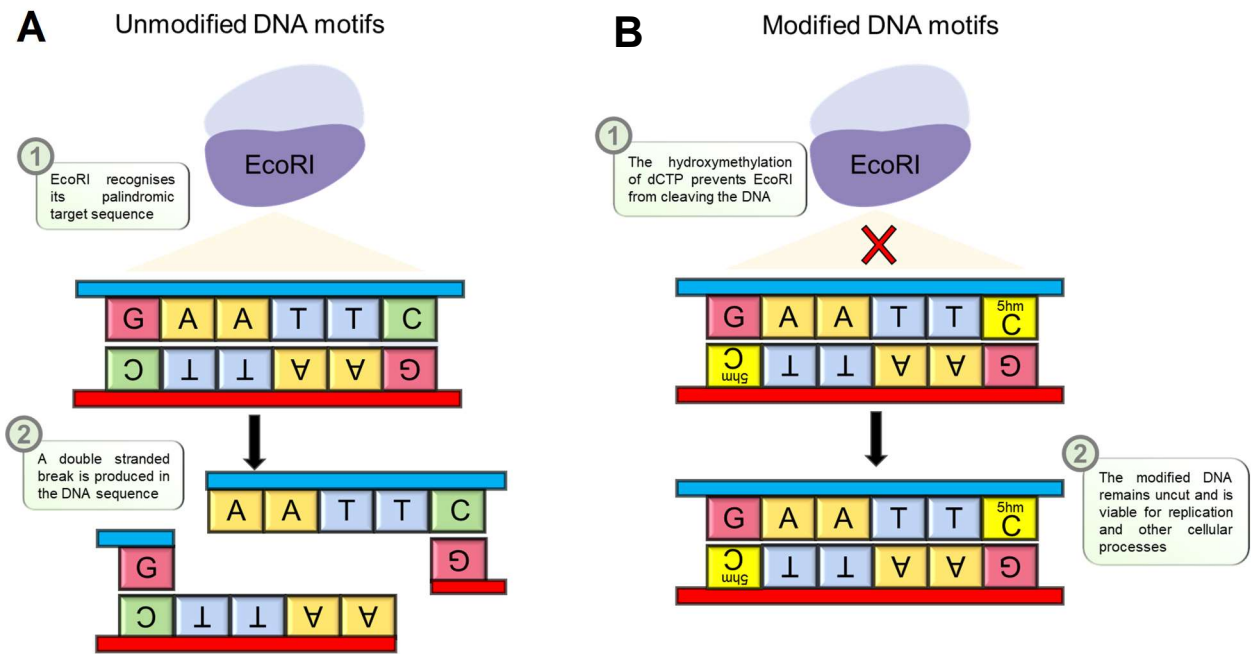


Figure 1.10: Modification of DNA motifs with 5hmC*. Recognition of target DNA sequences by a type II restriction endonuclease, EcoRI, and the evasion of restriction by the incorporation of a non-canonical DNA base. A) Incorporation of canonical cytosine residues allows EcoRI to introduce a double stranded break. B) Incorporation of 5hmC protects DNA against restriction.*

1.4.9 CRISPR-Cas; discovery, functionality and applications

The initial identification of clustered regularly interspaced short palindromic repeats (CRISPRs) was in 1987 by Ishino et al. whilst investigating *iap* driven alkaline phosphatase isoenzyme conversion (Interthal, Pouliot, and Champoux 2001). Secondary structures formed within CRISPR regions due to complementation of the palindromic sequences, making it difficult to read the sequence. It was due to the formation of these DNA hairpins that attention was given to the nature of CRISPR regions. With the advances of DNA sequencing, CRISPRs were appearing regularly in bacterial and archaeal genomes, and in 2005 Mojica et al. showed that non-repeating CRISPR sections originated from mobile elements such as plasmids or bacteriophages (Mojica et al. 2005). CRISPR associated (*cas)* genes were discovered to associate with CRISPR regions and the protein products of *cas* genes formed an adaptive immune response to viral infections (Makarova and Koonin 2015). The *cas* genes were initially discovered via genome comparisons of CRISPR containing genomes and have been categorised. Cas1 and Cas2 families sharing no known domains with other proteins. Cas3 was found to contain multiple conserved motifs of the superfamily 2 helicases, and Cas4 showed homology to RecB, an exonuclease involved in RecBCD homologous recombination (Ishino, Krupovic, and Forterre 2018).

1.4.10 CRISPR-Cas phage resistance

The CRISPR-Cas defence system functions by incorporating phage DNA sequences as spacers within CRISPR regions, which when expressed form crRNA sequences (Figure 1.11). crRNA sequences hybridise to their complementary phage DNA sequences, so that future phage infections can be targeted. Cas effector proteins bind tracrRNA sequences, which are guided by the crRNAs, localising the Cas nucleases to specific DNA motifs (Dy, Rigano, and Fineran 2018).

As a result, phage DNA is targeted by Cas nucleases in a sequence specific manner, which also prevents Cas nucleases from targeting host DNA (Figure 1.12).

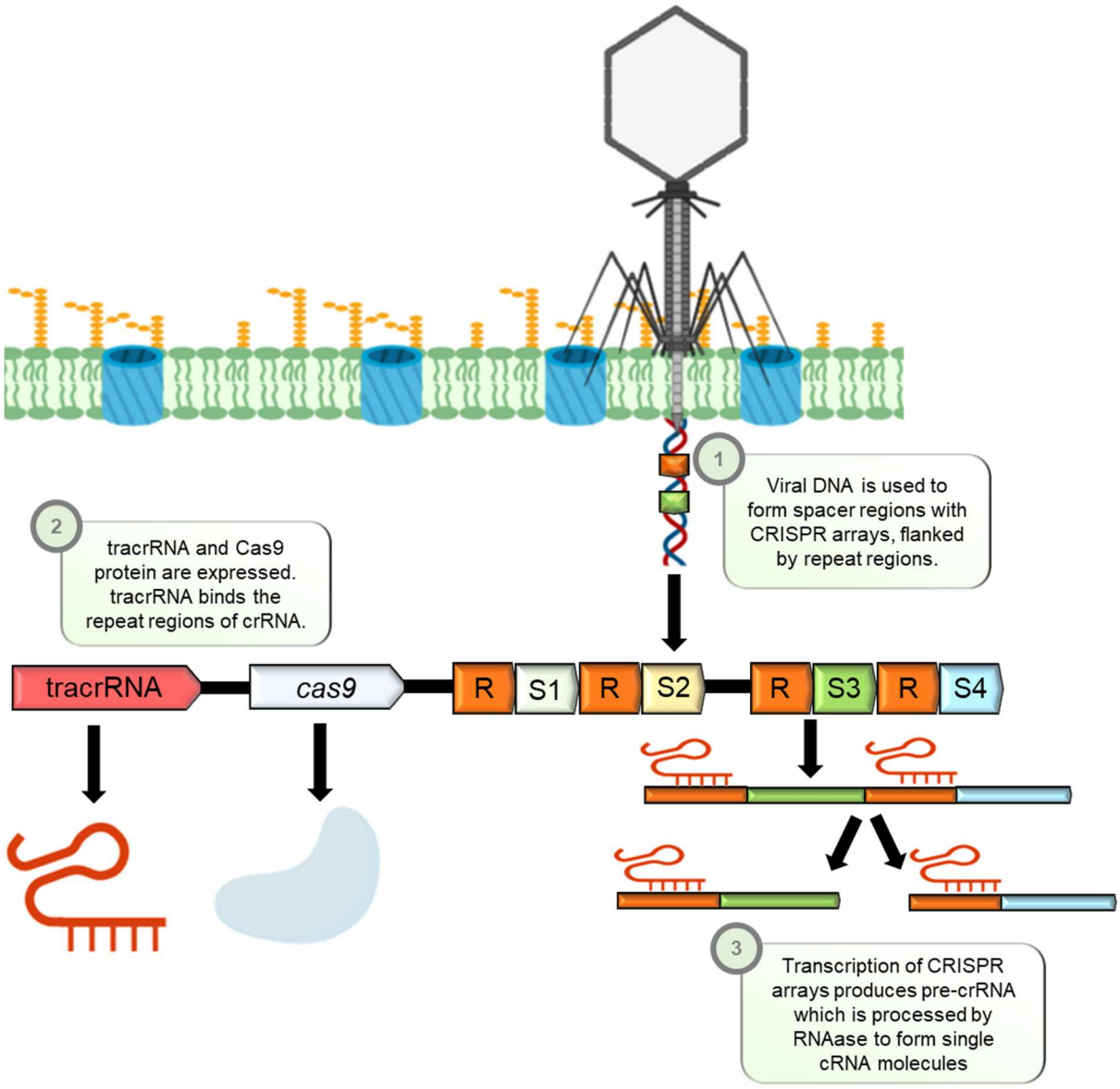


Figure 1.11: Production of CRISPR array from viral DNA. *Viral DNA is inserted into the host cell, and small spacer regions are added to the CRISPR array, flanked by direct repeating elements. Transcription and processing by RNAase of these regions leads to production of mature crRNA, which in complex with tracrRNA associates with the Cas9 nuclease (Jiang and Doudna 2017).*

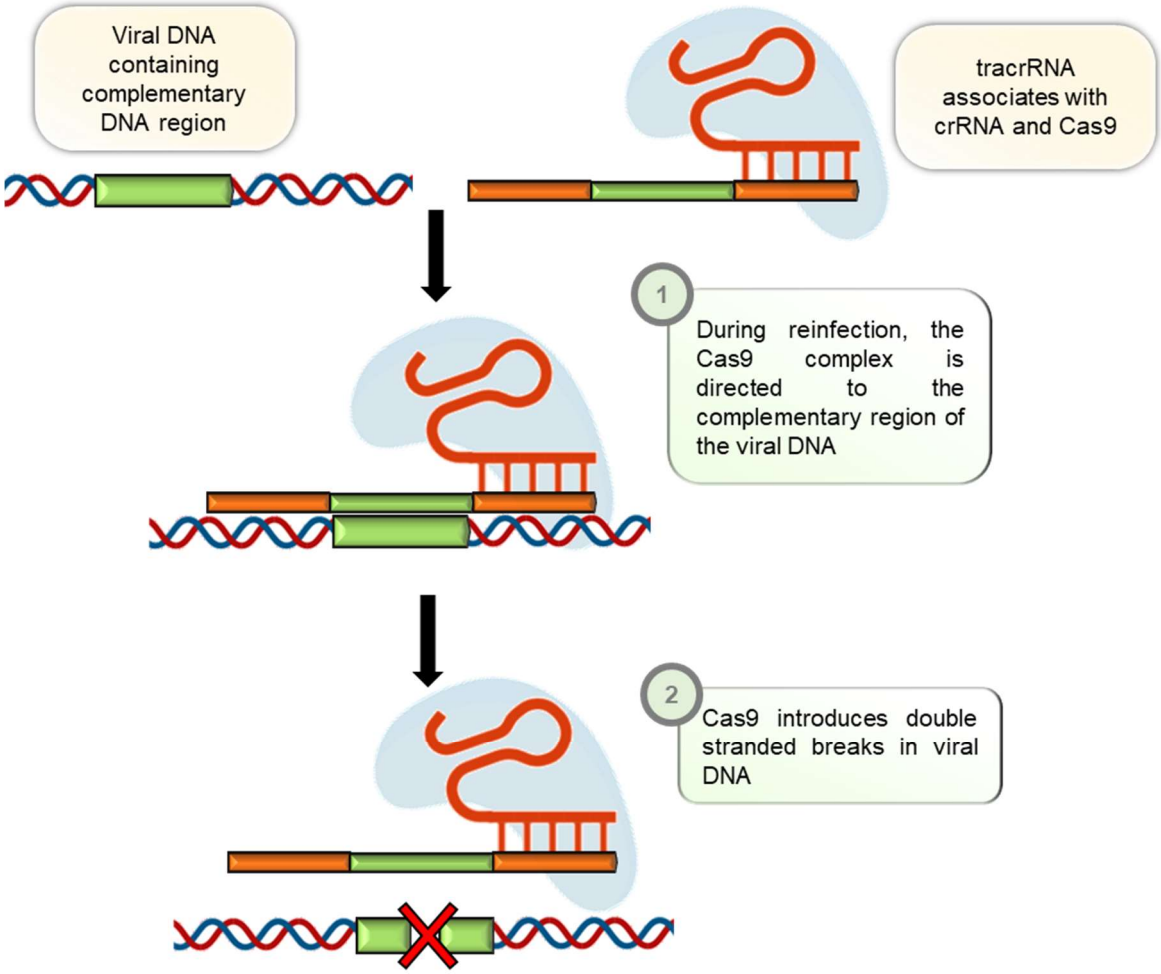


Figure 1.12: Activity of Cas9 nuclease. *Subsequent viral infections with genomes containing the region incorporated into the spacer regions of the CRISPR array are hybridised with the Cas9 complex. Cas9 forms a double stranded break in the viral DNA (Jiang and Doudna 2017).*

1.4.11 Biotechnological applications of CRISPR-Cas

Aside from the biological role in phage-resistance systems, CRISPR-Cas has been widely adopted for biotechnological purposes. The most heavily utilised CRISPR gene manipulation system is the type II CRISPR-Cas9, commonly from *Streptococcus pyogenes* (Gong et al. 2019). In this *S pyogenes* system, crRNA is hybridised with tracrRNA, directing the Cas9-RNA complex to the complementary viral sequence. Cas9 incorporates an HNH nuclease domain which cleaves the complementary target DNA strand, and a RuvC-like nuclease that cleaves the non-hybridised strand (Rodríguez-Rodríguez et al. 2019). The complex can be targeted to any 18-24 nt DNA motif to produce blunt end cuts. This revolutionised genome editing in eukaryotes due to the utilisation of non-homologous end joining in eukaryotic systems. Genome editing in a wide range of prokaryotes was already feasible due to the array of molecular processes for editing such as λ-red recombineering. Beyond using Cas9 for targeted DNA cleavage, mutant Cas proteins can be engineered to bind, but not cleave, target sequences. This can be used to effectively block other proteins from interacting with the target DNA sequence such as transcription factors. Cas9 can be fused to reporter elements such as GFP to allow for specific location labelling (Ishino, Krupovic, and Forterre 2018).

1.4.12 Evasion of CRISPR/Cas

Following a phage infection, a subset of bacterial and archaeal hosts incorporate spacer sequences of the phage genome. These spacer sequences are transcribed and cleaved to form small CRISPR RNAs. These crRNAs are complementary to phage genetic material and hybridise with incoming genetic material. These crRNAs deliver nucleases to the target phage genome for degradation (Stern and Sorek 2011b). Phages encoding non-canonical bases within their CRISPR recognition sites can evade crRNA targeting. Phages may also mutate or lose their spacer sequence in between the first and second round of infection (Barrangou et al. 2007).

Anti CRISPR (Acr) proteins are utilised by phages and mobile genetic elements to inhibit DNA recognition and cleavage by host CRISPR-Cas immune systems. These proteins can directly interact with the Cas nuclease to block DNA recognition, such as the mechanism of AcrIIA4 which binds the Pam recognition domain of Cas9 (Marino et al. 2020). The Cas9 nuclease can also be inhibited by AcrIIC1, which prevents DNA cleavage by associating with the HNH nuclease domain (Rauch et al. 2017). Other mechanisms of blocking CRISPR-cas activity including preventing crRNA loading and blocking effector complexes from being built.

1.4.13 Classification of restriction enzymes

Restriction modification (RM) systems operate based on the methylation of specific DNA motifs, facilitated by a methyltransferase. The cognate endonuclease typically degrades unmethylated DNA as the addition of methyl groups to host DNA is used to determine self from non self. Unmethylated phage DNA is susceptible to these endonucleases, preventing an infection from being established. Decades before the application of CRISPR-Cas systems, RM systems were demonstrated to cut specific DNA sequences, which revolutionised DNA cloning (Cohen et al. 1973; Morrow and Berg 1972). The endonucleases that comprise the cleavage activity of restriction modification (RM) systems can be classified into 4 main types (Type I-IV) characterised by the composition of their subunits, position at which they cleave relative to the recognition motif, cofactor requirements and substrate specificity (Loenen et al. 2013). There are subclasses that further divide these enzyme classes. Type I restriction endonucleases (REases) are typically grouped on their homology, type II REases are grouped on their catalytic properties, type III enzymes are combined REase and MTase complexes, and type IV are enzymes that are methylation dependent. Type IV enzymes are not classed as restriction modification systems due to the lack of a cognate MTase and their activity against methylated

DNA (Loenen et al. 2013; Pingoud et al. 2005). Bacteria employ a wide range of these systems in order to mount a thorough defence against phage infection.

1.4.14 Type I restriction enzymes

The earliest class of restriction enzymes were first introduced in the 1950s, when non-hereditary DNA modifications were discovered as a host-controlled variation (Bertani and Weigle 1953; Luria and Human 1952). Type I systems are present in roughly half of prokaryotic genomes and form a significant defence against introduction of foreign DNA, greatly reducing the success of transformation, conjugation and transduction. This produces a barrier to horizonal gene transfer (HGT) as unmethylated foreign DNA is readily degraded (Cooper et al. 2017). Evidence for the degradation and methylation of DNA was provided in the 1960s, adding insight into the mechanisms by which reversible changes to viral DNA affected progeny phage abilities to grow within their bacterial hosts (Dussoix and Arber 1962). RM systems comprising type I REases were demonstrated to consist of an element responsible for methylation of DNA, and another element responsible for the cleavage of DNA that did not have this methyl modification (Loenen et al. 2014). A key characteristic of type I systems is the translocation along DNA facilitated by the hydrolysis of ATP, initiated by the recognition of foreign DNA, driving endonuclease activity and supressing methyltransferase activity (van Noort et al. 2004). In the absence of divalent cations, protein-DNA complexes show disrupted hydrogen bonding networks and DNA contortion by the enzyme is not observed. DNA cleavage typically occurs as the result of the enzyme complex stalling against another complex (McClelland and Szczelkun 2004). As a result, there are no distinct cutting patterns produced by these enzymes.

Type I REases are formed of several subunits denoted Hsd, for host specificity. The recognition sequences for Type I enzymes are formed of two asymmetric parts, separated by a non-specific spacer region of 6 to 8 bp. For example, the REase EcoKI recognises the bipartite DNA sequence $AACN_6GTGC$ (Murray 2000). The EcoKI system can significantly reduce the success rate of infecting phages, allowing fewer than 1 in 10^5 to avoid degradation (Cooper et al. 2017). The genes encoding for each of the subunit components are denoted host specificity genes (*hsd).* The assembled protein complex has a stoichiometry of 1:2:2 comprised of HsdS, HsdM and HsdR respectively (Jindrova et al. 2005; Bourniquel and Bickle 2002).

HsdS is responsible for the recognition of the target sequence and as a monomer forms the core of the system. The majority of HsdS subunits are roughly 50 kDA and contain two target recognition domains (TRDs), with each domain recognising a different part of the target sequence. The 3′ end of the target sequence is recognised by the C-terminal TRD and the 5′ end is recognised by the N-terminal TRD (Fuller-Pace and Murray 1986; Dryden, Murray, and Rao 2001; Seidel et al. 2008). However, there are systems utilising smaller HsdS proteins that have been shown to contain only one TRD. Single TRD HsdS proteins are able to dimerise in order to recognise symmetrical target sequences (MacWilliams and Bickle 1996). HsdS can form a complex with HsdM as M_2S_1 which has methyltransferase activity and will methylate hemi-methylated target sequences (Kennaway et al. 2009). HsdM subunits recognise the methylation status of target sequences, transferring a methyl group from S-Adenosyl-L-methionine to adenine residues to form N6-methyl adenine (N6mA). This allows for the semi-conservative replication of DNA to produce two hemi-methylated products, containing a methylated parental strand, and a non-methylated daughter strand which can in turn be methylated again (Kozdon et al. 2013). Despite adenine being the most commonly methylated base via HsdM activity, more recently cytosine residues have been shown to undergo methylation to form N4-methyl cytosine (N4mC) (Morgan et al. 2016).

HsdR subunits are responsible for the restriction of unmethylated DNA and bind either side of the M_2S_1 complex to HsdM. The hydrolytic activity of HsdR is only activated when the full complex encounters unmethylated target sequences (Kennaway et al. 2009). HsdR subunits contain a PDE(x)K nuclease domain and an SF2 helicase domain (McClelland and Szczelkun 2004). The MTase core binds to the target sequence, and in the absence of methylation at key residues, each HsdR subunit translocates DNA resulting in loops of dsDNA being formed. ATP is hydrolysed in order to process the DNA and cleavage typically occurs when it can translocation is blocked by contact with another complex (Seidel et al. 2008).

It has been shown that the Ocr (obstruction of classical restriction) protein mimics DNA and confers protection to the T7 phage against type I restriction endonucleases, and also protects against BREX (Figure 1.13). Ocr binds PglX specifically, allowing T7 to circumvent BREX defences. The presence of the T7 0.3 gene that encodes Ocr allows the phage to form progeny phage and lyse its bacterial host, whereas T7 Δ*0.3* mutants cannot. Ocr was shown to bind PglX via affinity chromatography pull-downs, suggesting that the interaction of Ocr within the DNA binding domain on PglX is responsible for the inhibition of BREX. Only a slight reduction of methylation of BREX motifs was observed in Ocr+ BREX+ strains compared to Ocr- BREX+ (Isaev et al. 2020). Ocr is the first protein to be produced during T7 infection, inhibiting host nucleases, allowing the remainder of the T7 genome to be replicated and transcribed without restriction (Roberts et al. 2012). Similarly, coliphage P1 masks its restriction sites by injecting DarA and DarB simultaneously to its genome. DarA and B are packaged into the phage head during assembly and prevent type I RM systems from binding (Iida et al. 1987).

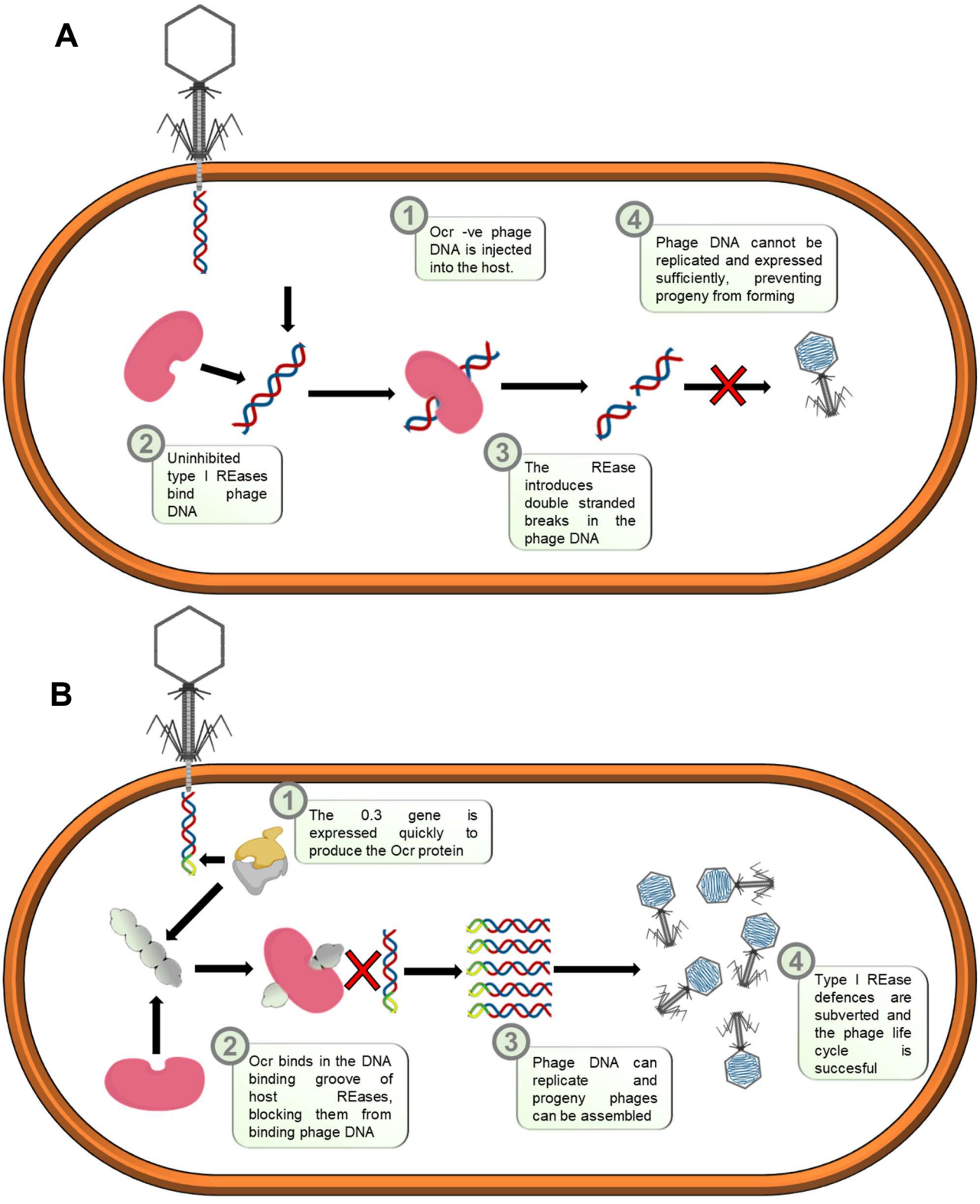


Figure 1.13: Mechanism of Ocr inhibition of type I REases*. A) Ocr -ve phages are susceptible to restriction and phage infection is prevented. B) Ocr successfully inhibits the activity of the host endonuclease allowing phage progeny to form following a successful infection.*

1.4.15 Type II restriction enzymes

The discovery and development of type II REases has largely shaped the field of molecular biology today. The key characteristic of type II enzymes is that DNA cleavage occurs at defined sites, either directly within the target sequence or close by. Commercially, this class of REases have significant biotechnological uses in laboratory practise for cloning and producing distinct DNA fragments (NEB. 2020). Type II REases exhibit significant heterogeneity compared to other classes of REases as they are classified on their catalytic properties, rather than sequence homology (Pingoud, Wilson, and Wende 2014). These enzymes produce DNA fragments with distinct, reproducible properties allowing for routine analysis via DNA gel electrophoresis. Similarly to type I REases, they typically require a divalent metal cation as a nuclease cofactor to coordinate interactions with target DNA. Certain enzymes such as EcoRV can still bind DNA non-specifically in the absence of metal ions, but exhibits no hydrolytic activity (Zahran et al. 2011). EcoRI can bind DNA in the presence of Ca^{2+} but is catalytically inactive without another metal ion present such as Mn^{2+} or Mg^{2+}. The activity of these enzymes is typically modulated by a methyltransferase, either as a separate entity or as a domain within the same protein chain (Pingoud, Wilson, and Wende 2014).

1.4.16 Further classification of type II restriction enzymes

Type II REases are further categorised into subclasses based on their features. Type II enzymes vary significantly in sequence and structure, and were originally categorised to include enzymes such as EcoRI and HindII, which cleave at palindromic sites and require divalent cations for hydrolysis (Boyer 1971). Enzymes of this type that were discovered early such as EcoRI and EcoRV became the basis for type II REase research until more enzymes were discovered and shown to exhibit large differences. Grouping these enzymes on phylogenetic proximity made little sense due to the lack of similarity observed between them. Instead, the

phenotypic properties of these enzymes form the classification structure (Niv et al. 2007; Pingoud et al. 2005; Roberts et al. 2003). These REases are divided into groups A-C, E-H, M, P, S and T, and many enzymes exist within more than one group (Loenen et al. 2013). Class II REases and their recognition sequences have been shown below (Table 1.1).

Enzyme	Classification	Recognition Sequence
SapI	Type IIA	5`-GCTCTTC(N)$_1$▼-3` 3`-CGAGAAG(N)$_4$▼-5`
BcgI	Type IIB	5`-▼(N)$_{10}$CGA(N)$_6$TGC(N)$_{12}$▼- 3` 3`-▼(N)$_{12}$GCT(N)$_6$ACG(N)$_{10}$▼- 5`
NgoMIV	Type IIF	5`-G▼CCGGC- 3` 3`-CGGCC▼G- 5`
MmeI	Type IIG	5`-TCCRAC(N)$_{20}$▼-3` 3`-AGGYTG(N)$_{18}$▼-5`
AhdI	Type IIH	5`-GACNNN▼NNGTC- 3` 3`-CTGNN ▼NNNCAG- 5`
DpnI	Type IIM	5`-GA▼ TC- 3` 3`-CT ▼AG- 5`
EcoRI	Type IIP	5`-G▼AATTC- 3` 3`-CTTAA▼G- 5`
FokI	Type IIS	5`-GGATG(N)$_9$▼-3` 3`-CCTAC(N)$_{13}$▼-5`

Table 1.3: Type II REases and their recognition sequences. *Cut sites are denoted by ▼. Nucleotides in RED have N6-methylation (Pingoud et al. 2005).*

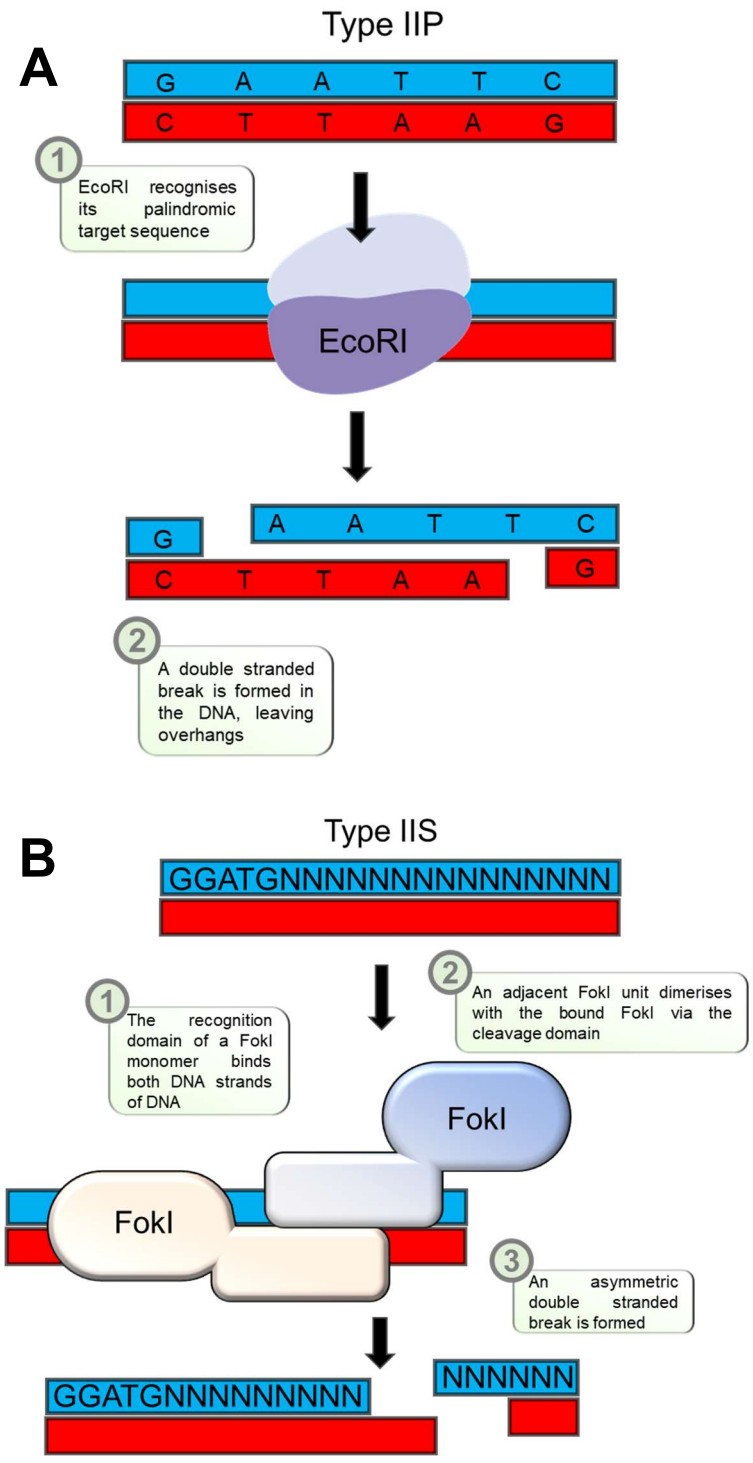


Figure 1.14: Mechanism of action for type II restriction enzymes. A) Cohesive end formation by a type IIP REase. *Introduction of staggered double stranded breaks in DNA by type IIP REase EcoRI. B)* Introduction of staggered double stranded breaks by a type IIS REase. *FokI is required to dimerise with an adjacent FokI unit in order to cleave its target.*

Type IIP REases are a generic collection of enzymes that recognise palindromic sequences and cleave at symmetrical sites within the target sequence or directly beside it (Roberts et al. 2003) (Figure 1.14). Type IIP represents the group that are routinely used in laboratories for cloning and gene analysis, such as EcoRI and HindIII. Type IIP enzymes typically act as homodimers, binding either side of a DNA duplex and cleave both strands simultaneously producing a double stranded break (Brickner and Chmielewski 1998). However, some type IIP enzymes operate as monomers and homotetramers. Monomeric REases of this class cut one strand of DNA at a time, producing nicks before the double stranded break is formed (Pingoud, Wilson, and Wende 2014). The arrangement of the active site within the enzyme subunits dictates the properties of the fragment terminus. For example, EcoRI recognises the palindromic sequence 5′-G^AAATC-3′ and cuts between G_1 and A_2 (denoted by ^), producing a staggered cut leaving a 3′ overhand of 4 bases. Both subunits of the EcoRI homodimer cleave their respective DNA strands at this site, leading to complementary overhands that are capable of hydrogen bonding to each other (Griffiths AJF 2000). Other enzymes such as EcoRV arrange their active sites symmetrically producing a straight cut within the target sequence, producing no overhang (Zhao et al. 2013). EcoRV recognises the sequence 5′-GAT^ATC-3′, cleaving between T_3 and A_4. The cognate methyltransferase M.EcoRV also recognises the G**A**TATC and methylates the first adenine within the sequence. The addition of this methyl group to A_2 at the N_6 position protects the DNA from restriction by sterically hindering the binding of EcoRV (Jurkowski et al. 2007). EcoRI adenine methyltransferase acts in a similar manner by methylating the second adenine within the GA**A**TTC. What made type IIP REases so important in the development of molecular biology was the manipulation of these controlled cleavage reactions to clone foreign DNA sections for genetic analysis. Restriction enzyme cloning allowed proteins from otherwise infeasible organisms to be overexpressed within a host such as *E. coli* and purified for analysis.

Type IIS REases such as FokI recognise non-palindromic, asymmetric sequences and cleave outside of their recognition target (Loenen et al. 2013). It requires two sites for DNA cleavage and dimerises in the presence of certain divalent metal ions to achieve this (Vanamee, Santagata, and Aggarwal 2001). Type IIS enzymes such as FokI typically bind DNA as monomers, however they cooperate with other FokI molecules to achieve dimerization, hence the necessity for more than one restriction site. Unlike the mechanism of type I REases, type IIP cut sites are fixed at a determined number of bases downstream or upstream of the recognition sequence. FokI recognises the target sequence 5′-GGATGN_{14}-3′, cleaving the top strand after N_9 and the bottom strand at N_{13} as shown below (Schierling et al. 2012). The reach of these enzymes is determined by the helical rotation of the bound DNA and the spatial arrangement of the recognition and cleavage domains.

Type IIA REases share a similarity to type IIS in that they recognised asymmetrical target sequences. As a result, FokI can be classified as both types due to its cleavage site being located outside of the asymmetrical recognition sequence (Loenen et al. 2013). Another type IIA enzyme, SapI, recognises the sequence GCTCTTCN_5, cutting as shown below (Pingoud, Wilson, and Wende 2014).

Type IIB REases cleave on both sides of the recognition sequence, producing two double stranded breaks within their DNA substrate (Marshall, Gowers, and Halford 2007). These enzymes typically have asymmetric recognition sequences that are bipartite, similar to type I REases, however they cut at distinct sites rather than at random. BcgI cuts double stranded DNA upstream and downstream of its recognition site, and requires Mg^{2+} and AdoMet for activity (Kong et al. 1993). Type IIB enzymes also require two recognition sites, typically preferring to bind *in cis* to another site on the same DNA molecule rather than *in trans* to a site on a separate DNA molecule. For this reason, enzymes such as BcgI exhibit much lower

activity on substrates containing only one site, due to the inability to crosslink to a separate site on another DNA chain (Marshall, Gowers, and Halford 2007). BcgI operates as a complex of two polypeptide chains, with subunit A containing domains for both methylation and restriction, and subunit B directing the complex to the recognition sequence. BcgI forms a complex of A and B in a 2:1 ratio, however many other type IIB enzymes have different subunit stoichiometries such as BplI with operates at a ratio of 1:1 (Vitkute et al. 1997; Kong et al. 1993).

Type IIC REases exist as a single polypeptide comprising both the endonuclease and methyltransferase as individual domains within a fusion protein. These enzymes typically have the PD-D/EXK nuclease domain situated at the N-terminus followed by an NPPY/F/W methylation domain (Pingoud, Wilson, and Wende 2014). HaeIV is an example of a single chain R-M system which recognises the sequence 5'-GAPyNNNNNPuTC-3′, cleaving either side of the target sequence and excising it (Piekarowicz et al. 1999). Due to the nature of this cleavage mechanism producing two DSBs, HaeIV also belongs to the IIB subclass of REases too.

Type IIE REases require more than one recognition site in order to cleave DNA. However, unlike other subclasses of type II REases, type IIE comprises a group of enzymes that bind multiple sites but only cleaves one. For example, EcoRII contains a domain for DNA cleavage that recognises the target sequence 5′-CCWGG-3′ at the C terminus, and two more at the N-terminus that recognise the same sequence but do not cleave it (Tamulaitis et al. 2006). The N-terminal domains act as allosteric activators that facilitate the hydrolytic activity of the C-terminal domain (Golovenko et al. 2009).

Type IIF REases also require more than one recognition site in order to cleave DNA, however they must form larger oligomeric structures such as homotetramers. Type IIF REases such as NgoMIV, Cfr10I and SfiI also belong to the PD-E(E/D)XK superfamily and form a dimer of dimer, with each primary dimer binding one DNA molecule each (Gasiunas et al. 2008). SfiI recognises the bipartite symmetrical sequence GGCCNNNNNGGCC, and the spacer sequence in between specific bases can largely affect the enzyme's activity rate (Williams and Halford 2001).

Type IIG encompasses a subclass of REases that recognise a variety of symmetric and asymmetric targets and are primarily grouped based on their requirement for AdoMet. Type IIG enzymes have evolved to combine the REase, MTase and subunit for substrate specificity within a single protein chain (Shen et al. 2011). There is significant crossover between enzymes of the IIB, IIC and IIG subclasses. MmeI is an example of a type IIG REase that consists of an N-terminal endonuclease domain, a C-terminal substrate recognition domain and is bridged by a methyltransferase domain in the middle of the chain. The MTase domain of MmeI only methylates one strand on its target DNA which is a significant enough modification to prevent restriction. MmeI is only sensitive to top strand N6mA methylation, and the methylation status of the bottom strand does not confer protection again resistance (Morgan et al. 2008). However, this would mean that during semi conservative replication, the daughter unmethylated DNA is vulnerable to restriction. How MmeI prevents itself from cleaving the unmethylated host DNA is unclear, however MmeI-like R-M systems are widespread and this potential toxicity is somehow negated (Morgan et al. 2009). MmeI cuts a full two rotations downstream from its asymmetric recognition site, with a spacer region of roughly 20 unspecified bases (Morgan et al. 2008). The long reach of this enzyme is likely facilitated by the positioning of the cleavage domain far away from the target recognition domain, separated by a helical spacer aligning the cleavage domain 20 bases away from the target sequence (Callahan et al. 2016)

Type IIH REases have separate methyltransferase and restrictions units, with the methyltransferase encompassing a similar organisation to a type I MTase. AhdI represents a type IIH REases with an endonuclease typical of a type II system, but with a type I like MTase. The recognition sequence of AhdI ($GACN_5GTC$) is bipartite making it typical of type I systems but exhibits rotational symmetry which is characteristic of type II systems (Marks et al. 2003).

Type IIM enzymes are grouped together as they recognise and cleave methylated DNA sequences. DpnI is a well utilised example and is routinely used within laboratories to remove methylated template DNA from reactions such as PCR (Krishnamurthy et al. 2015). DpnI has a short recognition sequence of Gm6ATC and cleaves one strand at a time, existing as a monomer (Pingoud, Wilson, and Wende 2014).

Type IIT represents a group of heterodimeric enzymes in which the majority are composed of two separate protein chains. Each protein chain contains a catalytic site and as a complex recognises asymmetric target sequences. Each subunit cleaves its own DNA strand, and so disruption of a single subunit results in DNA nicking due to cleavage of only one strand (Shen et al. 2018). This has led to the development of 'nickases', allowing for controlled nicking of dsDNA that can be used to target either the top or bottom strand depending on which enzyme subunit is disrupted (Wei et al. 2016). Some type IIT enzymes exist as a fusion protein on a single chain, but still act as a heterodimer functionally due to the difference within the TRD.

Prior to the subclassification system, all type II enzymes were grouped based on the premise that they cleave near to their recognition site. However, it is clear that within this class of REases, significant catalytic variation is observed. Before the discovery of type II REases, restriction assays were performed by measuring viscosity. However, type II enzymes were found to produce distinct fragments allowing for more precise analysis via gel electrophoresis

(Sharp, Sugden, and Sambrook 1973). This later led to the technology that is widely referred to as DNA fingerprinting (Jeffreys, Wilson, and Thein 1985).

1.4.17 Type III restriction modification systems

Type III REases are multisubunit complexes encoded by *res* and *mod* genes, and require two recognition sites to be present organised in an inverted repeated orientation (Butterer et al. 2014). Type III REase complexes such as EcoPI and PstII typically assemble with a stoichiometry of M_2R_1, however some assemble as M_2R_2. These enzymes target unmethylated DNA sequences, require ATP for hydrolysis, and AdoMet for methylation (Rao, Dryden, and Bheemanaik 2014). What separates them from other classes such as type I is that they have fixed cut sites close to the recognition sequence.

Type III REases recognise short asymmetric sequences (5-6 bp) and cleave downstream (25-27 bp) leaving 5′ overhangs. The Mod subunits are N6 adenine methyltransferases, and methylate only one strand of their target DNA, which is sufficient to prevent restriction (Sistla and Rao 2004). The presence of magnesium is required for enzyme activity, and DNA cleavage is dependent on ATP hydrolysis (Butterer et al. 2014). There are significantly fewer type III enzymes characterised than type II, as despite being an area of interest biochemically, as they do not currently provide as much of a biotechnological output as type II.

1.5 Type IV Restriction Enzymes

The most distinguishing property of type IV REases is the recognition and cleavage of modified DNA substrates. This class of REases contains a wide variety of enzymes that recognise base modifications such as methylcytosines (5mC), hydroxymethylcytosines (5hmC) and methyladenine (N6mA) (Xu et al. 2011; Lepikhov et al. 2001). Type IV REases also act on DNA substrates that have been further modified such as glucosyl-hydroxymethylcytosine (glc-5hmC). The addition of a glucose moiety to 5hmC can often protect a DNA substrate that would be susceptible to restriction in its absence (Loenen and Raleigh 2014). Similarly to type II REases, they are clustered based on their catalytic activity rather than by sequence or structural homology, however many share conserved domains. Before discussing details of type IV REases it is important to examine the types of DNA modifications that might be encountered.

1.5.1 DNA modifications in eukaryotes, prokaryotes and phages

DNA modifications can exist as both covalent binding of methyl groups directly to a DNA base such as cytosine, and changes to DNA ultrastructure such as acetylation of lysine residues of histone-DNA complexes (Ali et al. 2018) (Kumar, Chinnusamy, and Mohapatra 2018). This review will focus on the reversible covalent modification of DNA bases in the context of restriction and epigenetics.

1.5.2 Covalent DNA modifications in eukaryotes

The major modification of eukaryote DNA is the major epigenetic marker 5mC and is one of the most characterised epigenetic systems. Other less common modifications include 5hmC, 5-formylcytosine (5fC) and carboxylcytosine (5caC) (Jin and Liu 2018). Patterns of 5mC

modifications typically occur within DNA promoter regions, with the majority of CpGs in the human genome being methylated. Plants are the most highly methylated eukaryotic organisms, exhibiting up to 50% of all their cytosines being methylated (Zabet et al. 2017). CpG refers to any cytosine followed directly by a guanosine in the 5′>3′ direction. CpGs in the human genome are typically clustered within CpG island located within promoter regions and typically repress transcription, effectively reducing gene expression (Lim and Maher 2010). This achieved by preventing transcription factors to bind, or by recruiting repressors (Moore, Le, and Fan 2013). 5hmC differs significantly from 5mC in humans in that it has a contrasting permissive effect on gene expression (Mellén et al. 2012). It is distributed less at random that 5mC and its absence in certain tissues provides a potential biomarker for cancer diagnostic due to low levels within tumour tissues (Song et al. 2017). Type IV enzymes provide a potential method for identifying DNA modifications. By substituting the required nucleotide cofactor such as GTP for a non-hydrolysable analogue, type IV enzymes will bind their targets but will be unable to undergo hydrolysis. This presents an effective opportunity for localising and tagging motifs (Irizarry et al. 2008).

1.5.3 Covalent DNA modifications in prokaryotes

Whilst DNA methylation serves to primarily control gene expression in eukaryotes, it has other key roles in prokaryotes such as modulation R-M systems. Modification of G**A**TC motifs in γ-proteobacteria is modulated by DNA adenine methylase (Dam), which is referred to as an orphan enzyme in that it does not exist as part of an R-M system (Adhikari and Curtis 2016). Dam transfers a methyl group from AdoMet to the N6 position of adenine produced N6-methyladenine (Payelleville et al. 2018). Analogous to the effects of 5mC methylation in eukaryotes, N6mA modifications within G**A**TC motifs serves to control gene expression by repressing transcription. One such example of regulation in *E. coli* by N6mA modification is found within the pyelonephritis associated pili (*pap)* region (Stephenson and Brown 2016). This phase variation of expressing or repressing *pap* is modulated by the methylation state of the intergenic region of *papIB* (Hernday, Braaten, and Low 2003).

In *E. coli,* 5mC is also utilised. It is produced by the transfer of a methyl group from AdoMet to the C5 position of cytosine and is a pre-replicative modification. DNA cytosine methylase (Dcm) recognises the sequences 5′-C**C**WGG-3′, modifying the highlighted internal cytosine residues (Beaulaurier, Schadt, and Fang 2019). Dcm activity results in almost every CCWGG motif being methylated with very few remaining unmethylated (Casadesus and Low 2006). In prokaryotes, this modification typically occurs to protect against host nucleases that contain cytosine residues in their target sequence (Takahashi et al. 2002).

Phages also incorporate DNA modifications, in order to evade their bacterial host's restriction systems that would otherwise cleave the phage genome. All unmodified cytosines are replaced by 5hmC on T-even phage genomes, preventing restriction enzymes that have cytosines within their recognition sequence from targeting them (Wyatt and Cohen 1953). Within T-even coliphages such as T4, hydroxymethylcytosine modifications occur pre-replication, substituting the unmodified base for the methylated or hydroxymethylated counterpart. To produce 5hmC, unmethylated dCTP is dephosphorylated forming dCMP which is then hydroxymethylated by dCMP hm-transferase to form hydroxymethylcytidine monophosphate (5hm-CMP). Kinases add on two phosphate groups to produce 5hm-CTP which is then incorporated into the daughter DNA strand during replication (Weigele and Raleigh 2016). The dNTP synthesis complex (DSC) is provided with additional dCTP as a result of phage encoded DenA and DenB nucleases which degrade the host genome (Souther, Bruner, and Elliott 1972).

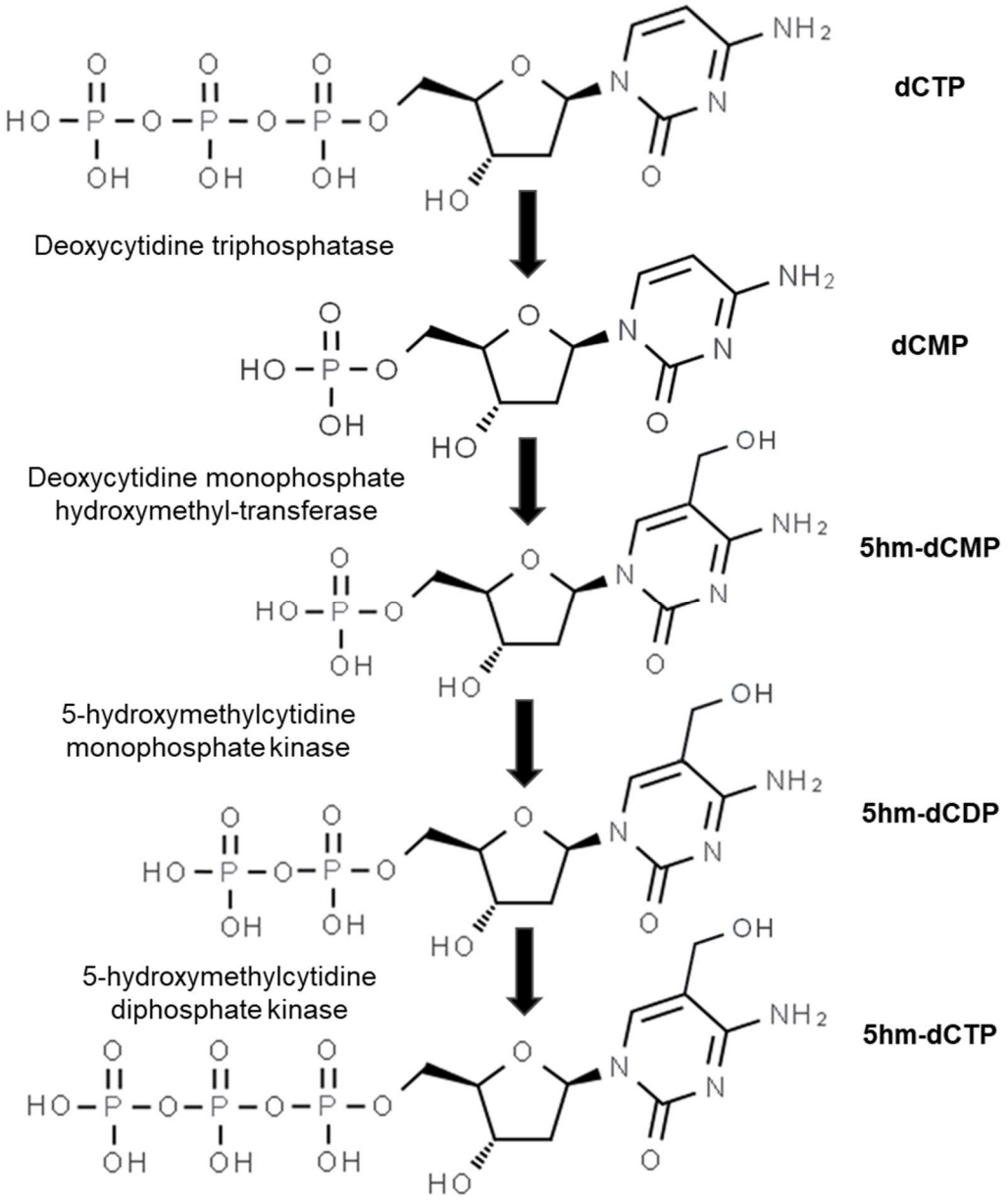


Figure 1.15: Biosynthesis of 5hm-dCTP. *Structural formula showing systematic production of 5-hydroxymethyl deoxycytidine triphosphate to be incorporated into phage genomes via DNA polymerase.*

1.5.4 DNA hypermodifications and other non-canonical bases in phage genomes

Phage genomes exhibit the widest variety of DNA base modifications of all organisms, and many of these secondary modifications occur post replication (Weigele and Raleigh 2016). DNA hypermodification refers to further changes of non-canonical bases such as 5hmC by the addition of groups such as glucose. Hypermodifications primarily serve to render methylated DNA that is susceptible to type IV REases unable to be restricted due to the addition of a moiety that changes the composition of that base. In T-even phages, 5hmC residues are glycosylated by glucosyl-transferases that transfer a glucose moiety from UDP-glucose to the hydroxyl group of 5hmC (Sommer, Depping, Piotrowski, and Ruger 2004).

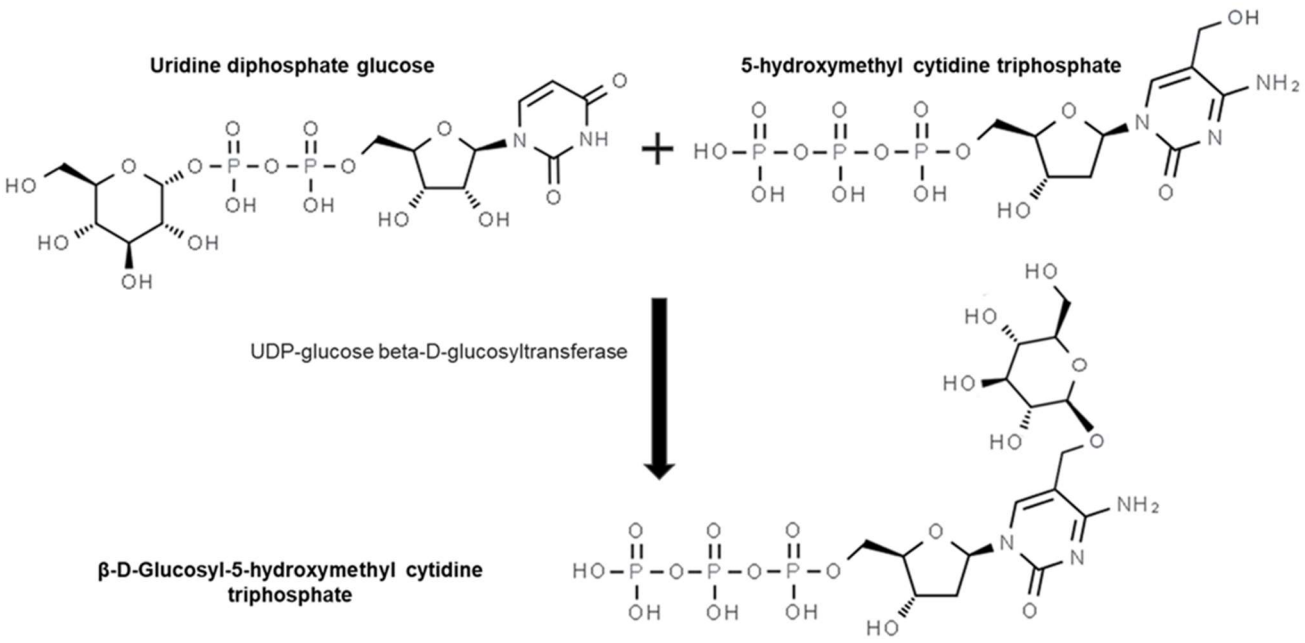


Figure 1.16: Modification of 5hm-dCTP by β-glucosyltransferase. *Condensation of uridine diphosphate glucose and 5-hydroxymethyl cytidine triphosphate to produce a hypermodified DNA base.*

1.5.5 Carbamylation

The *E. coli* phage μ incorporates N6-carbamoyl-methyladenine into its genome, accounting for ~15% of all adenine bases. This hypermodification of N6-methyladenine confers significant resistance to a wide array of REases that are active against non-hypermodified genomes (Allet and Bukhari 1975). This reaction is catalysed by the Mom protein which transfers a formamide moiety to N6mA from a coenzyme-A carrier (Kaminska and Bujnicki 2008).

1.5.6 Substitution with 2,6-diaminopurine

The *Synechococcus* phage S-2L contains 100% substitution of adenine for 2,6-diaminopurine. 2,6-diaminopurine nucleosides undergo several rounds of phosphorylation to produce 2,6-diaminopurine nucleotides which are incorporated into the genome during replication (Khudyakov et al. 1978) (Weigele and Raleigh 2016).

1.5.7 Guanine modifications

The *Shigella* phage DDVI contains 7-methyguanine, in which the nitrogen at position 7 has been methylated and accounts for ~1% of all guanine-based residues. (Nikolskaya, Lopatina, and Debov 1976; Weigele and Raleigh 2016). Guanine methylase is induced by a DDVI infection and transfer a methyl group from AdoMet to guanine. The incorporation of 7-methylguanine aids protection against restriction in B-type *E. coli* strains (Nikolskaya et al. 1979). 7-methylguanine is an essential 5′ DNA and RNA modification in eukaryotes allowing for the efficient translation of mRNA transcripts (Cowling 2009).

1.5.8 Unmethylated 5-hydroxycytosine

The N-17 phage of *Shigella flexneri* substitutes all cytosine residues for 5-hydroxycytosine (5hC), where hydroxylation of the 5th carbon occurs directly, rather than to the methyl group observed with 5hmC (Kchromov et al. 1980). *Rhizobium* phage RL38JI incorporates 5hC in its genome and also incorporates glycosylated 5hC (Weigele and Raleigh 2016).

1.5.9 Thymine and uracil derivatives

The *Bacillus* phage PBS1 completely substitutes thymidine for uridine in its genome, facilitated by increasing the dUTP availability in contrast to dTTP availability. Uracil DNA glycosylase is inhibited by UDG inhibitor (UGI), preventing uracil from being removed from the genome by DNA repair mechanisms (Wang and Mosbaugh 1989). The concentration of dTTP is reduced by the activity of a phosphorylase that removes the γ-phosphate, leaving only dTDP. The concentration of dUTP is increased by the activity of dCTP deaminase which ensures an equilibrium of dUTP and dCTP availability (Price 1974).

Simarly to PBS1, another *Bacillus* phage SPO1 substitutes 100% of its thymidine bases for 5-hydroxymethyldeoxyuracil (5hmdU) instead. SPO1 also uses cytidine deaminase to produce dUTP, which is then hydroxylated to produce 5hmdU. In phages such as SPO10, 5hmdU can be further modified via the addition of pyrophosphate to the 5-hydroxyl group, which can be further modified to produce hypermodified bases such as α-glutamylthymine and α-putrescinylthymine (Weigele and Raleigh 2016).

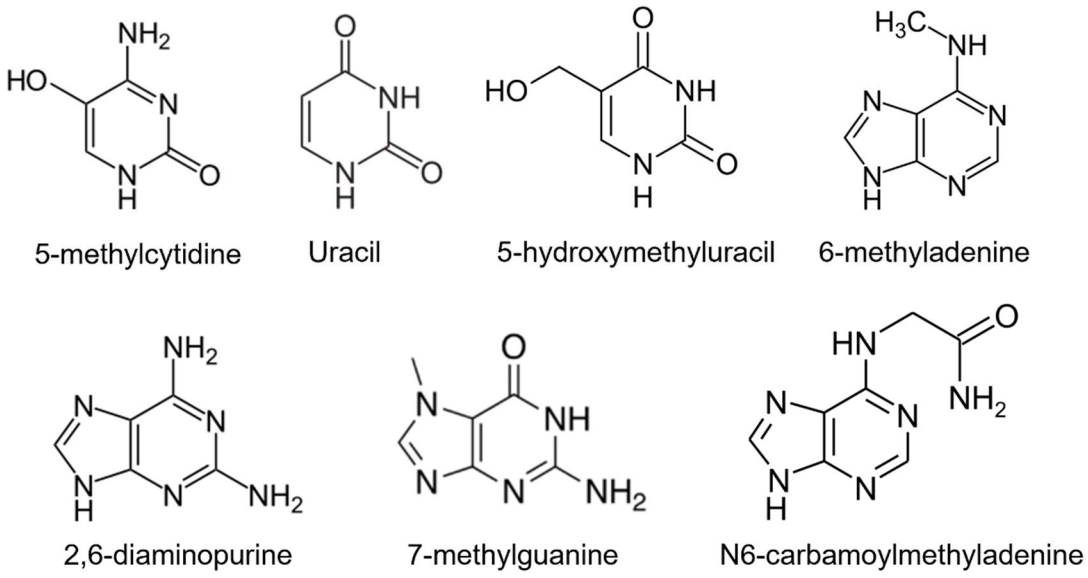


Figure 1.17: Structural formulae of non-canonical bases. *Non-canonical bases incorporated in phage DNA genomes as a method of evading host defences.*

1.5.10 DNA modification dependent restriction endonucleases

As previously discussed in this chapter, type IV REases recognise and cleave methylated DNA sequences. Type IIM REases such as DpnI share this characteristic, and therefore Loenen et al advocate including type IIM enzymes within the type IV family (Loenen and Raleigh 2014). The restriction endonuclease database (REBASE) clusters type IIM and type IV enzymes separately (Roberts et al. 2015).

1.5.11 Original studies of modified DNA hydrolysis

The initial identification of a type IV restriction REase was of the RgIB (restricts glucoseless phage) system and its activity against T-even phage glucosyl transferase mutants. T-even phages incorporate glc-5hmC as a result of glucosyl transferase (*gt)* activity, which protects the phage genome from restriction. T-even *gt* mutants were found to be susceptible to cleavage from RglB whereas WT T-even phages were not (Revel and Georgopoulos 1969). The *rgl*

genes were renamed *mcrA* for the prophage-like element, and *mcrBC* for the two genes encoding the restriction system. The McrBC system requires GTP rather than ATP for hydrolysis, as well as Mg^{2+}, and binds specifically to DNA containing methylated cytosine residues (Stewart et al. 2000b). Cleavage via McrBC is conditional on the presence of two recognition sites, with optimal cleavage occurring between a spacer element of 55-103 bp. Recognition elements were shown to have no requirement to be situated on the same face of the DNA helix as changing the spacer size by 4 bp has no effect on hydrolytic activity of McrBC (Stewart and Raleigh 1998). McrBC primarily forms tetradecameric complex in the presence of GTP, and translocates DNA until a collision with a secondary complex occurs. Upon this collision, DNA is hydrolysed approximately 30 bp away from the recognition sequence (Panne et al. 2001). The translocation of DNA and the stimulation of cleavage resulting from complex collision draws many comparisons to the activity of type I REases. Later, other systems such as Mrr (modified rejection and restriction) were shown to restrict N6mA and 5mC modified DNA (Zheng et al. 2010; Waite-Rees et al. 1991).

1.5.12 Diversity within type IV systems

Type IV systems are not clustered based on their catalytic activities in the way that type II systems are. For example, the type IIS are clustered on their characteristic of recognising asymmetric sequences and cleaving outside of their recognition target. Type IV REases vary significantly regarding the sequence nature that they recognise, the methods in which they cleave and the modifications in which they recognise. Therefore, the type IV classification covers a broad spectrum of enzymes that recognise DNA containing modifications. There is little sequence homology collectively between type IV enzymes, with a variety of DNA binding and endonuclease domains present, as well as significant variability in nucleotide and cofactor requirements (Loenen and Raleigh 2014).

1.5.13 Endonuclease domains within type IV systems

A common nuclease domain within type IV REases is the HNH domain, grouping them to the HNH superfamily along with colicins, transposases and DNA packaging factors (Cymerman et al. 2006). A loop region connects two anti-parallel β-sheets within the core of the HNH domain, flanked by an α-helix. The first histidine residue located at the terminus of the first β-sheet, is involved in DNA cleavage and the second histidine binds a metal cation. The asparagine residue coordinates the β-sheets together (Flick et al. 1998). HNH endonuclease domains are found within the type IV system of McrA.

PD-(D/E)XK phosphodiesterases are also commonly incorporated into type IV systems. Outside of RM systems, this domain is utilised by systems for tRNA splicing, DNA repair, Holliday junction resolving and DNA polymerase termination (Steczkiewicz et al. 2012; Aravind, Makarova, and Koonin 2000). The conserved structural core of the domain is comprised of four β-sheets, flanked by an α-helix at each side. This family of proteins exists in type I, II and II RM systems as well. The D/E and K residues are located within the active site of the domain in the Y-shaped bend of the 2nd and 3rd β-sheets (Feder and Bujnicki 2005).

The *S. aureus* SauUSI contains a phospholipase D (PLD) domain located within its N-terminal region (Loenen and Raleigh 2014). The PLD superfamily is another highly diverse collection of enzymes and other proteins involved in DNA hydrolysis, toxicity, viral envelopes and membrane metabolism (Interthal, Pouliot, and Champoux 2001). There are other domains incorporated by type IV REases that are less characterised than HNH and PD-(D/E)XK domains.

The nuclease domain of GmrD (sometimes expressed as a GmrSD fusion) contains a DUF1524. Despite a lack of further characterisation, DUF1524 has been implicated in DNA restriction in GmrSD homologues.

1.6 Phage Defence Evasion Mechanisms

The incorporation of modified DNA bases by phages leads to effective evasion of host defence systems. However, beyond the evasion strategies already described for BREX, CRISPR-cas and RM systems, there are many other evasion systems used to inhibit host defences.

1.6.1 Manipulating host receptors binding and access

Phage λ typically binds to the *Escherichia* maltose pore LamB, which is also the receptor for phages such as K10 and TP1 (Roa 1979; Chatterjee and Rothenberg 2012). As a defence mechanism against phage absorption, *E. coli* can suppress *lamB* expression, reducing the number of receptors available to the invading phage. A pattern of four mutations within the distal domain of the λ protein J allows the phage to interact with OmpF of the host (Samson et al. 2013). A trimeric β-barrel pore is formed by both OmpF and LamB which may account for why the mutant receptor binding protein J can interact with both.

Phages that interact with host lipopolysaccharide (LPS) as their receptor can often be prevented access by the modification of the LPS. *E. coli* strains that express modified LPS however can still be susceptible to T7 phages that encode mutations within the tail tubular proteins A and B (Qimron et al. 2006). *E. coli* O18:K1:H7 strains expressing the K1 capsular polysaccharide exhibit increased virulence. The sialic acid capsule reduces human complement activation and opsonisation, providing resistance to the innate immune system (Abreu and Barbosa 2017). The K1 capsule also forms a layer above host receptors for phage RBPs. Coliphages such as K1F and T7 encode endosialidases that actively degrade the capsule, allowing the phage to access the buried host receptors.

1.6.2 Negation of abortive infection systems

Abortive infection systems serve to prevent the spread of an invading phage by shutting down cellular responses in a controlled host death. This prevents the formation of progeny phages and the spread of the infection further within the host's community. Abortive infection can be mediated by toxin-antitoxin systems (Figure 1.18). In *Pectobacterium atrosepticum,* a type III TA system consisting of a protein toxin ToxN, and an RNA antitoxin ToxI, form a heterohexameric complex in a 3:3 stoichiometry (Blower, Short, et al. 2012). ToxI is transcribed to produce multiple repeats of the same sequence. ToxN is a sequence specific endoribonuclease of the Kid family that has cellular targets as well as ToxI. ToxN cleaves ToxI transcripts to produce 36 nucleotide length units, which bind the active site of ToxN to form a triangular complex that prevents toxicity (Short et al. 2018). The *P. atrosepticum* phage TE genome has been shown to incorporate short sequences homologous to the host's ToxI, TE phages that were ToxIN sensitive were shown to have 1.5 repeats of these pseudo-ToxI sequences, whereas ToxIN resistant escape phages were found to contain 4.5 to 5.5 repeats (Blower, Evans, et al. 2012). The phage encoded pseudo-ToxI mimics the activity of ToxI and forms a pseudoknot with ToxN, preventing it from cleaving its cellular targets and inducing host death. This allows the invading phage to propagate and subvert a toxin-antitoxin abortive infection system.

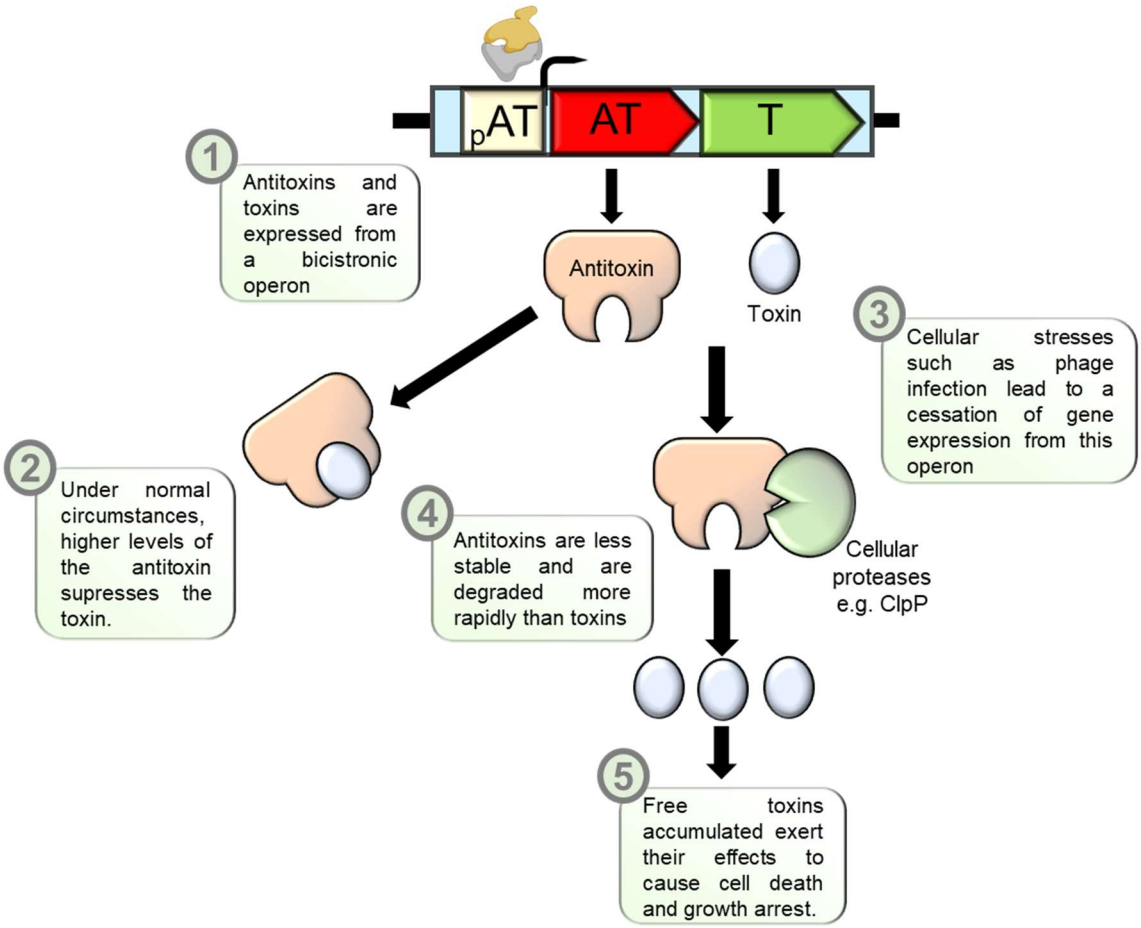


Figure 1.18: Interaction of toxin-antitoxin systems. *Antitoxins suppress the toxic effects of the toxin until infection occurs, resulting in unbound toxins exerting their effects to prevent phages from hijacking host machinery and producing progeny.*

1.6.3 Expression of inhibitory components for restriction enzymes

Phages encode a variety of systems designed to inhibit bacterial host defence systems. The T4 genome is heavily modified due to the incorporation of 5hmC, which is further modified by the glycosylation of these residues. In order to target these modified genomes, bacteria have evolved restriction systems that target genomes that contain these modified cytosine residues. As a counter-resistance measure, T4 phages produce the IPI protein, which inhibits the activity of the type IV REase GmrSD (Rifat et al. 2008b). IPI is expressed as a late gene and package as an internal head protein of progeny phages, which protects phage gDNA from restriction by GmrSD family enzymes during DNA injection.

1.7 Clustering of Phage Defence Systems

Phage defence systems target all aspects of the phage life cycle, from the point of phage interaction at the outer membrane in the form of adsorption prevention, to prevention of phage assembly later in the life cycle. Phage defence systems such as RM, and those closely associated such as BREX or DISARM, rely on the recognition of foreign DNA and its subsequent degradation. These systems rely on the activity of a methyltransferase that distinguishes host from foreign DNA. They share similarities with CRISPR-Cas as both systems employ a nuclease to degrade foreign DNA. However, the mechanism in which Cas enzymes identify their target differs by their use of crRNA components that direct the nuclease to the target, rather than the nuclease identification of methylation states. Abortive infection systems may involve the recognition of foreign DNA in order to trigger their activation, however the effector mechanisms employed are much diverse. As a result, strains that employ multiple phage defence systems can effectively muster a multi-level defence against phages, reducing the efficiency of DNA injection, degrading DNA that enters the cytoplasm and inducing host cell death in the event of a failure in first line defences. There has been a large proliferation in phage defence systems being discovered due to their association with other systems. BREX was first detected by Goldfarb et al when screening for genes associated with the Pgl system (Goldfarb et al. 2015). Similarly, the DISARM system was discovered by the identification of a DUF1998 protein within defence islands. This domain was predicted to have antiphage activity as it was found to be enriched within prokayotic defence islands (Ofir et al. 2018). The recently discovered PARIS phage defence system was identified a two component toxin-antitoxin system, due to its proximity to P4-like prophage elements in *E. coli* (Rousset et al. 2021). Our initial interest in the *Escherichia fergusonii* plasmid pEFER was due to the presence of a type I BREX system, however upon further characterisation it is clear that BREX is not the only phage defence system encoded by pEFER.

The clustering of defence systems provides prokaryotic hosts with a significant advantage, mediated by a multi-faceted approach to preventing phage proliferation. For example, phages that can evade a type II RM system by incorporating non-canonical bases may still be susceptible to an Abi system. RM systems are typically employed during the earlier stages of the phage life cycle, and if successful, negates the requirement for host cell growth arrest or death as a result of abortive infection. The association of these systems within defence islands may indeed make it easier to disseminate to new hosts, and provide a collectively greater selective advantage.

1.8 Project Aims

This project aimed to broaden our understanding of the mechanisms utilised by bacteria to protect against viral infection. The novel BREX resistance system contains uncharacterised components that exhibit different functionalities to previously characterised systems.

The primary aims of this study were:

1) Quantify the resistance potential of BREX against a variety of phages and ascertain which BREX components are essential for function.
2) Understand the regulatory systems of BREX.
3) Functionally characterise a novel type IV restriction endonuclease both *in vivo* and *in vitro*, demonstrating its phage resistance phenotype and capacity for specific DNA cleavage.
4) Obtain a 3D macromolecular structure for this type IV restriction endonuclease via X-ray crystallography.

Chapter 2: Materials and Methods

2.1 Media, Reagents and Solutions

Growth media, antibiotics (and other culture supplements) and solutions used during this study are detailed in Tables 2.1, 2.2 and 2.3 respectively. All growth media was sterilised by autoclaving at 121 °C for 20 min. Where appropriate, solutions were sterilised by vacuum filtration through 0.22 μm filters (Merck).

2.2 Bacterial Strains and Growth Conditions

All bacterial strains used in this study are detailed in Table 2.4. Unless otherwise stated, *E. coli* and *S. enterica* strains were grown at 37 °C, either on agar plates or shaking at 220 rpm for liquid cultures. Luria broth (LB) was used as the standard growth media for liquid cultures, and was supplemented with 0.35% w/v or 1.5% w/v agar for semi-solid and solid agar plates, respectively. Growth was monitored using a spectrophotometer (WPA Biowave C08000) measuring optical density at 600 nm.

Table 2.1: Growth media used in this study

Medium	Components per litre
Luria-broth	10 g Casein digest peptone 5 g Yeast extract 5 g NaCl For semi-solid agar, add 3.5 g agar For solid agar, add 15 g agar
2x YT	16 g Casein digest peptone 10 g Yeast extract 5 g NaCl
SM Complete Medium (Molecular Dimensions)	21 g/L Base medium 5 g/L Nutrient mix (added post-autoclave)

Table 2.2: Antibiotics and supplements used in this study

Solution	Stock concentration	Concentration (µg/ ml unless stated)
Antibiotics		
Ampicillin	1000 x stock, 100 mg/ml in dH_2O	100
Chloramphenicol	1000 x stock, 20 mg/ml in dH_2O	50
Gentamycin	1000 x stock, 20 mg/ml in dH_2O	20
Kanamycin	1000 x stock, 50 mg/ml in dH_2O	50
Spectinomycin	1000 x stock, 50 mg/ml in dH_2O	50
Tetracycline	1000 x stock, 10 mg/ml in 50% EtOH (v/v)	10
Supplements		
D-glucose	20% stock (v/v), 40 g in 200 ml dH_2O	0.20%
IPTG	1000 x stock, 1 M in dH_2O	1 mM
L-arabinose	20% stock (v/v), 40 g in 200 ml dH_2O	0.10%
Methionine inhibitory feedback mix	100 mg L-selenomethionine, L-lysine, L-threonine, L-phenylalanine 50 mg L-leucine, L-isoleucine, L-valine Add dH_2O to 10 ml	0.10%
X-gal	1000 x stock, 40 mg/ml in DMF	40

Table 2.3: Solutions used in this study

Solution	Components
Phage work	
Phage buffer	10 mM Tris pH 7.4 10 mM MgSO4 0.01% gelatin (w/v)
Protein work	
Lysis buffer (A500)	500 mM NaCl 20 mM Tris HCl pH 7.9 5 mM Imidazole pH 8.0 10% Glycerol (v/v)
Ni-NTA elution buffer (B500)	500 mM NaCl 20 mM Tris HCl pH 7.9 250 mM Imidazole pH 8.0 10% Glycerol (v/v)
Ni-NTA low salt elution buffer (B100)	100 mM NaCl 20 mM Tris HCl pH 7.9 250 mM Imidazole pH 8.0 10% Glycerol (v/v)
FPLC low salt buffer (A100)	100 mM NaCl 20 mM Tris HCl pH 7.9 5 mM Imidazole pH 8.0 10% Glycerol (v/v)
FPLC high salt buffer (C1000)	1000 mM NaCl 20 mM Tris HCl pH 7.9 10% Glycerol (v/v)
FPLC sizing column buffer	500 mM KCl 50 mM Tris HCl pH 7.9 10% Glycerol (v/v)
Sample storage buffer	500 mM KCl 50 mM Tris HCl pH 7.9 70% Glycerol (v/v)
Crystallisation buffer	200 mM NaCl 20 mM Tris HCl pH 7.9 2.5 mM DTT
10x DPMG buffer	0.2 M Tris pH 7.5 0.5 M Potassium acetate 0.1 M Magnesium acetate

Table 2.3 continued: Solutions used in this study

Solution	Components
DNA work	
50x TAE buffer (1 L)	242 g Tris base 57.1 mL Glacial acetic acid (17.4 M) 18.61 g EDTA, disodium salt Adjusted to pH 8.0
6x Loading dye	Thermo 6x DNA loading dye
Agarose gels	1x TAE 500 mg/ L EtBr
Solution A	9.9 ml 1M $MnCl_2$ 49.5 ml 1M $CaCl_2$ 198 ml 50 mM MES 742.6 ml MQ H_2O (For Solution A + glycerol, add 100 ml glycerol 642.6 ml H_2O)
SDS-PAGE	
10x Running buffer (1 L)	30.2 g Tris base 141 g Glycine 0.1 L 10% SDS (w/v)
Resolving gel (for 4 BioRad mini-gels)	6 mL 40% Acrylamide 6 mL 1.5 M Tris pH 8.8 240 µl 10% SDS (w/v) 240 µL 10% Ammonium persulphate 24 µl TEMED 11.5 ml dH20
Stacking gel (for 4 BioRad mini-gels)	1.5 mL 40% Acrylamide 3.75 mL 0.5 M Tris pH 6.8 240 µl 10% SDS (w/v) 240 µL 10% Ammonium persulphate 24 µl TEMED 9.4 ml dH20
4x SDS loading dye (10 ml stock)	2 ml 1 M Tris pH 6.8 0.8 g SDS (w/v) 4 ml Glycerol 8 mg Bromophenol blue 3 ml of 1 M DTT 4.2 ml dH20

Table 2.4: Bacterial strains used in this study

Strain	Relevant characteristic(s)	Source
Escherichia coli		
BL21 (DE3)	K-12 F- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3)	Invitrogen
DH5α	K-12 F- φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR17*(rK–, mK+) *phoA supE44* λ– *thi-1 gyrA96 relA1*	Invitrogen
DH10B	K-12 F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(*ara-leu*)7697 *galU galK* λ– *rpsL*(StrR) *nupG*	Invitrogen
ER2566	K-12 F- *λ- fhuA2 [lon] ompT lacZ::T7p07 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm]*	New England Biolabs
ER2796	K-12 F- *λ- fhuA2 Δ(lacZ)r1 glnV44 mcr-62 trp-31 dcm-6 zed-501::Tn10 hisG1 argG6 rpsL104 dam-16::Kan xyl-7 mtlA2 metB1 (mcrB-hsd-mrr) 114::IS10*	New England Biolabs
Salmonella enterica		
D23580*ΔΦ*	ST313 D23580 *ΔBTP1 ΔGifsy-1 ΔGifsy-2 ΔST64B ΔBTP5*	Owen et al, 2017
D23580*ΔΦΔBREX*	ST313 D23580 *ΔBTP1 ΔGifsy-1 ΔGifsy-2 ΔST64B ΔBTP5 ΔSTM4491-STM4498*	This study

2.3 Bacteriophage Work

All bacteriophages detailed in this study (Table 2.5) were isolated during the duration of this project from various water sources in Durham, UK.

2.3.1 Isolation of environmental bacteriophages

All *E. coli* bacteriophages listed in Table 2.5 were isolated from freshwater sources in Durham city centre and the surrounding areas. Water sampling and the first round of phage isolation was performed as an undergraduate laboratory practical session ran by Dr Blower. A 10 ml water sample was filtered through a 0.22 µm filter to remove debris. Filtrates were supplemented with 10 ml of LB, and then inoculated with 1 ml of an overnight culture of DH5α. For *S. enterica* phages, sewage effluent was collected from a sampling site in Durham, courtesy of Northumbrian Water Ltd. Filtrates were supplemented with 10 ml of LB, and inoculated with 10 ml of D23580*ΔφΔBREX*. Cultures were grown for 3 days before a 1 ml aliquot was transferred to sterile microcentrifuge tube and centrifuged at 12000 g for 5 min at 4 °C. The supernatant was transferred to a new microcentrifuge tube and 100 µl of chloroform was added to kill any remaining bacteria. Samples were vortexed before a serial dilution series was prepared in phage buffer, ranging from factors of 10^{-1} to 10^{-8}. 10 µl of serial dilutions and 200 µl of DH5α was added to 4 ml of semi-solid LB agar in a sterile 7 ml plastic bijou. Samples were gently mixed and poured on to LB agar plates. Plates were incubated overnight and individual plaques were harvested the following morning using sterile toothpicks into 200 µl phage buffer. Serial dilutions were re- prepared, and a series of agar plates were poured as detailed above. These steps were repeated if necessary, to ensure a single plaque morphology was observed for all dilutions. The semi-solid agar layer of a plate exhibiting syncytial plaque formation was scraped off into a 50 ml centrifuge tube. 3 ml of phage buffer and 500 µl of chloroform was added before vortexing for 2 min.

Table 2.5: Phages used in this study

Phage	Relevant characteristic(s)	Source
Escherichia coli		
AL25	Isolated from freshwater sources in Durham, UK	This study
Alma	Isolated from freshwater sources in Durham, UK	This study
Bam	Isolated from freshwater sources in Durham, UK	This study
Baz	Isolated from freshwater sources in Durham, UK	This study
BB1	Isolated from freshwater sources in Durham, UK	This study
BGP	Isolated from freshwater sources in Durham, UK	This study
BHP	Isolated from freshwater sources in Durham, UK	This study
CP	Isolated from freshwater sources in Durham, UK	This study
CS16	Isolated from freshwater sources in Durham, UK	This study
EH2	Isolated from freshwater sources in Durham, UK	This study
EL	Isolated from freshwater sources in Durham, UK	This study
Geo	Isolated from freshwater sources in Durham, UK	This study
Jura	Isolated from freshwater sources in Durham, UK	This study
Mak	Isolated from freshwater sources in Durham, UK	This study
Mav	Isolated from freshwater sources in Durham, UK	This study
NP	Isolated from freshwater sources in Durham, UK	This study
NR1	Isolated from freshwater sources in Durham, UK	This study
PATM	Isolated from freshwater sources in Durham, UK	This study
Paula	Isolated from freshwater sources in Durham, UK	This study
QOTSP	Isolated from freshwater sources in Durham, UK	This study
SAP	Isolated from freshwater sources in Durham, UK	This study
Sipho	Isolated from freshwater sources in Durham, UK	This study
Solly	Isolated from freshwater sources in Durham, UK	This study
Some	Isolated from freshwater sources in Durham, UK	This study
SPSP	Isolated from freshwater sources in Durham, UK	This study
TB34	Isolated from freshwater sources in Durham, UK	This study
TB36	Isolated from freshwater sources in Durham, UK	This study
TB37	Isolated from freshwater sources in Durham, UK	This study
Titus	Isolated from freshwater sources in Durham, UK	This study
Trib	Isolated from freshwater sources in Durham, UK	This study
Salmonella enterica		
AWAQ	Isolated from sewage effluent, NWL site, Durham UK	This study
DA1	Isolated from sewage effluent, NWL site, Durham UK	This study
DB1	Isolated from sewage effluent, NWL site, Durham UK	This study
KMP	Isolated from sewage effluent, NWL site, Durham UK	This study
LTE	Isolated from sewage effluent, NWL site, Durham UK	This study
SB58	Isolated from sewage effluent, NWL site, Durham UK	This study
SGP	Isolated from sewage effluent, NWL site, Durham UK	This study
SB58	Isolated from sewage effluent, NWL site, Durham UK	This study

Samples were incubated at 4 °C for 30 min before being centrifuged at 4000 g for 20 min. The supernatant was transferred to sterile glass bijou and 500 µl of chloroform was added. Phage lysates were stored at 4 °C.

2.3.2 Efficiency of plating assays

Phage titres were calculated by making a serial dilution of lysates in phage buffer. A series of agar plates were prepared as detailed in section 2.3.1 and incubated overnight at 37 °C. Visible plaques were counted to determine the plaque forming units per ml (pfu/ml). If calculated phage titres were $<10^8$ pfu/ml, another round of lysate preparation was performed to increase phage titres. Phage stocks with titres $>10^8$ pfu/ml were selected for EOP assays (Ellis and Delbrück 1939). Serial dilution series were made of phage lysates and semi solid LB agar plates were poured using the relevant bacterial strain and incubated overnight at 37 °C. Titres were calculated and EOP values were obtained by dividing the titre from the test host over the titre of the control host.

2.3.3 Purification of phage genomic DNA

Phage gDNA was prepared from high titre lysates via phenol-chloroform extraction. 500 µl of phage lysates was transferred to a 1.5 ml microcentrifuge tube and 2 µl of DNase I (Thermo) was added to remove bacterial DNA. Samples were incubated at 37 °C for 30 min before 20 µl 1 M EDTA was added. Samples were incubated at 75 °C for 30 min to inactivate DNase I. 500 µl of a 25:24:1 phenol:chloroform:isoamyl alcohol mixture was added. Samples were vortexed and incubated at 4 °C for 10 min before centrifugation at 12000 g for 10 min. Supernatants were transferred to a fresh tube and an equal volume of 24:1 chloroform:isoamyl alcohol was added. Samples were centrifuged and supernatants were transferred to a fresh tube. Residual organic material was removed by repeating the above step if necessary.

Phage gDNA was purified using ethanol precipitation and gDNA pellets were dried using vacuum centrifugation at room temperature. Pellets were resuspended in sterile MQ water and incubated at room temperature of a rocking shaker for 1 hour. Samples were incubated at 4 °C for 24 hours to ensure samples had fully dissolved before use. gDNA purity and integrity was analysed via agarose gel electrophoresis in 1% agarose TAE at 120 V for 45 min. gDNA content was quantified via nanodrop (Cytiva).

2.3.4 Growth assays during phage infection

A single colony was used to inoculate an overnight culture of LB incubated at 37 °C. The overnight culture was used to inoculate 200 µl of LB at an OD_{600} of 0.01 in the corresponding well of a sterile 96-well plate. Cultures were infected with phage stocks to an MOI of 0.001. Plates were incubated at 37 °C within a SPECTROstar Nano plate reader (BMG Labtech) shaking at 400 rpm. OD_{600} was measured at 10-min intervals and data was averaged from 3 independent cultures to plot growth curves in Figure 3.5.

2.4 Recombinant DNA Work

All molecular biology performed in this study were done using standard methods unless otherwise stated (Sambrook 2001). All oligonucleotide primers were obtained from Integrated DNA Technologies (IDT) and are detailed in Table 2.6. Plasmids used in this study are detailed in Table 2.7.

2.4.1 Bacterial plasmid and genomic DNA extraction

Plasmid DNA was isolated using the NEB Monarch series of plasmid mini-prep spin kits according to the manufacturer's instructions. Bacterial genomic DNA was prepared via phenol chloroform extraction. 500 µl of overnight liquid culture was centrifuged at 12000 g for 10

min. Pellets were resuspended in 0.1 M Tris pH 7.4 and added to a 500 µl mixture of 25:24:1 phenol:chloroform:isoamyl alcohol. Samples were vortexed and incubated at 4 °C for 10 min before centrifugation at 12000 g for 10 min. Supernatants were transferred to a fresh tube and an equal volume of 24:1 chloroform:isoamyl alcohol was added. Samples were centrifuged and supernatants were transferred to a fresh tube. Residual organic material was removed by repeating the above step if necessary. Bacterial gDNA was purified using ethanol precipitation and gDNA pellets were dried using vacuum centrifugation at room temperature. Pellets were resuspended in 10 mM Tris pH 7.4, 1 mM EDTA and incubated at room temperature of a rocking shaker for 1 hour. Samples were incubated at 4 °C for 24 hours to ensure samples had fully dissolved before use. gDNA purity and integrity was analysed via agarose gel electrophoresis in 0.7% agarose TAE at 120V for 45 min. gDNA content was quantified via nanodrop (ThermoFisher). PacBio sequencing was facilitated by Dr Rick Morgan, NEB and Dr Darren Smith, Northumbria University.

2.4.2 DNA purification from enzymatic reactions

DNA was extracted from agarose gels using the NEB Monarch range of DNA gel extraction kits according to the manufacturer's instructions. PCR and restriction digest products that were not analysed via gel electrophoresis were also purified using NEB Monarch range of DNA gel extraction kits. These products were mixed with an equal volume of DNA binding buffer (NEB) prior to column purification.

2.4.3 Polymerase chain reaction

All PCRs were performed using Q5 polymerase (NEB). Reactions were incubated in a Nexus GX2 thermocycler (Eppendorf). Primer annealing temperatures were calculated using the NEB Tm Calculator. Reactions were set up using the parameters below.

Q5 DNA Polymerase PCR

Component	Volume (µl)
10x Q5 reaction buffer	5
10 µm FW primer	2.5
10 µm RV primer	2.5
100 ng/µl template DNA	0.2
2 mM dNTPs	5
Q5 DNA polymerase	0.2
MQ H_2O	34.6

	Step	Temperature (°C)	Time (s)
	Initial denaturation	95	60
30-40 cycles	Denaturation	95	15
	Annealing	Variable*	30
	Extension	72	30 per kb
	Final extension	72	300
	Hold	10	∞

*Annealing temperature calculated using NEB Tm calculator

Table 2.6: Primers used in this study

Primer	Sequence	Description
ΔB-D23-DelF	TTTTGCGAGCTGCATAGCAGTCTTCAGAGGGTAG ATATGACGATGGTGTAGGCTGGAGCTGCTT	D23580 BREX deletion FW
ΔB-D23-DelR	AGTTTGACCCACTCCTCGATATATTTAGTTTACTC CCAACGCCTTCATATGAATATCCTCCTTAG	D23580 BREX deletion RV
ΔB-S23-SeqF	ATGAAAGCCTGGATGACCCC	D23580 BREX seq FW
ΔB-S23-SeqR	TGCGTTAGTGGAAAACTATGAGC	D23580 BREX seq RV
M13 FW	GTAAAACGACGGCCAG	puC19 MCS FW
M13 RV	CAGGAAACAGCTATGAC	pUC19 MCS RV
PF46	GCAAGGACGAGAATTTCCC	pRW50 MCS FW
PF47	CTGGCATTTGGTCAGCG	pRW50 MCS RV
PF138	CACACTTTGCTATGCCATAG	FWD pBAD30 MCS
PF139	GCTACTGCCGCCAGG	REV pBAD30 MCS
TRB740	CTGGAATGTGCTCGTCTTCA	pEFER 700c
TRB741	GCTTCCGATAAGATCGCCTA	pEFER 1400c
TRB742	AGCAGTTCATACTTGGCTGGA	pEFER 2100c
TRB743	CCAAAGAAGCCAGAAATCCA	pEFER 2800c
TRB744	GCATCAATGTCTGCCTGAGA	pEFER 3500c
TRB745	AATCAGCAATTGGCGTTCA	pEFER 4200c
TRB746	GAAGATCCCGTCCAATCTGA	pEFER 4900c
TRB747	CACTGTCTTTTTCTAACGCAGG	pEFER 5600c
TRB748	GATCTTATTGTTCCACCCGC	pEFER 6300c
TRB749	TAGGACAAAGTAGTGCCGCC	pEFER 7000c
TRB750	CTTTGGGGTAAATGTGGTCC	pEFER 7700c
TRB751	GCCTTTAGGGTTGTCTGGGT	pEFER 8400c
TRB752	CTATGTCCAGGCCGTAGAGG	pEFER 9100c
TRB753	AACTGGATTAGCGTTGAGATAA	pEFER 9800c
TRB754	CTTTTGAGTTCAATATTGCTGC	pEFER 10500c
TRB755	TGAGACATAGTGATTCGGCG	pEFER 11200c
TRB756	TTGCCCATATTGAGGATGGT	pEFER 11900c
TRB757	CAACTGTTGGCTAAGCACCTT	pEFER 12600c
TRB758	GTATAACTCGGCAAAACGCC	pEFER 13300c
TRB759	CCGGTTGTTTCTCATGTTTTG	pEFER 14000c
TRB760	GCTTAACGCTTACTTTGTCGATG	pEFER 14700c
TRB761	TCACCAGTGTCTCAACGCTT	pEFER 15400c
TRB762	GATCAGAAAACTTCGCCGAC	pEFER 16100c
TRB763	CACTTCACATCATCGCCATC	pEFER 16800c

Table 2.6 continued: Primers used in this study

Primer	Sequence	Description
TRB845	CAACAGCAGACGGGAGGTATGAATATAAAAGAATATTTA	*brxA* for pSAT1-LIC FW
TRB846	GCGAGAACCAAGGAAAGGTTATTATATTGTGCACTCCATGACCTC	*brxA* for pSAT1-LIC RV
TRB847	CAACAGCAGACGGGAGGTATGAAGGCTCTCCAGACCAGG	*brxB* for pSAT1-LIC FW
TRB848	GCGAGAACCAAGGAAAGGTTATTAATTTTTTTTACCGTCTTCAGG	*brxB* for pSAT1-LIC RV
TRB849	CAACAGCAGACGGGAGGTATGAAAATTGAACAGATTTTT	*brxC* for pSAT1-LIC FW
TRB850	GCGAGAACCAAGGAAAGGTTATTACTTAATTCTTACTCGGTCGCC	*brxC* for pSAT1-LIC RV
TRB851	CAACAGCAGACGGGAGGTATGTATCAAGCGGGTGGAACA	*brxU* for pSAT1-LIC FW
TRB852	GCGAGAACCAAGGAAAGGTTATTATTCTGTCTCTACGCTGCTTGC	*brxU* for pSAT1-LIC RV
TRB853	CAACAGCAGACGGGAGGTATGAATACGAACAACATGAAA	*pglX* for pSAT1-LIC FW
TRB854	GCGAGAACCAAGGAAAGGTTATTAGTTAACCTCCGGGGCCGAACC	*pglX* for pSAT1-LIC RV
TRB855	CAACAGCAGACGGGAGGTATGCAGATTAGCGAACTGGCA	*pglZ* for pSAT1-LIC FW
TRB856	GCGAGAACCAAGGAAAGGTTATTAGAAAAAATCATCTTGGAATGC	*pglZ* for pSAT1-LIC RV
TRB857	CAACAGCAGACGGGAGGTATGACAACTGATTTTTCACAA	*brxL* for pSAT1-LIC FW
TRB858	GCGAGAACCAAGGAAAGGTTATTAATTAACCCCAAGCGCCTTAAA	*brxL* for pSAT1-LIC RV
TRB865	TTTGAATTCAGGAGATATCTTATGCACCATCACCATCACCATGGAATGTATCAAGCGGGTGGAACA	*brxU* EcoRI FW
TRB866	TTTTCTAGATTATTATTCTGTCTCTACGCTGCTTGC	*brxU* XbaI RV
TRB873	TTAATGCAGCTGATTAATACG	pSAT1-LIC seq primer FW
TRB874	AATCAATGAAACAGACACACC	pSAT1-LIC seq primer RV
TRB875	TACTCAAGCTTATGCATGC	pSAT1-LIC hSUMO2 seq FW
TRB876	TTGAATTCAGGAGATATCTTATGCACCATCACCATCACCATGGACAAGACAACACTGGACTTGAAG	*brxR* EcoRI
TRB877	TTAAGCTTTTATTACGATTCGCTATATCCAGGAGCG	*brxR* HindIII
TRB878	CAACAGCAGACGGGAGGTCAAGACAACACTGGACTTGAAG	*brxR* for pSAT1-LIC FW
TRB879	GCGAGAACCAAGGAAAGGTTATTACGATTCGCTATATCCAGGAGCG	*brxR* for pSAT1-LIC RV
TRB904	TTGAATTCGTTATGGCTGGATCACAGC	pEFER 12400 EcoRI FW
TRB905	TTGAATTCCAATGGCTTGAGATGGCATG	pEFER 13400 EcoRI FW
TRB906	TTGAATTCGCTGTATTGATAGACTACGC	pEFER 13599 EcoRI FW
TRB907	TTAAGCTTGCCAACGGATTGTTCTGGC	pEFER 13989c HindIII RV
TRB908	TTGAATTCGGATTTTACTAAACACCCGC	pEFER 23105 EcoRI FW
TRB909	TTGAATTCAAACTTGGATTTCGTCGCCG	pEFER 23621 EcoRI FW
TRB910	TTAAGCTTTGCTCTGGGTCGTACCAAAA	pEFER 24203c HindII RV
TRB950	CGAAATTAATACGACTCACTATAGG	pET T7 promoter FW
TRB951	ATGCTAGTTATTGCTCAGCGGT	pET T7 terminator RV

Table 2.6 continued: Primers used in this study

Primer	Sequence	Description
TRB1146	TTTGGATCCGGTCTCGGGAGGTTATGGCTGGATCACAGC	GGA F1 BamHI FW
TRB1147	TTTCTGCAGGGTCTCGACCTCAATCAAATCTTCCCG	GGA F1 PstI RV
TRB1148	TTTGGATCCGGTCTCGAGGTCATGGAGTGCACAATATGAAGG	GGA F2 BamHI FW
TRB1149	TTTCTGCAGGGTCTCGCTGACATTTCAACGTCAGTG	GGA F2 PstI RV
TRB1151	TTTCTGCAGGGTCTCCGTTTCTTCAATGCGCTGTCTGC	GGA F3 PstI RV
TRB1154	TTTGGATCCGGTCTCGGACATTCCAAAGATGATTCTGGAGAA	GGA F5 BamHI FW
TRB1155	TTTCTGCAGGGTCTCGCAAGATCAAACTTTCCTAGCTG	GGA F5 PstI RV
TRB1156	TTTGGATCCGGTCTCCCTTGAGACGTCACTCTGGGCATATCTGAG	GGA F6 BamHI FW
TRB1157	TTTCTGCAGGGTCTCCTAGAAAAAATCATCTTGGAATGCCAAATC	GGA F6 PstI RV
TRB1158	TTTGGATCCGGTCTCCTCTAAGGCGCTGAGGTGA	GGA F7 BamHI FW
TRB1159	TTTCTGCAGGGTCTCCATGGAGGACAGAAACTGTCTACCG	GGA F7 PstI RV
TRB1160	TTTCTGCAGGGTCTCCATGGCAATCAAATCTTCCCG	GGA F1-Control RV
TRB1161	CTGCAGGAAGGTTTAAACGCATTTAGG	pGGA seq FW
TRB1162	TAATACGACTCACTATAGGGAGACGC	pGGA seq RV
TRB1203	TTTCTGCAGGGTCTCGGCAATGAACTCACGTTCCGGTAACTG	GGA F3 BamHI FW
TRB1204	TTTGGATCCGGTCTCCTTGCCGTCCTAGATGGAC	GGA F4 BamHI FW
TRB1209	TTTCTGCAGGGTCTCGTGTCACGAGAGCGGAAACCAC	GGA F4 PstI RV
TRB1211	GCTGCAATGATACCGCGTGACCCACGCTCACCGG	Quickchange to remove BsaI from pUC19 FW
TRB1212	CCGGTGAGCGTGGGTCACGCGGTATCATTGCAGC	Quickchange to remove BsaI from pUC19 RV
TRB1307	TAATAACCGGAACGTGAGTTC	Δ*brxU* pUC19-F4 FW
TRB1308	TTACTTAATTCTTACTCGGTCG	Δ*brxU* pUC19-F4 RV
TRB1309	TAATAAGGTTTCCGCTCTCG	Δ*pglX* pUC19-F5 FW
TRB1310	TTATTCTGTCTCTACGCT	Δ*pglX* pUC19-F5 RV
TRB1358	TTTGGATCCGGTCTCGGGAGGCCGCTTACGAGCGTGTA	GGA F0 BamHI FW
TRB1359	TTTCTGCAGGGTCTCGGTTTGGTTAGACGCCTCGGTAGT	GGA F0 PstI RV
TRB1434	ATTTTTAATTTAAAAGGAACTAGGTGAAGATC	pUC19 FW synthetic BrxU substrate
TRB1435	TAATTTTTAAATCAATCTAAAGTATATATGAGTAA	pUC19 RV synthetic BrxU substrate

Table 2.7: Plasmids used in this study

Plasmid	Description	Source
pKD46	λ red expression vector, L-ara inducible, temperature sensitive, Ap^R	Daksenko and Wanner (2000)
pBAD30	L-ara inducible expression vector, 6His-yORF, repressed by glu, Ap^R	Guzman et al (1995)
pCP20-Gm	Expression vector for FLP recombinase, Gm^R	Doublet et al (2008)
pEFER	*E. fergusonii* ATCC 35469 pEFER, encodes BREX and *brxU,* Ap^R, Tc^R, Sm^R	ATCC
pEFER-Km5	pEFER with Tn5 insertion in *brxA,* Ap^R, Tc^R, Sm^R, Km^R	This study
pEFER-Km7	pEFER with Tn5 insertion in spectinomycin resistance gene, Ap^R, Tc^R, Km^R	This study
pKD4	Template plasmid for λ red recombination, Km^R	Daksenko and Wanner (2000)
pRW50	Promoterless *lacZ* downstreak of MCS for promoter activity assays, Tc^R	Blower Lab
pSAT1-LIC	ITPG inducible expression vector, 6His-hSUMO2-yORF, Ap^R	Blower Lab
pTRB443	pSAT1-LIC-*brxA,* IPTG inducible, Ap^R	This study
pTRB444	pSAT1-LIC-*brxB,* IPTG inducible, Ap^R	This study
pTRB445	pSAT1-LIC-*brxC,* IPTG inducible, Ap^R	This study
pTRB446	pSAT1-LIC-*brxU,* IPTG inducible, Ap^R	This study
pTRB447	*pSAT1-LIC-pglX,* IPTG inducible, Ap^R	This study
pTRB449	pSAT1-LIC-*pglZ,* IPTG inducible, Ap^R	This study
pTRB450	pSAT1-LIC-*brxL,* IPTG inducible, Ap^R	This study
pTRB451	pBAD30-*brxR,* IPTG inducible, Ap^R	This study
pTRB452	pRW50-S4, Tc^R	This study
pTRB454	pRW50-S1, Tc^R	This study
pTRB455	pRW50-S3, Tc^R	This study
pTRB464	pRW50-S2, Tc^R	This study
pTRB465	pRW50-S5, Tc^R	This study
pTRB466	pRW50-S6, Tc^R	This study
pTRB479	pUC19 derivative, 4x mutagenesis to remove BsaI sites	This study
pTRB498	pTRB479 containing GGA fragment 1, 5′ fuses with pGGA, Ap^R	This study

Table 2.7 continued: Plasmids used in this study

Plasmid	Description	Source
pTRB499	pTRB479 containing GGA fragment 2, Ap^R	This study
pTRB500	pTRB479 containing GGA fragment 3, Ap^R	This study
pTRB501	pTRB479 containing GGA fragment 4, Ap^R	This study
pTRB502	pTRB479 containing GGA fragment 5, Ap^R	This study
pTRB503	pTRB479 containing GGA fragment 6, Ap^R	This study
pTRB504	pTRB479 containing GGA fragment 7, 3′ fuses with pGGA, Ap^R	This study
pTRB505	pTRB479 containing GGA control fragment, Ap^R	This study
pTRB506	pBREX GGA assembly, fragments 1-7, Cm^R	This study
pTRB507	pTRB479 containing GGA fragment 7, 5′ and 3′ fuse with pGGA, Ap^R	This study
pTRB514	pBREX-*ΔpglX,* Cm^R	This study
pTRB519	pBAD30:*brxU*, Ap^R	This study
pTRB520	pBREX-*ΔbrxU,* Cm^R	This study
pTRB563	pBrxXL GGA assembly, fragments 0-7, Cm^R	This study
pTRB564	pBrxXL-*ΔpglX* GGA assembly, fragments 0-7, Cm^R	This study
pTRB565	pBrxXL-*ΔbrxU* GGA assembly, fragments 0-7, Cm^R	This study
pTRB566	pBrxXL-*ΔbrxUΔpglX* GGA assembly, fragments 0-7, Cm^R	This study
pTRB578	pTRB479 containing GGA fragment 1, 5′ fuses with fragment 0, Ap^R	This study
pTRB579	pTRB479 containing GGA fragment 0, 5′ fuses with pGGA, Ap^R	This study
pTRB603	pBAD30-*brxU* mutant D474A	This study
pTRB604	pBAD30-*brxU* mutant E528A	This study
pTRB605	pBAD30-*brxU* mutant H475A	This study
pTRB606	pBAD30-*brxU* mutant N485A	This study
pTRB607	pBAD30-*brxU* mutant N519A	This study
pTRB608	pBAD30-*brxU* mutant Q43A	This study
pTRB609	pBAD30-*brxU* mutant Q101A	This study
pTRB610	pBAD30-*brxU* mutant R46A	This study
pTRB611	pBAD30-*brxU* mutant R102A	This study
pTRB612	pBAD30-*brxU* mutant S50A	This study
pTRB613	pBAD30-*brxU* mutant S50DA	This study
pTRB614	pBAD30-*brxU* mutant R38A, S50D	This study
pTRB615	pBrxXL-*ΔbrxS* GGA assembly, fragments 0-7, Cm^R	This study
pTRB616	pBrxXL-*ΔbrxT* GGA assembly, fragments 0-7, Cm^R	This study
pUC19	Cloning vector, Ap^R	New England Biolabs

2.4.4 Restriction enzyme cloning with DNA ligase

DNA digestion for cloning was performed using restriction enzymes from NEB and ThermoFisher, using the manufacturer's instructions. Digest products were purified either by PCR clean up or by agarose gel electrophoresis and subsequent extraction (Section 2.4.2). For ligation, a ratio of 3:1 insert: vector was used and reactions were set up as detailed below prior to transformation of heat-shock competent DH5α. Positive colonies of transformed DH5α were selected for on LB agar containing the relevant antibiotic.

Q5 DNA Polymerase PCR

Component	Volume (µl)
10x T4 DNA ligase buffer	2
T4 DNA ligase	1
Cut vector	1 (variable)
Cut insert	3 (variable)
MQ H_2O	3 (variable)

2.4.5 Heat-shock competent *E. coli* preparation

Heat shock competent *E. coli* were prepared by inoculating 5 ml of LB with a single colony (and relevant antibiotic where necessary) and grown overnight at 37 °C. Overnight cultures were used to inoculate 25 ml to an OD_{600} of 0.05 and were supplemented with 375 µl of 1M $MgCl_2$. Cultures were grown to an OD_{600} of 0.6-0.8 before being pellet by centrifugation at 4000 g for 10 min at 4 °C. Cell pellets were washed by resuspension in 10 ml ice-cold Solution A (Table 2.3) and incubated on ice for 20 min. Cells suspensions were centrifuged at 12000 g for 10 min at 4 °C and another wash step with Solution A was performed. Cell pellets were then resuspended in Solution A + glycerol (Table 2.3). Competent cells were either used

immediately or flash frozen in liquid nitrogen as 50 or 100 µl aliquots and stored for up to six months at -80 °C.

2.4.6 Transformation of heat-shock competent *E. coli*

For transformation of *E. coli* with supercoiled plasmid DNA, 1 µl of plasmid DNA (50 ng/µl) was added to a 50 µl cell aliquot and incubated on ice for 30 min. For cloning, the entire ligation mix was added to a 100 µl cell aliquot. After 30 min, cells were incubated for 45 seconds in a 42 °C water bath and placed on ice for 2 min. Cells were supplemented with 1 ml of 2x YT and recovered for 1 hour at 37 °C. Cells were plated on LB agar containing the relevant antibiotic. When transforming cells with pBAD30 constructs, recovery cultures and LB agar were supplemented with 0.2% D-glucose to repress transcription of the target gene.

2.4.7 Ligation independent cloning

Ligation independent cloning (LIC) was used to generate N-terminal 6His-tagged constructs for protein overexpression using pSAT1-LIC. Immediately following the 6His-tag is the hSUMO2 gene, allowing expression of a 6His-hSUMO2-yORF fusion protein. Further vector and expression details can be found in Section 4.3. 1 µg of pSAT1-LIC was digested with StuI making a single cut, blunt ended, linear product which was purified via gel extraction. Inserts were amplified via PCR using pEFER as a DNA template and primers shown in Table 2.6. Primers were designed to include an 18-21 bp complementary sequence to the target region, with an upstream LIC site to allow interaction with LIC treated linear pSAT1-LIC. Both vector and insert were treated by T4 DNA polymerase to produce extended overhangs. T4 DNA polymerase exhibits 3'-5' exonuclease activity in the absence of nucleotides. Insert LIC sites are designed to exclude adenines, and vector LIC sites exclude thymine. As a result, T4 DNA ligase can create extended cohesive ends. With the addition of dTTP to vector reactions and

dATP to insert reactions, T4 DNA polymerase will stall upon encountering the first dTTP (for vectors) or dATP (for inserts), allowing for control over the length of the generated overhangs (Aslanidis and de Jong 1990). Following treatment with T4 DNA polymerase, a 10 µl 1:1 vector: insert mixture was prepared and left overnight at room temperature to allow for annealing to occur prior to transformation into heat-shock competent DH5α. Positive colonies of transformed DH5α were selected for on ampicillin containing LB agar.

2.4.8 Golden Gate Assembly

Golden gate assembly was performed to produce a plasmid containing the BREX locus without pEFER background DNA. Quikchange mutagenesis was performed to remove BsaI sites from pUC19 to produce pTRB479. BREX locus was partitioned into multiple ~3-4 kb regions and amplified using primers (Table 3.4) with flanking BsaI sites with unique overhang sequences. GGA design is detailed in Section 3.8. pTRB479 and inserts were digested with BsaI and ligated with T4 DNA ligase to create GGA donor constructs. A 3:1 insert: vector ratio was used in a golden gate assembly reaction containing BsaI and T4 DNA ligase. Positive colonies of transformed DH5α were selected for on ampicillin containing LB agar.

2.4.9 Construction of D23580*ΔφΔBREX* via λ-red recombination

λ-red recombination was used to delete all genetic material including and between STM4491 and STM4498 in D23580*ΔφΔBREX*. Primers were designed to contain ~40 bp complementary regions immediately followed λ-red recombination sites. pKD4 was used as a DNA template which encodes a Km resistance cassette to allow positive selection, flanked by FRT sites to allow its removal (Datsenko and Wanner 2000). Following PCR, samples were resolved via gel electrophoresis and extracted. As pKD4 cannot replicate without the host expressing the λ-

pir protein, there is no requirement to treat PCR products with DpnI prior to PCR clean up. For the preparation of electrocompetent D23580*Δφ*-pKD46, firstly electrocompetent D23580*Δφ* was prepared (Section 2.4.10). pKD4 PCR products were concentrated by ethanol precipitation and 2 µl was added to electrocompetent D23580*Δφ*-pKD46 cells. Following electroporation, cells were supplemented with 1 ml 2x YT and incubated at 37 °C for 1 hour to remove the temperature sensitive pKD46. Km^R mutants were transformed via another round electrocompetent cell preparation and electroporation with pCP20-Gm. Following electroporation, samples were plated on LB agar containing gentamycin and incubated overnight at 30 °C. Gm^R mutants were selected and grown at 43 °C to remove pCP20-Gm. Mutants were screened for both kanamycin and gentamycin sensitivity to confirm the removal of both the Km^R resistance cassette from pKD4 and pCP20-Gm. gDNA was prepared from Km^S Gm^S mutants and sequenced to confirm the removal of the BREX locus as well as the Km^R cassette.

2.4.10 Electrocompetent *S. enterica* preparation for λ-red recombination

A single colony of D23580*Δφ* was used to inoculate 10 ml LB and cultures were grown at 37 °C degrees until an OD_{600} of 0.6-0.8. Cells were pelleted via centrifugation at 4000 g for 10 min at 4 °C and resuspended in ice-cold MQ H_2O. Cells were kept on ice at all times. Cell washing was repeated twice to remove salts before a final washing step in ice-cold 10% glycerol. D23580*Δφ* was incubated with 1 µl of pKD46 (20 ng/µl) for 10 min. Samples were electroporated using a Gene Pulser (BioRad) and immediately supplemented with 1 ml of 2x YT media. Samples were incubated shaking at 37 °C for 1 hour before plating on LB agar containing ampicillin. A second round of electrocompetent cell preparation was performed with D23580*Δ*-pKD46 by inoculating an overnight culture of 10 ml of LB supplemented with

ampicillin and grown at 30 °C. The overnight culture was used to inoculate 10 ml of fresh LB containing ampicillin and 0.1% L-ara. Cultures were grown at 30 °C until an OD_{600} of 0.6-0.8 was reached. Washing steps were performed as previously described. D23580*Δ*-pKD46 Km^R mutants were selected for on LB agar containing kanamycin.

2.4.11 Transposon mutagenesis

Transposon mutagenesis was performed using EZ::Tn5^{TM} <Kan-2> according to the manufacturer's instructions (Cambridge Bioscience). Insertion events were mapped using random-primed PCR and subsequent DNA sequencing (Zou et al. 2003).

2.4.12 DNA sequencing

All DNA sequencing in this study was performed by DBS Genomics, Durham University. Sequences were confirmed using Chromas software (Technelysium).

2.4.13 Site directed mutagenesis

Site directed mutants were designed and then Genscript were commissioned to generate the constructs.

2.5 Protein Production and Purification

2.5.1 1D SDS-PAGE gel electrophoresis

SDS-PAGE gels were prepared as detailed in Table 2.3 using BioRad mini-gel casts. 5 ml of resolving gel was poured per gel and 250 µl of isopropanol was gently layered on top to ensure a smooth surface was formed and all bubbles were dissipated. Resolving layers were left to set for up to 1 hour depending on acrylamide content before isopropanol was decanted. The

remaining gel volume was filled with stacking gel mix and 10-well or 15-well combs were fitted gently to prevent bubble formation. Gels were left to set for 1 hour before use. Protein samples were prepared by mixing with 4x SDS loading buffer and denatured at 95 °C for 5 min. For 10-well gels, 10 µl sample volumes were loading and for 15-well gels, 5 µl sample volumes were loading. Gels were ran at 180 V for 100 min unless otherwise stated. Gels were rinsed in dH_2O before being stained using InstantBlue Coomassie protein stain (AbCam).

2.5.2 Overexpression of native proteins

For expression of target proteins, BL21 (DE3) was transformed with pSAT1-LIC or pBAD30 expression constructs (Table 2.7). Expression strain ER2566 has been used as a substitute for BL21 (DE3), however no significant changes in yield were observed for the proteins expressed in this study. A single colony was used to inoculate 25 ml LB containing ampicillin and grown overnight at 37 °C. Overnight cultures were used to inoculate six 2 L baffled flasks containing 1 L 2x YT supplemented with ampicillin. Cultures were grown at 37 °C shaking at 180 rpm until an OD_{600} of ~0.6 and which point protein expression was induced. For pSAT1-LIC constructs, expression was induced by the addition of IPTG to a final concentration of 1 mM and cultures were grown overnight at 18 °C. For pBAD30 constructs, expression was induced by the addition of L-arabinose to a final concentration of 0.1% w/v and grown for 4 hours at 37 °C. Cultures were pelleted at 5000 g for 30 min at 4 °C using an Avanti JXN-26 series centrifuge in a JLA-8.1 rotor (Beckman). Cell pellets were resuspended in 50 ml of ice-cold A500 (Table 2.3) and used immediately or were flash frozen in liquid nitrogen and stored at -80 °C.

2.5.3 Overexpression of selenomethionine-labelled protein

SM-labelled protein expression was performed using the SM Medium (Molecular Dimensions). pSAT1-LIC expression constructs were used to transform BL21 (DE3) and a single colony was used to inoculate a 5 ml SM Complete Medium (Table 2.3) starter culture containing ampicillin, which was grown for 36 hours at 37 °C. Starter cultures were used to inoculate 1 L of SM Complete Medium in a 2 L baffled flask to an OD_{600} of 0.05 and cultures were grown at 37°C until an OD_{600} of 0.6-0.8, at which point 10 ml of 100x methionine inhibitory feedback mix was added (which contains a total of 100 mg SM). 30 min after the addition of 100x methionine inhibitory feedback mix, IPTG was added to a final concentration of 1 mM to induce protein expression by T7 RNA polymerase. Cultures were grown overnight at 18 °C and were pelleted as described in Section 2.5.2. Once purified, SM incorporation was quantified by ES^+ TOF MS (Figures 5.3 and 5.13).

2.5.4 IMAC protein purification

Samples and buffers were kept on ice or at 4 °C at all times during purification. Cell pellets suspended in A500 were sonicated for a total of 2 min. Sonication was performed in 4x 30 sec bursts with samples cooled on ice in-between bursts. Cell lysates were centrifuged at 25000 g for 30 min using an Avanti JXN-26 series centrifuge in a JA-25.50 rotor (Beckman). Clarified cell lysates were transferred to a pre-chilled flask and loaded through a 5 ml pre-packed NI-NTA His-Trap HP column (Cytiva) using a peristaltic pump with a flow rate of 1 ml/min. Following lysate loading, 50 ml of A500 was used to wash the column to remove unbound contaminants. For the expression of mutant pBAD30-*brxU* constructs (Figure 4.22), an additional wash step using A500 supplemented to a final imidazole concentration of 50 mM was used. Target proteins were eluted were using high imidazole buffer B500 which contains

250 mM imidazole. Fusion proteins expressed from pSAT1-LIC constructs were further purified via removal of the affinity 6His-tag and FPLC.

2.5.5 Affinity tag removal with Sentrin protease

200 µl of 0.25 mg/ml 6His-SenP was added to fusion proteins that had been purified via IMAC (Section 2.5.4) and dialysed into A100 (Table 2.3) overnight at 4 °C. Samples were then loaded onto a clean 5 ml His-Trap column equilibrated in A100 at 1 ml/min using a peristaltic pump and the flow-through fraction was collected. Cleaved affinity tags as well as the 6His-tagged SenP are retained by the column in these conditions. The flow-through was loaded onto a pre-packed 5 ml Hi-Trap Q column (Cytiva) at 1 ml/min. Flow-through was collected and stored as a precautionary measure in the event that column binding efficiency was low.

2.5.6 Anion exchange chromatography

A pre-loaded Hi-Trap Q column (Section 2.5.5) was loaded onto an ÄKTA Pure purification system (Cytiva) equilibrated in A100. C1000 was titrated in until a final NaCl concentration of 500 mM was reached over a 40 ml elution period. Fractions were dispensed in 2 ml aliquots in 96-well deep trays and analysed by SDS-PAGE for purity and content. An example chromatogram for anion exchange during the purification of BrxR is shown in Figure 2.1. Fractions containing target protein were pooled together and concentrated to be used in size exclusion chromatography.

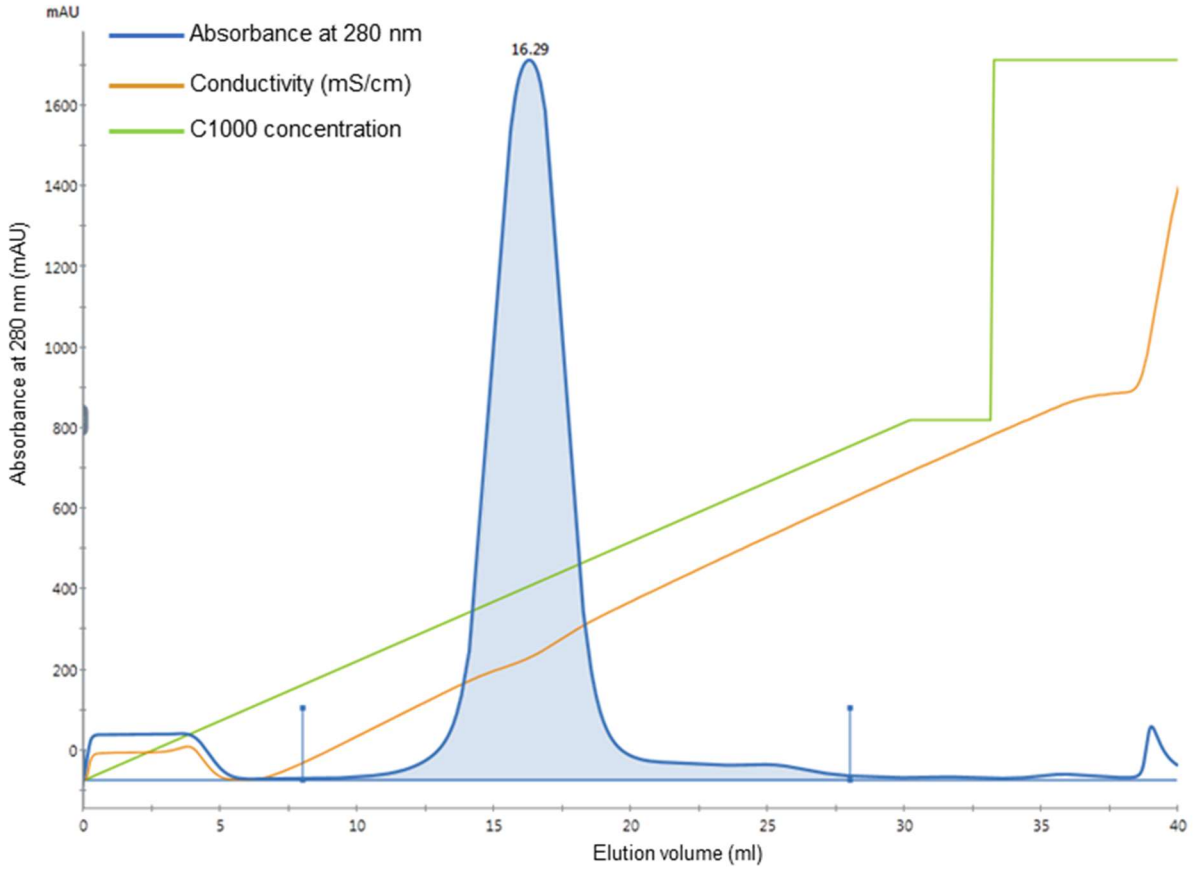


Figure 2.1: Anion exchange chromatogram showing the elution of native mature BrxR. *Fractions were collected in 96-well plates in 2 ml volumes. Conductivity increases as increasing concentrations of C1000 are titrated in. Conductivity reaches its max value of 44.05 mS/ cm at an elution volume of 40 ml, corresponding to 100% C1000.*

2.5.7 Preparative size exclusion chromatography

A HiPrep S-300 HR 16/60 (Cytiva) was used for resolving protein samples with a mass (in solution) greater than 100 kDA and a Hi Prep S-200 16/60 (Cytiva) was used for samples below 100 kDA. Pooled fractions following anion exchange chromatography were concentrated and loaded on to the ÄKTA Pure system. A total protein volume of 2 ml was used primarily, however 5 ml volumes were occasionally used in the event of a high protein yield, to avoid precipitation. Columns were equilibrated in FPLC sizing column buffer (Table 2.3) by running 2 column volumes through at 0.5 ml/min. Samples were manually loaded into storage loops and were then injected onto columns to be resolved at a flow rate of 0.5 ml/min. Fractions were collected in 2 ml volumes and analysed via SDS-PAGE. Fractions containing only purified products were pooled. For storage, samples were concentrated to 1 mg/ml and 0.5 volumes of sample storage buffer (Table 2.3) was added. Samples were aliquoted in 20 µl volumes prior to flash freezing in liquid nitrogen and storage at -80°C. For crystallography, samples were dialysed into crystallisation buffer (Table 2.3) and concentrated. For longer term storage, samples were flash frozen in liquid nitrogen for storage at -80 °C.

2.6 Protein crystallography

2.6.1 Protein crystallisation

Following concentration, samples were kept on ice for short term storage and either a Mosquito Xtal3 robot (SPT Labtech) was used to set automated crystallisation trials, or hand drops were made using Cryschem M plates (Hampton Research). For a detailed description of protein crystallisation screening and condition optimisation, see Sections 5.2-5.8. It should be noted that a single freeze-thaw cycle of BrxU before crystallisation resulted in significantly poorer X-ray diffraction, with only 1 in 20 crystals diffracting beyond 4 Å.

2.6.2 Protein crystal harvesting and data collection

A cryo-solution of 50 mM Tris pH 7.9, 150 mM NaCl and 80% w/v glycerol was prepared and mixed in a 1:1 ratio with crystal drop mother liquor to form the cryoprotectant. A 2 µl drop of cryoprotectant was placed on a glass microscope slide. Protein crystals were harvested from drops using cryo-loops and were placed in the cryoprotectant drop for 10 sec. Mounted crystals within cryo-loops were immediately transferred to a unipuck in liquid nitrogen for storage and shipment to Diamond Light Source (DLS).

2.6.3 Data collection and structure determination

X-ray data collection is detailed in Tables 5.1 and 5.2. Diffraction data were processed with iSpyB (DLS) and then AIMLESS from CCP4 (Winn et al. 2011) was used to corroborate the spacegroups. The crystal structure of SM-BrxU was solved using single anomalous dispersion (SAD), by providing the SHELX suite (Sheldrick 2008) in CCP4 with the SM-BrxU anomalous dataset. The solved starting model for SM-BrxU was then built in REFMAC (Vagin et al. 2004) and BUCCANEER (Cowtan 2006). The refined model for SM-BrxU was then used as a search model to solve the native BrxU structure by molecular replacement using Phaser (McCoy et al. 2007). The initial native BrxU model was built in BUCCANEER, then iteratively refined and re-built using PHENIX (Adams et al. 2010) and COOT respectively. The quality of the final model was assessed using PHENIX, COOT and the wwPDB validation server (Emsley and Cowtan 2004). Structural figures were generated using PyMol (Schrödinger).

2.7 Assays for Protein Characterisation

2.7.1 β-galactosidase reporter assays

Reporter assay designs were based on the original Miller assays (Miller 1972). Regions of the *E. fergusonii* BREX operon were cloned into the pRW50. DH5α was transformed with pRW50 constructs (Section 4.6) and transformants were selected on LB agar containing tetracycline. DH5α containing pBAD30-*brxR* was transformed with the same constructs. A single colony was used to inoculate an overnight culture at 37 °C shaking at 180 rpm. The overnight culture was used to inoculate 15 ml volume of LB containing tetracycline and 0.1% L-arabinose, and grown until mid-log phase at 37 °C with shaking at 180 rpm. OD_{600} was recorded for each culture. 120 µl of β-galactosidase reaction buffer was added to a 96-well plate followed by 80 µl of culture. The plate was incubated at 30 °C for 30 min and the reaction was stopped by the addition of β-galactosidase stopping buffer. Absorbance was measured at OD_{420} and OD_{520} using a SPECTROstar Nano plate reader (BMG Labtech). Miller units were calculated as instructed (Schaefer et al. 2016).

2.7.2 DNA hydrolysis assays

Phage gDNA was isolated using a phenol-chloroform extraction method and ethanol purification. 2 µl of 100 ng/µl gDNA was added to a mixture of 2 µl 10x DPMG buffer, 2 µl 10 mM ATP, 5 µl of 2 µM BrxU and 9 µl of nuclease free water. Nuclease free water was used for negative controls in place of ATP. When additional metals were used (Figure 4.13), these were added to replace some of the water volume. For these experiments, a version of DPMG buffer was used that did not contain $MgSO_4$. Samples were incubated at 37 °C for 30 min and reactions were terminated by incubating at 75 °C for 10 min. 4 µl of 6x TriTrack loading dye (Thermo Scientific) was added and thoroughly mixed. 12 µl was loaded on to a 0.9% agarose

gel containing EtBr at a concentration of 0.5 µg/ml. 2 µl of 1 kb GeneRuler (ThermoFisher) was loaded as a ladder. Agarose gels were resolved at 120 V for 45 min in 1x TAE buffer containing 0.5 µg/ml EtBr. Gels were visualised with BioRad Image Lab software. Band intensity was calculated using ImageJ. Enzyme activity was quantified by dividing the band intensity of sample by the band intensity of the negative control without ATP.

2.7.3 Analytical Gel Filtration

A Superdex 200 Increase GL 5/150 (Cytiva) was connected to an ÄKTA Pure system (Cytiva) and equilibrated by running through 2 column volumes of filtered MQ water and 5 column volumes of FPLC sizing column buffer at 0.15 ml/min. A 50 µl sample was prepared containing 5 µl of 5 µM BrxU and 5 µl 10x DPMG buffer, and made up to 50 µl with MQ water. 5 µl of 10 mM nucleotide was used to replace 5 µl of MQ water in nucleotide +ve samples. Samples were incubated at 37 °C for 15 min and were loaded into a 10 µl loop via a 50 µl Hamilton syringe. 2 column volumes of analytical sizing buffer was run through the sample loop directly on to the column at 0.15 ml/min. Absorbance at 280 nm was measured corresponding to the concentration of protein. Experiments were run in triplicate.

2.7.4 Phosphate production assays

Phosphate detection assays were performed according to the BIOMOL green (Enzo Life Sciences) protocol. A 96-well plate format was used and columns 1 and 2 were used for phosphate standard serial dilutions to obtain a slope intercept form to allow for calculation of experimental results. Experimental wells were set up in 50 µl total volumes. Each well contained 5 µl 10x DPMG buffer, 5 µl NTP and 5 µl protein, made up to 50 µl with MQ H_2O. 5 µl was added for each additional reagent (EDTA, DNA) in place of equal volumes of MQ H_2O . All experiments were incubated at 37 °C for 30 min shaking at 400 rpm and reactions

were terminated by the addition of 100 µl of BIOMOL green reagent (Cambridge Bioscience) and immediately transferred to a SPECTROstar plate reader. Plates were incubated at 30 °C shaking at 400 rpm whilst the BIOMOL green reagent developed. Absorbance at 620 nm was measured after 30 min. Values were averaged from 3 independent wells.

Chapter 3: Investigating BREX within Phage Defence Islands

3.1 Introduction

The origins of the BacteRiophage EXclusion (BREX) system trace back to 2015 when it was first described by Goldfarb et al., and shown to confer resistance to phage infection. The *Bacillus cereus* BREX system is a type 1 BREX system comprised of 6 core BREX genes. Type 1 BREX systems are encoded within both Gram-positive and Gram-negative bacteria (Goldfarb et al. 2015), including pathogens such as *Escherichia fergusonii* and *Salmonella enterica* Serovar Typhimurium. *E. fergusonii* is an emerging human and animal pathogen, closely related to extraintestinal pathogenic *E. coli* (ExPEC) (Adesina et al. 2019). *E. fergusonii* ATCC 35469 encodes a 55.15 kb plasmid, pEFER, that contains a type 1 BREX locus and represents an extrachromosomal model system for studying BREX. The second Gram-negative pathogen example, *S. enterica* sv Typhimurium is typically considered as a non-typhoidal serovar, primarily causing gastrointestinal infections and arising from consumption of infected foods (Mead et al. 1999). Historically, this serovar has been widely regarded to cause self-limiting enterocolitis, however multiple lineages of invasive non-typhoidal *Salmonella* (NTS) have recently been shown to cause invasive, systemic clinical syndromes (Rodwell et al. 2021). *S.* Typhimurium ST313, notably strain D23580, has been extensively linked to invasive NTS infections in sub-Saharan Africa where it is responsible for 50,000 deaths every year (Stanaway et al. 2019). Studying both of these loci allows us to consider whether BREX is active in Gram-negative bacteria and how it may differ according to whether it is encoded on a plasmid, or the chromosome.

In addition to potentially providing phage-resistance, the *E. fergusonii* plasmid pEFER also encodes three antibiotic resistance genes conferring resistance to ampicillin, streptomycin/spectinomycin and tetracycline. The products of these genes form a beta-

lactamase, a streptomycin kinase and a tetracycline exporter pump respectively, providing a significant survival advantage to its host. As a result, pEFER represents a mobile genetic element that might provide resistance to both antibiotics and phages. Plasmid pEFER encodes the six BREX genes homologous to those from *B. cereus*, from *brxA* to *brxL*. However, the pEFER BREX locus also appears to include multiple genes upstream of *brxA*. Furthermore, the additional gene *brxU* is inserted between *brxC* and *pglX*. Similarly, the *Salmonella* BREX system included genes STM4494 and STM4493 inserted between *pglX* and *pglZ*. As there is increasing evidence that phage-resistance systems are often encoded within extensive phage defence islands (Rousset et al. 2021; Zhang et al. 2019), it was hypothesised that both of these BREX loci might include multiple phage-resistance mechanisms. Our initial aim was therefore to study BREX activity in Gram-negative hosts, as part of investigating the overall impact of phage defence islands.

3.2 Open reading frames encoded by pEFER and predicted functions

The 58 ORFs of pEFER were annotated with gene size, orientation, position, and predicted function (Table 3.1). The canonical type I BREX locus consists of *brxA, brxB, brxC, pglX, pglZ* and *brxL* (Table 3.1). The putative defence island also includes the inserted *brxU* gene, and the immediate upstream gene, *brxR*, which had a predicted function as a transcriptional regulator. As a result of later experiments, it became apparent that a further two genes upstream of *brxR*, named *brxS* and *brxT*, were also required for BREX function, and have been added here for clarity. The full pEFER phage defence island is shown in bold (Table 3.1). A scale diagram of pEFER and its notable features is shown in Figure 3.1. The putative pEFER BREX locus encodes proteins with a range of predicted functions (Table 3.1).

BrxR was predicted to contain a WYL domain with a SH3 beta-barrel fold. The autorepressor sll7009 from Synechocystis sp also contains a WYL domain and has been shown to act as a negative regulator of CRISPR-Cas (Makarova et al. 2014). Genes *brxS* and *brxT* were originally excluded from the BREX locus and were not included in the initial plasmid constructs produced. However, they were later found to be required for phage defence, and that PglX activity was dependent on the presence of both *brxS* and *brxT*. The initial BREX experiments and the construction of plasmids for assaying BREX activity are detailed first in this chapter, followed by analyses including *brxS* and *brxT*. No function could be predicted for BrxA and BrxB, although BrxA has been predicted to be an inner membrane protein. BrxC is predicted to be an ATPase of the AAA superfamily and contains a phosphate binding loop. BrxU has been predicted to be an endonuclease due to its DUF1524 and DUF262 domains, and is a homologue of the fused type IV restriction enzyme GmrSD (He et al. 2015), sharing 13% sequence identity. PglX has previously been characterised as a methyltransferase (Goldfarb et al. 2015). PglZ is predicted to belong to the alkaline phosphatase family, which form dimers and require both zinc and magnesium for their activity (Orhanović and Pavela-Vrančič 2003). BrxL is a homologue of the protease LonA, which is an ATP-dependent serine peptidase (Wang et al. 1993).

Outside of the BREX locus, EFER_p0036 has been identified as the beta-lactamase, with EFER_p0037 to EFER_40 constituting a tetracycline resistance cassette and EFER_p0047 to EFER_p0049 encoding a streptomycin/ spectinomycin resistance cassette. EFER_p0053 and EFER_p0054 may encode a type II TA system of the VapC family of toxins. VapC toxins are ribonucleases that stall protein translation by degrading tRNAs (Walling and Butler 2018). A significant portion of pEFER is comprised of insertion elements and transposons with limited predicted functions. EFER_p0001 is predicted to be involved in replication initiation, although the copy number of pEFER is unknown.

Due to its size, pEFER is predicted to be low copy number and likely uses EFER_p0004/5, homologues of ParAB partition systems (Yamaichi and Niki 2000), to ensure daughter cells receive at least one copy during cell division.

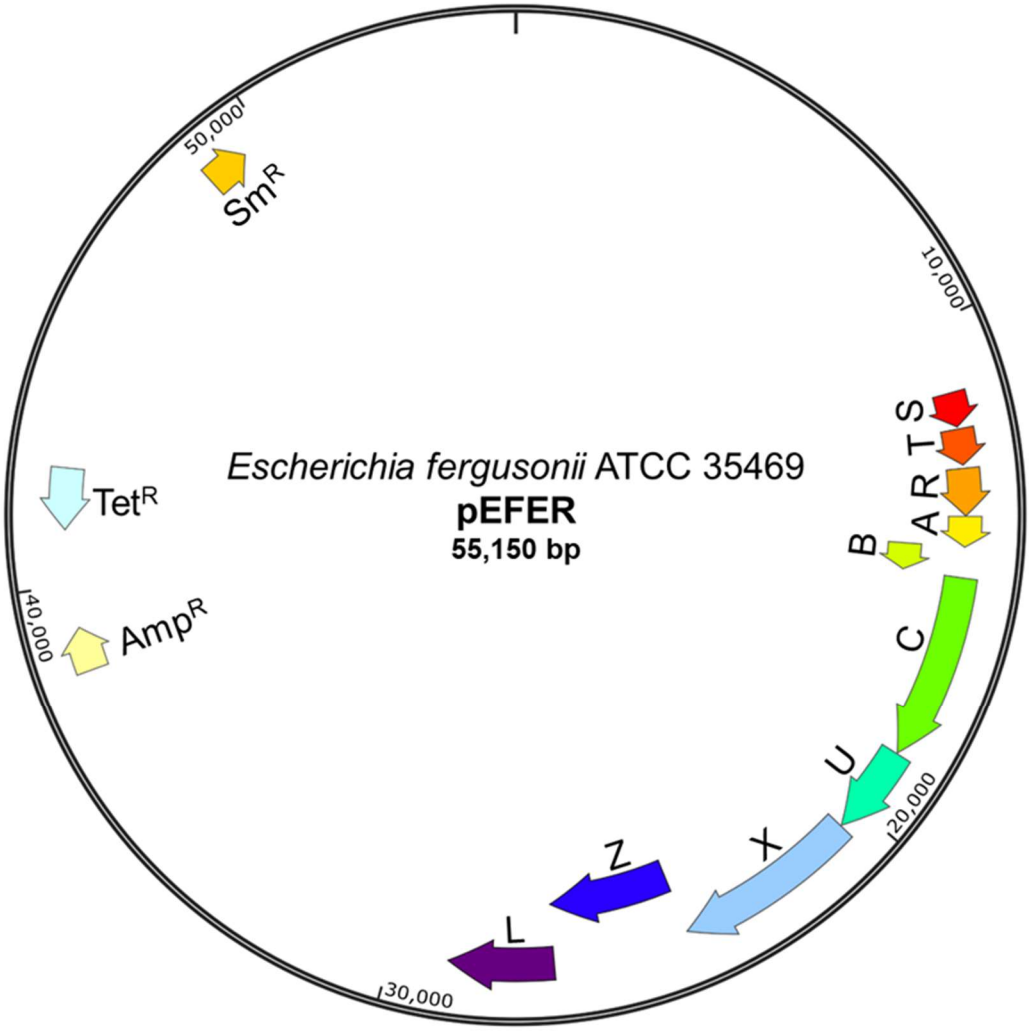


Figure 3.1: Plasmid map of* Escherichia fergusonii *ATCC 35469 pEFER with ORFs to scale. *Additional ORFs (outside of BREX) not shown. ORFs coding for antibiotic resistance cassettes are shown as Amp^R, Tet^R and Sm^R.*

Gene	Size (bp)	Orientation	Start	End	BLASTp Output and Predicted Function
EFER_p0001	866	Fw	450	1316	Replication inititation protein
EFER_p0002	1218	Fw	1637	2855	Hypothetical protein
EFER_p0003	167	Fw	1840	2007	AAA ATPase
EFER_p0004	1205	Fw	2249	3454	Plasmid partition protein ParA
EFER_p0005	977	Fw	3451	4428	Plasmid partition protein ParB
EFER_p0006	1271	Fw	4510	5781	Y family DNA polymerase SamB
EFER_p0007	707	Fw	5781	6488	ImpA
EFER_p0008	167	Fw	6370	6537	Hypothetical protein
EFER_p0009	752	Fw	6557	7309	RetA, putative reverse transcriptase
EFER_p0010	416	Rv	6614	7030	Hypothetical protein
EFER_p0011	134	Fw	7054	7188	Putative reverse transcriptase
EFER_p0012	275	Fw	7461	7736	IS1 family transposase
EFER_p0013	503	Fw	7655	8158	IS1 family transposon
EFER_p0014	299	Fw	9008	9307	Hypothetical protein
EFER_p0015	968	Fw	9620	10588	IS5 family transposase
EFER_p0016	218	Fw	10781	10999	Hypothetical protein
EFER_p0017	**671**	**Fw**	**11390**	**12061**	**IS3 family transposase BrxS**
EFER_p0018	179	Rv	12093	12272	Hypothetical protein
EFER_p0019	**704**	**Fw**	**12109**	**12813**	**Hypothetical protein BrxT**
EFER_p0020	**926**	**Fw**	**12882**	**13808**	**WYL-domain transcriptional regulator BrxR**
EFER_p0021	**599**	**Fw**	**13818**	**14417**	**Putative DUF1819 inner membrane protein BrxA**
EFER_p0022	**593**	**Fw**	**14405**	**14998**	**Hypothetical protein BrxB**
EFER_p0023	**3677**	**Fw**	**15017**	**18694**	**AAA ATPAase BrxC**
EFER_p0024	**1763**	**Fw**	**18711**	**20474**	**DUF262 DUF1524 restricition endonuclease BrxU**
EFER_p0025	**3653**	**Fw**	**20490**	**24143**	**Methyltransferase PglX**
EFER_p0026	**2657**	**Fw**	**24143**	**26800**	**Alkaline phosphatase PglZ**
EFER_p0027	**2075**	**Fw**	**26817**	**28892**	**ATP dependent Lon-like protease**
EFER_p0028	746	Rv	28885	29631	IS110 family transposase
EFER_p0029	179	Rv	29000	29179	IS3 transposase
EFER_p0030	2885	Rv	29749	32634	Tn3 family transposase TnpA
EFER_p0031	623	Fw	32760	33383	Recombinase
EFER_p0032	296	Fw	33393	33689	Transposase element YahA
EFER_p0033	665	Fw	33686	34351	IS3/ IS1133 family transposase
EFER_p0034	3164	Rv	34376	37540	Tn3 family transposase TnpA
EFER_p0035	593	Fw	37509	38102	Tn3 resolvase TnpR
EFER_p0036	860	Fw	38285	39145	Broad sprectrum beta-lactamase
EFER_p0037	593	Fw	39261	39854	IS4 family transposase
EFER_p0038	416	Rv	39864	40280	Tetracycline efflux protein TetD
EFER_p0039	668	Fw	40293	40961	Tetracycline efflux protein TetC
EFER_p0040	1253	Rv	41074	42327	Tetracycline efflux protein TetA
EFER_p0041	626	Fw	42358	42984	Tetracycline repressor protein TetR
EFER_p0042	686	Rv	42962	43648	Transcriptional regulator ArsR
EFER_p0043	497	Rv	43656	44153	Hypothetical protein
EFER_p0044	50	Rv	44305	44355	Antibiotic monoxygenase
EFER_p0045	1205	Fw	44799	46004	Glutamate permease JemA
EFER_p0046	1208	Rv	46370	47578	IS4 family transposase
EFER_p0047	182	Fw	47886	48068	Streptomcyin kinase StrA
EFER_p0048	803	Fw	47864	48667	Streptomcyin kinase StrB
EFER_p0049	182	Fw	47886	48068	Streptomcyin kinase StrB
EFER_p0050	566	Fw	49475	50041	IS110 family transposase
EFER_p0051	242	Rv	49963	50205	AAA ATP-binding transposase
EFER_p0052	587	Rv	50260	50847	IS3 family transposase
EFER_p0053	416	Rv	51383	51799	Type II TA family toxin VapC
EFER_p0054	230	Rv	51796	52026	Type II TA family antitoxin AbrB
EFER_p0055	350	Fw	52600	52950	Hypothetical Protein with helix turn helix domain
EFER_p0056	743	Fw	53001	53744	Hypothetical Protein
EFER_p0057	776	Rv	53741	54517	Phage integrase
EFER_p0058	290	Rv	54575	54865	Hypothetical Protein

Table 3.1: Annotation of pEFER ORFs with predicted protein function. *BREX locus shown in* ***bold***.

3.3 pEFER and *S. enterica* as model systems for studying BREX

Having identified the putative phage defence island from pEFER, it can be compared to the chromosomal phage defence island in *Salmonella* D23580 (Figure 3.2). The D23580 BREX locus shares the six core type 1 BREX genes in the pEFER and *B. cereus* systems, with two additional ORFs located between *pglX* and *pglZ,* denoted STM4494 and STM44943 (Figure 3.2). This cassette is encoded on the antisense strand of the D23580 genome. Bioinformatic analysis of these additional ORFs indicates that they function as a standalone phage defence system, homologous to the two component anti-anti-restriction system, PARIS (phage anti-restriction-induced system). As detailed by Rousset et al, PARIS consists of a AAA+ ATPase and a DUF4435 protein and confers resistance to a broad range of phages (Rousset et al. 2021). In D23580, STM4494 is predicted to be an AAA+ ATPase and STM4493 contains DUF4435, indicating that a PARIS system is embedded within the BREX locus. As a result, both the pEFER and D23580 BREX loci constitute phage defence islands, with pEFER encoding a type IV restriction system and D23580 encoding PARIS. D23580 provides a model chromosomal system in a pathogenic strain of *S. enterica* whilst pEFER constitutes a plasmid-borne BREX system, allowing ease of mutagenesis.

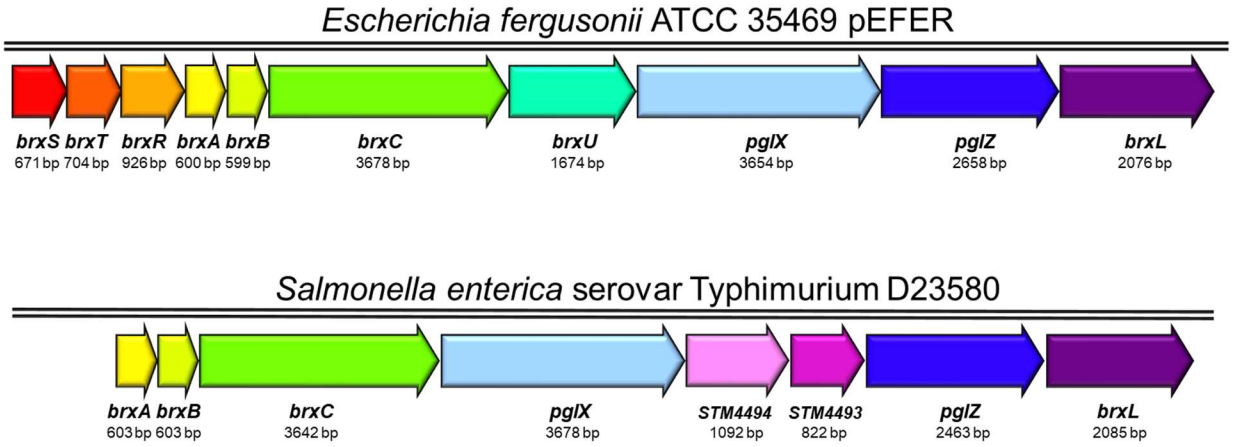


Figure 3.2: BREX phage defence islands in* E. fergusonii *pEFER and* S. enterica *D23580. *Homologues are shown in conserved colours.*

3.4 Isolation of environmental phages and initial testing against pEFER.

Having mapped out our putative phage-resistance systems, we needed suitable phages for testing. Phages used in this study were isolated from freshwater sources within Durham, UK, as part of demonstrations performed for third-year undergraduate practical laboratory sessions. Over two years, thirty environmental coliphages were isolated and purified. Based on transmission electron microscopy undertaken by the students, all phages isolated in this study are dsDNA phages belonging to the Myoviridae, Siphoviridae and Podoviridae families. Using this diverse collection, it was then possible to perform Efficiency Of Plating (EOP) assays to determine anti-phage activity. Plasmid pEFER was used to transform *E. coli* DH5α, selecting for ampicillin resistance via the native beta-lactamase encoded on pEFER. The EOPs obtained using DH5α pEFER, against a DH5α control, demonstrate that pEFER reduced plating for 22 of the 30 phages (Table 3.2). The scale of effect varied, with phages such as CP and Paula unable to form plaques on DH5α pEFER, whereas phages such as TB34 and Trib had a 100-fold reduction (Table 3.2).

In order to attribute this anti-phage activity to the BREX locus, transposon mutagenesis was performed on pEFER to produce knockout mutants. Positive clones from the mutagenesis were screened for (i) loss of anti-phage activity against phage Paula or (ii) loss of streptomycin resistance. Following screening, mutant plasmids were isolated and the insertion sites were mapped in pEFER using random-primed PCR (Zou et al. 2003). Clone Km5 was identified by screening for sensitivity to phage Paula, and had a transposon insertion within *brxA*, so was therefore considered to be a BREX knock-out mutant (Figure 3.3).

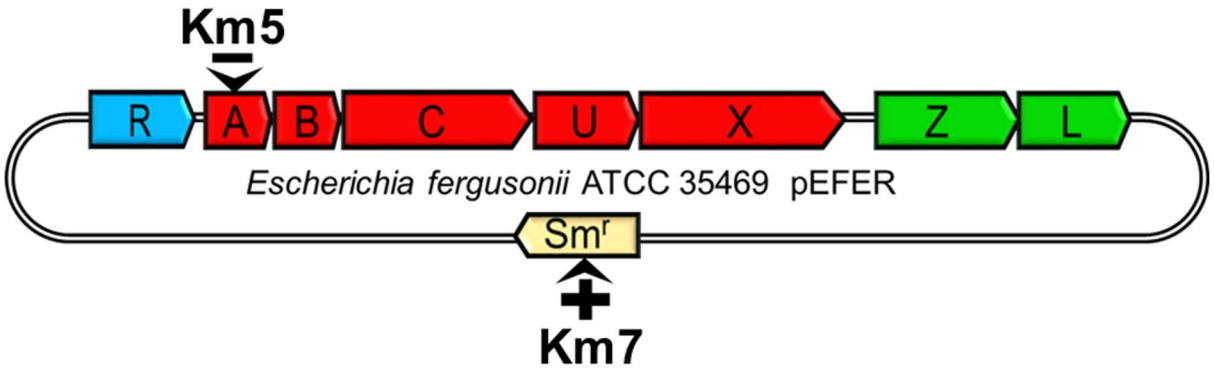


Figure 3.3: Plasmid map of pEFER showing Tn5 insertion locations. *Km5 is the result on Tn5 inserting into brxA (EFER_p0021), and Km7 is the result of Tn5 inserting into the streptomycin resistance gene StrB (EFER_p0048).*

Clone Km7 was identified by screening for streptomycin sensitivity, and had a transposon insertion within the streptomycin gene (Figure 3.3). Both of these mutant plasmids were tested against the suite of thirty phages, again using DH5α as a common control strain (Table 3.2). Km5 had significantly reduced impact on phage plating compared to pEFER wild type (WT) (Table 3.2). Despite the reduction in anti-phage activity, phages Bam, CS16, and Titus were still impacted by Km5, though to a lesser extent than by pEFER WT (Table 3.2). Interestingly, the Km5 mutation had no impact on the phage-resistance seen against phages Mak and Mav (Table 3.2), which could imply a further phage-resistance system on the pEFER plasmid but outside of the identified phage defence island. In contrast, the Km7 mutant had broadly the same phage-resistance profile as for pEFER WT (Table 3.2).

Φ	pEFER	Km5	Km7
AL25	1.03	0.9	0.86
Alma	0.73	0.97	0.96
Bam	5.75×10^{-4}	1.01×10^{-3}	1.50×10^{-4}
Baz	0.73	1.4	0.35
BB1	0.48	1.32	0.62
BGP	4.00×10^{-2}	0.14	9.09×10^{-2}
BHP	1.21	1.29	1.34
CP	$<1.36 \times 10^{-8}$	0.17	1.36×10^{-7}
CS16	5.95×10^{-8}	2.73×10^{-3}	$<8.26 \times 10^{-8}$
EH2	$<3.44 \times 10^{-7}$	1.94	$<3.44 \times 10^{-7}$
EL	$<3.31 \times 10^{-8}$	0.18	$<3.31 \times 10^{-9}$
Geo	$<2.19 \times 10^{-8}$	2.06	$<2.19 \times 10^{-8}$
Jura	1.07	1.22	1.28
Mak	2.43×10^{-3}	3.73×10^{-3}	7.06×10^{-4}
Mav	$<2.37 \times 10^{-8}$	3.43×10^{-7}	$<2.05 \times 10^{-8}$
NP	$<4.11 \times 10^{-8}$	0.17	$<4.11 \times 10^{-8}$
NR1	$<3.26 \times 10^{-8}$	0.13	$<3.2 \times 10^{-8}$
PATM	$<4.17 \times 10^{-8}$	1.06	$<3.08 \times 10^{-8}$
Paula	$<2.86 \times 10^{-8}$	1.19	$<2.87 \times 10^{-8}$
QOTSP	4.92×10^{-7}	0.11	$<1.50 \times 10^{-8}$
SAP	$<7.95 \times 10^{-8}$	1.74	$<7.95 \times 10^{-8}$
Sipho	2.04×10^{-3}	3.12	1.38×10^{-4}
Solly	4.21×10^{-7}	0.11	$<8.51 \times 10^{-7}$
Some	$<1.40 \times 10^{-7}$	1.97	$<1.40 \times 10^{-7}$
SPSP	1.26	1.07	0.63
TB34	7.37×10^{-2}	0.69	0.12
TB36	$<1.69 \times 10^{-7}$	0.30	$\leq 3.47 \times 10^{-8}$
TB37	1.78	1.07	1.12
Titus	4.71×10^{-3}	3.00×10^{-2}	3.01×10^{-3}
Trib	2.49×10^{-2}	0.45	9.46×10^{-3}

Table 3.2: EOP values for pEFER-based assays. *Values are mean EOPs from triplicate data. Values with < extended below the range of this assay and formed no plaques.*

EOP ≥ 0.1
0.1 > EOP ≥ 0.01
0.01 > EOP ≥ 0.001
0.001 > EOP ≥ 0.0001
EOP < 0.0001

3.5 Characterisation of *Salmonella* BREX and generation of BREX knockout

Having confirmed that pEFER provides phage resistance against a wide range of environmental coliphages, a strategy was devised to similarly investigate the Salmonella BREX system. As *Salmonella* ST313 strain D23580 encodes BREX within its chromosome, in order to isolate any phages that might be susceptible to BREX, it was first necessary to delete the BREX locus. This was performed using λ-red recombination (Datsenko and Wanner 2000). The deletion mutant would then allow potential BREX sensitive ST313 phages to be enriched from environmental samples, prior to isolation and testing against ST313 WT. The ST313 strain D23580 is highly virulent due its repertoire of 5 prophages within its chromosome and poses a significant health hazard. D23580Δφ is a mutant with all prophage regions removed and can be safely worked on within a containment level 2 laboratory (Owen et al. 2017). D23580Δφ will from here on be referred to as D23. The progenitor D23580, which contains the prophage regions, is not used in this study. The BREX locus was removed from D23 using scar-less λ red recombination, as described in Blank et al, 2011 (Blank, Hensel, and Gerlach 2011). Flanking primers were designed to remove all ORFs from STM4498 (*brxA)* to STM4491 (*brxL*) (Figure 3.4).

pKD4 was used as a template for PCR using pKD4 ΔBREX FW/RV to generate a 1.4 kb fragment with homologous terminal regions for the target sites (STM4498/91). D23 was electroporated with this cassette, and the kanamycin resistance cassette was removed with FLP recombinase (Derous et al. 2011). Following removal of the kanamycin cassette, the deletion site was confirmed by sequencing of amplicons generated across the deletion site, showing an inframe fusion of the upstream and downstream regions flanking the BREX locus. This produced the strain D23850ΔφΔBREX, which will be referred to as D23ΔB.

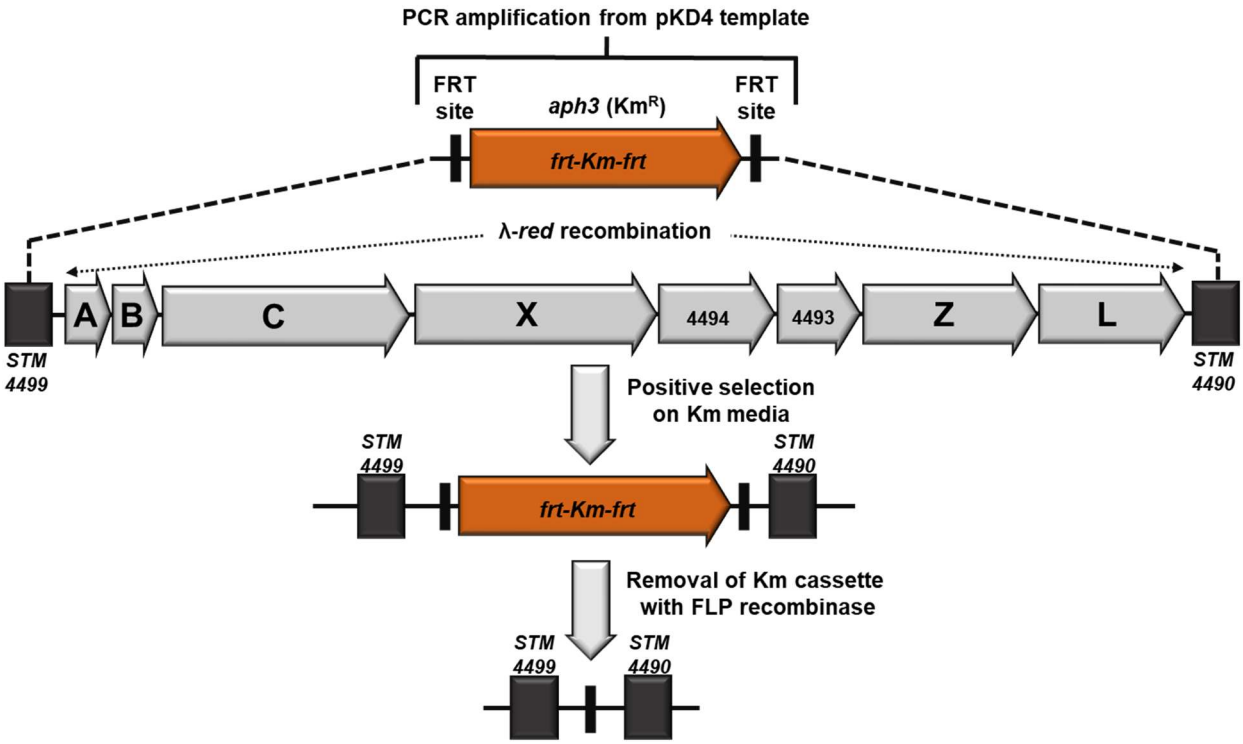


Figure 3.4: Generation of BREX knockout using lambda red recombination. *All genetic material between* STM4498 *and* STM4991 *was removed and the kanamycin resistance cassette was cured.*

3.6 Isolation of ST313 Phages

Attempts were made to isolate *Salmonella* phages from freshwater sources within Durham, UK, using the D23ΔB strain. Unfortunately, no phage plaques were obtained following enrichment. Sewage effluent was then obtained with assistance from Northumbrian Water and this was used for phage enrichment on D23ΔB. A range of plaques were obtained following these enrichments, and eight phage lysates were prepared from visibly distinct plaques. EOP assays were performed for the eight phage isolates, testing the ability of the phages to plaque on D23, with D23ΔB as the control (Table 3.3).

Phages KMP, SB58 and SL2K were sensitive to the *Salmonella* phage defence island, with a 100-fold reduction in EOP (Table 3.3). Phage DB1 was weakly affected, with an EOP of 0.13 (Table 3.3). The remaining four phages appeared unaffected. These data confirm that the phage defence island within D23 is active, and open to further study.

Phage	EOP
KMP	3.82×10^{-2}
LTE	1.12
AWAQ	1.14
SB58	4.23×10^{-2}
SL2K	3.03×10^{-2}
SGP	0.71
DA1	0.98
DB1	0.13

EOP ≥ 0.5
0.5 > EOP ≥ 0.1
EOP < 0.1

Table 3.3: EOP values for phages against D23*. These are mean EOP values from triplicate data.*

3.7 Growth curves with pEFER, Km5, Km7 and WT.

Having demonstrated that both the model phage defence islands from *E. fergusonii* and *Salmonella* were indeed active, the *E. fergusonii* model was selected to pursue due to the greater range of observed effects versus phages. To continue to characterise the pEFER system, the growth of DH5α and DH5α pEFER was monitored, with and without the addition of phages (Figures 3.5 and 3.6). Phages Pau ($p = 7.3x10^{-13}$) and PATM ($p = 1.4x10^{-15}$) drastically reduced the optical density of DH5α cultures after 100 minutes, followed again by recovery at around 300 minutes, whereas the DH5α pEFER continued to grow throughout (Figure 3.5A and B). In contrast, phage Baz ($p = 0.16$) was unaffected by pEFER (Table 3.2) and for both DH5α and DH5α pEFER, optical density is reduced in Baz-infected cultures, with growth recovery occurring around 300 minutes post-infection, likely due to the selection of phage-resistant mutants (Figure 3.5C). Phage Trib ($p = 4.6x10^{-14}$) also reduced growth of the DH5α culture, and the DH5α pEFER culture continued to grow, but it reached a lower stationary density than the cultures infected with PATM and Paula (Figure 3.5D). This might be due to pEFER providing lower resistance to Trib than to PATM and Paula (Table 3.2 and Figure 3.6). The DH5α and DH5α pEFER cultures grew similarly when not challenged with phages ($p = 0.09$) (Figure 3.5E). The reduced initial growth rate observed in the DH5α pEFER uninfected culture can be attributed to the addition of chloramphenical in the growth media which was absent from the DH5α uninfected culture.

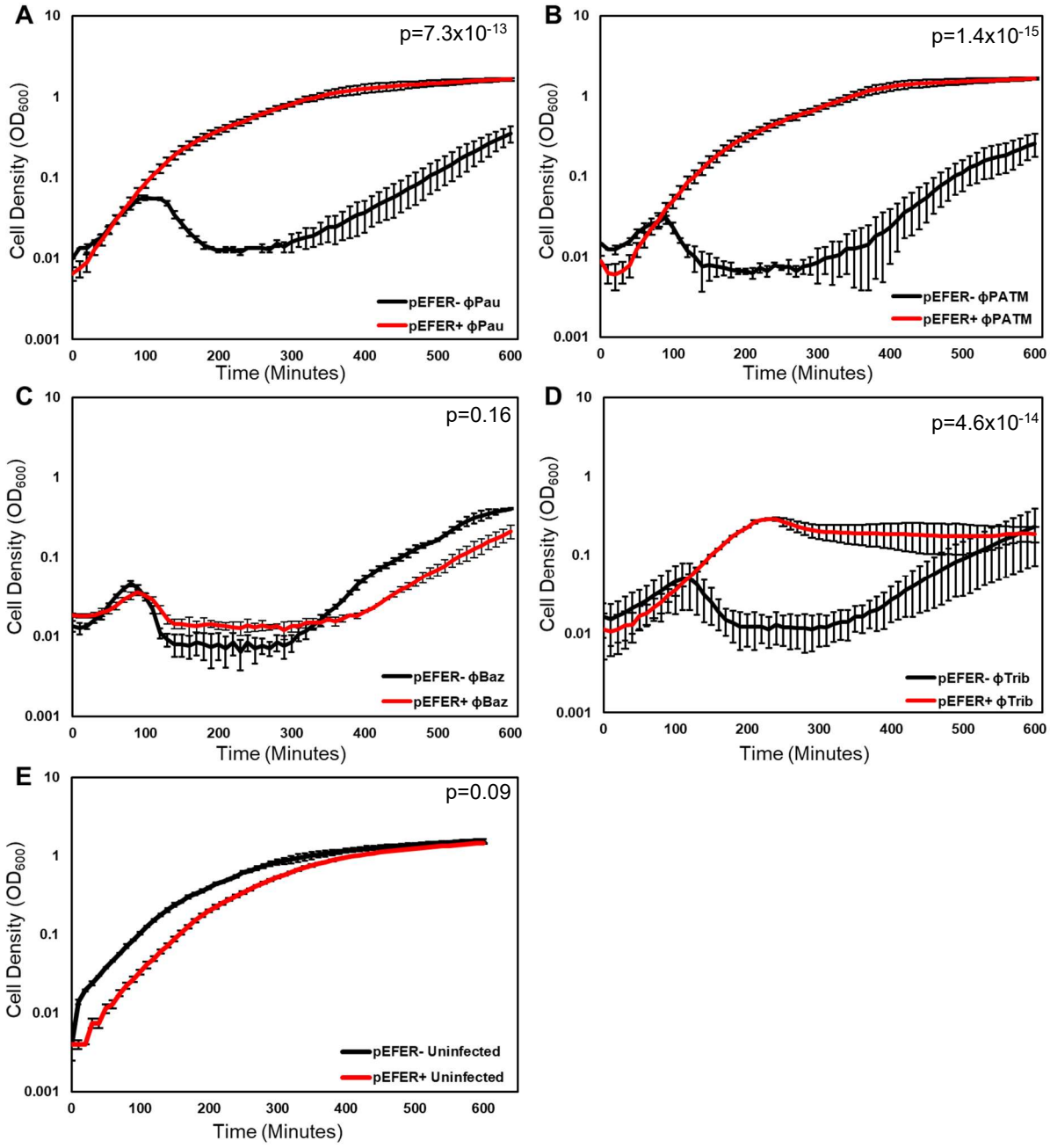


Figure 3.5: Culture dynamics of phage infected DH5α (black) and DH5α containing pEFER (red). *A+B) No significant reduction in culture is observed in pEFER+ cultures when infected with Pau and PATM. C) Reduction in culture density is observed for pEFER+ cultures when infected with Baz. D) pEFER confers significant resistance to Trib, however culture density is reduced in comparison to uninfected controls. Bacterial cultures were grow3n to mid exponential phase and diluted to a starting OD_{600} of 0.05. Cultures were inoculated at time point 0 minutes and cell density was measured every 10 minutes (A-D). Experiments were run in triplicate and the mean values are plotted. Error bars represent standard deviation of replicates. An uninfected control is shown in E which has not been inoculated with phage. A one-way ANOVA was performed for each dataset to indicate statistical significance between pEFER- and pEFER+ strains.*

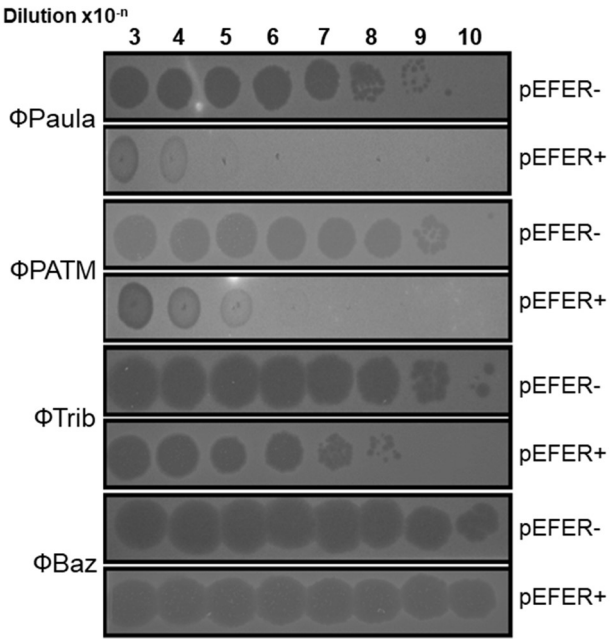


Figure 3.6: Representation of EOP assays for phages used in growth assays. *Phages are pipetted onto bacterial soft agar lawns. Phage stock serial dilutions were prepared from 10^{-1} to 10^{-8} and 10 ul was added. Baz shows no reduction in EOP against pEFER. Paula, PATM and Trib are observed to have reduced EOP values.*

3.8 Golden Gate Assembly of pGGA-BREX

Due to the observation of additional anti-phage activity encoded by pEFER outside of the identified phage defence island (Table 3.2), it was decided to prepare a plasmid containing only the target region (Figure 3.3). Initial designs did not take genes *brxS* and *brxT* into account, and instead focussed on sub-cloning from *brxR* to *brxL* (Figure 3.7 A). Due to the large size of this region (18 kb), a single PCR fragment could not be obtained, and efforts to produce an intact locus via overlap PCR were unsuccessful. A new approach was prepared, based on Golden Gate Assembly (GGA) (Engler, Kandzia, and Marillonnet 2008; NEB 2020). The BREX locus was split into 7 fragments of ~3 kb sizes, encompassing the entire locus and upstream promoter regions from nucleotide positions 12400 to 28950 (Figure 3.7 A). Each fragment was amplified via PCR with primers containing BamHI/PstI restriction sites for cloning into pTRB479. pTRB479 is a pUC19 derivative prepared by Dr Tim Blower, which has undergone mutagenesis to remove additional BsaI sites that would otherwise interfere with GGA reactions. To prepare suitable fragments for GGA cloning, primers were designed to

include ~20 bp terminal regions encoding upstream BsaI sites, and a further restriction cloning site for cloning into pTRB479 (Table 3.4). Fragments were cloned into pTRB479 and confirmed via sequencing with M13 Fw and Rv. An additional primer (F7_C_Rv) was used to produce a GGA construct containing only fragment 1 for use as a negative control plasmid in phage assays.

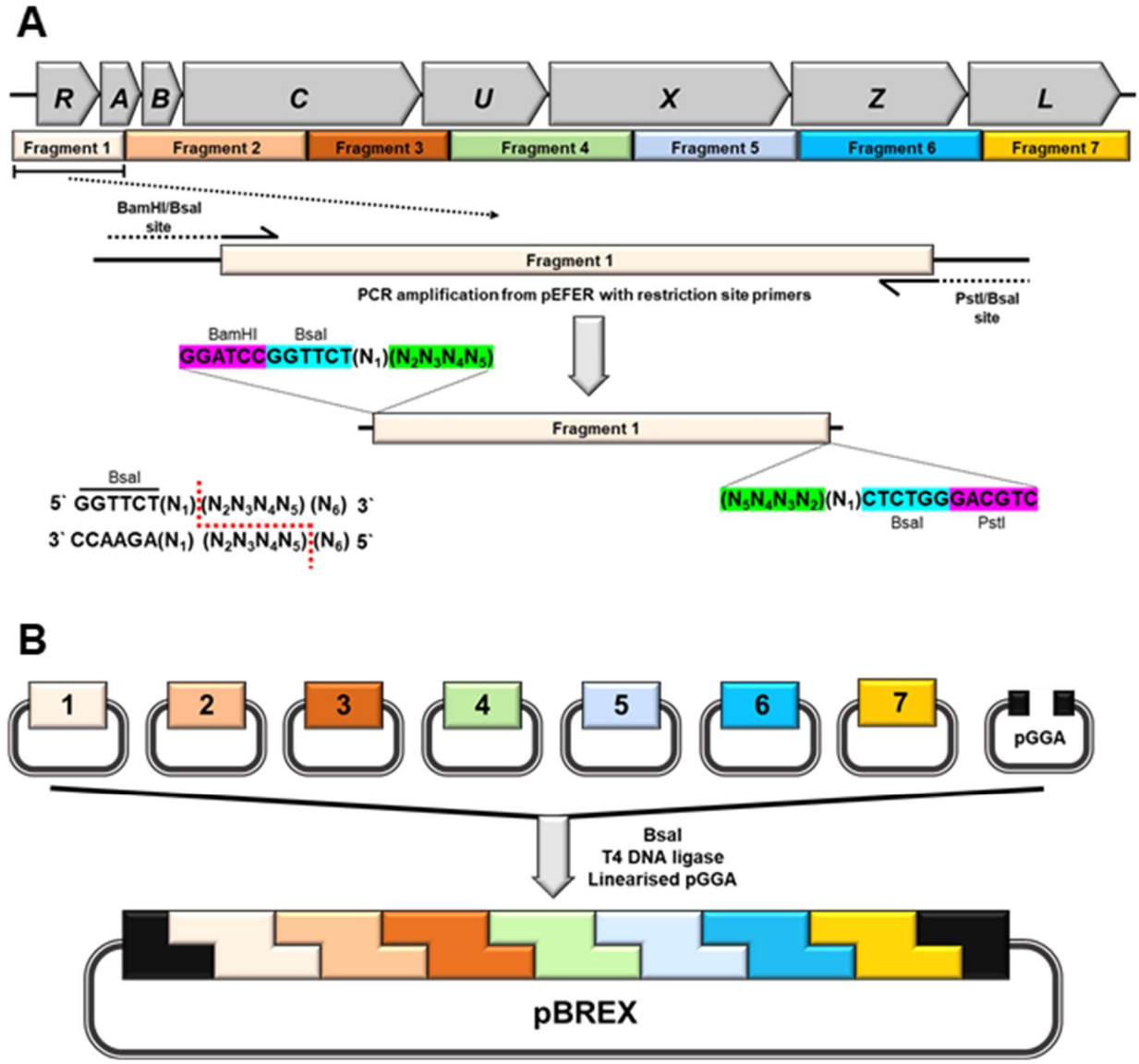
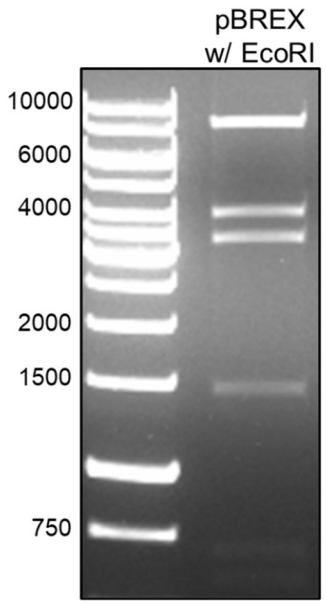


Figure 3.7: Synthesis pathway for pBREX via GGA*. A) The BREX locus is divided into 7 sections and each individual fragment is amplified and cloned in pTRB479. B) pTRB479 donor plasmids containing BREX fragments are digested with golden gate reaction mix. BsaI cleaves donors and T4 DNA ligase fuses the cohesive assembly.*

Primer	Sequence
F1_Fw	TTTGGATCCGGTCTCGGGAGGTTATGGCTGGATCACAGC
F1_Rv	TTTCTGCAGGGTCTCGACCTCAATCAAATCTTCCCG
F2_Fw	TTTGGATCCGGTCTCGAGGTCATGGAGTGCACAATATGAAGG
F2_Rv	TTTCTGCAGGGTCTCGCTGACATTTCAACGTCAGTG
F3_Fw	TTTGGATCCGGTCTCCTCAGAAGTCTCGAGCAAGCTTTCAG
F3_Rv	TTTCTGCAGGGTCTCCGTTTCTTCAATGCGCTGTCTGC
F4_Fw	TTTGGATCCGGTCTCGAAACCTCTTCAATACCACAAATTATC
F4_Rv	TTTCTGCAGGGTCTCGTGTCACGAGAGCGGAAAC
F5_Fw	TTTGGATCCGGTCTCGGACATTCCAAAGATGATTCTGGAGAA
F5_Rv	TTTCTGCAGGGTCTCGCAAGATCAAACTTTCCTAGCTG
F6_Fw	TTTGGATCCGGTCTCCCTTGAGACGTCACTCTGGGCATATCTGAG
F6_Rv	TTTCTGCAGGGTCTCCTAGAAAAAATCATCTTGGAATGCCAAATC
F7_Fw	TTTGGATCCGGTCTCCTCTAAGGCGCTGAGGTGA
F7_Rv	TTTCTGCAGGGTCTCCATGGAGGACAGAAACTGTCTACCG
F7_C_Rv	TTTCTGCAGGGTCTCCATGGAGGACAGAAACTGTCTACCG

Table 3.4: Primer sequences for donor fragment PCR. *Bold text highlights BamHI restriction sites in **RED**, PstI restriction sites in **PINK**, BsaI recognition sites in **GREEN** and BsaI cut sites containing are highlighted in **CYAN** and shown in **bold**.*

Equimolar concentrations of each of the seven donor plasmids were used in a GGA reaction with pGGA. As show in in Table 3.4 in cyan, each primer contains a unique 4 nucleotide sequence downstream of the BsaI recognition sequence. This results in the generation of unique cohesive ends, ensuring that fragments are ligated in the correct order and orientation. In Figure 3.7A, fragment amplification is shown with flanking regions containing the BsaI recognition and cleavage sites. In Figure 3.7B, individual donor plasmids can be seen undergoing digestion and ligation to form pBREX. Transformants were selected on chloramphenicol containing agar. Extracted plasmids were digested with EcoRI and the digestion pattern was compared to the predicted fragment lengths (Figure 3.8). The final resulting plasmid was named pBREX (Figure 3.7). A second control plasmid, pGGA-C, was constructed using primers F1_Fw and F7_C_Rv (Table 3.4), and therefore contains the pGGA backbone and a single fragment of the BREX locus.



#	Coordinates	Length (bp)
1	6205-14619	8415
2	2116-6204	4089
3	14620-18022	3403
4	634-2115	1482
5	18023-18687	665
6	34-633	600
7	1-33	33

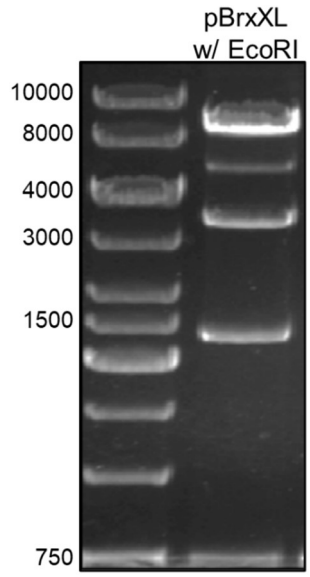
Figure 3.8: Restriction digest of pBREX with EcoRI. *500 ng of plasmid DNA was incubated with EcoRI for 3 hours at 37 °C and resolved via agarose gel electrophoresis at 120 V for 60 minutes in 1% agarose in 1x TAE. Fragment sizes were predicted using NEBcutter V2.0.*

Having obtained the sub-clones of pEFER, EOP assays were performed with DH5α pBREX, using DH5α pGGA-C as a negative control. Unexpectedly, phages that were previously shown to be affected by pEFER were not inhibited by pBREX, with Paula and PATM being observed to have EOP values of 1.21 and 1.43, respectively. To ascertain whether the BREX locus in pBREX is active, we undertook PacBio methylome sequencing of genomic DNA extracts from *E. coli* strain NEB2796 containing pEFER or pBREX, in an attempt to observe BREX-dependent DNA modifications as per the *B. cereus* system (Goldfarb et al. 2015). This work was done in collaboration with Dr Rick Morgan (New England Biolabs) and Dr Darren Smith (Northumbria University). Whilst modifications were observed with pEFER (see section 3.11), no modifications were observed with pBREX. This suggested that the BREX system was not active in pBREX. It was hypothesised that two further ORFs upstream of *brxR* in pEFER, denoted *brxS* and *brxT,* might be required for BREX activity as they could potentially be

operonic. Upon closer inspection, these two ORFs were found to be conserved in other γ-proteobacterial BREX systems.

3.9 Addition of *brxS* and *brxT* to sub-cloned BREX constructs

Two additional pTRB479 based donor plasmids were produced, with F_0 containing the *brxS* and *brxT* regions, and a new F1 that could form cohesive ends to F_0 in a GGA reaction. The resulting GGA plasmid constructs were digested with EcoRI and the digest pattern was analysed (Figure 3.9).



#	Coordinates	Length (bp)
1	7158-15572	8415
2	2116-7157	5042
3	15572-18975	3403
4	634-2115	1482
5	18976-33	698
6	34-633	600

Figure 3.9: Restriction digest of pBrxXL with EcoRI. *500 ng of plasmid DNA was incubated with EcoRI for 3 hours at 37 °C and resolved via agarose gel electrophoresis at 120 V for 60 minutes in 1% agarose in 1x TAE. Fragment sizes were predicted using NEBcutter V2.0.*

The additional of *brxS* and *brxT* successfully increased the size of fragment 2 by 953 bp, allowing easy identification of correct assembly. This larger plasmid was named pBrxXL. pGGA and internal BREX primers were used for sequencing to confirm the construct assembly was correct. Preliminary EOP testing showed pBrxXL conferred resistance to pEFER-susceptible phages, which indicated this construct was a suitable minimal template for further mutagenesis studies, prior to full EOP assays (Section 3.12).

3.10 Golden gate assembly of mutant BREX constructs

In order to assess the role of individual components of the BREX locus, gene deletion donor plasmids were prepared. pTRB479-based donor plasmids were produced using truncated PCR fragments, removing the ribosomal binding site (RBS) and beginning of individual target genes. GGA reactions were then successfully performed as previously described to generate *ΔbrxS, ΔbrxT, ΔbrxU, ΔpglX* single mutants, and a *ΔbrxUΔpglX* double mutant. Deletion donor plasmids were produced for *ΔbrxR, ΔbrxL, ΔbrxC*, *ΔbrxB* and *ΔpglZ* however no correctly assembled GGA constructs were obtained for these mutants.

3.11 Identification of PglX target site and host methylome sequencing

Further runs of PacBio sequencing was used to confirm the target sequence of the pEFER methyltransferase PglX, and the impact of the generated mutations. Strain ER2796 is a methyltransferase deficient *E. coli* that can be used to unambiguously deduce methylation patterns from introduced methyltransferases (Anton et al. 2015). ER2796 was individually transformed with pBREX, pBrxXL, pBrxXL-*ΔbrxU,* pBrxXL-*ΔpglX* and pBrxXL-*ΔbrxUΔpglX*. ER2796 could not be transformed with pEFER, for reasons that remain unclear. DH5α was used as a host for pEFER. Genomic DNA was prepared from each strain, and also ER2796 and DH5α without plasmids. These samples, alongside *E. fergusonii* ATCC 35469 genomic DNA (purchased from ATCC), were sent for PacBio sequencing, once again performed by Dr Rick Morgan (New England Biolabs).

The motif GCGA**A**T was observed to be methylated at the fifth position of DH5α gDNA when pEFER was present, and this was not seen when pEFER was absent. This pattern had previously also been observed in earlier samples of *E. coli* NEB2796 pEFER genomic DNA (see section 3.8). The *E. fergusonii* genomic DNA also showed this methylation motif, amidst

other modifications. This motif is distinct from both the original PglX motif identified for the *B. cereus* BREX system (Goldfarb et al. 2015) and the *E. coli* BREX system (Gordeeva et al. 2019), but all three have N6-methylation of adenines at the 5th position of the 6 bp non-palindromic motif (Figure 3.10). As non-palindromic sequences are only hemi-methylated, methylation is not observed on the complementary strand (Marinus and Casadesus 2009; Adhikari and Curtis 2016). Methylation at GCGA**A**T was also observed in gDNA prepared from ER2796 in the presence of pBrxXL and pBrxXL-Δ*brxU*, which suggests that BrxU is not required for methylation. In contrast, this methylation pattern was not observed for gDNAs prepared from pBrxXL-Δ*pglX* and pBrxXL-Δ*brxUΔpglX,* indicating that it is PglX-dependent.

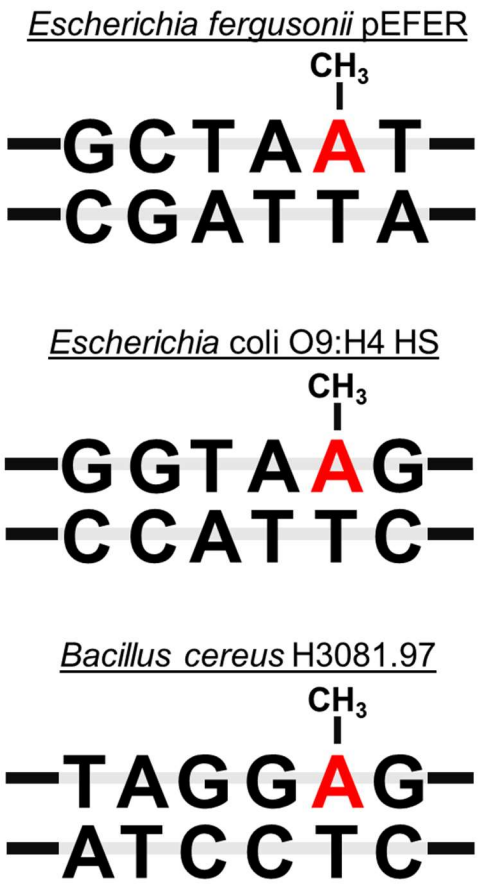


Figure 3.10: Target sequences of the BREX methyltransferase in different species. *Non-palindromic DNA sequences are hemimethylated at the adenine at the fifth position shown in* ***red.***

3.12 Efficiency of Plating assays with GGA constructs

The full library of 30 phages were tested against four of the GGA constructs, using DH5α as the host and pGGA-C as a control. pBrxXL retained a similar phage-resistance profile to pEFER (Table 3.5). pBrxXL provided resistance against 18 of the phages, with a range of effects, from prevention of plaque formation (eg. Paula), to a modest 100-fold decrease in EOP (eg. BGP). Phages such as Bam, Mak, Mav and Titus were not affected by pBrxXL, which further supports the earlier hypothesis that there is another phage-resistance system elsewhere on pEFER, as shown by EOP data with Km5 (Table 3.2). When tested against BrxXL-*ΔpglX*, phages CS16, PATM, Paula, Sipho and Trib were no longer reduced in EOP (Table 3.5). This indicates that these phages are targeted by a PglX-dependent mechanism, therefore BREX. Surprisingly, these data demonstrate that the more extensive range of phage resistance is BREX-independent. EOP values on the pBrxXL-*ΔbrxU* strain show that resistance against phages BGP, CP, EH2, EL, Geo, NP, NR1, QOTSP, SAP, Solly, Some, TB36 and TB37 is dependent on BrxU (Table 3.5). TB37 was shown to be resistant to pEFER, but was found to be susceptible to pBrxXL. A repeat of TB37 will be performed as it is possible the lysate preparation at the time of testing with pEFER was of poor quality. All phage resistance is removed in the double mutant (Table 3.5). Collectively, these data show that BrxU is the primary source of phage-resistance in the pEFER phage defence island.

In order to better understand the BREX mechanism, attempts were made to isolate 'escape mutant' phages for BREX-sensitive phages PATM and Paula. In order to obtain plaques for PATM and Paula, the maximum phage concentration used in plating assays was increased 10-fold. Individual plaques that had formed on pBrxXL-*ΔpglX* containing hosts were isolated and fresh phage lysates were prepared EOP assays were repeated with the putative escape mutants, however values remained in line with previous EOP assays.

Φ	pEFER	pBrxXL	pBrxXL *ΔpglX*	pBrxXL *ΔbrxU*	pBrxXL *ΔbrxU ΔpglX*
AL25	1.03	0.83	0.22	1.42	0.88
Alma	0.73	0.11	1.33	1.02	1.7
Bam	5.75×10^{-4}	0.65	1.02	0.64	1.18
Baz	0.73	1.49	2.01	1.5	1.92
BB1	0.48	0.33	1.12	0.79	1.72
BGP	4.00×10^{-2}	1.14×10^{-2}	5.32×10^{-2}	0.87	1.79
BHP	1.21	0.74	2.01	0.32	2.11
CP	$<1.36 \times 10^{-8}$	$<6.00 \times 10^{-8}$	$<4.92 \times 10^{-8}$	0.76	3.45
CS16	5.95×10^{-8}	6.47×10^{-3}	1.73	5.32×10^{-3}	1.45
EH2	$<3.44 \times 10^{-7}$	$<2.46 \times 10^{-9}$	$<1.67 \times 10^{-9}$	1.42	5.3
EL	$<3.31 \times 10^{-8}$	$<1.58 \times 10^{-7}$	$<4.26 \times 10^{-7}$	0.99	0.81
Geo	$<2.19 \times 10^{-8}$	$<1.70 \times 10^{-7}$	$<3.03 \times 10^{-7}$	0.92	0.73
Jura	1.07	0.68	0.63	1.51	1.14
Mak	2.43×10^{-3}	1.24	0.98	2.01	1.91
Mav	$<2.37 \times 10^{-8}$	0.87	0.63	0.77	0.68
NP	$<4.11 \times 10^{-8}$	$<1.07 \times 10^{-7}$	$<1.07 \times 10^{-7}$	1.2	1.32
NR1	$<3.26 \times 10^{-8}$	2.24×10^{-7}	1.02×10^{-7}	1.32	1.21
PATM	$<4.17 \times 10^{-8}$	$<7.76 \times 10^{-7}$	0.7	$<7.76 \times 10^{-7}$	1.29
Paula	$<2.86 \times 10^{-8}$	$<2.54 \times 10^{-7}$	1.53	$<2.54 \times 10^{-7}$	2.14
QOTSP	4.92×10^{-7}	$<8.33 \times 10^{-8}$	$<8.33 \times 10^{-8}$	2.73	0.91
SAP	$<7.95 \times 10^{-8}$	$<1.71 \times 10^{-7}$	$<4.52 \times 10^{-7}$	1.1	0.82
Sipho	2.04×10^{-3}	8.33×10^{-3}	0.58	6.30×10^{-3}	2.08
Solly	4.21×10^{-7}	$<2.48 \times 10^{-7}$	$<3.07 \times 10^{-7}$	0.73	1.82
Some	$<1.40 \times 10^{-7}$	$<7.14 \times 10^{-9}$	$<2.86 \times 10^{-8}$	0.32	3.86
SPSP	1.26	1.21	0.46	1.51	0.61
TB34	7.37×10^{-2}	1.9	1.51	0.51	1.6
TB36	$<1.69 \times 10^{-7}$	$<2.07 \times 10^{-8}$	$<2.07 \times 10^{-8}$	2.04	0.84
TB37	1.78	8.99×10^{-3}	2.52×10^{-2}	1.43	0.72
Titus	4.71×10^{-3}	1.58	0.78	2.61	0.93
Trib	2.49×10^{-2}	8.56×10^{-3}	1.17	2.10×10^{-2}	1.21

Table 3.5: EOP values for with Golden gate assembled plasmid constructs. *Values presented with < extended beyond the range of this assay and formed no plaques. Values shown are calculated mean average of three individual experiments.*

$X \geq 0.1$
$0.1 > X \geq 0.01$
$0.01 > X \geq 0.001$
$0.001 > X \geq 0.0001$
$0.001 > X \geq 0.0001$

3.13 Understanding the requirement of *brxS* and *brxT* by GGA mutagenesis

As *brxS* and *brxT* are an unexpected requirement for BREX activity, beyond the systematic mutagenesis performed on the *E. coli* BREX system (Gordeeva et al. 2019), the previously obtained mutant GGA constructs (section 3.10) were used to transform DH5α for EOP assays, using pGGA-C as a control (Table 3.6). Tables are shown with standard deviation values in supplementary material .Phages BGP, CP, Geo, NR1 and TB37 are all susceptible to BrxU, whereas Paula is BREX-sensitive. pBrxXL-*ΔbrxS* had no impact on phage resistance, but resistance was abolished for all phages by the deletion of *brxT* (Table 3.6). This implies that BrxS is dispensable for phage resistance, but BrxT is required for resistance by both BrxU and BREX. Alternatively, the *brxT* deletion may have somehow affected regulation of locus expression. Further experiments are required to explore the contribution of *brxT* and assess why its deletion results in deactivation of both BREX and BrxU in their entirety.

Φ	pBrxXL *ΔbrxS*	pBrxXL *ΔbrxT*
BGP	2.46×10^{-3}	1.79
CP	$<3.71 \times 10^{-8}$	1.93
Geo	$<2.93 \times 10^{-8}$	4.71
NR1	$<4.30 \times 10^{-8}$	2.31
Paula	$<4.51 \times 10^{-8}$	0.92
TB37	8.11×10^{-4}	1.04

X ≥ 0.1
0.1 > X ≥ 0.01
0.01 > X ≥ 0.001
0.001 > X ≥ 0.0001
0.001 > X ≥ 0.0001

Table 3.6: EOP values with pBrxXL-*ΔbrxS *and pBrxXL-*ΔbrxT *GGA constructs. *Values presented with < extended beyond the range of this assay and formed no plaques. Values shown are calculated mean average of three individual experiments.*

3.14 Discussion

Combining multiple phage defence systems produces a phage defence island that increases the likelihood a bacterium will be protected from phage infection (Bernheim and Sorek 2020). Phages that incorporate modified DNA bases in order to evade classical type IIP REases (Bryson et al. 2015) or BREX (Gordeeva et al. 2019) may be susceptible to other systems, such as type IV restriction enzymes. The initial interest in the pEFER BREX system was due to the ease of genetic manipulation as it represented a system from a close relative of *E. coli* that was plasmid-borne. However, upon investigating the effects of BREX via EOP assay, it was found that the once deprioritised *brxU* ORF situated between *brxC* and *pglX* was encoding an effective type IV restriction enzyme.

A BREX deletion mutant was generated in D23580 removing the entire BREX locus. This was performed to permit the enrichment and isolation of BREX sensitive phages, and has resulted in a collaborative publication (Rodwell et al. 2021). As mentioned earlier in section 3.5, a recent pre-publication article has detailed a new phage defence system, PARIS. PARIS involves the detection of phage encoded anti-restriction components and consists of a AAA+ ATPase and a DUF4435 protein. At the time of mutant generation, STM4493 and STM4494 were predicted to have accessory roles within BREX. However, they have since been identified as a standalone two-component phage defence system. As a result, the EOP values shown in Table 3.3 detail effect of a combined BREX and PARIS defence island. Further mutations are to be generated by removing STM4493 and STM4494 to assess the effects of BREX. A BREX knockout mutant could be obtained via a single round of λ-red recombination, removing the upstream BREX region of STM4494. Deletion of *pglX* has been shown in this chapter, as well as by Gordeeva et al, to inactivate BREX activity (Gordeeva et al. 2019). Therefore, it is unnecessary to remove the remaining *pglZ* and *brxL* as they do not confer phage resistance in

the absence of *pglX.* However, deletion of both regions would provide a "cleaner" BREX deletion mutant.

BREX was initially discovered by Goldfarb et al, who identified conserved genes clustered with *pglZ*. Here, two additional phage defence systems have been identified that are encoded within BREX systems. Within the pEFER BREX locus, BrxU and BREX work simultaneously to prevent against phage infection. From the 30 phages tested within this study, they target distinctly different phages, with no phage demonstrating susceptibility to both systems (Table 3.2 and 3.5). Our interest shifted from investigating pEFER as a means to study BREX to analysing the plasmid as a mobile, multi-system phage defence island. The acquisition of pEFER by a bacterial host transforms it into a multi-drug, multi-phage resistant strain. Bacterial communities that are modulated by the constant pressures of phage infection have the potential to expand beyond the phage-host equilibrium upon acquisition of pEFER. Beyond BREX and BrxU, pEFER also encodes an additional unidentified phage resistance system (Table 3.1). Bioinformatic analysis of the additional genes encoded within pEFER reveal a number of potential Candidates. For instance, *pEFER_p0052* and *pEFER_0053*, which likely encode for a VapC toxin and putative antitoxin, respectively. The VapBC family of TA systems constitutes the largest type II TA system families in bacteria and archaea, accounting for ~40% of all type II TA systems (McKenzie et al. 2012; Pandey and Gerdes 2005). The VapC toxin functions primarily as a ribonuclease and degrades tRNA to stall protein translation. Whilst VapC toxins have not yet been reported to abort phage infections, multiple TA systems are involved in phage defence (Song and Wood 2020). pEFER also contains multiple insertion sequence (IS) elements, indicating several occurrences of transposase uptake. It is possible that one or more of the IS elements integrated into pEFER might a role in phage defence but there is no proposed mechanism. Effectively, pEFER represents a phage defence island with multiple defence systems working synergistically. Phages that incorporate modified bases may be resistant to

restriction by type II endonucleases, however they in turn become susceptible to the type IV endonuclease, BrxU. Likewise, phages that are resistant to multiple classes of restriction endonucleases may be susceptible to BREX. This ensures a multi-strategy approach to preventing phage infection for the host.

Using transposon mutagenesis, the BREX locus has been inactivated in Km5. The Tn5 transposon contains a kanamycin resistance cassette, flanked by hairpin forming regions which act as transcriptional terminator regions. Along with identifying the potential presence of an additional phage defence system encoded outside of the BREX locus, it allowed for the identification of BREX/BrxU-sensitive phages (Table 3.2). The production of GGA assembly constructs allowed for the targeted study of components within the BREX locus of pEFER. It was shown that BrxU activity accounted for reduced EOP values for the majority of phages tested, however BREX was the active system against PATM, Paula, Sipho and Trib. Deletion of *brxU* did not have an effect on EOP values for BREX-sensitive phages, indicating that there is no interaction between BrxU and BREX components. Similarly, deletion of *pglX* did not hinder BrxU activity. Interestingly, the BREX-induced EOP was significantly lower for PATM and Paula than for Sipho and Trib. This difference could be explained via future comparison of the two viral genomes. It is possible that PATM and Paula contain a higher number of GCTAAT motifs and are therefore more readily recognised by BREX. Alternatively, Sipho and Trib might have counter-defence mechanisms that reduce the impact of BREX. There are a number of important BREX components that remain to be tested by generating deletion mutants via GGA. Gordeeva et al have shown the requirement of *brxB, brxC, pglX, pglZ* and *brxL* for an active BREX system (Gordeeva et al. 2019). It is predicted that *brxR* is a required component of pEFER BREX, however the construct pBrxXL-*ΔbrxR* has not yet been made. It is expected that the observations following deletion of individual pEFER BREX genes will align with Gordeeva et al, however this remains to be tested.

Deletion of *brxS* from pBrxXL had no significant effect on EOP values for the 6 phages tested and is likely uninvolved in phage-resistance. However, deletion of *brxT* resulted in the inactivation of both BREX and BrxU. The requirement for BrxT could be easily tested by separately expressing BrxU and testing for anti-phage activity. If BrxT is not required, at least for BrxU, the role of BrxT might be more focussed on regulation and expression of the BREX locus. In order to fully assess the role of *brxT,* additional experiments are required. Introduction of a point mutation that prevents translation of *brxT*, but allows transcription of the rest of the operon would show whether the abolished phenotype for pBrxXL-*ΔbrxT* (Table 3.6) is due directly to BrxT activity, or due to the interference in regulatory regions when generating pBrxXL-*ΔbrxT.*

With the pEFER defence island established as containing independent BREX and type IV restriction endonuclease phage defence systems, the next chapter will detail the biochemical investigation of BREX components and BrxU. Inducible constructs have been generated for the overexpression of individual proteins for biochemical analysis. With the majority of phages tested in this chapter being sensitive to BrxU, a large portion of Chapter 4 will focus on this system. A number of biochemical assays have been developed for assessing BrxU activity, such as FPLC gel filtration, DNA hydrolysis and inorganic phosphate production. As Chapter 3 has focused on the *in vivo* characteristics of BREX and BrxU, Chapter 4 details the *in vitro* experiments performed on these systems.

Chapter 4: Functional Characterisation of BREX and BrxU Phage Resistance Mechanisms

4.1 Introduction

Chapter 3 has shown that the BREX locus of pEFER encodes two phage defence systems that confer resistance to different subsets of phages. This was demonstrated primarily through deletion mutagenesis of *pglX* and *brxU* and quantifying the effect on phage plaquing. The *in vivo* impact of both BREX and BrxU was determined using an array of 30 phages. In this chapter, both systems are investigated biochemically, primarily through the overexpression and purification of each individual protein, to investigate the activity *in vitro*. Due to the identification of DUF262 and DUF1524 domains, it was predicted that BrxU would function similarly to GmrSD (He et al. 2015). The presence of the DUF262 domain indicated that nucleotide hydrolysis was required for activity, and the DUF1524 domain indicated that BrxU's main function was to cleave DNA. He et al. have shown that GmrSD targets phages with modified genomes, allowing multiple assays to be developed for the functional characterisation of BrxU. Using this information, assays have been developed to assess the enzymatic activities of key BREX proteins and BrxU. Individual BREX proteins have been overexpressed as soluble proteins and purified. FPLC gel filtration analysis has then been used to determine the multimeric states of each protein. The BrxU endonuclease has also been extensively characterised, with the identification of its target substrate, reaction conditions, multimeric state, cofactor requirements and catalytic properties. This allows a better model to be developed as to the complementary nature of BREX and BrxU within the pEFER phage defence island.

4.2 Phyre2 protein modelling

Functional predictions were made for each protein of the pEFER phage defence island using Phyre2 (Kelley et al. 2015) to perform protein modelling and identify potentially related protein structures (Table 4.1). Hits were obtained for all proteins except BrxT, which could not be modelled due to a lack of suitable templates (Table 4.1). As Phyre2 modelling relies solely on the input target protein sequence, the hits have to be carefully considered. Nevertheless, identification of conserved domains can be used to develop appropriate assays to test protein activity. Structural models are shown in Figure 4.1 overleaf.

Protein	Top Hit (PDB Code)	Sequence ID (%)	Coverage (%)	Confidence (%)
BrxS	MMTV integrase, catalytic core (5CZ1)	31	65	99.8
BrxT	N/A	N/A	N/A	N/A
BrxR	Proteasome accessory factor b/c (6SJ9)	14	84	100
BrxA	Protein of unknown function from *M. magneticum* (3BHW)	29	98	100
BrxB	Isomerase (3ODP)	17	49	76
BrxC	Replicative helicase (5UDB)	19	35	19
BrxU	SspE nuclease (6JIV)	12	79	100
PglX	MmeI Type IIG RM system (5HR4)	16	71	100
PglZ	Alkaline phosphatase (1EI6)	20	30	98.7
BrxL	Lon protease (6ON2)	18	82	100

Table 4.1: Phyre2 output parameters for pEFER phage defence island proteins.

65% of BrxS was modelled with 99.8% confidence using the catalytic core domain of the MMTV integrase (PDB: 5CZ1). Retroviral integrases catalyse the insertion of viral DNA into the host DNA (Ballandras-Colas et al. 2016). It is possible that BrxS is involved in the integration of the BREX locus and has a role in horizontal gene transfer. 84% of BrxR was modelled using PafBC (Müller et al. 2019), identifying a winged helix-turn-helix (wHTH) domain. This strongly suggested that BrxR might be a DNA-binding protein, and therefore a transcriptional regulator of the phage defence island. BrxA was modelled using a protein of unknown function from *Magnetospirillium magneticum*, and BrxB was modelled using an isomerase, though the confidence score was low (Table 4.1). Neither of these hits gives much indication of biochemical activity. BrxC was modelled against a replicative helicase Mcm2-7

with 35% coverage at the N-terminus (Table 4.1). Though the scores were low for the BrxC model, this did identify BrxC as a potential AAA family ATPase (Yuan et al. 2017). BrxU was modelled against SspE, a DNA nicking enzyme involved in the phosphorothioation-sensing phage defence system (Xiong et al. 2020). PglX was modelled against MmeI, a type IIG restriction enzyme expected to have methyltransferase activity (Callahan et al. 2016). This matches the previously identified role of PglX (Gordeeva et al. 2019; Goldfarb et al. 2015), though the hit is based on only 71% of PglX, leaving the C-terminal domain unmodelled. Both PglZ and BrxL also produced hits matching expected activities, alkaline phosphatase, and lon protease, respectively (Table 4.1).

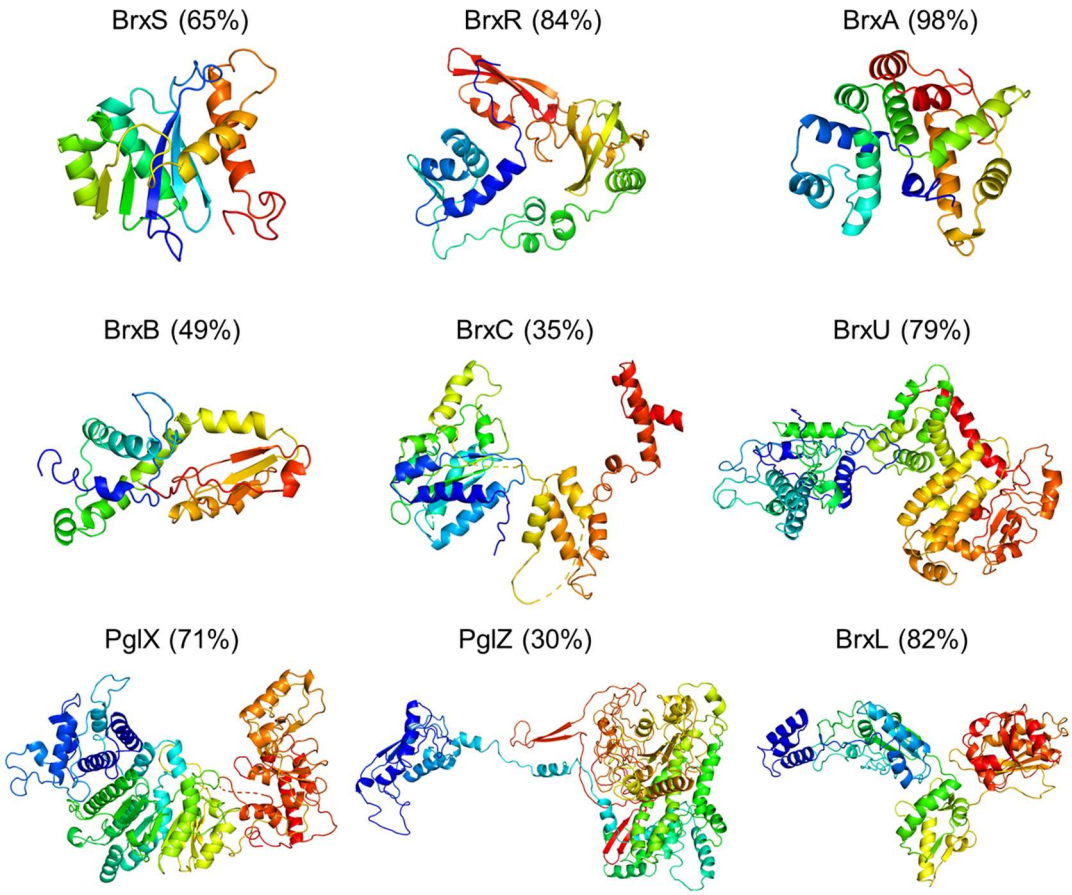


Figure 4.1: Phyre2 structural models for pEFER phage defence island proteins. *Models are coloured in rainbow format with the N-terminus in **BLUE** and C-terminus in **RED**. Coverage (%) is shown beside labels.*

4.3 Cloning of individual BREX genes into an inducible expression system

In order to study the properties of individual BREX proteins, each gene was cloned into pSAT1-LIC, an IPTG-inducible plasmid under the control of the *lac* repressor. Ligation independent cloning (LIC) (Aslanidis and de Jong 1990) was used to create constructs for each BREX gene. pSAT1-LIC is a 4.7 kb plasmid that can be linearised by the restriction enzyme StuI, forming blunt ends at the cloning site (Figure 4.2).

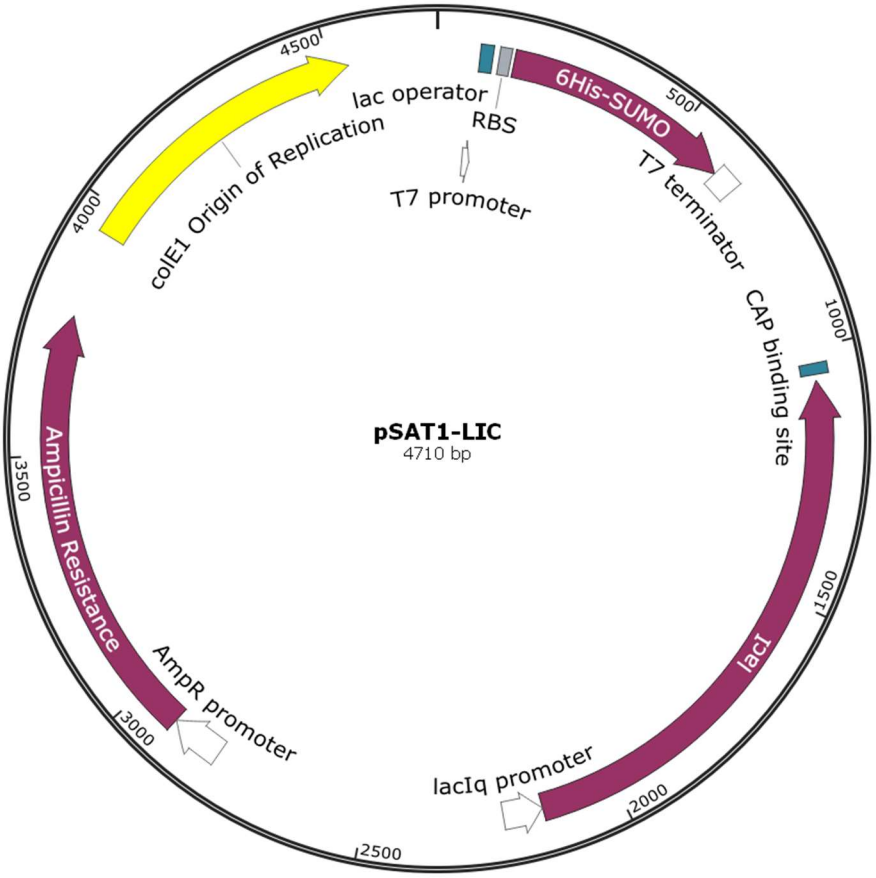


Figure 4.2: Scaled plasmid map of pSAT-LIC. *T7 promoter is directly upstream of the ribosomal cloning site (RBS). Presence of IPTG causes LacI to dissociate from the T7 promoter, allowing T7 RNA polymerase to transcribe a 6His-hSUMO2-Protein fusion.*

Following blunt-end cleavage, the 3′ to 5′ exonuclease activity of T4 DNA polymerase was used to create 5′ overhangs. Complementary LIC regions were engineered into PCR primers to allow the generated overhangs between vector and insert to anneal. A full protocol for LIC is detailed in Section 2.47. The DNA sequence immediately upstream of the LIC cloning site encodes a hexahistidine (6His) tag fused to the hSUMO2 gene. Proteins cloned into the LIC site are expressed with an N-terminal 6His-hSUMO2 tag that can be removed via cleavage with Sentrin protease (SenP) to yield a mature, scarless protein (Figure 4.3).

Figure 4.3: Target protein construct expressed from pSAT1-LIC. *Protein of interest is immediately preceded by the SenP cut site, allowing scarless cleavage of the mature protein from the 6His-hSUMO2 tag.*

The removal of the hexahistine tag ensures there is no downstream interference and allows for a two-step immobilised metal affinity chromatography (IMAC) purification (Section 2.47). PCR products and linearised pSAT1-LIC were treated with T4 DNA polymerase and annealed samples were used to transform DH5α (Figure 4.4). Constructs were confirmed via sequencing. Constructs for all BREX genes were produced including *brxS* and *brxT* (data not shown for *brxS* and *brxT*; Table 2.7).

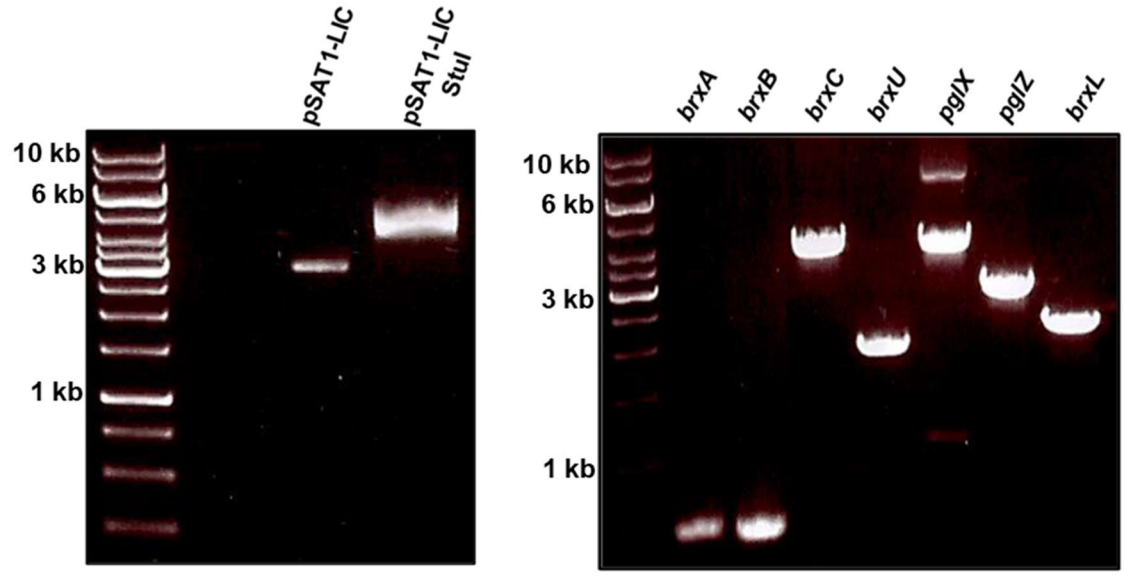


Figure 4.4: Production of phage defence island gene expression vectors*. Agarose gel electrophoresis in 1% agarose with 1x TAE. Samples were resolved for 1 hour at 120 V. A) Linearised pSAT1-LIC migrates slower following digestion with StuI than super-coiled pSAT1-LIC. B) PCR products for individual genes using pEFER as a template and primers detailed in Materials and Methods (Section 2.43 and Table 2.6).*

4.4 Protein overexpression and purification

All proteins expressed from pSAT1-LIC were purified using the same workflow. Following over-expression and bacterial cell lysis, clarified cell extracts were subjected to a single (IMAC) step. The eluted fusion proteins were then cleaved to remove the 6His-hSUMO2 tag, then subjected to another round of IMAC. The flow-through samples containing the cleaved target protein were then further purified by anion exchange chromatography, and finally by size exclusion chromatography. Using this throughput, soluble protein was obtained for BrxR, BrxA, BrxB, BrxC, BrxU, PglX, PglZ and BrxL (Figure 4.5). BrxS and BrxT have not yet been expressed and purified.

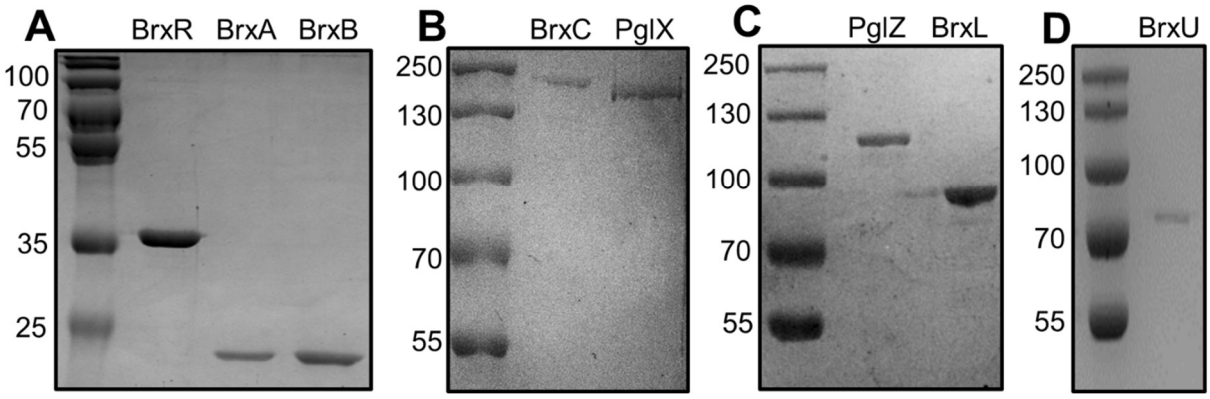


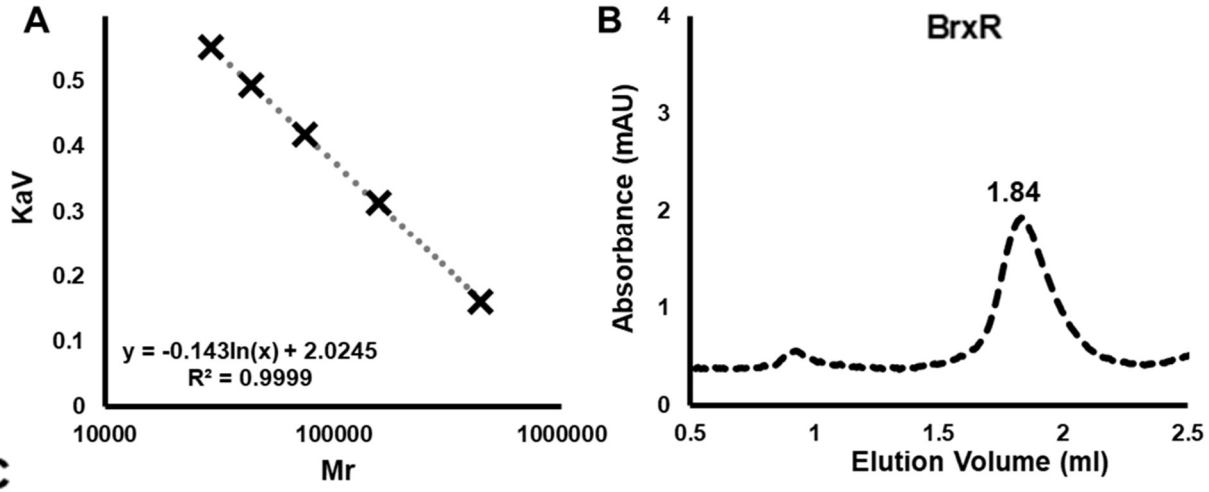
Figure 4.5: SDS-PAGE of purified BrxU and BREX proteins. *Following purification, final samples were separated by SDS-PAGE and ran for 100 min at 180 V. Molecular weights shown in kDa. 10 µl of samples were loaded at a concentration of 0.1 mg/ml. A) BrxR, BrxA and BrxB were resolved in 15% acrylamide gels. B and C) BrxC, PglX, PglZ and BrxL were resolved in 8% acrylamide gels. D) BrxU was resolved in 10% acrylamide gels.*

4.5 Analytical size exclusion chromatography of BREX proteins

To determine the multimeric state of each phage defence island protein, size exclusion chromatography was performed on purified samples using a Superdex 200 increase GL 5/150 (S200i; Cytiva). The S200i is a small gel filtration column with a bed volume of 3 ml for rapid analytical size exclusion chromatography. The elution volume (eV) of each protein was compared to eV of known protein standards and a calculated complex mass was obtained for each protein (Figure 4.6). Blue dextran was used a control standard to calculate the void elution volume of 0.99 ml. The calculated protein mass is calculated using the gradient and Y intercept in Figure 4.6. KaV is calculated with the following formula: $\frac{\text{(Sample Elution Volume – Void Volume)}}{\text{(Total Column Volume – Void Volume)}}$

KaV can then be used to calculate the molecular mass: $\ln\left(\frac{\text{(KaV – 2.0245)}}{\text{(-0.143)}}\right) = \text{Calculated Mr}$

BrxR was observed to elute from the S200i with a peak eV of 1.84 ml, giving it a KaV value of 0.43. This KaV value was used to obtain a calculated Mr of 67404 daltons, which is 1.97x its theoretical Mr. This data shows that BrxR forms a dimer in solution. This supports the Phyre2 model wherein BrxR was predicted to contain a wHTH domain, common in dimeric transcriptional regulators (Sarkar-Banerjee et al. 2018). The remaining BREX proteins were analysed in the same manner (Figure 4.6C).



C

Protein	Mr (Theoretical)	KaV	log Mass	Mr (Calculated)	Ratio vs Mr	State
BrxR	35294	0.430	11.15	69404	1.97	Dimer
BrxA	22741	0.496	10.69	43800	1.93	Dimer
BrxB	22770	0.587	10.05	23157	1.02	Monomer
BrxC	139553	0.096	13.48	718268	5.15	Pentamer
PglX	138657	0.273	12.25	208008	1.50	Monomer
PglZ	100707	0.289	12.14	187046	1.86	Monomer/ Dimer
BrxL	77488	0.334	11.82	136004	1.76	Monomer/ Dimer

Figure 4.6: Analytical gel filtration indicates protein multimeric states. *A) Calibration KaV values against known molecular masses. The y = mx + c equation is used for all calculations of molecular mass. B) Chromatogram of BrxR elution on the S200i. X-axis value shown for main peak corresponding to BrxR. C) Calculation of protein mass from elution volume. Deductions for multimeric state are made by comparing the calculated mass from the eV to the theoretical mass calculated from the protein sequence.*

BrxA was observed to form a dimer with a calculated mass at 1.93x its Mr, and BrxB was found to remain as a monomer (Figures 4.6C and 4.7). BrxC was observed to have the lowest eV of all proteins tested and was calculated to form a pentamer (Figure 4.7). This fits with the Phyre2 identification of BrxC as a AAA family ATPase (Table 4.1), as these have been reported to sometimes form higher order structures (Hilbert et al. 2015). Surprisingly, the calculated mass of PglX was exactly 1.5x greater than its theoretical Mr (Figure 4.6C). It is most likely that PglX is a monomer in solution and that its non-globular shape causes it to migrate slower than expected through the S200i. This would be consistent with the overall shape of the type IIL restriction endonuclease MmeI (Callahan et al. 2016), which was identified as the top hit for PglX by Phyre2 (Table 4.1). MmeI cleaves DNA as a dimer, however it requires at least two recognition sequences to function (Callahan et al. 2016). It is plausible that PglX exists as a monomer catalytically, or that it forms dimers when in complex with target DNA similarly to MmeI (Callahan et al. 2016). Both PglZ and BrxL were found to form dimers (Figure 4.6C). As PglZ was predicted to have an alkaline phosphatase domain (Table 4.1), this is consistent with other studied alkaline phosphatases (Orhanović and Pavela-Vrancic 2003). The major peak for BrxL is at 1.58 ml (Figure 4.7) and so BrxL was calculated to be a dimer in solution (Figure 4.6C). An additional smaller peak was observed at 1.20 ml which would mean BrxL had assembled into an octameric structure (Figure 4.7). BrxL was modelled against Lon protease (Table 4.1), which forms hexameric rings (Vieux et al. 2013). It is therefore possible that BrxL forms large multimeric complexes such as octamers, albeit it is also possible that BrxL has migrated more slowly due to an asymmetric overall shape, and that the peak observed at 1.72 ml corresponds to monomeric BrxL.

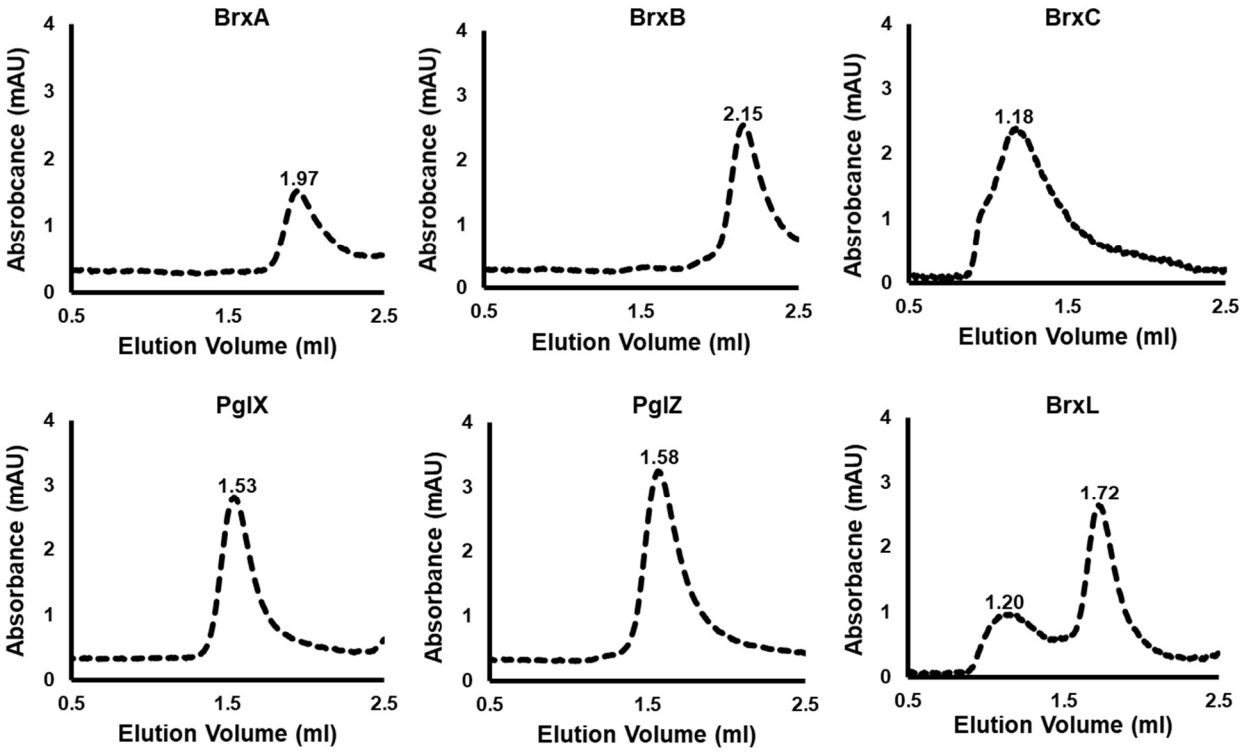


Figure 4.7: Gel filtration chromatograms of individual protein elutions via S200i*. A 10 µl sample at 500 nM was resolved at 0.175 ml/min. Trace shown is representative of three experimental repeats. X-axis value shown for major peaks.*

Having calculated the individual multimeric states of each BREX, it is important to consider that this might not correspond to their final assembly as part of an active BREX complex. Further investigation is required to see how the BREX components interact, what sub-complexes might form and whether the multimeric state of an individual BREX protein alters.

4.6 BrxR is a transcriptional regulator and an autorepressor

BrxR was predicted to have a role in transcriptional control of the pEFER BREX locus due to its position directly upstream of BrxA, the modelling of a wHTH domain by Phyre2 (Table 4.1) and its dimeric state in solution (Figures 4.6C and 4.7). Goldfarb et al. predicted that that the BREX locus in *B. cereus* contained two major promoter regions. They suggested that *ABCX* were transcribed as one operon and that *ZL* formed a separate operon (Goldfarb et al. 2015). To explore this model further, whole transcriptomic data for *Salmonella* was analysed as part of SalComMac (*Salmonella* Compendium Mac) (Srikumar et al. 2015). Using data from *S. enterica* ST4-74, primary transcriptional start sites were identified directly upstream of STM4498 (*brxA)* and STM4491 (*pglZ*), consistent with the RACE sequencing in Goldfarb et al. 2015. ST4-74 is very closely related to strain ST313 that was investigated within Chapter 3, and the BREX phage defence islands have identical sequences. Based on these observations, it was predicted that pEFER BREX would operate in a similar manner, though due to the larger size of the pEFER phage defence island, there might potentially be an additional promoter region upstream of *brxR* (Figure 4.8). It is important to point out that the following BrxR analysis was performed prior to identification of the need for *brxS* and *brxT*. These two genes and any preceding promoters are therefore not in the analysis, and those experiments would be needed in future. Working with the region *brxR* to *brxL*, six regions within the pEFER phage defence island were selected for further analysis (Figure 4.8).

The six identified regions were cloned into pRW50 and were confirmed via sequencing. pRW50 contains a tetracycline resistance cassette and a β-galactosidase reporter gene, *lacZ*. *lacZ* is downstream of the multiple cloning site (MCS), and in its native form pRW50 does not encode a promoter for *lacZ*. Introduction of an active promoter sequence into the MCS of pRW50 results in quantifiable transcription of *lacZ*, using variants of the Miller assay (Schaefer et al. 2016).

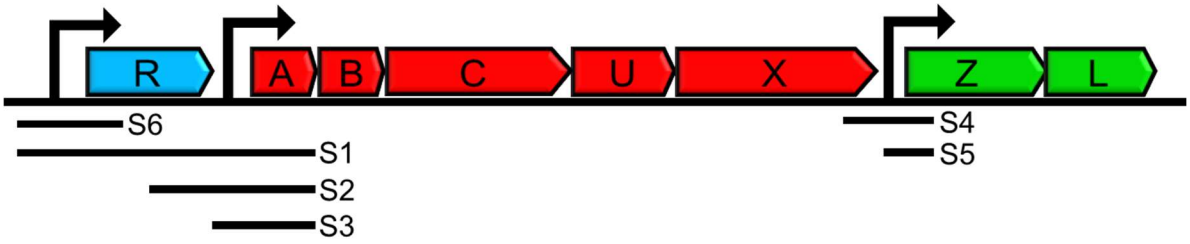


Figure 4.8: Identification of promoter regions within the BREX locus. *Genes are group into predicted operons and coordinated by colour.*

As BrxR was predicted to function as a transcriptional regulator of the pEFER phage defence island operon, a pBAD30-*brxR* expression construct was prepared. DH5α was transformed with either pBAD30 as a control or with pBAD30-*brxR*. These strains were transformed further with one of the 6 pRW50 constructs, or an empty pRW50 plasmid as a promoterless negative control. In the absence of *brxR*, regions S1, S2, S3 and S6 displayed promoter activity beyond that of the negative control strain (Figure 4.9). Region S6 had the strongest promoter, but activity was relatively lower in S1 then S2 and then S3. In the absence of BrxR, regions S4 and S5 showed no additional promoter activity (Figure 4.9). When BrxR was expressed, however, the promoter activity from S6, S1 and S2 was reduced (Figure 4.9). In contrast, the activity from S3 increased, and region S4 now also showed activity, though S5 remained inactive (Figure 4.9). This indicates that BrxR can act both as a transcriptional repressor, and as a

transcriptional activator, differentially controlling gene expression from the phage defence island.

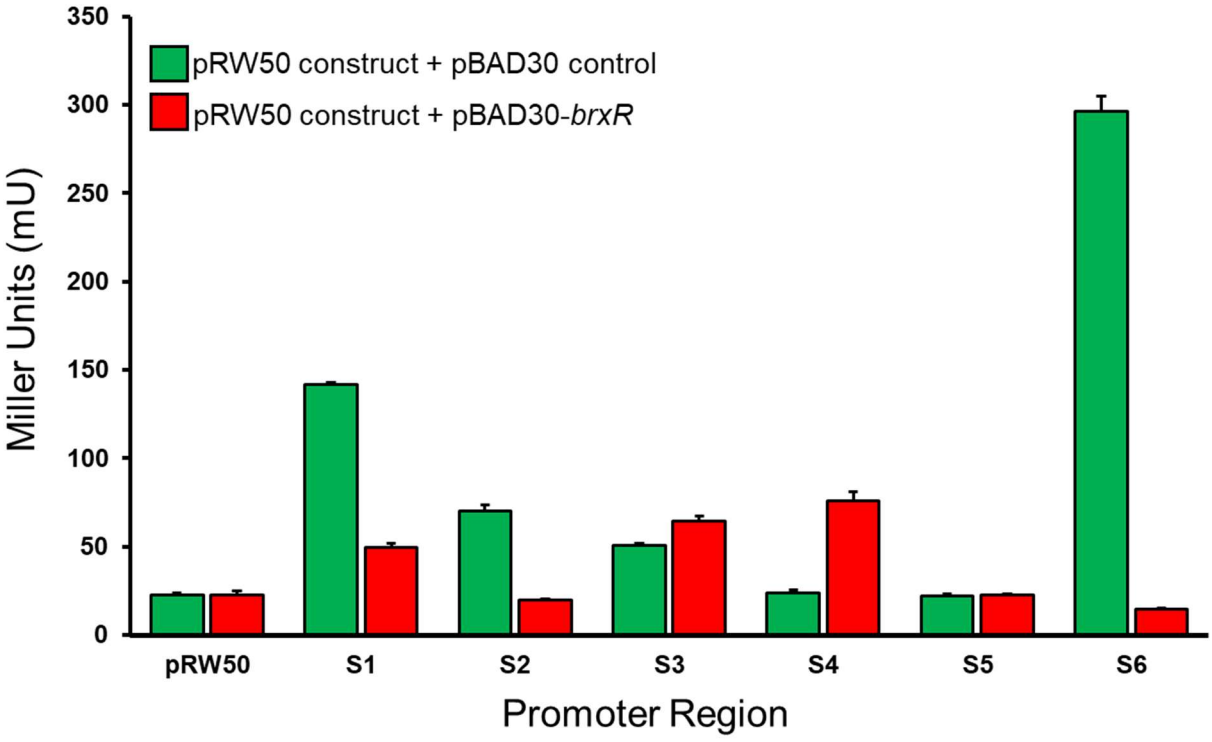


Figure 4.9: Phage defence island promoter activity. *LacZ activity was quantified measuring absorbance at 420 nm and Miller units were calculated. Presented data are the mean and standard deviations from triplicate data.*

4.7 Inorganic phosphate detection assay development for BrxC, PglZ and BrxL

BrxC was predicted to be an AAA family ATPase by Phyre2 (Table 4.1) and encodes a Walker box motif. PglZ was found to belong to the alkaline phosphatase-like family which non-specifically remove terminal phosphates, and BrxL was modelled to be structurally similar to Lon protease that also belongs to the AAA superfamily of ATPases. Each of these proteins had a predicted role in phosphate removal, either as an ATPase or phosphatase. A single assay was developed to test if each protein was able to hydrolyse ATP to produce inorganic phosphate (Pi). The commercial reagent BIOMOL green was used, as it undergoes a colour change when Pi is present. The amount of Pi production was calculated against a standard of known phosphate concentrations. Titrations of each protein were tested for production of Pi (Figure 4.10).

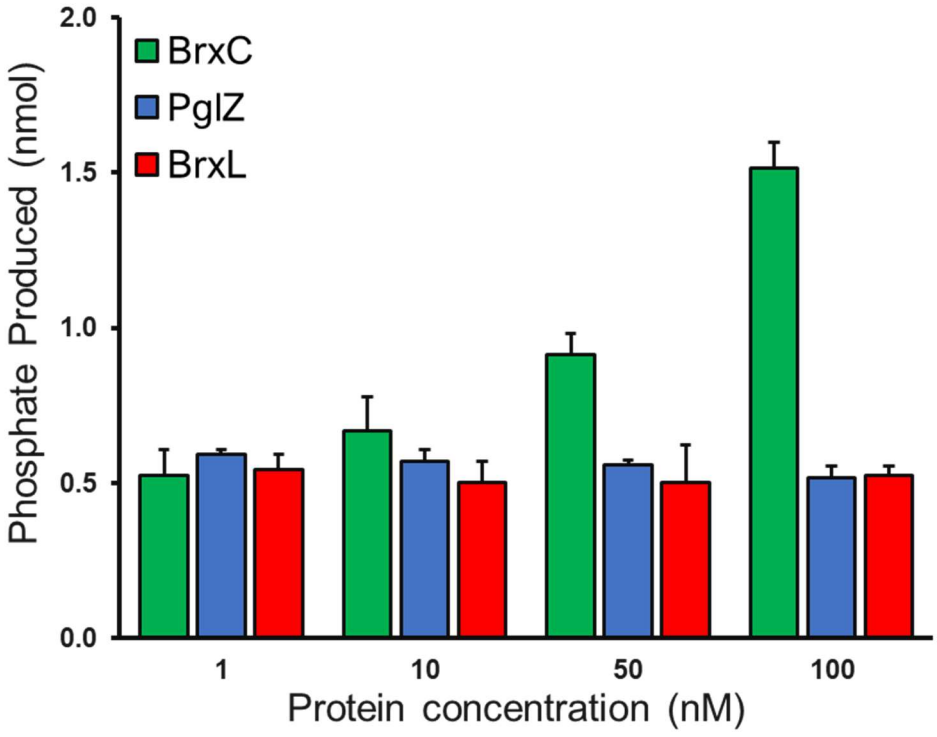


Figure 4.10: BIOMOL green inorganic phosphate detection assays*. BrxC was observed to produce inorganic phosphate. Presented data are the mean and standard deviations from triplicate data.*

BrxC was observed to have ATPase activity that increased in a concentration dependent manner. 1.51 nmol of Pi was detected at 100 nM, decreasing to 0.91 nmol at 50 nM BrxC (Figure 4.10). At the lower BrxC concentrations (10 nM and 1 nM), Pi production could not be distinguished against negative controls. No increase in Pi could be detected for both PglZ and BrxL, indicating that these proteins did not have activity in the conditions tested. It is possible that both of these proteins exhibit ATPase activities in the presence of a required cofactor, such as a divalent metal cation, a partner protein or targeted substrate, which would have been missing from this assay. The commercial alkaline phosphatase FastAP (Thermo) at 0.1 U/µl was used as a control, producing 2.49 nmol of Pi (data not shown).

4.8 Identification of BrxU as a putative endonuclease

BrxU has been identified as a phage-defence system operating independently of the pEFER BREX system. Bioinformatic analysis of BrxU using BLASTp (Altschul et al. 1990) identified DUF262 and DUF1524 domains, which are also encoded by the single polypeptide type IV restriction enzyme GmrSD (Machnicka et al. 2015) (He et al. 2015). Direct sequence alignment of BrxU and GmrSD nevertheless shows a low shared sequence identity of 24.8%, using EMBOSS Stretcher (Madeira et al. 2019). Members of the DUF262 family exhibit variable architectures, often resembling the ParB DNA binding domain and contain the DGQQR nucleotide-binding motif (Chen et al. 2015). DUF1524 proteins contain a conserved DHxxP motif which is found in the histidine-metal finger nucleases (Jablonska et al. 2017). As a result, the GmrSD family has been extensively characterised bioinformatically, identifying multiple conserved motifs (Machnicka et al. 2015). Using PROMALS3D (Pei, Kim, and Grishin 2008) to perform a sequence alignment of BrxU with GmrSD, these motifs appear to be present in BrxU (Figure 4.11).

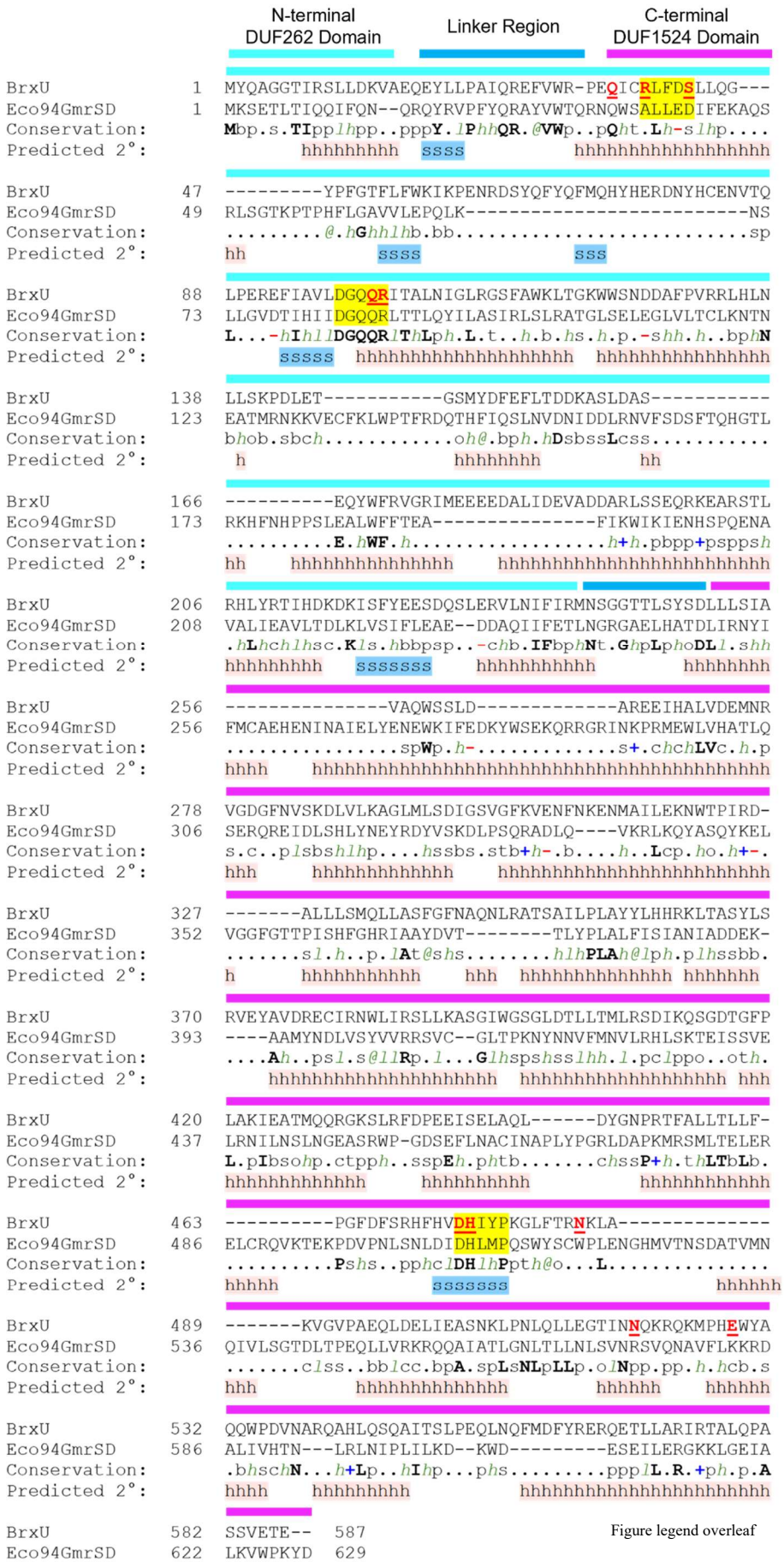


Figure legend overleaf

Figure 4.11: Sequence alignment of BrxU and GmrSD. *BrxU and Eco94GmrSD were aligned using PROMALS3D. Predicted domain architectures are indicated. RLFDS, DGQQR and DHIYP motifs are highlighted in yellow. Residues selected for mutation (Fig. 5) are shown in red and underlined. Conservation of residues is denoted using the following key: conserved amino acid residues are shown in bold and uppercase letters; conserved aliphatic residues (I, V, L), shown as l; conserved aromatic residues (Y, H, W, F), shown as @; conserved hydrophobic residues (W, F, Y, M, L, I, V, A, C, T, H), shown as h; conserved alcohol residues (S, T), shown as o; conserved polar residues (D, E, H, K, N, Q, R, S, T), shown as p; conserved "tiny" residues (A, G, C, S), shown as t; conserved small residues (A, G, C, S, V, N, D, T, P), shown as s; conserved bulky residues (E, F, I, K, L, M, Q, R, W, Y), shown as b; conserved positively charged residues (K, R, H), shown as +; conserved negatively charged residues (D, E), shown as -; conserved charged residues (D, E, K, R, H), shown as c. Secondary structure (2°) prediction denoted using the following key; α-helices, h, β-strands, s.*

The two key motifs highlighted by Machnika et al. are highlighted in Figure 4.11. Both GmrSD and BrxU contain a DGQQR motif in the N-terminal domain, highlighted in Figure 4.11 in cyan, reflecting the signature sequence of NTP binding domains of sulfiredoxin (Srx) (Machnicka et al. 2015). Additionally, both proteins contain a DHxxP domain, highlighted in green in Figure 4.11. This motif is predicted to form the active site of the endonuclease. BrxU has a DHIYP motif, whereas GmrSD has DHLMP. The conservative substitution of isoleucine for leucine is unlikely to have a significant effect on this motif as they both contain non-polar, aliphatic side chains however this is yet to be tested (Creixell et al. 2012). Substitution of tyrosine for methionine is a larger substitution, but as both BrxU (this study) and GmrSD (He et al. 2015) have been shown to degrade phage gDNA and confer resistance *in vivo*, the difference at this residue does not affect enzyme activity.

4.9 BrxU provides resistance against phages *in vivo*

To confirm that BrxU can act independently to reduce phage plaquing, *brxU* was cloned into pBAD30. EOP assays were performed against the suite of 30 coliphages and compared to values obtained for pBrxXL-*ΔpglX* (Table 4.2). The overall trends of EOP values obtained with pBAD-*brxU* and pBrxXL-*ΔpglX* were highly consistent. This further reinforces that BrxU is a standalone phage defence system embedded within the pEFER phage defence island, functioning entirely independent of BREX. The majority of BrxU-sensitive phages were completely prevented from forming plaques, such as phage CP (Table 4.2). In contrast, BGP and TB37 were the only BrxU sensitive phages to show modest impact.

Φ	pBrxXL-ΔX	pBAD30-*brxU*
AL25	0.22	1.71
Alma	1.33	1.41
Bam	1.02	0.91
Baz	2.01	1.17
BB1	1.12	1.22
BGP	5.32×10^{-2}	6.17×10^{-2}
BHP	2.01	1.63
CP	$<4.92 \times 10^{-9}$	$<5.11 \times 10^{-9}$
CS16	1.73	1.03
EH2	$<1.67 \times 10^{-9}$	$<4.02 \times 10^{-9}$
EL	$<4.26 \times 10^{-7}$	$<3.43 \times 10^{-7}$
Geo	$<3.03 \times 10^{-7}$	$<2.81 \times 10^{-7}$
Jura	0.63	1.21
Mak	0.98	0.65
Mav	0.63	1.49
NP	$<1.07 \times 10^{-7}$	$<2.98 \times 10^{-7}$
NR1	1.02×10^{-7}	$<7.82 \times 10^{-8}$
PATM	0.7	1.42
Paula	1.53	1.61
QOTSP	$<8.33 \times 10^{-8}$	$<9.62 \times 10^{-8}$
SAP	$<4.52 \times 10^{-7}$	$<2.15 \times 10^{-7}$
Sipho	0.58	1.11
Solly	$<3.07 \times 10^{-7}$	$<7.89 \times 10^{-8}$
Some	$<2.86 \times 10^{-8}$	$<8.24 \times 10^{-8}$
SPSP	0.46	1.32
TB34	1.51	0.67
TB36	$<2.07 \times 10^{-8}$	$<5.28 \times 10^{-8}$
TB37	2.52×10^{-2}	2.44×10^{-2}
Titus	0.78	1.41
Trib	1.17	1.85

EOP ≥ 0.1
0.1 > EOP ≥ 0.01
0.01 > EOP ≥ 0.001
0.001 > EOP ≥ 0.0001
EOP < 0.0001

Table 4.2: EOP values for pBrxXL-ΔX in comparison with pBAD30-*brxU. *EOP values align indicated BrxU is the only active phage defence component of pBrxXL-ΔX. Values with < extended below the range of this assay and formed no plaques. Data shown are mean values from triplicate experiments.*

4.10 BrxU hydrolyses sensitive phage genomes into non-specific fragment lengths

Due to its activity against a range of phages (Table 4.2), and its predicted function as a nuclease due to the presence of DUF262 and DUF1524 domains (Figure 4.11), BrxU was tested for its ability to cleave DNA substrates. Phage gDNA was purified via phenol-chloroform extraction and used in DNA digest assays. Initial DNA hydrolysis trials were performed using gDNA from Geo (Figure 4.12). gDNA was not hydrolysed in the absence of ATP, however addition of ATP to a final concentration of 1 mM resulted in the complete degradation of Geo gDNA (Figure 4.12A). gDNA hydrolysis was observed to be dependent on the presence of suitable divalent cation such as Mg^{2+} (Figure 4.12B). Addition of EDTA prevented BrxU from binding Mg^{2+}, resulting in no gDNA hydrolysis (Figure 4.12B).

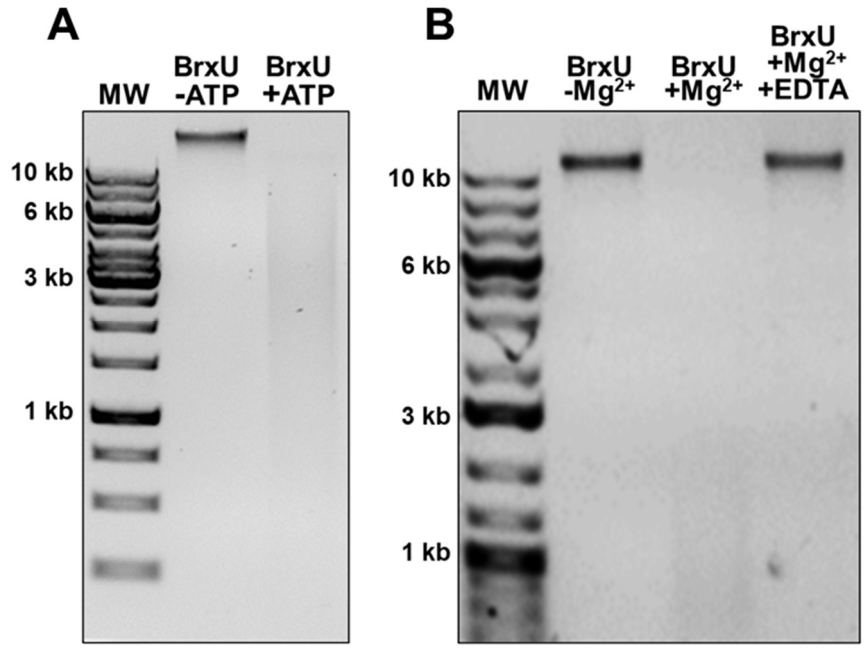


Figure 4.12: Initial trial BrxU DNA hydrolysis assays. *A) Geo gDNA is hydrolysed upon the addition of 1mM ATP. Each sample contains 500 nM BrxU and 10 mM $MgSO_4$. Samples resolved in 1% agarose TAE at 120 V for 45 min. B) gDNA hydrolysis is dependent on the addition of 10 mM $MgSO_4$. Addition of 10 mM EDTA prevents BrxU from binding Mg^{2+}. Samples resolved in 0.8% agarose TAE at 120 V for 45 min.*

Next, gDNAs from all 30 coliphages, and with the addition of phage λ, were tested for BrxU sensitivity *in vitro*, in the absence and presence of ATP (Figure 4.13). The gDNAs from phages that showed reduced EOP with pBAD-*brxU* (Table 4.2), were degraded by BrxU (Figure 4.13). Whilst phages BGP and TB37, which both had modest reductions in EOP, were not fully degraded (Figure 4.13), phage NR1, with a very low EOP (Table 4.2), was also not fully degraded. Despite this discrepancy with NR1, overall, the trends were consistent where BrxU degraded the DNA at a level matching the impact on EOP. λ gDNA was not hydrolysed, and would therefore be predicted to be resistant to pBAD-*brxU* in EOP assays. Whilst it could be predicted that some phages may encode proteins to inhibit type IV restriction enzymes, such as T4 IPI (Rifat et al. 2008a), none of the phages showed gDNA degradation despite a high EOP value, indicating our suite of coliphages does not encode inhibitors of BrxU.

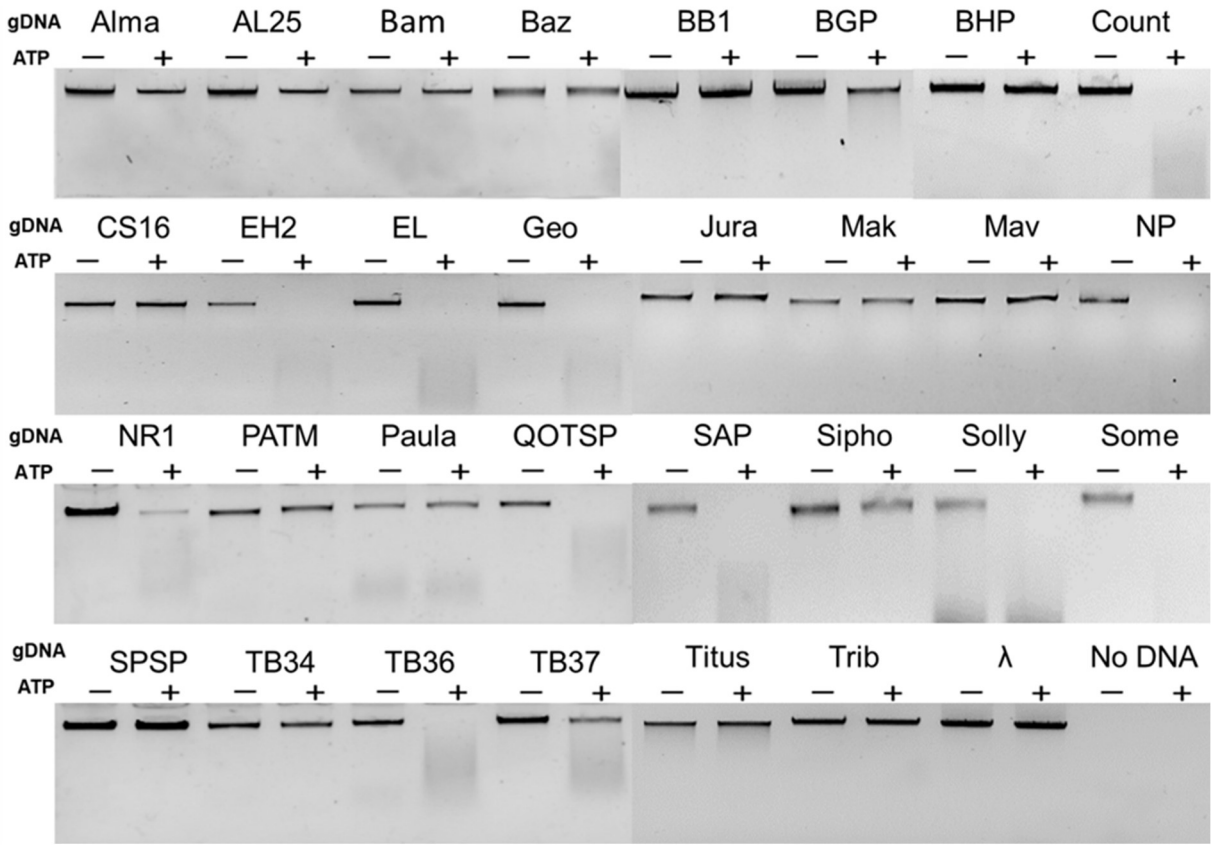


Figure 4.13: Phage gDNA hydrolysis by BrxU. *Phages with reduced EOP values in Table 4.2 correspond with cleaved gDNA. All samples contain 500 nM BrxU. Sample ATP content shown above lanes and denote by +/-. Samples resolved in 1% agarose TAE at 120 V for 45 min.*

4.11 BrxU requires a divalent metal cofactor for hydrolysis

Early DNA hydrolysis assays were attempted using a standard protein storage buffer with no additional metals added, and no DNA hydrolysis was observed (data not shown). Addition of $MgSO_4$ to a final concentration of 10 mM stimulated BrxU activity, and this could be blocked by the addition of 10 mM EDTA (Figure 4.12B). In order to assess the metal cofactor requirements of BrxU, a range of metals were titrated in DNA hydrolysis assays, using concentrations from 10^{-8} M to 10^{-2} M and Geo gDNA as substrate (Figure 4.14). The amount of remaining Geo gDNA was enumerated by measuring band intensity using ImageJ (Figure 4.15). Mg^{2+} had been demonstrated to induce hydrolysis at 10^{-2} M in previous assays (Figure 4.9). Mg^{2+} was observed to induce activity as low as 10^{-4} M resulting in only 4% of substrate remaining (Figures 4.14 and 4.15).

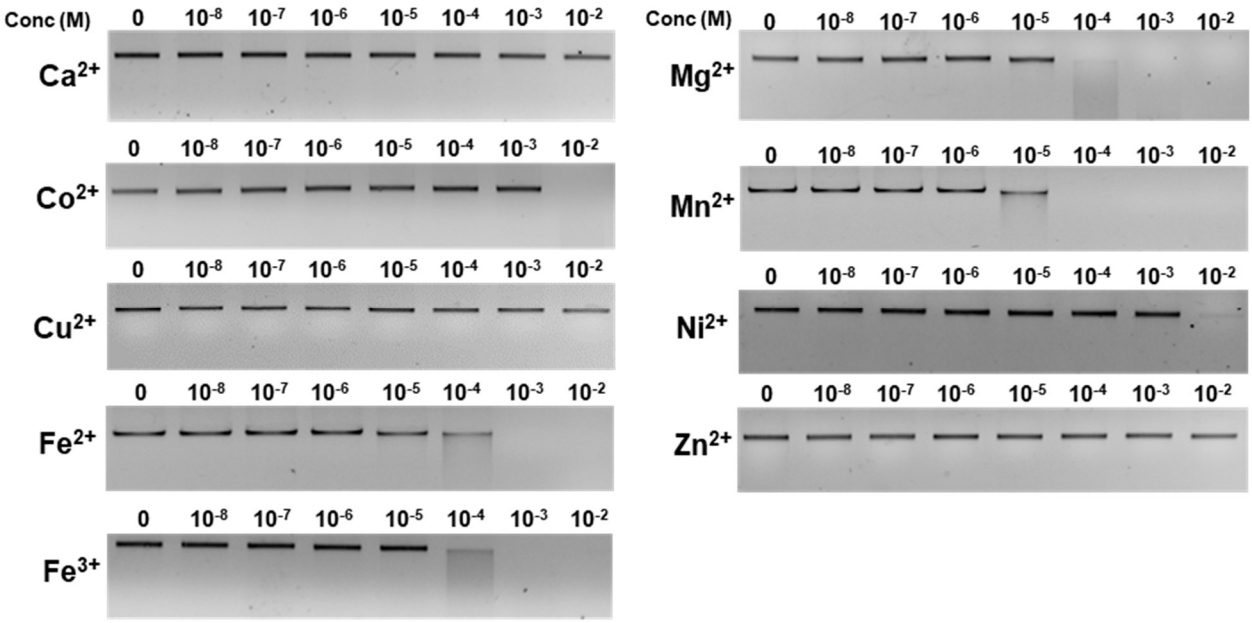


Figure 4.14: Metal-dependent DNA hydrolysis by BrxU. *BrxU can utilise Mg^{2+}, Mn^{2+}, Ni^{2+}, Fe^{2+} and Fe^{3+}. All samples contain 500 nM BrxU and 1 mM ATP. Data shown is representative of triplicate experiments. Samples were resolved in 1% agarose TAE at 120 V for 45 min.*

No activity was observed at 10^{-5} M indicating a minimum threshold between these two values is required for BrxU activity with Mg^{2+}. Mn^{2+} was observed to induce activity as low as 10^{-5} M resulting in 30% of substrate being hydrolysed. Fe^{3+} was found to induce DNA hydrolysis at 10^{-4} M resulting in only 31% of substrate remaining. Fe^{2+} also induced DNA hydrolysis at 10^{-4} M, however significantly greater substrate (65%) was left remaining. (Figure 4.15). Fe^{2+} was observed to induce weak activity as low as 10^{-5} M, with 86% of substrate remaining. BrxU can utilise both oxidation states of iron, however activity is higher with Fe^{3+} than Fe^{2+}. Ni^{2+} and Co^{2+} induced hydrolysis at 10^{-2} M however no activity was observed with lower concentrations. Cu^{2+}, Zn^{2+} and Ca^{2+} did not induce any BrxU activity even at the highest concentration tested (Figure 4.14).

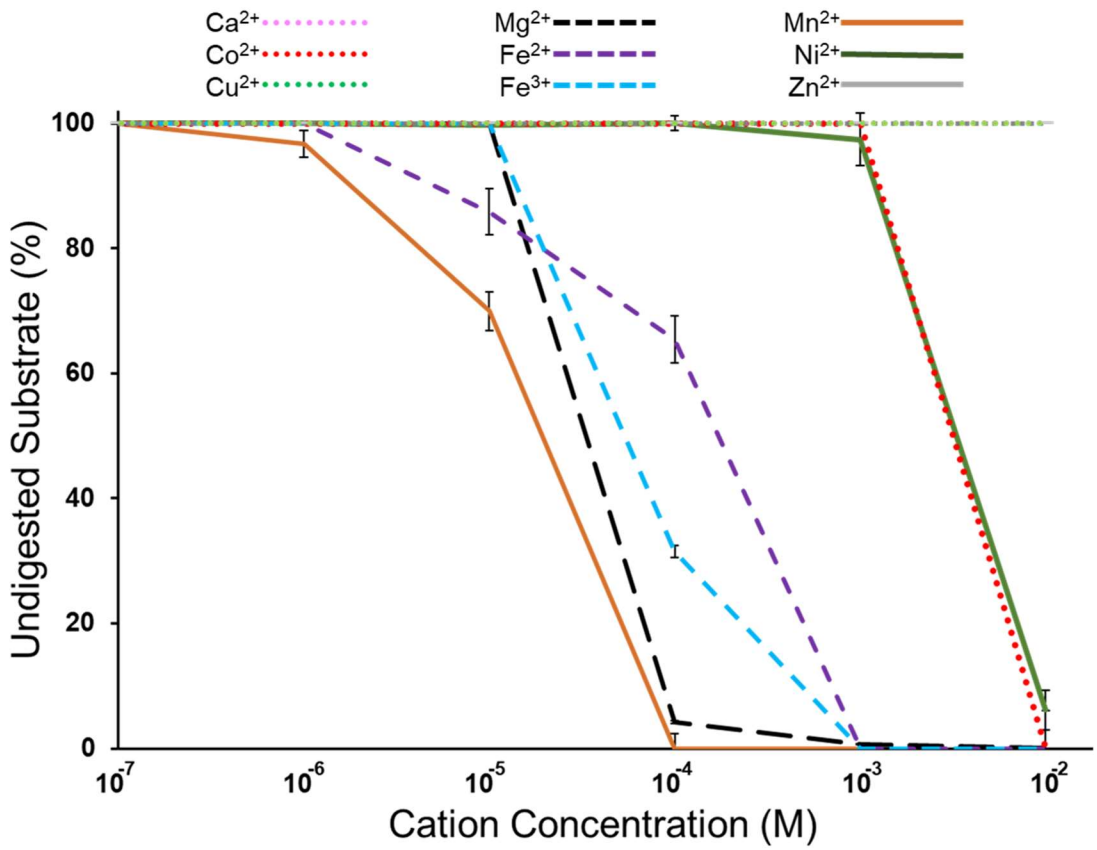


Figure 4.15: Quantification of BrxU metal specificity. *BrxU can utilise manganese at lower concentrations than any other divalent cations tested. Data shown are mean values from triplicate experiments. Error bars represent standard deviation.*

4.12 BrxU is most active with ATP as the nucleotide cofactor

BrxU was found to be active when 1 mM ATP was present (Figure 4.12A). To assess the optimal conditions for BrxU, hydrolysis assays were repeated with a range of nucleotide cofactors, at a range of concentrations from 0.1 mM to 10 mM (Figure 4.16). Geo gDNA was digested into smeared fragments no larger than 1 kb at 1 mM and 10 mM ATP. Smeared fragments extended as far as 2 kb for 0.1 mM ATP, however no intact gDNA remained at all concentrations.

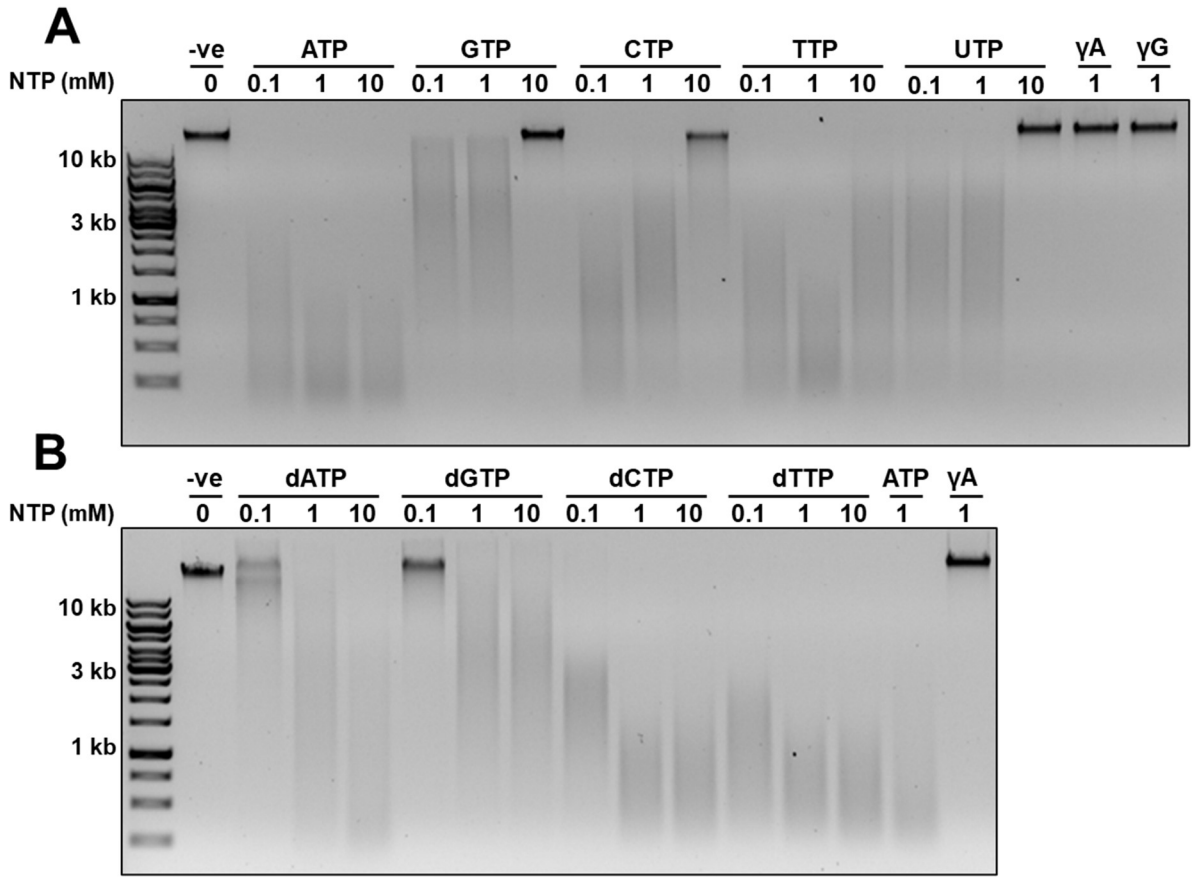


Figure 4.16: Nucleotide specificity DNA hydrolysis assays show BrxU has a preference for ATP. *A) rNTPs B) dNTPs. In both gels, each sample contains 500 nM BrxU and 10 mM $MgSO_4$. Nucleotide concentration shown above each lane. Gels are representative of triplicate data. Samples were resolved in 1% agarose TAE at 120 V for 45 min. Non-hydrolysable analogues γA (ATP-γ-S) and γG (GTP- γ-S) are shown as controls and cannot be utilised by BrxU for DNA hydrolysis.*

The optimal concentration for ATP was assessed to be 1-10 mM as the smeared fragment sizes are at their smallest indicating more complete hydrolysis (Figure 4.16). At 10 mM, GTP did not induce BrxU activity, however at lower concentrations Geo gDNA was partially digested. A similar pattern was observed with CTP with the highest activity observed at 0.1 mM, however BrxU activity was reduced at 1 mM and was found to cause partial hydrolysis at 10 mM. TTP induced hydrolysis at all concentrations, however peak activity was observed at 1 mM. UTP induced equal activity levels at 0.1 mM and 1 mM and BrxU was inactive at 10 mM. Addition of the non-hydrolysable ATP analogue, ATP-γ-S, resulted in no hydrolysis of Geo gDNA, which was also observed with GTP- γ-S (Figure 4.16). Deoxynucleotides could also be used by BrxU (Figure 4.16B), demonstrating BrxU has wide substrate promiscuity. Collectively, these data show that hydrolysis of the γ-phosphate of a nucleotide is required for DNA to be cleaved, and that binding of a nucleotide is not sufficient.

In order to quantify the amount of nucleotide hydrolysis by BrxU, an assay was designed, again using BIOMOL green, to detect Pi production. Nucleotides and deoxynucleotides were tested, as well as the requirement for substrate DNA and magnesium for BrxU NTPase activity (Figure 4.17). The base reaction conditions were set up using a final concentration of 500 nM BrxU, 0.1 mM NTP, 10 mM $MgS0_4$ and 100 ng of BGP gDNA. NTPase activity was the highest when ATP was used, producing 2.9 nmol of Pi. This activity remained unchanged when BGP gDNA was substituted for resistant λ gDNA, highly sensitive Geo gDNA (Figure 4.17) or when gDNA was absent (Figure 4.17). This indicates that the NTPase domain of BrxU is not stimulated by the presence of target DNA and that hydrolysis of NTPs occurs independently. Of the remaining rNTPs, TTP produced the next highest amount of Pi at 1.97 nmol, followed by UTP (1.63 nmol), CTP (1.42 nmol) and GTP (1.21 nmol).

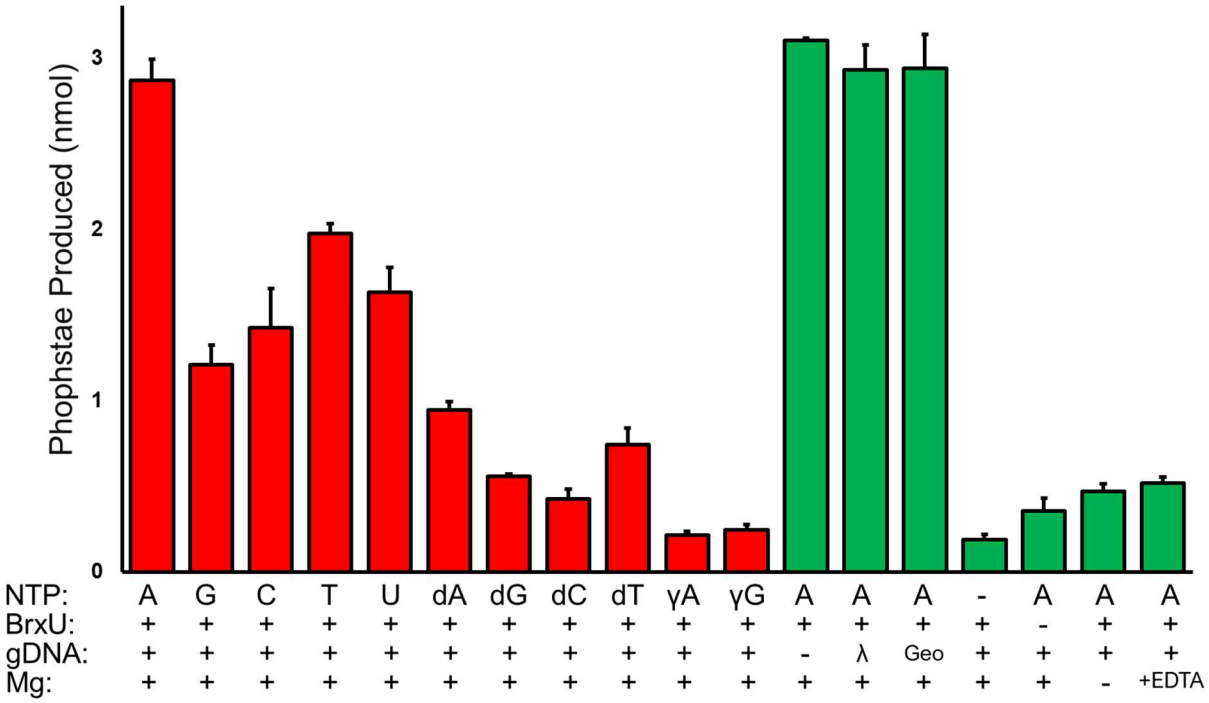


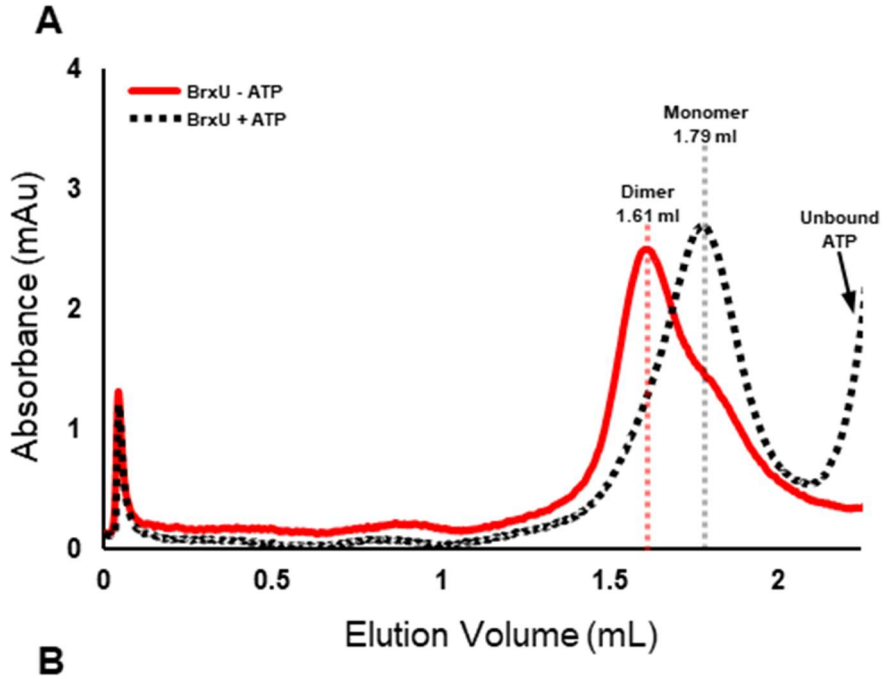
Figure 4.17: BIOMOL green phosphatase assays with BrxU identify ATP as the most readily hydrolysed NTP substrate. *BGP gDNA was the default DNA substrate. Data shown are mean and standard deviation from triplicate data.*

All rNTPs produced higher levels of Pi compared to the negative controls; in the absence of BrxU, only 0.36 nmol of Pi was detected, and in the absence of any NTP, only 0.19 nmol was detected (Figure 4.17). 0.19 nmol was recorded as the background Pi level, likely present due to the slight absorbance of the BIOMOL reagent at 620 nm. The increase from 0.19 nmol to 0.36 nmol observed in the BrxU negative control could be accounted for by non-enzymatic hydrolysis of ATP due to the conditions of the assay. Phosphate production was decreased when dNTPs were used indicating that deoxyribose sugars as part of the dNTP were less suitable for the BrxU NTPase. A slight correlation was observed between rNTPs and dNTPs, with ATP and dATP producing the highest levels of Pi in each class, with TTP and dTTP producing the second highest. This suggests a direct preference for adenine and thymine bases, with a further preference for ribose over deoxyribose.

The non-hydrolysable analogues, ATP-γ-S and GTP-γ-S produced no increase in Pi production as expected. In the absence of magnesium, 0.47 nmol of Pi was produced, comparable with the no BrxU control. The addition of EDTA to samples containing magnesium prevented ATP from being hydrolysed, demonstrating that magnesium is required for nucleotide hydrolysis (Figure 4.17).

4.13 Gel filtration analysis shows dimeric BrxU dissociating upon nucleotide binding

Having analysed the BREX proteins by analytical gel filtration, BrxU was also assessed, in the presence and absence of ATP (Figure 4.18). BrxU was observed to have an elution volume of 1.61 ml when resolved using an S200i size exclusion column. When the sample was incubated with 1 mM ATP prior to loading, BrxU eluted at 1.79 ml indicating a significant change in multimeric state. This corresponds with a shift from a dimer to a monomer when calculated from a standard of known protein sizes (Figure 4.18). This shift suggests nucleotide binding causes BrxU dimers to dissociate into monomers.



B

Sample	Mr (Theoretical)	KaV	log Mass	Mr (Calculated)	Ratio vs Mr	State
BrxU - ATP	67908	0.314	11.962	156697	2.31	Dimer
BrxU + ATP	68415	0.405	11.325	82845	1.21	Monomer

Figure 4.18: BrxU undergoes a multimeric shift upon ATP binding. *A) BrxU elutes from the S200i at 1.61 ml in the absence of ATP, and at 1.79 ml in the presence of ATP. A 10 µl sample at 500 nM was resolved at 0.175 ml/min. Traces are representative of triplicate data. No data is shown beyond 2.2 ml due to significant absorbance from unbound ATP. B) Calculation of protein mass from elution volume. Deductions for multimeric state are made by comparing the calculated mass from the eV to the theoretical mass calculated from the protein sequence.*

It cannot be confidently stated that the active form of BrxU is monomeric, as higher order structures may form in the presence of target DNA. However, it is plausible to suggest that upon nucleotide binding, a shift in protein conformation occurs. To investigate the role of nucleotides on the multimeric state of BrxU, all nucleotides previously tested as supporting DNA cleavage, in the presence or absence of magnesium, were incubated with BrxU prior to analysis by gel filtration (Figure 4.19). BrxU elutes at 1.61 ml in both the presence or absence of magnesium, showing that the binding of metal cofactor alone is insufficient in inducing this change (Figure 4.19A). However, magnesium was found to have an important role in inducing this change for each nucleotide tested other than ATP and its derivative ATP-γ-S (Figure 4.19B and K). The shift observed with both ATP-γ-S and GTP-γ-S also indicates that nucleotide hydrolysis is not a requirement for BrxU to monomerise (Figure 4.19K and L). Whilst ATP was found to induce to dimer to monomer shift whether magnesium was present or not, for all other nucleotides tested, magnesium was required for a full shift from 1.61 ml to 1.79 ml. A partial shift was observed for all nucleotides in the absence of magnesium indicated by a 'shoulder' peak at 1.79 ml (Figure 4.19C-J and L). When magnesium is added, the major peak shifts to 1.79 ml for all nucleotides. CTP with magnesium induces this shift, however a significantly larger second peak remains at 1.6 ml indicating that at the concentration tested, addition of CTP results in reduced BrxU monomer formation. dNTPs were also tested and did not fully reflect the profiles observed when NTPs were used (Figure 4.19). As for the NTPs, in the presence or absence magnesium, all dNTPs were able to induce a shift to the monomer state to some degree. However, in the absence of magnesium, dATP (unlike ATP) only induced a small subset of BrxU molecules to form monomers, producing a smaller 'shoulder' peak at 1.79 ml, with the largest peak at 1.61 ml. (Figure 4.19G). This further shows that whilst being promiscuous for nucleotide substrates, BrxU has a preference for NTPs over dNTP, and that nucleotide binding induces a large conformational change from a dimeric to monomeric state.

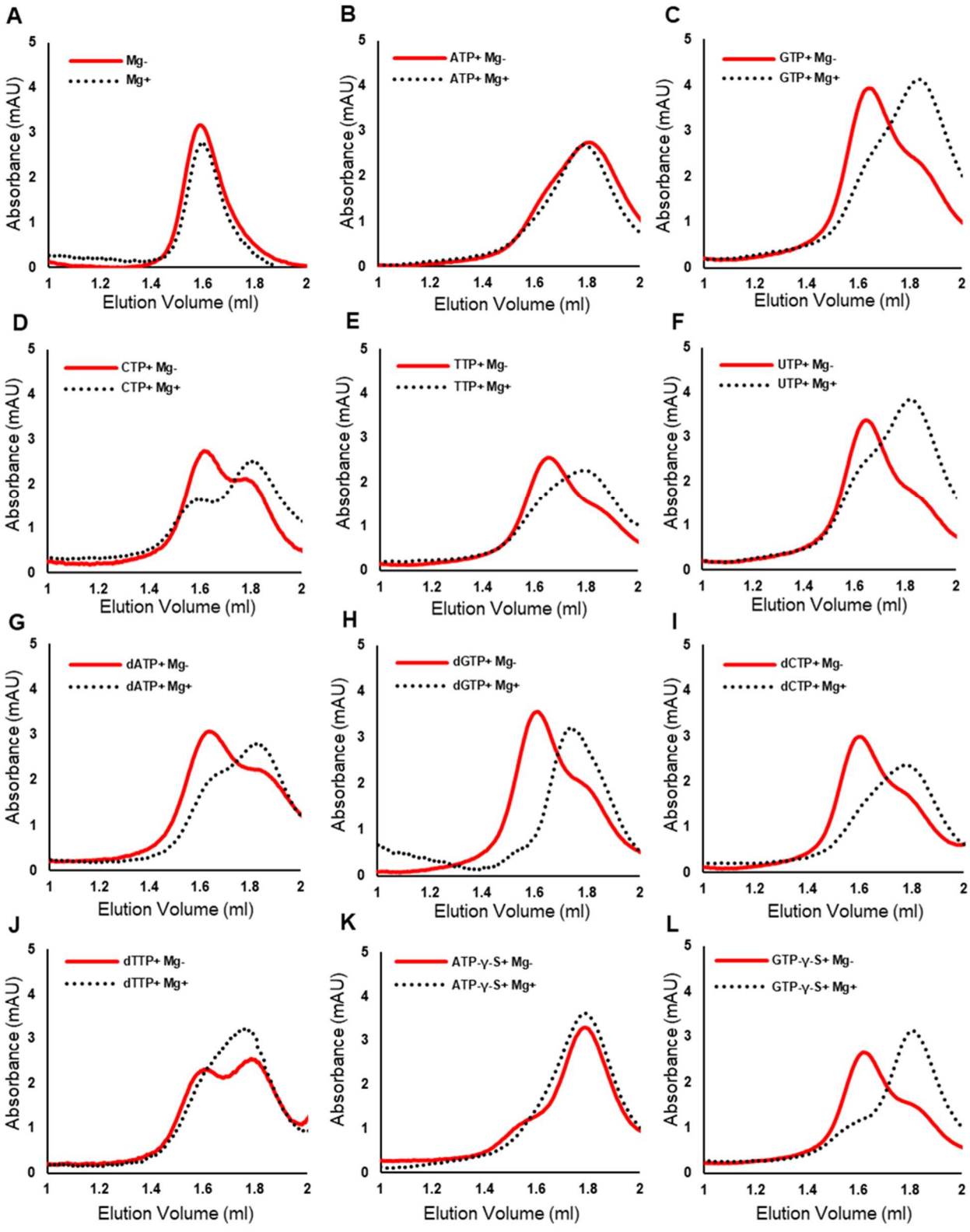


Figure 4.19: Gel filtration analysis shows ATP is the optimal nucleotide. *Presence of Mg^{2+} aids monomer formation. Traces in black (Mg+) represent samples that were preincubated with 10 mM $MgSO_4$. Traces in red (Mg-) represent samples that do not contain $MgSO_4$ prior to loading. 10 µl sample at 500 nM was resolved at 0.175 ml/min. Traces are representative of triplicate data.*

To further assess the role of the phosphate groups within ATP, BrxU was analysed using ADP and AMP (Figure 4.20). ATP induces BrxU monomer formation regardless of Mg^{2+} presence (Figure 4.20A), however BrxU remains dimeric when ATP is substituted for AMP (Figure 4.20B). ADP induces monomer formation, however in the absence of Mg^{2+}, dissociation is only part achieved (Figure 4.20C). This indicates a direct role for the both the β and γ-phosphate of ATP. As AMP does not induce monomer formation, a direct interaction of BrxU and the β-phosphate of ATP is required. As Mg^{2+} aids monomer formation for ADP (Figure 4.20C), it is likely that the β-phosphate is stabilised by Mg^{2+} in place of the γ-phosphate which is absent in ADP. Mg^{2+} has been shown to bind ADP in this manner in other ATPase domains (Thomsen and Berger 2012).

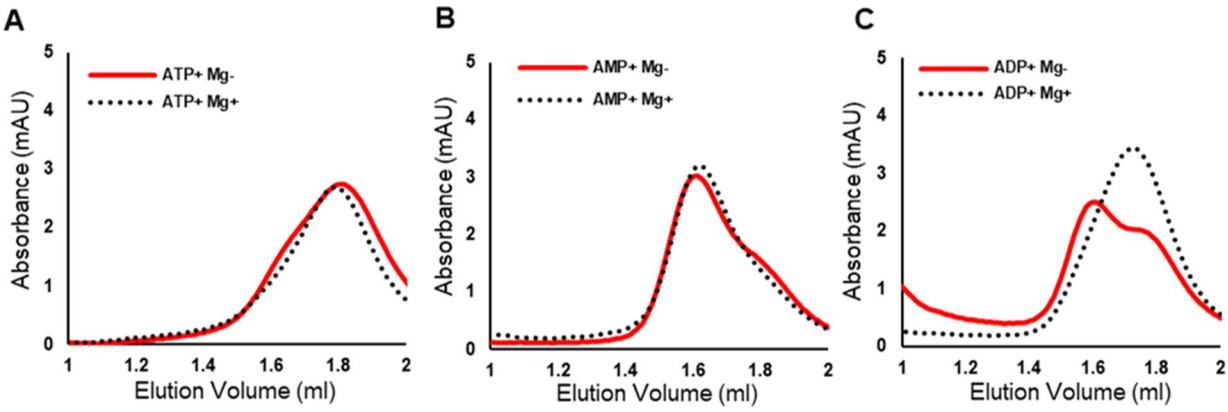


Figure 4.20: BrxU dimers undergo dissociation when bound to ATP and ADP, but not AMP. *Traces in black represent samples that were preincubated with 10 mM $MgSO_4$. Traces in red represent samples that do not contain $MgSO_4$ prior to loading. 10 µl sample at 500 nM was resolved at 0.175 ml/min. Traces are representative of triplicate data.*

4.13 BrxU targets phage genomes that incorporate modified cytosines

Phages such as T4 incorporate 5hmC into their genome during DNA replication instead of unmodified cytosine (Kutter and Wiberg 1969; Vlot et al. 2018). Modified bases such as 5hmC provide resistance against restriction enzymes with cytosine within their recognition sequence. Modified T4 genomes containing 5hmC are further modified by the addition of glucose moieties, catalysed by glucosyltranferases, yielding glc-5hmC containing genomes (Sommer, Depping, Piotrowski, and Rüger 2004). This large glucosylation modification functions along with cytosine methylation to circumvent restriction by causing increased steric hindrance to restriction enzymes. Type IV restriction enzymes such as McrBC have evolved to target modified DNAs, as a response to the phage counter-defence (Stewart et al. 2000a). To test if BrxU targets DNA with modified cytosines, artificial substrates were generated via PCR. PCR was performed using pUC19 as a template with primers amplifying a 2.7 kb fragment, with dCTP substituted with 5mC and 5hmC (Figure 4.21A). Some of the 5hmC modified sample was treated with T4 β-glucosyltransferase to generate glc-5hmC. Hydrolysis assays were performed with 1 mM ATP and 10 mM $MgSO_4$. BrxU hydrolysed all samples that contained modified cytosines, regardless of the nature of modification (Figure 4.21B). No hydrolysis of PCR products containing dCTP was observed (Figure 4.21B). Interestingly, 5mC containing substrates were hydrolysed, which contrasts with the more limited substrate range of the homologue GmrSD (He et al. 2015). To test if BrxU exhibited endonuclease or exonuclease activity, PCR product containing 5hmC from Figure 4.21A was phosphorylated with polynucleotide kinase, and ligated with T4 DNA ligase to form a circular product. Due to a limited yield of ligated DNA, only a faint band can be observed in Figure 4.21C representing circular DNA. Nevertheless, both linear and circular DNAs were degraded by BrxU, showing that BrxU is an endonuclease (Figure 4.21C).

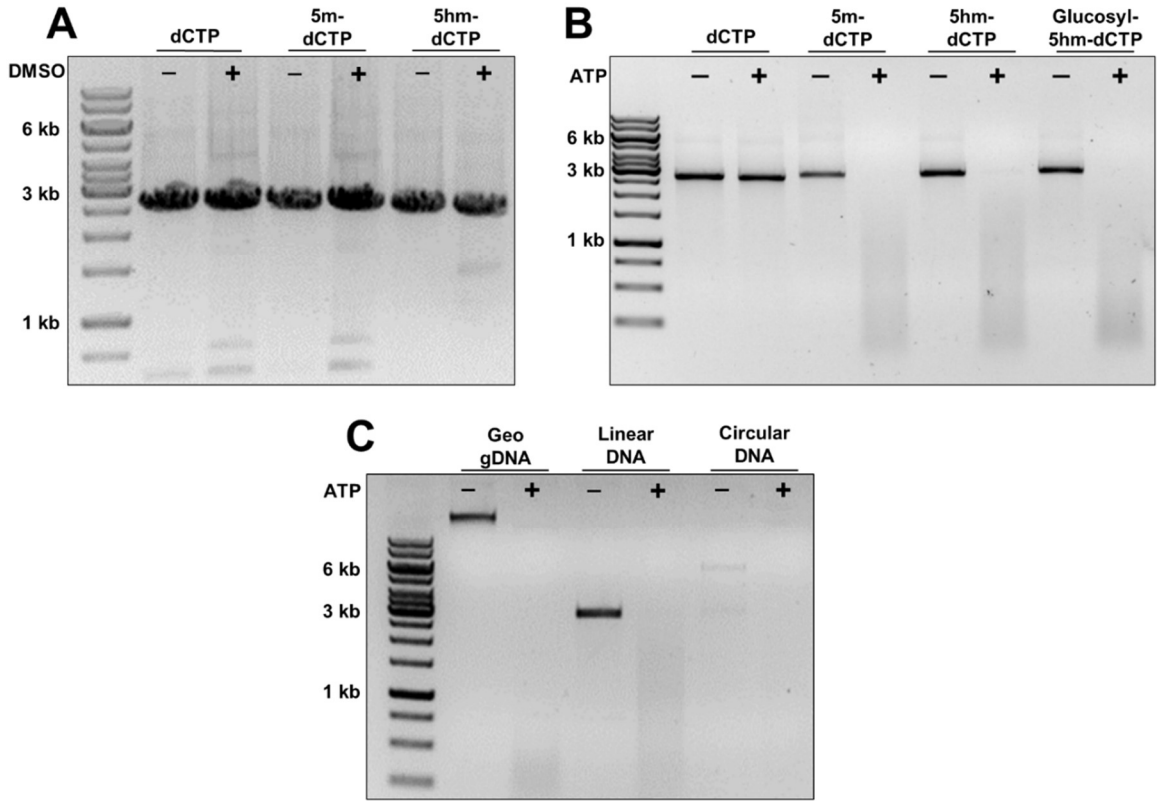


Figure 4.21: BrxU is an endonuclease that targets DNA substrates that contain modified cytosine residues. *A) PCR amplification of artificial BrxU substrate following substitution of dCTP for modified cytosine bases. B) DNA hydrolysis assays show BrxU degrades DNA containing 5mC, 5hmC and glc-5hmC. BrxU does not degrade DNA that does not incorporate modified cytosines. BrxU at 500 nM. C) BrxU has endonuclease activity as it degrades both linear and circular DNA. PCR product from Figure 4.21A was phosphorylated and ligated to produce a circular product which is degraded. Geo gDNA is shown as a positive control. A faint band representing circular DNA can be seen at ~5 kb. All samples were resolved in 1% agarose TAE at 120 V for 45 min.*

4.14 Mutation analysis of BrxU identifies key residues for enzyme activity

Following consideration of the identified amino acid sequence motifs in the GmrSD homologue, and alignment to BrxU (Figure 4.11), 11 single point mutants and 1 double mutant of BrxU were selected for further analysis. To reduce the steps required for purification, and to allow analysis of EOP within the same DH5α host strain as for previous assays, a 6His-BrxU construct was designed for mutation and cloning into pBAD30 (Guzman et al. 1995). These constructs were made commercially by Genscript. The models presented in Machnicka et al were used extensively to aid identification of mutational targets. In GmrS, the DGQQR motif is proposed as an equivalent of the signature sequence (F(G/S)GCHR) of the NTP binding motif of Srx (MACH REF).

In the N-terminal DUF262 domain, the mutants Q35A, R38A, S42A, S42D, Q101A and R102A were generated, along with a double mutant R38A/S42D. In the C-terminal DUF1524 domain, the mutants D474A, H475A, N485A, N519A and E528A were generated. It was predicted that the R38 and S42 of the N-terminal domain were involved in binding the nucleotide base (Machnicka et al. 2015), and that mutating these residues to alanine would prevent nucleotide binding. Generation of the S42D and R38/S42D double mutant was chosen due to the hypothesis that substituting a negatively charged residue could change nucleotide preference from ATP. Q101 and R102 are part of the highly conserved DGQQR motif of DUF262, aligning with the GmrSD residues Q86 and R87 that were modelled to interact with the phosphate chain of ATP (Machnicka et al. 2015). Using the model from Machnicka et al, Q35 of BrxU (Q34 of GmrS) was predicted to interact with the α-phosphate of ATP, with Q101 and R102 binding the β and γ-phosphate respectively.

D507, H508 and N522 were found to be key catalytic residues in GmrSD and that alanine substitution removed its endonuclease activity (He et al. 2015). D474 and H475 of BrxU align with GmrSD D507 and H508 and were chosen as a similar result was expected. It was predicted that N519 was a key residue and formed part of the HNH nuclease motif of DUF1524.

Mutant	EOP
WT	$<4.85 \times 10^{-8}$
Q35A	$<5.94 \times 10^{-8}$
R38A	$<3.97 \times 10^{-8}$
R38A/ S42D	0.89
S42A	1.08
S42D	1.19
Q101A	0.95
R102A	0.67
D474A	0.72
H475A	0.97
N485A	$<3.97 \times 10^{-8}$
N519A	0.36
E528A	0.13

EOP ≥ 0.5
0.5 > EOP ≥ 0.1
EOP < 0.1

Table 4.3: EOP values for 6His-*brxU *and mutants against phage Geo. *Data shown is mean calculated from triplicated experiments. Values with < extended below the range of this assay and formed no plaques.*

The ability for 6His-BrxU to confer resistance against phage Geo matched values for the untagged pBAD30-*brxU* construct (Tables 4.2 and 4.3), demonstrating that the N-terminal 6His tag had no impact on BrxU activity. Counter-resistance activity was abolished in S42A, S42D, R38A/S42D, Q101A, R102A, D474A and H475 (Table 4.3). EOP values for Q35A, R38A and N485A remained constant with WT indicating no changes to the activity of BrxU (Table 4.3). EOPs of 0.36 and 0.13 were observed for N519A and E528A, respectively, showing that these mutations have greatly reduced BrxU activity, but it was not entirely abolished.

4.15 Expression and purification of BrxU mutants

BrxU can be expressed from pBAD30-*6His-brxU* in both DH5α and BL21 DE3 strains. A higher yield was observed in protein expression with BL21 DE3 and so all mutants, and the WT, were expressed in this strain. No toxicity was observed during any preparation for all mutants. Expression of BrxU from this construct yields an N-terminal 6His-BrxU that can be purified via one-step IMAC. Contaminating proteins can be removed using 50 mM imidazole in wash buffers, and BrxU can be eluted at 250 mM imidazole. Purified protein was obtained for all mutants, with similar expression levels (Figure 4.22).

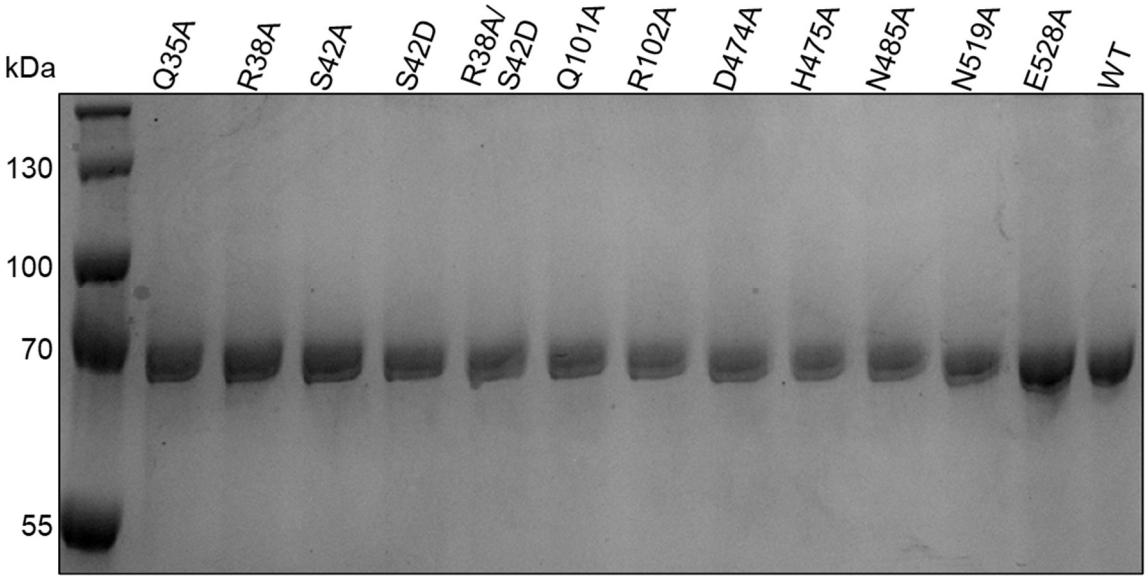


Figure 4.22: SDS-PAGE showing the purity of IMAC purified BrxU constructs expressed from pBAD30-*brxU**. *Samples resolved in 10% acrylamide at 180 V for 100 min.*

4.16 BrxU mutant DNA digest assays

In order to investigate the effect of introduced mutations on the ability of BrxU to cleave phage gDNA, DNA hydrolysis assays were performed using purified protein for each mutation. The aim of these experiments was to semi-quantify the capacity for each BrxU mutant to hydrolyse phage Geo gDNA, using a gradient of protein concentration (Figure 4.23).

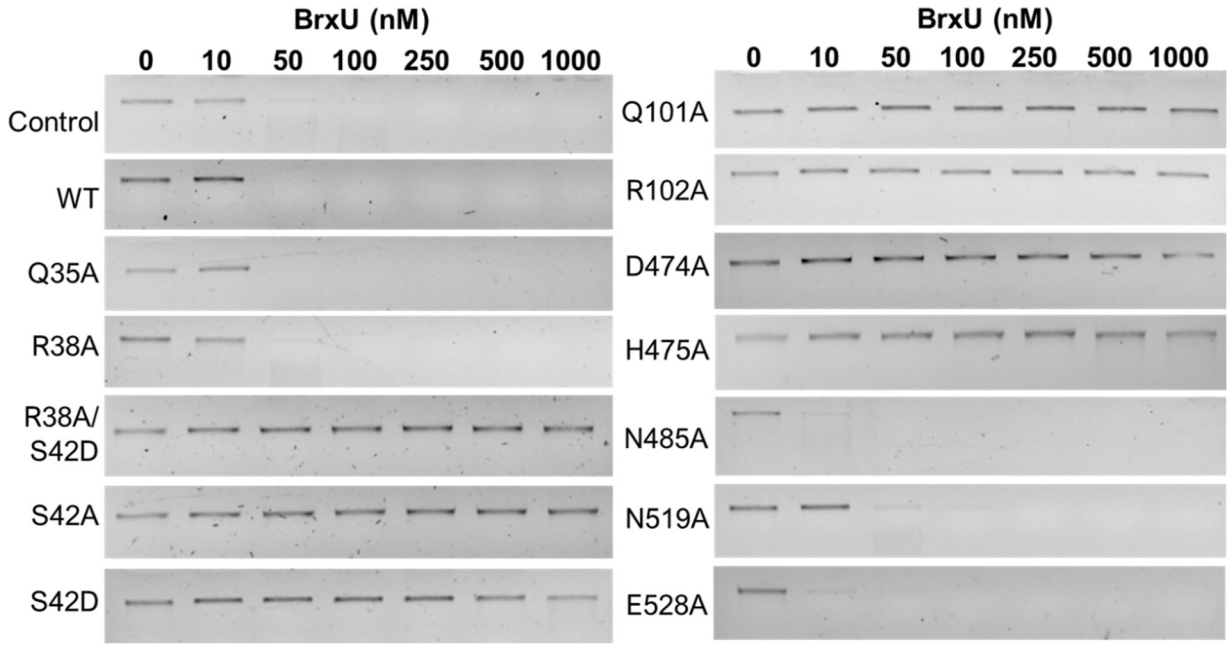


Figure 4.23: DNA hydrolysis assays with BrxU mutants. *100 ng of Geo gDNA was incubated with a gradient of BrxU concentrations at 37 °C for 60 min and resolved via agarose gel electrophoresis at 120 V for 45 min. All samples contain 10 mM $MgSO_4$ and 1 mM ATP. Control sample is untagged BrxU expressed from pSAT1-LIC-*brxU*. WT and mutants are expressed from pBAD30-*brxU *and its derivatives. Data shown are representative of triplicate experiments.*

WT BrxU expressed from pBAD30-*6His-brxU* was observed to be as active as the positive control expressed from pSAT1-LIC-*brxU* (Figure 4.23). WT BrxU degraded the gDNA substrate at 50 nM but not at 10 nM. Q35A, R38A and N519A were found to be as active as WT (Figure 4.23). R38A/S42D, S42A, S42D, Q101A, R102A, D474A and H475A were observed to be inactive even at protein concentrations of 1000 nM, indicating a complete inactivation of the enzyme (Figure 4.23). In contrast, N485A and E528A were observed to be more active than WT, degrading gDNA at 10 nM, however substrates were only partially degraded at this concentration. The observed phage-resistance and gDNA digestion phenotypes correlate for Q35A, R38A and N485A (Table 4.3). In addition, whilst N519A and E528A have reduced impact on EOP, these data show they are still able to digest gDNA (Table 4.3 and Figure 4.23).

4.17 ATPase activity is elevated in S42A

NTPase assays were repeated using each purified BrxU mutant. Assays conditions were 500 nM BrxU, 0.1 mM ATP, 10 mM $MgSO_4$ and 100 ng Geo gDNA for test samples. Untagged BrxU expressed from pSAT1-LIC-*brxU* was used as a positive control and found to show no significant difference in activity compared to 6His-BrxU WT from pBAD30-6His-*brxU* (Figure 4.24). Mutants that were previously shown to provide phage-resistance and digest modified phage gDNA, Q35A, R38A, N485A, N519A and E528A, all had similar activity to WT (Figure 4.24). Surprisingly, whilst showing no phage resistance or ability to cleave phage gDNA, S42A had increased NTPase activity, producing 5.57 nmol of Pi (Figure 4.24). In contrast. a reduction in activity was observed for inactive mutants S42D, R38A/S42D, Q101A and R102A (Figure 4.24). Whilst mutants D474A and H475A were shown to not restrict phage, and were unable to digest phage gDNA, they were observed to have WT levels of NTPase activity (Figure 4.24). That could suggest that these mutants were somehow deficient for DNA substrate cleavage and recognition, rather than NTP binding and hydrolysis.

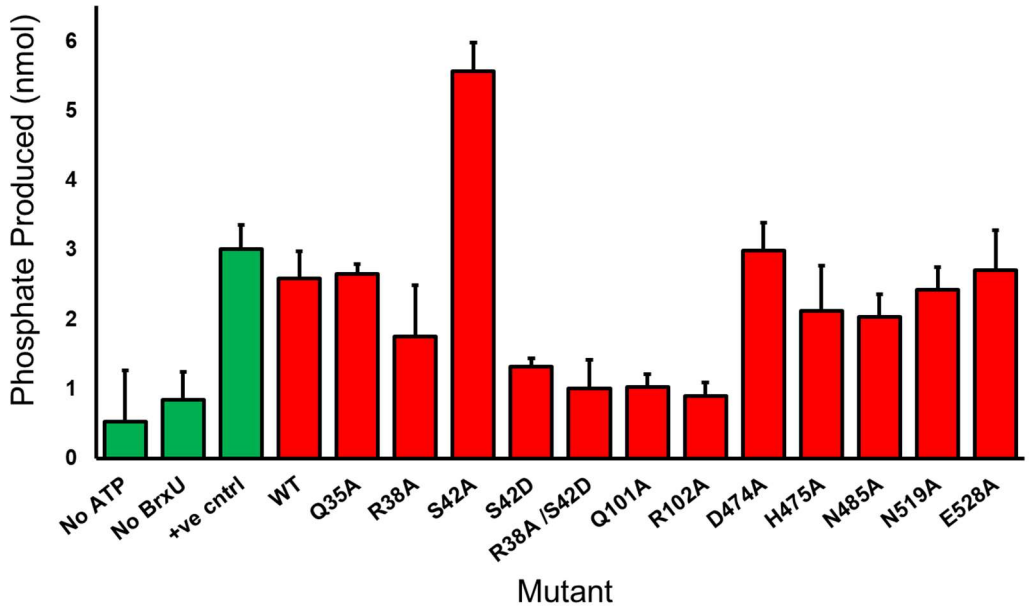


Figure 4.24: BIOMOL green phosphate detection assays with BrxU mutants. *Elevated phosphate levels are detected for S42A. Presented data are the mean and standard deviations from triplicate data.*

4.18 Identification of key residues in BrxU dimerisation

Next, the mutants were tested to examine their impact on switch from BrxU dimers to monomers. All experiments were run with 10 mM $MgSO_4$. WT BrxU expressed from pBAD30-6His-*brxU* was compared to BrxU expressed from pSAT1-LIC-*brxU* (Figure 4.25). The addition of 1 mM ATP induced a shift from a dimer to monomer for both BrxU ("Control"; Figure 4.25A) and 6His-BrxU ("WT", Figure 4.35B), confirming the 6His tag does not alter BrxU behaviour in these experiments

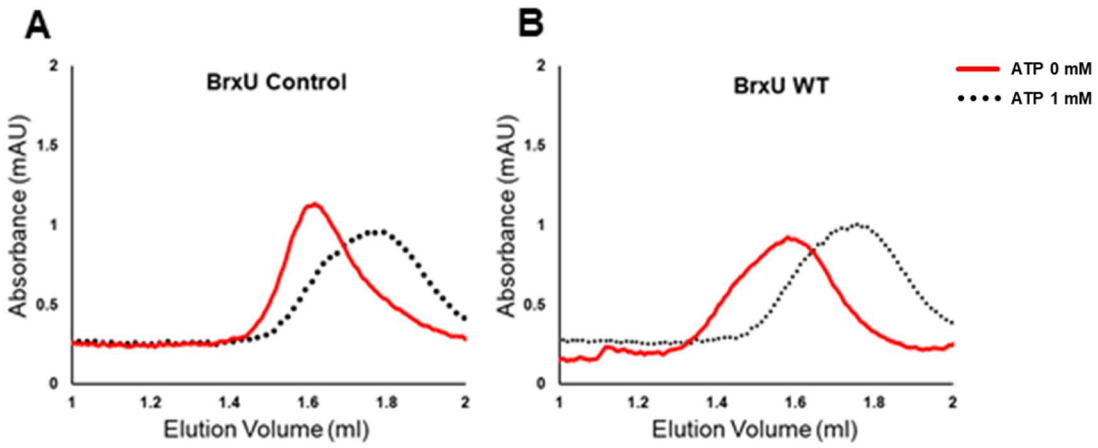


Figure 4.25: Gel filtration analysis confirms 6His-tagged BrxU undergoes monomerisation upon ATP binding. *Traces in black represent samples that were incubated with 1 mM ATP prior to loading, traces in red were not. 10 µl sample at 500 nM was resolved at 0.175 ml/min. Traces are representative of triplicate data.*

With data from Figure 4.25 confirming that WT 6His-BrxU expressed from pBAD30:*brxU* undergoes the same multimeric shift as untagged BrxU, mutant 6His-BrxU proteins were analysed in a similar manner (Figure 4.26). All mutations in the C-terminal domain (D474A to E528A) were found to behave similarly to WT, with the addition of 1 mM ATP inducing a full shift to the monomer form of BrxU (Figure 4.26H-L).

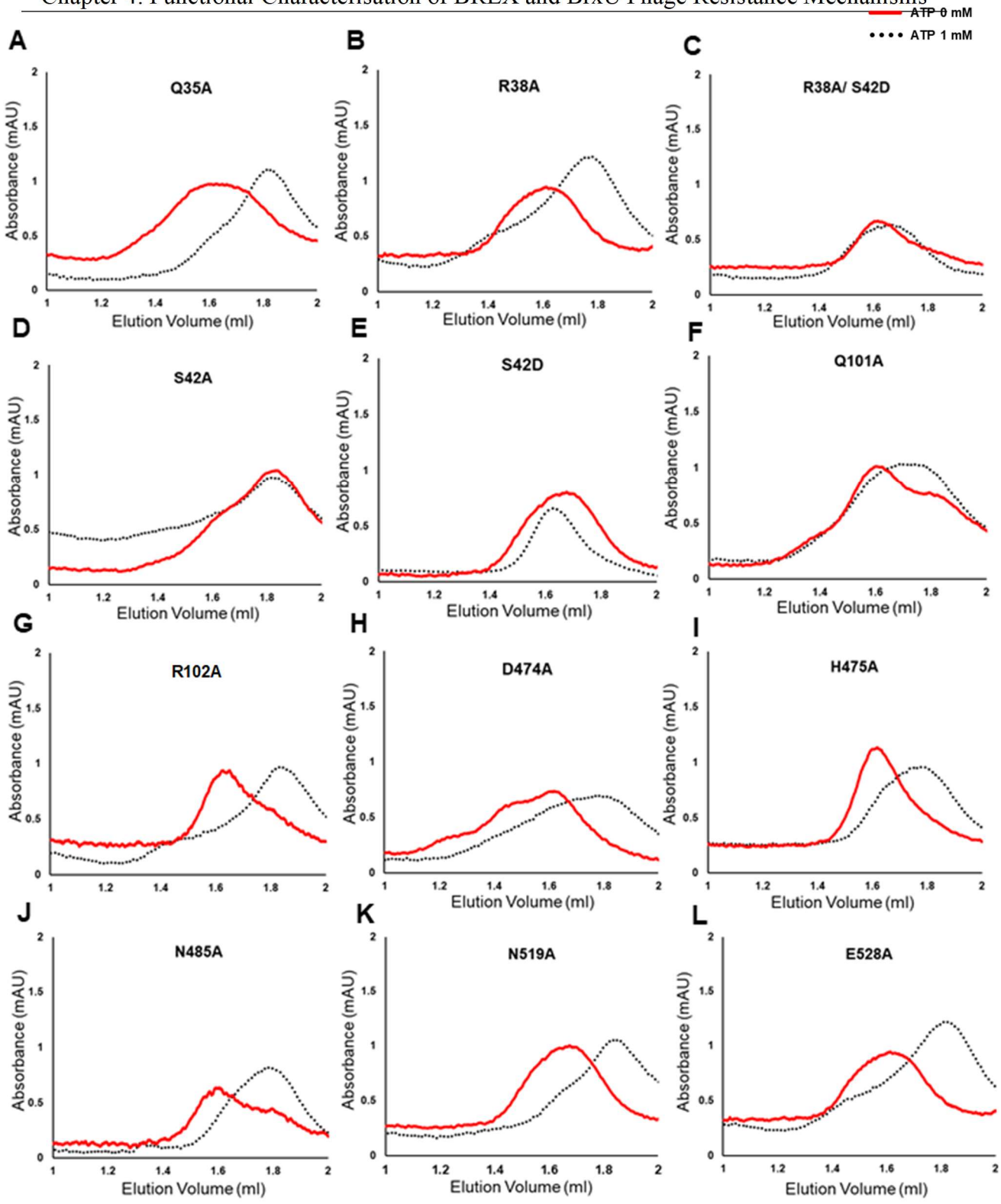


Figure 4.26: Gel filtration analysis of BrxU mutants reveals S42 is involved in dimer formation. *Traces in black represent samples that were incubated with 1 mM ATP prior to loading, traces in red represent samples that contained 0 mM ATP. 10 µl sample at 500 nM was resolved at 0.175 ml/min. Traces are representative of triplicate data.*

This suggests that D474A and H475A are not compromised for ATP binding, hydrolysis, and conformational changes, so the observed lack of activity for these mutants against phages might instead be down to substrate DNA recognition and cleavage. Within the N-terminal domain, Q35A, R38A and R102A acted as WT. Q101A was also observed to undergo a shift in the absence of ATP, with a larger shoulder peak at 1.8 ml. However, in the presence of ATP, Q101A elutes over a wider peak at 1.71 ml, indicating equal levels of dimer and monomer (Figure 4.26F). S42A resulted in an elution volume of 1.8 ml even in the absence of ATP, indicated that the mutation results in BrxU permanently forming monomers (Figure 4.26D). This corresponds with the increased NTPase activity (Figure 4.24). However, in both R38A/S42D and S42D mutants (Figure 4.24C and E), BrxU elutes as a dimer in both the absence and presence of ATP, which again correlates with the reduced ATPase activity of these mutants (Figure 4.24). S42 therefore appears to be an important residue in NTP recognition and hydrolysis, which then impacts conformational shifting of BrxU.

4.19 Discussion

4.19.1 Individual BREX protein roles

All BREX proteins with the exception of BrxS and BrxT have been expressed as soluble proteins and can be purified. BrxR has been identified as a transcriptional regulator and has been shown to act as an autorepressor using β-galactosidase assays (Figure 4.9). Assessment of the 6 promoter regions investigated in section 4.6 shows that a strong promoter region lies direct upstream of *brxR*, indicating that *RABCUX* are transcribed as one polycistronic operon. It is possible that *ZL* may also be transcribed as part of this operon, however expression of BrxR was shown to also increase transcription of a promoter upstream of *ZL*. This indicates that the BREX locus in pEFER may be transcribed as a two-step system, with *RABCUX* being expressed first. The presence of BrxR prevents further transcription of this operon, whilst an

increase in activity in the *ZL* promoter suggests that during this autorepression at *RABCUX*, *ZL* is activated and levels of PglZ and BrxL increase. BrxR showed significant homology to PafBC which upregulates response genes when DNA damage is detected (Müller et al. 2019). It is predicted that BrxR interacts with DNA similarly to PafBC, as Phyre2 modelling has predicted the presence a wHTH domain. This would involve the recognition helices of the wHTH domain binding into the major groove of DNA stabilised by the interaction of the wings with the minor groove. Interestingly, BrxR is regularly found clustered with other phage defence genes. Bioinformatic analysis of BrxR performed by Joshua Lee (Unpublished MBiol thesis) identified 197 BrxR homologues that are found upstream of phage defence systems. It is likely that BrxR functions as a phage sensing element, activating transcription of BREX and *brxU* as a result of a phage infection. BrxR encodes a WYL-domain, that has been proposed to bind a yet-unidentified ligand (Müller et al. 2019). This ligand may well regulate BrxR-DNA binding. BrxR could therefore be a sensor of phage infection, regulating an appropriate response. Further analysis of BrxR should be undertaken by creating a deletion mutant in pBrxXL to determine if the phage defence island is still active without it, and performing electrophoretic mobility shift assays to determine sites of DNA binding. Furthermore, the promoter analysis should be expanded to *brxS* and *brxT*, and the region just upstream of these genes.

BrxC was predicted to have ATPase activity, which was demonstrated in Figure 4.10. BrxC was able to hydrolyse ATP in the absence of other BREX proteins, indicating that the ATPase domain of BrxC requires no additional protein factors to be active. It is predicted that the ATPase activity of BrxC is providing the energy for translocation of a BREX complex along target DNA, and so it is feasible that BrxC forms complexes with BREX effector proteins. Both PglZ and BrxL were observed to not hydrolyse ATP in the same assay. PglZ was also observed to not cleave another substrate, p-nitrophenol phosphate (data not shown). This suggests that if either BrxL or PglZ does function as a phosphatase, it is dependent on other factors to activate

it such as another BREX protein, or target substrate. Having purified the BREX proteins, it will be possible to mix them and separate by analytical sizing as part of future experiments. This will allow observation as to whether any BREX complexes form. These can then be tested for ATP hydrolysis activity.

PglX has been shown to be required for BREX activity as its deletion abolishes phage resistance and target DNA remains unmethylated (Table 3.5). A structural and biochemical analysis of PglX from *S. enterica* is being performed by another PhD student, Sam Went, within the Blower laboratory. PglX from pEFER can be expressed as a soluble protein and purified. Using purified PglX, it will be possible in future to test PglX and any BREX complexes for methyltransferase activity.

4.19.2 Gel filtration analysis of BREX proteins

Gel filtration analysis allows for the prediction of multimer formation based on the calculated size of protein eluates. BrxR was observed to form dimers which is supported its role as a transcriptional regulator. BrxA was found to be dimeric in solution and BrxB was found to be monomeric. With no further characterisation of each protein, it is difficult to suggest roles for each protein and how their multimeric state would affect this. The pentamer formed by BrxC is characteristic of many AAA ATPases. For instance, the large terminase, TerL, packages viral DNA into the procapsid and forms pentameric ring structures (Hilbert et al. 2015). The hydrolysis of ATP then powers DNA translocation into the procapsid. It is possible that the ATPase activity of BrxC provides the mechanical force for a complex of BREX proteins to translocate target DNA.

PglX is most likely to be monomeric in solution, although it is likely that it forms homodimers upon binding DNA. The PglX homologue, MmeI, forms dimers in complex with DNA with

subunits interacting with each strand (Callahan et al. 2016). Phyre2 models of PglX using MmeI as a template shows PglX as an elongated protein which likely migrates through the S200i slower than more globular proteins. PglZ and BrxL were both calculated to form dimers, however due to their predicted elongated shapes it is possible that they migrate slower and are monomeric. Interestingly, the smaller peak at 1.2 ml observed for BrxL may represent a small proportion of BrxL multimers that may assemble under the correct conditions. Addition of magnesium and ATP was performed however no change was observed.

4.19.3 BrxU confers resistance to BREX-resistant phages

BrxU was originally identified as a putative nuclease due to the presence of the C-terminal DUF1524 domain. DUF1524 has been implicated in phage defence systems such as GmrSD, which targets phages encoding modified cytosine residues in their genomes (He et al. 2015). *In vivo,* expression of BrxU from pBAD30-*6HisbrxU* conferred resistance to all phages susceptible to pBrxXL-*ΔpglX*, showing that BrxU functions independently of BREX (Table 4.2). DNA hydrolysis assays have shown that all phages that were susceptible *in vivo* are degraded by BrxU (Figure 4.13). BGP and TB37 were found to be sensitive to BrxU but to a much lesser extent that an all other sensitive phages. Both BGP and TB37 were only partially degraded and had EOP values that sat in between hypersensitive and resistant phages. It would be interesting to explore the mechanism behind the reduced activity against these phages and what distinguishes them from hypersensitive phages such as TB36.

4.19.4 Endonuclease activity of BrxU

One of the more intriguing aspects of BrxU is the generation of DNA ‘smears’ observable during hydrolysis assays (Figure 4.12 and Figure 4.16). Rather than generating distinct DNA fragments characteristic of type IIP enzymes, DNA smears appear indicating cleavage at non-

specific sites. GmrSD has been detailed to cut 17-23 bp away from the recognition sequence (He et al. 2015). It is likely that BrxU cleaves DNA at a similar distance away however this has not been investigated in this study, and would require future purification and sequencing of digested fragments. A key part of the biochemical characterisation of BrxU was to assess its cofactor requirements. BrxU was found to exhibit endonuclease activity in the presence of ATP, however it was able to utilise GTP, CTP, TTP and UTP, as well as all canonical dNTPs (Figure 4.16). It was through this assessment that the requirement for nucleotide hydrolysis was noted, as binding of non-hydrolysable nucleotide analogues resulted in DNA substrates remaining intact (Figure 4.16). With the putative NTPase activity situated in the N-terminal DUF262 domain, and the putative nuclease domain situated in the C-terminal DUF1524 domain, it is most likely that nucleotide hydrolysis is required for molecular motion and/or DNA manipulation rather than DNA hydrolysis. ATP was found to be the preferred nucleotide cofactor, generating the smallest DNA fragments. However, BrxU is able to utilise a wide range of nucleotide cofactors, indicating that the nucleotide binding site exhibits flexibility around the nitrogenous base. BrxU NTPase activity was demonstrated through colourimetric assays developed using BIOMOL green, indicating that ATP is more readily hydrolysed than other nucleotides, supporting the hypothesis that BrxU has preference for ATP (Figure 4.17). Interestingly, NTPase activity was not further stimulated by the addition of target DNA (Figure 4.17). It was predicted that the energy derived from the hydrolysis of ATP was utilised for DNA manipulation. It is still possible that the hydrolysis of ATP does drive a mechanical force and that DNA does not have to be bound for this to occur. Furthermore, it can be deduced that BrxU has at least one metal binding site as the removal of magnesium (or its sequestering with EDTA) prevented nucleotide hydrolysis (Figure 4.12B). BrxU was originally noted to be a Mg^{2+} dependent nuclease, however it has been found to be active in the presence of Mn^{2+}, Fe^{2+}, Fe^{3+}, Ni^{2+} and Co^{2+}. The promiscuity of BrxU and its coordinating metal ions suggests that the

nuclease active site may be functional in a range of conformations. Manganese was noted to be utilised at lower concentrations that magnesium, however this does not necessarily suggest a preference for manganese. Cellular manganese levels can increase significantly during phage infections as demonstrated in *S. enterica*, whilst levels of zinc, iron and magnesium decrease (Yousuf et al. 2020).

4.19.5 BrxU dissociates from a homodimer to a monomer upon nucleotide binding

Gel filtration analysis was used to investigate the effect of nucleotide binding on BrxU. In the presence of Mg^{2+}, all nucleotide cofactors tested were able to induce the dissociation, however only ATP and ATP-γ-S induced this shift without Mg^{2+}. It is hypothesised that binding of a metal cofactor correctly coordinates the nucleotide binding site, allowing for non-preferred nucleotides to bind and induce this dissociation. It is possible that BrxU forms higher order structures in the presence of DNA, with BrxU monomers binding individuals target sites and dimerising. McrBC forms tetradecameric multimers in the presence of Mg^{2+} and GTP (Panne et al. 2001), however no large structures were observed for BrxU. The dissociation observed may be due to the binding of nucleotides inducing a conformation change, allowing BrxU monomers to bind target DNA. The DNA binding regions of BrxU may be occluded when dimerisation occurs and may become exposed again during dissociation.

4.19.6 BrxU substrate identification

Type IV restriction enzymes recognise modified DNA substrates that contain non-canonical bases, such as 5hmC. GmrSD was found to hydrolyse DNA substrates that contained 5hmC and glc-5hmC but not 5mC (He et al. 2015). BrxU was observed to hydrolyse artificial DNA substrates that had been synthesised with dCTP substituted for 5mC, 5hmC and glc-5hmC. DNA glycosylation is utilised by phages such as T4 to evade restriction, conferring resistance

to defence systems such as CRISPR-Cas (Bryson et al. 2015). The addition of a large glucose moiety to target bases can cause steric hindrance to enzymes such as Cas9. However, BrxU is able to hydrolyse a wide range of modified DNA substrates, including glycosylated DNA. As BrxU is also able to digest DNA containing 5mC, it is likely that the DNA recognition apparatus of BrxU has a direct interaction with the C_5 of cytosines, and that the addition of a methyl group is sufficient to result in cleavage.

4.19.7 Identification of key catalytic residues in BrxU

Mutagenesis of BrxU has revealed the importance of several residues within both domains. In the N-terminal DUF262 domain, mutations were designed to prevent interactions with nucleotides. Both Q35A and R38A were found to remain active, indicating that nucleotide binding is less conditional on these residues that previously believed. Modifications of S42 resulted in BrxU antiphage activity being inactivated. Interestingly, the S42A mutant was found to be monomeric in the absence of ATP, suggesting this residue has an important role in the dimer interface and that its disruption results in a fixed state of BrxU. The binding of ATP inducing the monomerisation of BrxU has been hypothesised as an activation step for activity, and the disruption of S42 likely prevents nucleotide binding correctly. Strikingly, the S42A mutant was observed to hydrolyse ATP at an increased rate. It is currently unclear as to how this is facilitated. Conversely, S42D mutants were observed to be fixed as dimers even in the presence of ATP. This mutant was generated in order to assess if nucleotide preference could be altered by substituting for a negatively charged residue at this position. It is possible that the S42D mutant is still able to switch conformations when other NTPs such as GTP are present, however this has not been tested yet. Q101 and R102 are part of the highly conserved DGQQR motif, predicted to bind the β- and γ-phosphates of ATP, respectively. Both Q101A and R102A yielded inactive enzymes, illustrating the importance of these two residues. D474 was predicted

to be involved in metal binding, and its substitution to alanine prevented BrxU from cleaving DNA. ATPase activity of all C-terminal mutants was found to be unchanged as expected.

Having performed a range of biochemical assays with BrxU and assessing its required reaction conditions, nuclease activity, multimeric state and nucleotide interaction, as well as demonstrating its anti-phage activity *in vitro*, BrxU was selected as a candidate for structural characterisation. The next chapter will detail the progress made in BREX protein crystallisation, and will detail the crystallisation, structure determination and analysis of BrxU.

Chapter 5: BREX Protein Crystallisation and Structural Analysis of BrxU

5.1 Introduction

The pEFER phage defence locus contains multiple target proteins of interest for crystallographic studies. Phyre2 models gave insight into the potential function of each protein with partial modelling of certain domains, leaving other domains unmodelled. For example, modelling the winged HTH domain of BrxR (Figure 4.1) aided its identification as a transcriptional regulator, however the C-terminal domain was left unmodelled. Comparative bioinformatics hinted that the C-terminal domain may function as a sensor of phage infection, making BrxR an interesting target for structural characterisation. Similarly, BrxC, PglX, PglZ and BrxL were identified as interesting targets for crystallographic study as they represent novel enzymes in phage defence. Protein crystallisation was performed alongside functional characterisation. Following the characterisation of BrxU as an independent type IV restriction enzyme, efforts to solve the structure of BrxU were prioritised over other proteins such as BrxR. A structure for BrxU was obtained and further work is ongoing for BrxR. This chapter covers progress on each protein, detailing the advances in crystallisation for each before a full report on the structural characterisation of BrxU is presented.

5.2 Protein crystallisation preparation

Crystallisation trials were performed with purified protein samples (Figure 4.5) expressed from pSAT1-LIC constructs. Samples were dialysed to remove glycerol and reduce salt content, and concentrated. For initial screening, starting protein concentrations were 12 mg/ml for BrxR, BrxA and BrxU, and 10 mg/ml for BrxC, PglZ, BrxL. Screens were set using 96 well 2-drop vapour diffusion sitting drop plates (Swissci MRC) using either an Innovadyne or an STPlabtech Mosquito Xtal3 robot. Drops were set at 1:1 and 2:1 protein: precipitant ratios forming 200 nl and 300 nl final drops. A range of 10 commercially available screens for soluble proteins (Molecular Dimensions) were used; BCS Eco, Clear Strategy 1+2 Eco, JCSG+ Eco, LMB, MIDAS, Morpheus, Pact Premier Eco, SG1 Eco, Structure Screen 1 Eco and Structure Screen 2 Eco. A total of 1920 conditions were set for each protein. No crystals were obtained for BrxC.

5.3 BrxR protein crystallisation and data collection

BrxR formed crystals overnight in 21 screening conditions, mostly forming very thin, needle crystals unsuitable for mounting (Figure 5.1 R1-R3). Long rod crystals appeared in SG1 F8 (Figure 5.1 R6). The best crystals appeared in Pact Premier F8, comprised of mostly needle crystals and a number of shorter, thicker rods. This condition consisted of 0.1 M Bis-Tris propane pH 6.5, 0.2 M sodium sulphate and 20% w/v PEG 3350 (Figure 5.1 R4). This condition was selected for manual optimisation. Reduction of the starting PEG 3350 concentration to 19% yielded thicker rods that were suitable for mounting (Figure 5.1 R5). PEG 3350 concentration was reduced further in an attempt to slow crystal formation but no significant crystal growth was observed below 19%.

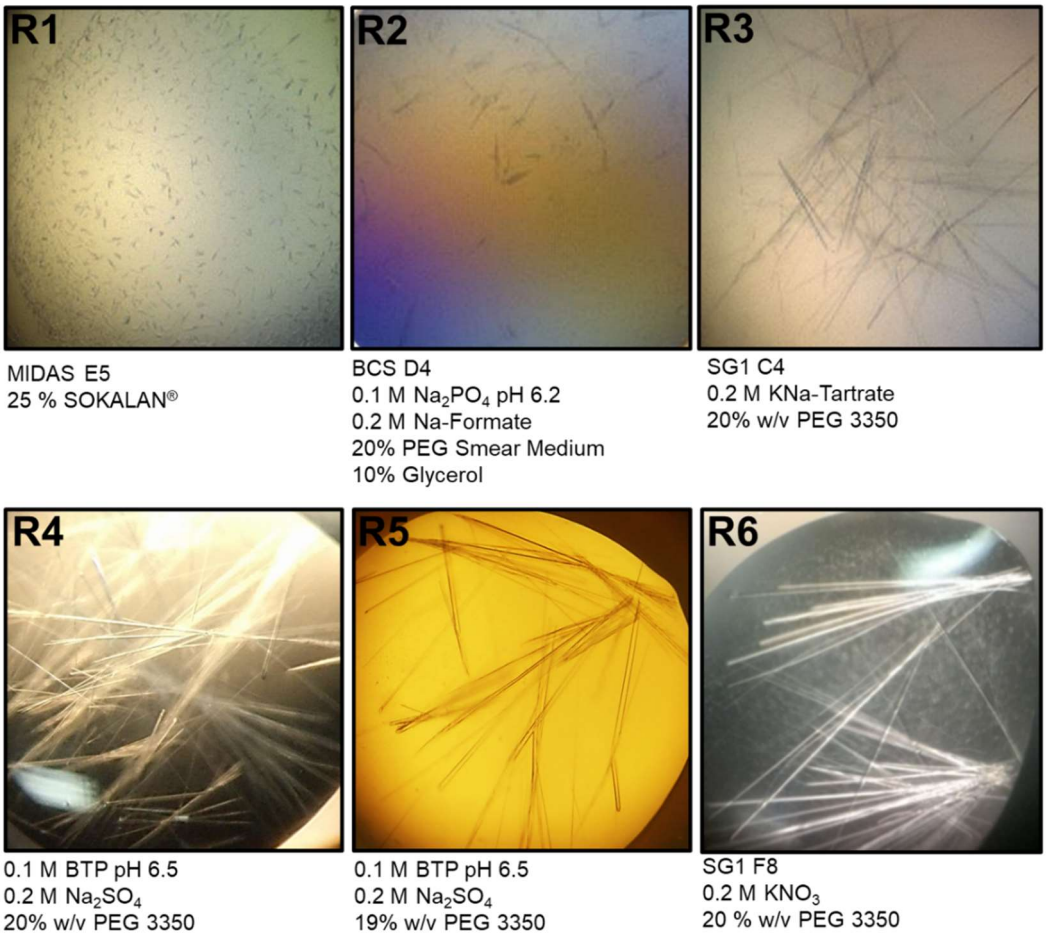


Figure 5.1: Native BrxR crystallisation conditions. *Not all conditions shown. Images are not to scale. Diffraction data for R5 can be seen in Figure 5.2. BrxR set at 12 mg/ml.*

These crystals were flash frozen in liquid nitrogen following cryoprotection and transported to the Diamond Light Source (DLS) synchrotron facility in Didcot, UK. The highest resolution dataset obtained for BrxR had a max resolution of 2.8 Å (Figure 5.2). Parameters for data obtained for this crystal are detailed in Table 5.1. No solution could be obtained for BrxR via molecular replacement due to a lack of a suitable model at this resolution.

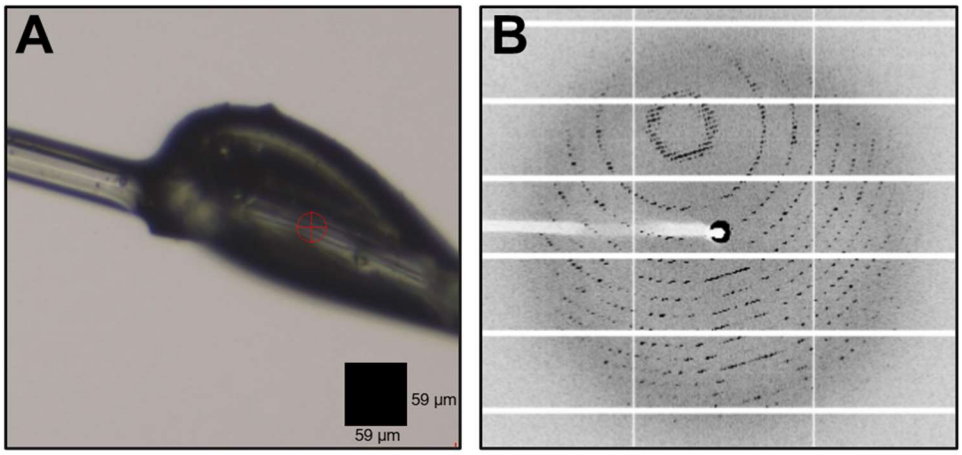


Figure 5.2: X-Ray data collection of native BrxR. *A) Mounted crystal from condition R5 of Figure 5.1. B) A single lattice is observed with BrxR diffracting to a max resolution of 2.8 Å.*

Parameter		Value
Crystal		R5
Crystallisation condition		0.1 M BTP pH 6.5 0.2 M Na_2SO_4 PEG 3350 w/v 19%
Wavelength (Å)		0.9763
Beamline		DLS I03
Space group		P 41 2 2
Unit cell dimensions	A	130.68
	B	130.68
	C	360.71
	α	90
	β	90
	γ	90
Resolution (Å)		2.8
Total reflections		146734
Unique reflections		77765
Multiplicity		1.9
Completeness (%)		99.9
Mean I/sigma(I)		4.3
Rmerge		0.084
Rmeas		0.118
CC1/2		0.994

Table 5.1: Crystallographic parameters for data collection of R5 in Figure 5.2. *A single dataset processed via DIALS to 2.8 Å was processed via AIMLESS.*

Strategies for obtaining phase data were discussed, with BrxR being identified as a suitable candidate for selenomethionine (SM) substitution. Including the N-terminal methionine, BrxR has a total of 4 methionine residues. As this protein was expressed as a tagged fusion construct, the N-terminal methionine is still present and has not been cleaved by methionine aminopeptidase. SM-BrxR was produced using a commercially available methionine-deficient nutrient mix (Molecular Dimensions). Following expression, SM-BrxR was purified by the exact procedure used for native BrxR. Positive electrospray time of flight (ES^+ TOF) mass spectrometry was used to determine SM incorporation (Figure 5.3). For each methionine residue substituted for SM, a mass change of +47 will be observed. Native BrxR was observed to have a mass of 33595 and SM-BrxR was observed to have a mass of 33735. The difference between the two species was calculated to be 140, which closely corresponds to 3 SM substitutions. A smaller peak was found in SM-BrxR at 33688, which closely corresponds to 2 SM substitutions. SM-BrxR was rescreened against the 10 commercial screens and was also manually set using a matrix of conditions based on Pact Premier F8, however no significant crystal formation was observed in the commercial screen. Crystallisation was observed within the optimised matrix but only small, thin needles were formed, which were unsuitable for mounting. Greater optimisation is required to proceed with BrxR structural studies.

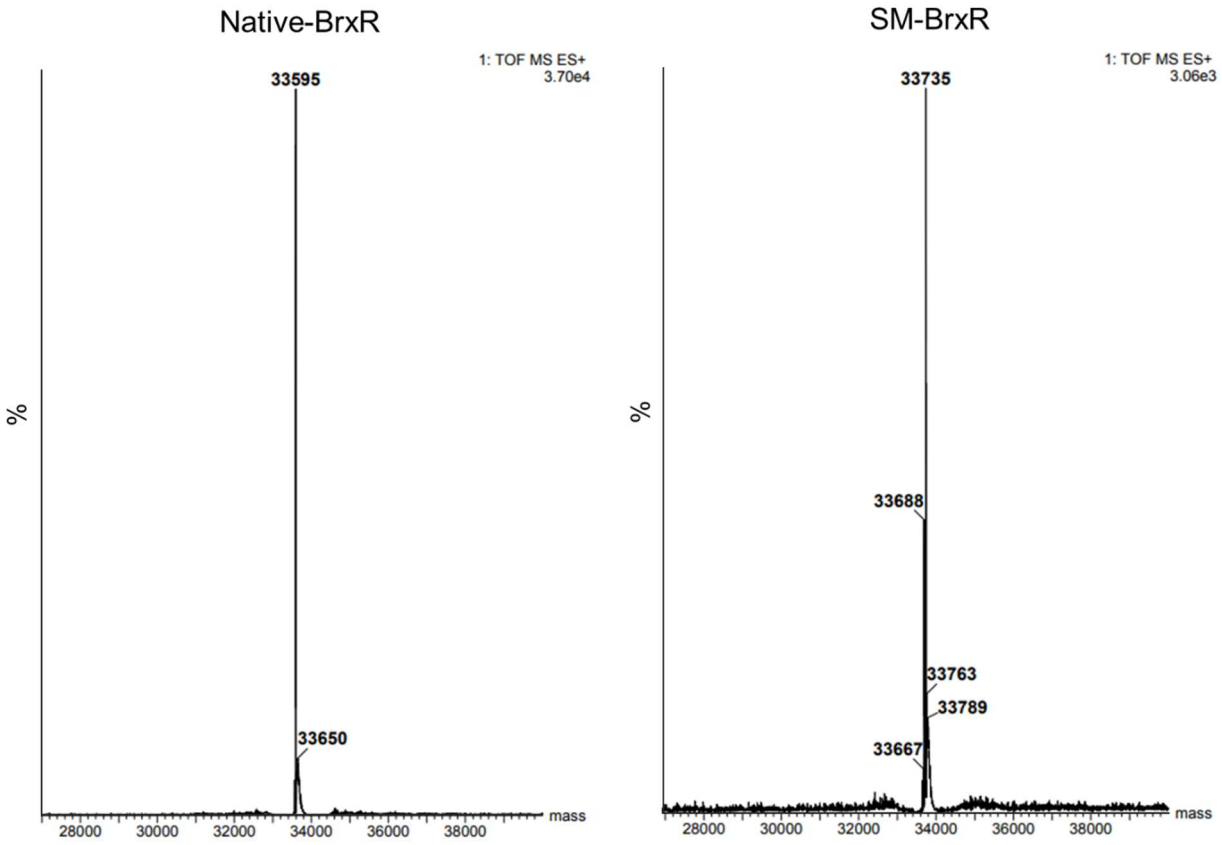


Figure 5.3: Mass spectrometry for native and selenomethionine substituted BrxR. *Positive electrospray time of flight mass spectrometry (ES+ TOF MS) shows a native mass increase of 140 for SM-BrxR (right) compared to native-BrxR (left), closely corresponding to 3 residues of SM being incorporated.*

5.4 Crystallisation trials of BrxA

BrxA crystals were found in MIDAS™ D6 forming small, plate-like crystals after 5 months (Figure 5.4). These crystals were mounted and flash frozen in liquid nitrogen, however they were damaged during transit to DLS due to a dewar malfunction and no X-ray diffraction data could be obtained. Due to the length of time required for crystal formation, and due to a solution for the NusB homologue already being deposited in the PDB, BrxA was designated as a low priority target and has yet to be further pursued .

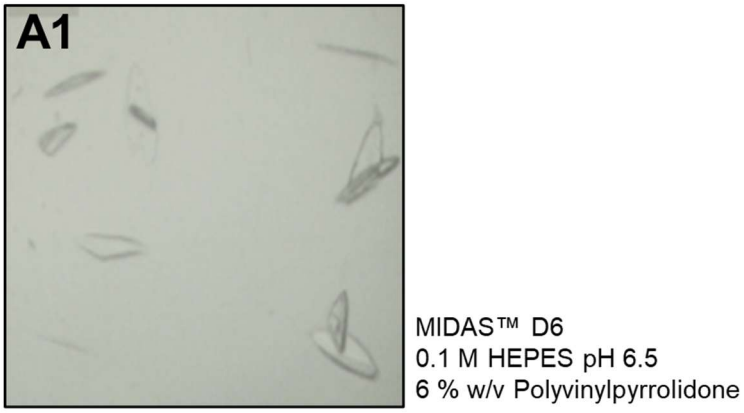


Figure 5.4: BrxA crystallisation condition. *Significant crystal growth was observed in a single condition. No X-ray diffraction was obtained for BrxA. BrxA set at 12 mg/ml.*

5.5 Crystallisation trials of PglZ

PglZ was observed to crystallise in the presence of 0.2 M KSCN, PEGs of mixed weights and at a pH of 8.5 (Figure 5.5). Crystals were observed to be thin, weak needles, typically unsuitable for mounting. Multiple replicates were performed and needle crystals from Z1-Z3 were mounted and frozen in liquid nitrogen. To assess if the crystals produced diffraction patterns prior to shipping to DLS, they were tested using a MetalJet D2 homesource beamline (Bruker), with the assistance of Dr Arnaud Basle (Newcastle University). Unfortunately, no diffraction could be detected.

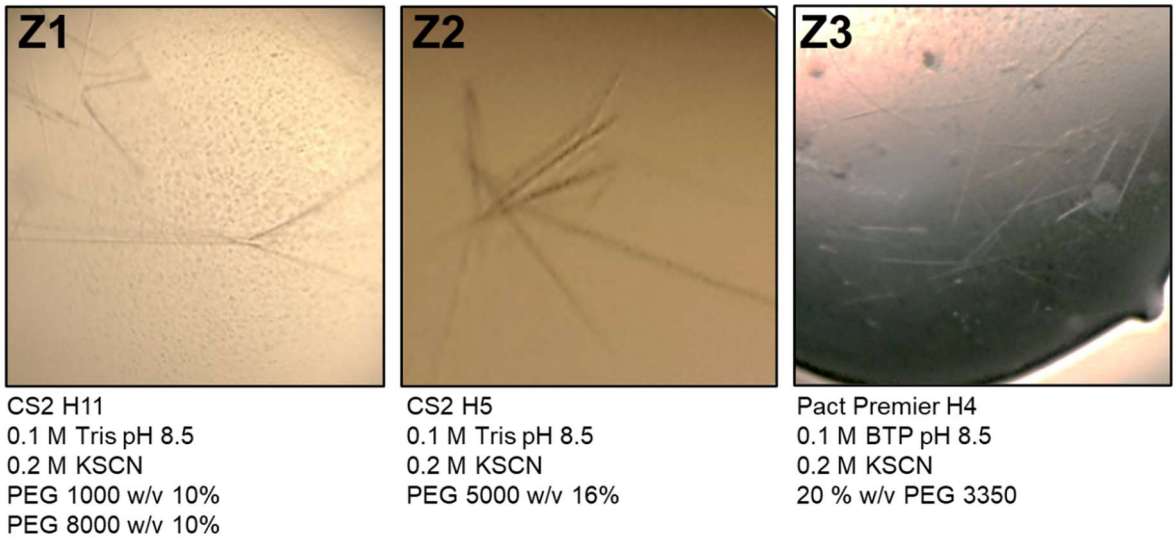


Figure 5.5: Crystallisation conditions of PglZ. *Needle crystal formation was observed in drops containing KSCN. No X-ray diffraction data was obtained for PglZ. PglZ set at 10 mg/ml.*

5.6 BrxL crystallisation and data collection

Crystal formation was observed for BrxL in Pact Premier B5, which produced small, bi-pyramidal crystals (Figure 5.6 L1). Drops were re-set manually (Figure 5.6 L2 and L3), which produced similar crystals. Higher protein: precipitant ratios caused fewer crystals to form. The crystals in L3 were mounted and frozen in liquid nitrogen before being shipped to DLS.

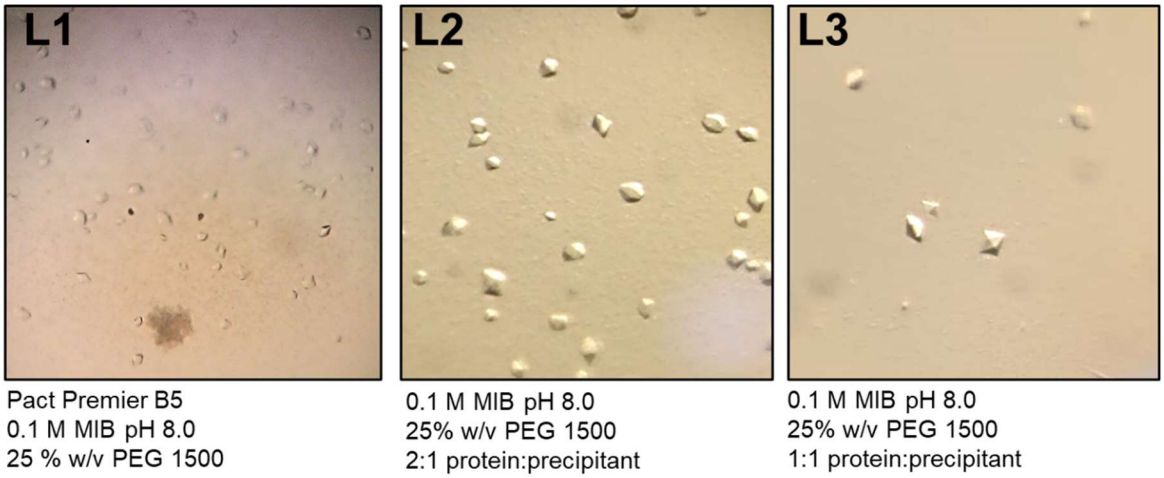


Figure 5.6: Crystallisation conditions for BrxL. *The crystallisation screening condition Pact Premier B5 was identified in L1. Manual replication of this condition is shown in L2 and 3 in larger drop sizes (3 µl and 2 µl, respectively) with varying drop ratios. BrxL set at 10 mg/ml.*

The best dataset diffracted to 3.50 Å, ruling out the possibility of salt crystals being harvested (Figure 5.7). BrxL has yet to be further optimised to improve crystal quality and no higher resolution data has been obtained. LonA has been identified as a homolog (Figure 4.1 and Table 4.1) and may be a suitable model for molecular replacement once higher resolution data is obtained for BrxL.

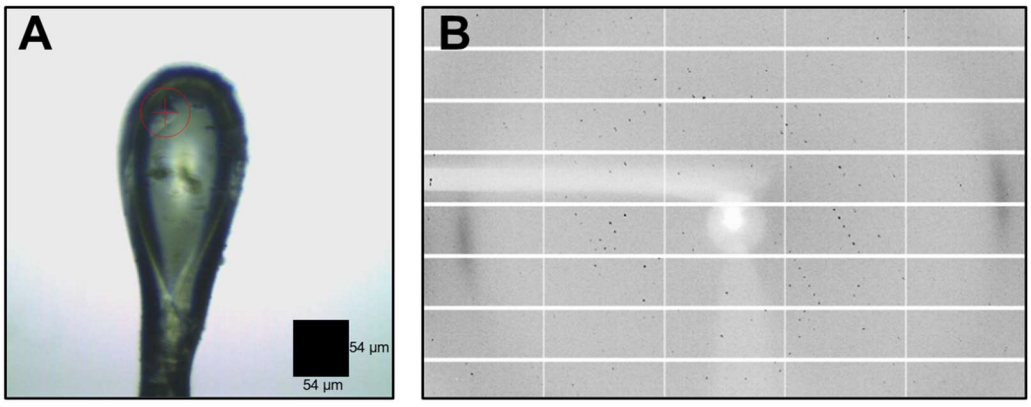


Figure 5.7: Data collection for BrxL crystals. *A) Mounted crystal from Figure 5.6. B) X-ray diffraction of BrxL extends to 3.5 Å.*

5.7 BrxU crystallisation trials and optimisation

BrxU was initially screened with no additional cofactors, producing crystals with poor morphologies. With developments in the functional characterisation of BrxU, it was hypothesised that BrxU may crystallise bound to a nucleotide ligand in the presence of a divalent cation. BrxU was rescreened with the addition of 2 mM AMP-PnP and 5 mM $MgSO_4$, yielding multiple new crystals (Figure 5.8). Crystal formation was observed in a total of 18 different screening conditions. Crystals from U5 and U6 were found to be salt crystals due to the distinct diffraction patterns observed during data collection at DLS. U2 was observed to form highly feathered crystals that were deemed unsuitable for mounting in cryoloops. U4 formed mostly long needles with fewer rod-like crystals. Significant improvements were observed during optimisation of U4 (Figure 5.8)

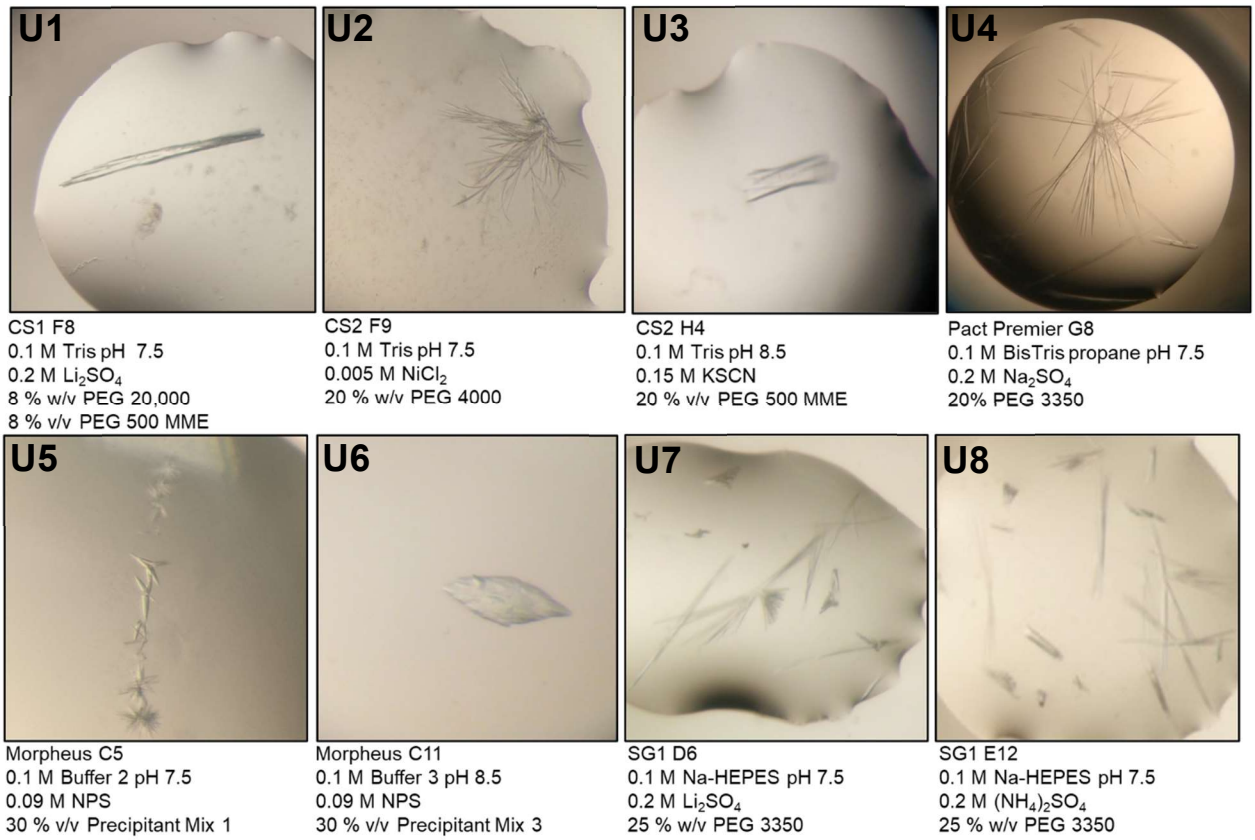


Figure 5.8: BrxU crystallisation trials with AMP-PnP. *N1-N8 show initial crystal hits from 96-well screening plates. N5-6 were noted to be salt crystals when subject to X-ray diffraction. N4 was selected for manual optimisation. BrxU set at 12 mg/ml, 2 mM AMP PnP, 5 mM $MgSO_4$.*

As $(NH_4)_2SO_4$ was found to be a common salt required for BrxU crystal formation across all 18 hits, the condition used for N4 was adjusted. Sodium sulphate was substituted for $(NH_4)_2SO_4$ and a manual optimisation screen was designed, altering pH, salt concentration and drop ratio. Drop setting at pH 8.0 resulted in weaker crystal formation with highly feathered ends making them unsuitable for mounting. Reduction in $(NH_4)_2SO_4$ to 0.15 M increased the number of crystals forming, however they were much smaller and also unsuitable for mounting. Increasing $(NH_4)_2SO_4$ concentration to 0.3 M prevented crystal formation. Crystal formation was observed at higher PEG 3350 concentrations (>24%) for these salt concentrations, however they were also deemed unmountable. In order to change multiple variables in a single drop, a matrix of drop ratios was designed. Drops were set with variable protein: precipitant ratios. Representative images of the best crystals formed at each ratio are shown below in Figure 5.9.

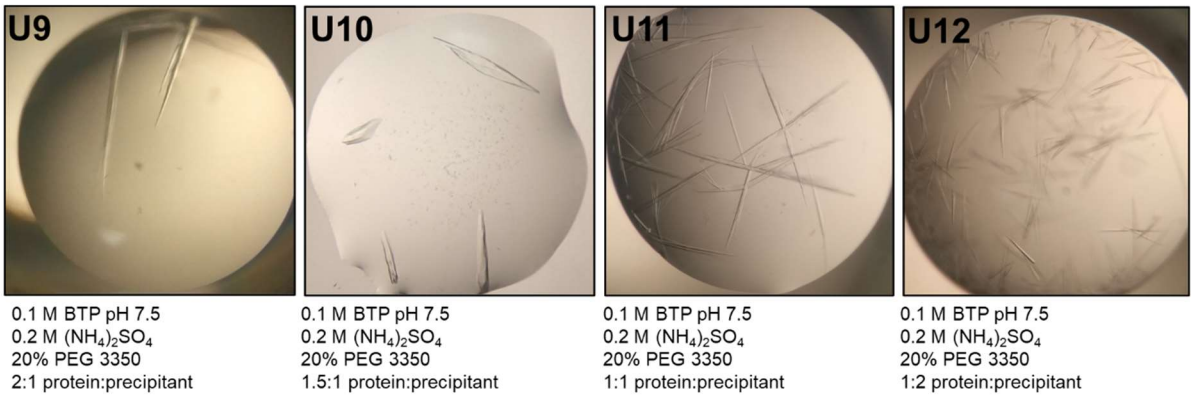


Figure 5.9: Manual optimisation of BrxU crystal formation with AMP-PnP. *Protein concentration decreases from right to left by increasing the ratio of precipitant to protein in the drop. BrxU set at 12 mg/ml, 2 mM AMP PnP, 5 mM $MgSO_4$.*

A general trend can be observed of larger, thicker, crystals appearing at greater protein: precipitant ratios (Figure 5.10). U12 contains higher ammonium sulphate and PEG 3350 concentrations, producing a large quantity of thin, needle crystals. U10 crystals appear thicker and are suitable candidates for mounting. In a number of replicates of U9 no crystals were seen, however replicates that formed were suitable candidates for mounting despite being much longer than those in U10. Previously, drops had been set using the non-hydrolysable analogue AMP-PnP. It was also hypothesised that crystals could be obtained bound to ATP in the absence of magnesium, allowing BrxU to interact with its preferred nucleotide whilst still remaining inactive and incapable of hydrolysing the γ-phosphate of ATP. Drops were set substituting AMP-PnP for ATP, but without magnesium. Unfortunately, no crystals formed. However, crystals were obtained when ATP was used with magnesium (Figure 5.10).

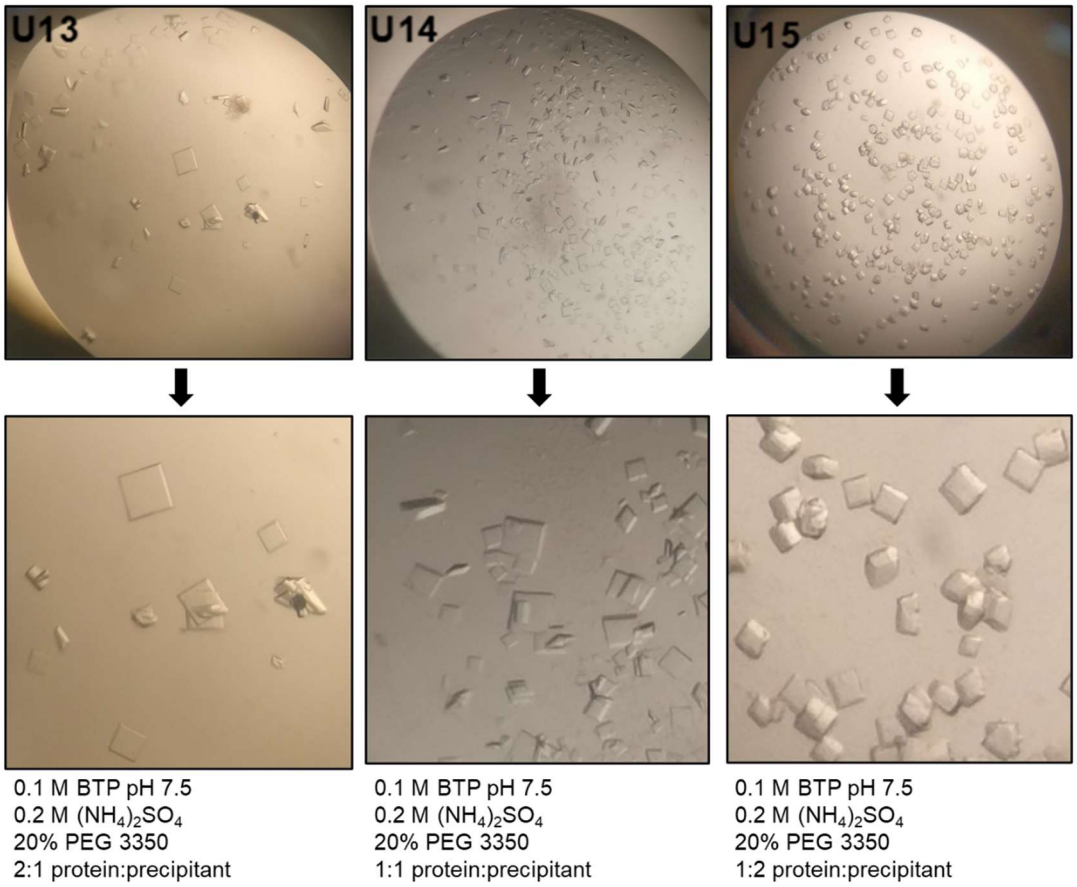


Figure 5.10: BrxU crystallisation with ATP. *Higher precipitant concentrations inducing more crystal nucleation. BrxU set at 12 mg/ml, 2 mM ATP, 5 mM $MgSO_4$.*

A similar pattern in the ATP with magnesium crystals is observed as when compared to Figure 5.9, with larger crystals forming at higher protein: precipitant ratios. Strikingly, crystals appeared in a significantly different form when set with ATP (Figure 5.10) compared to AMP-PnP (Figure 5.9), and were only observed in drops with magnesium. Crystals appear as long rods with AMP-PnP (Figure 5.9) but distinct flat squares or thicker cubic crystals with ATP (Figure 5.10). As BrxU was observed to undergo a multimeric shift when incubated with 1 mM ADP but not with 1 mM AMP, it was predicted that BrxU crystals would also form when AMP-PnP/ATP was substituted for ADP. Morphology of crystals U16-U18 (Figure 5.11) reflect the patterns observed for U13-U15 (Figure 5.10), with lower protein: precipitant ratios resulting in a greater number of less uniformly shaped, cubic crystals (Figure 5.11).

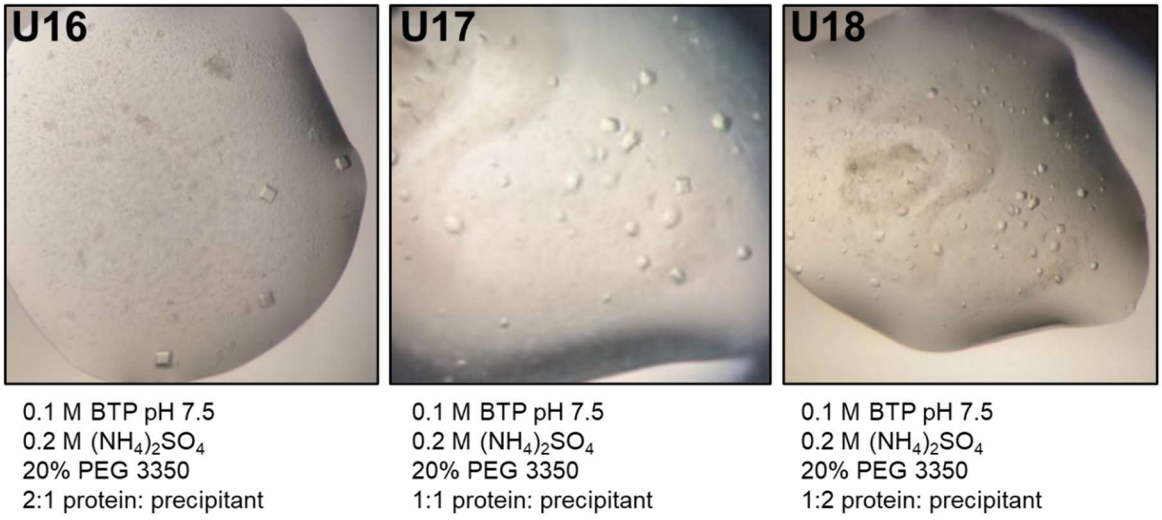


Figure 5.11: Crystallisation of BrxU with ADP. *Higher precipitant concentrations inducing more crystal nucleation. BrxU set at 12 mg/ml, 2 mM ADP, 5 mM $MgSO_4$.*

5.8 BrxU X-ray diffraction data collection

With crystals set in the absence of AMP-PnP and $MgSO_4$, the highest resolution dataset reached 5.05 Å (Figure 5.12A and B). In the presence of AMP-PnP and $MgSO_4$, the best crystal diffracted to a conservative estimate of 2.49 Å (Figure 5.12C and D). Two individual datasets were then collected on the same crystal (Figure 5.12C) and merged to create a dataset with a maximum resolution of 2.12 Å. All data collection parameters can be seen in Table 5.2. No suitable model could be obtained for structure solution via molecular replacement, so SM substituted protein was produced for BrxU.

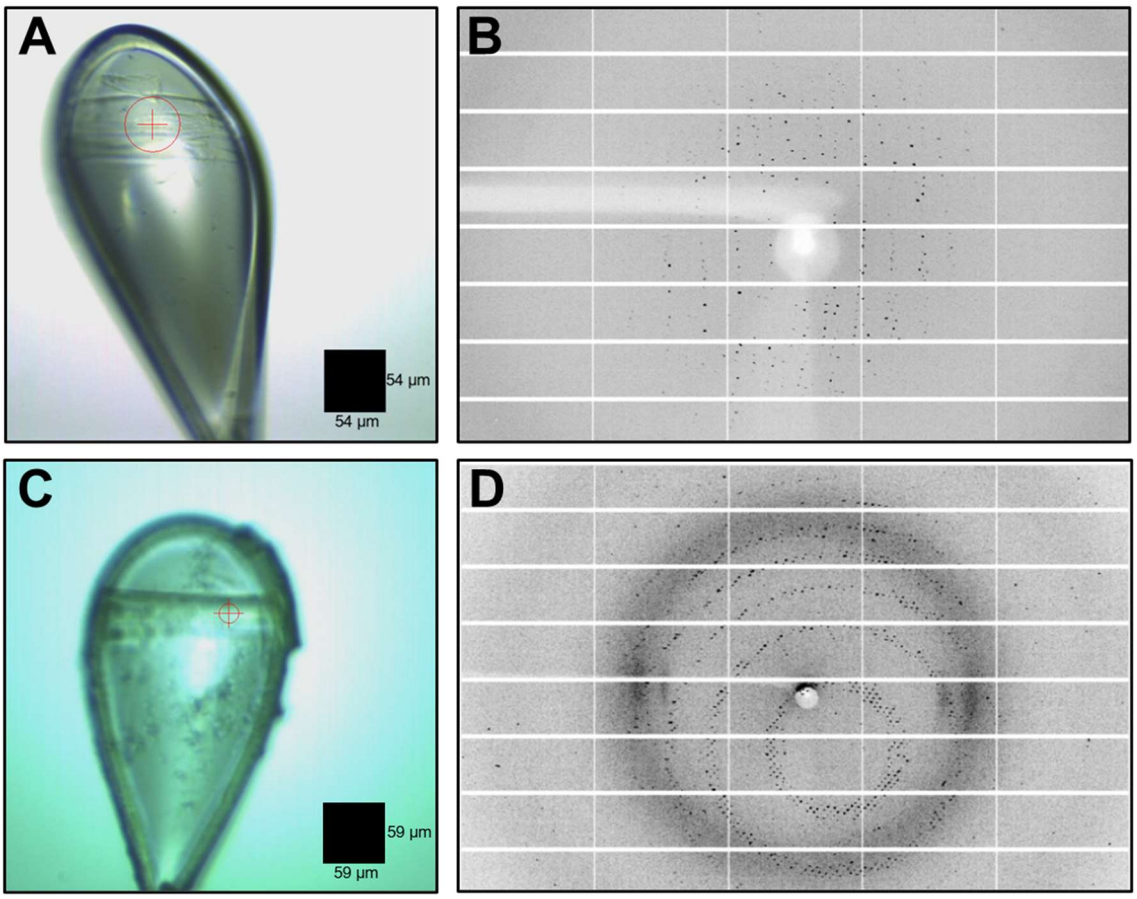


Figure 5.12: X-ray data collection of native BrxU. *A) Mounted BrxU crystal set without AMP-PnP or $MgSO_4$. B) Weak diffraction of BrxU to 5.05 Å. C) Mounted BrxU set with AMP-PnP. D) BrxU crystal set with AMP-PnP diffracting to 2.49 Å.*

5.9 Obtaining phasing data for BrxU via selenomethionine substitution

BrxU was determined as a suitable candidate for SM substitution as it encodes 13 methionine residues including the N-terminal methionine. An identical procedure was followed to section 5.3. BrxU has a mass of 67908 and ES^+ TOF MS showed native BrxU to have a mass of 67906, which is within the range of expected values. The major peak for SM-BrxU was at 68493 which corresponds to approximately 12 SM substitutions (Figure 5.13). This suggests the sample had sufficient SM incorporation for obtaining phasing data by anomalous dispersion.

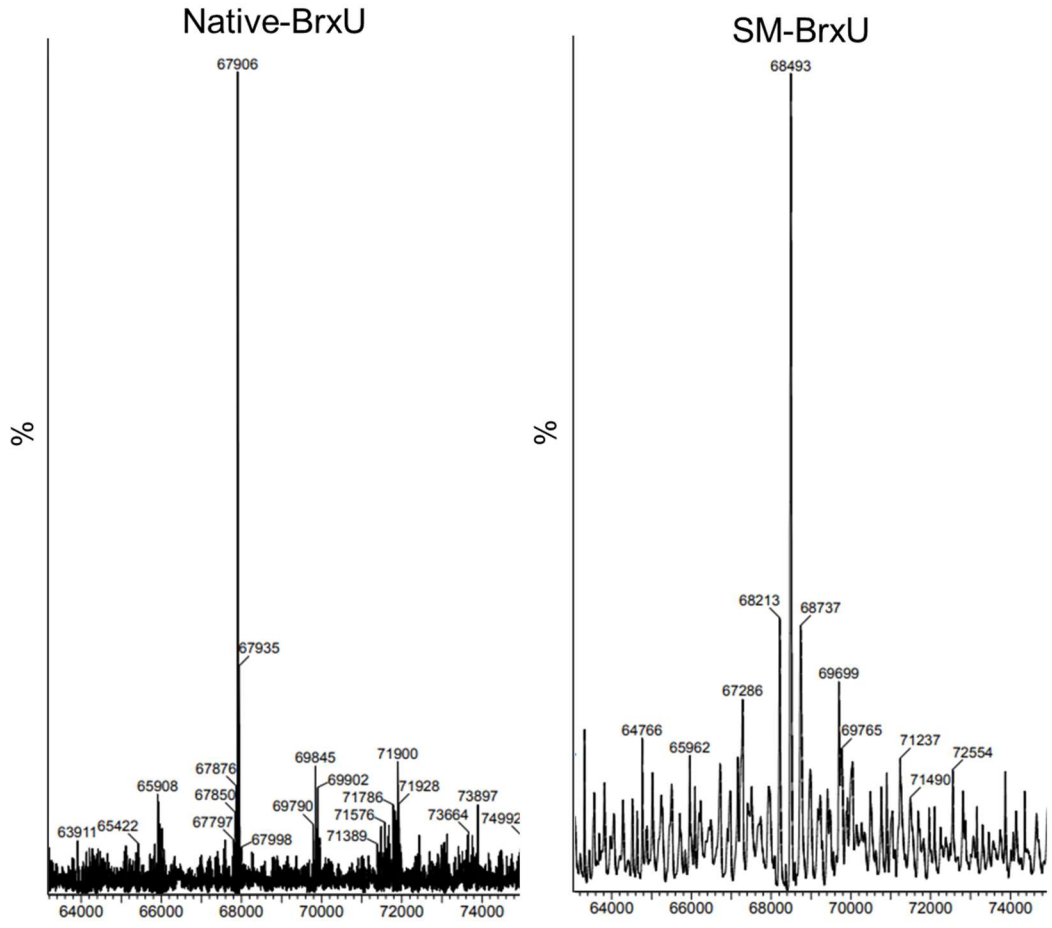


Figure 5.13: Mass spectrometry for native and selenomethionine substituted BrxU. *Positive electrospray time of flight mass spectrometry (ES+ TOF MS) shows a native mass increase of 587 for SM-BrxU (right) compared to native-BrxU (left) corresponding to approximately 12 residues of SM being incorporated.*

Due to a limited final yield for SM-BrxU (~0.6 mg), no further screening was performed, and trays were set based on the conditions used for U9-12. Crystal formation for SM-BrxU was observed to be comparable to BrxU, following a similar pattern to that shown in Figure 5.9. A strategy was designed to obtain datasets at a single wavelength for phasing via single-wavelength anomalous dispersion (SAD). The selenium K edge was identified via fluorescence scanning (Table 5.2). Due to the relatively weak anomalous signal detected, 15 datasets from across 4 individual crystals were merged into a final dataset at 2.7 Å. This dataset had sufficient anomalous signal to produce a solution using SHELX (Usón and Sheldrick 2018), and the initial output model was then built using BUCCANEER (Cowtan 2006).

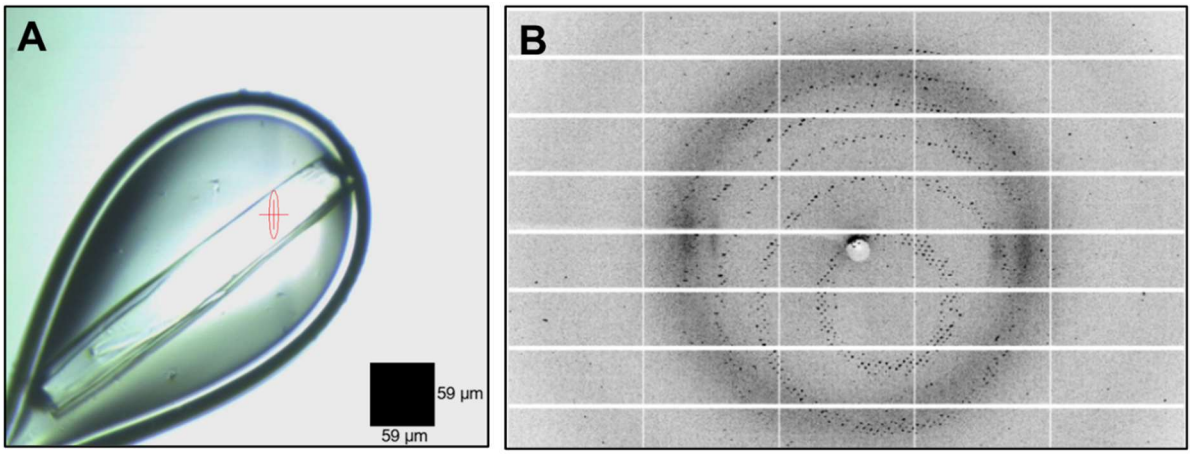


Figure 5.14: X-ray data collection of SM-BrxU. *A) Mounted SM-BrxU crystal set with AMP-PnP. B) X-ray diffraction pattern of BrxU extending to 3.06 Å. 15 datasets collected at 0.9786 Å were merged to produce a single file containing significant anomalous signal to be used in obtaining phasing data.*

Parameter		Value	
Sample		Native-BrxU	SM-BrxU
Crystallisation condition		0.1 M Tris pH 7.5 0.2 M Na2SO4 PEG 3350 w/v 20%	0.1 M Tris pH 7.5 0.2 M Na2SO4 PEG 3350 w/v 20%
Wavelength (Å)		0.9781	0.9786
Beamline		DLS I24	DLS I24
Space group		C 1 2 1	C 1 2 1
Unit cell dimensions	A	214.62	197.79
	B	67.256	68.34
	C	126.547	129.17
	α	90	90
	β	102.05	95.05
	γ	90	90
Resolution range (Å)		72.89-2.12	128.93-2.70
Total reflections		194994	4685930
Unique reflections		99627	47486
Multiplicity		2	98.7
Completeness (%)		99.4	99.5
Mean I/sigma(I)		5.4	14.5
Rmerge		0.078	0.492
Rmeas		0.111	0.495
CC1/2		0.994	0.995
Rwork		0.201	-
Rfree		0.248	-
No. of non-hydrogen atoms		9848	- -
Macromolecules		9464	-
Ligands		34	-
Solvent		350	-
Protein Residues		1160	-
RMSD (bonds, Å)		0.008	-
RMSD (angles, °)		0.9	-
Ramachandran favored (%)		96.44	- -
Ramachandran allowed (%)		3.56	- -
Ramachandran outliers (%)		0	- -
Average B-factor		55.55	-

Table 5.2: Crystallographic data for BrxU in both native and SM substituted form.

5.10 Protomeric structure of BrxU

A protomer model generated for SM-BrxU was used as a molecular replacement search model to solve the native-BrxU dataset at 2.12 Å (Table 5.2). The initial PHASER molecular replacement output model was updated with BUCCANEER and then iteratively improved using COOT and PHENIX.REFINE (Table 5.2). The final model contained two copies of BrxU in the asymmetric unit. Figure 5.15 shows a cartoon representation of a BrxU protomer in three orientations.

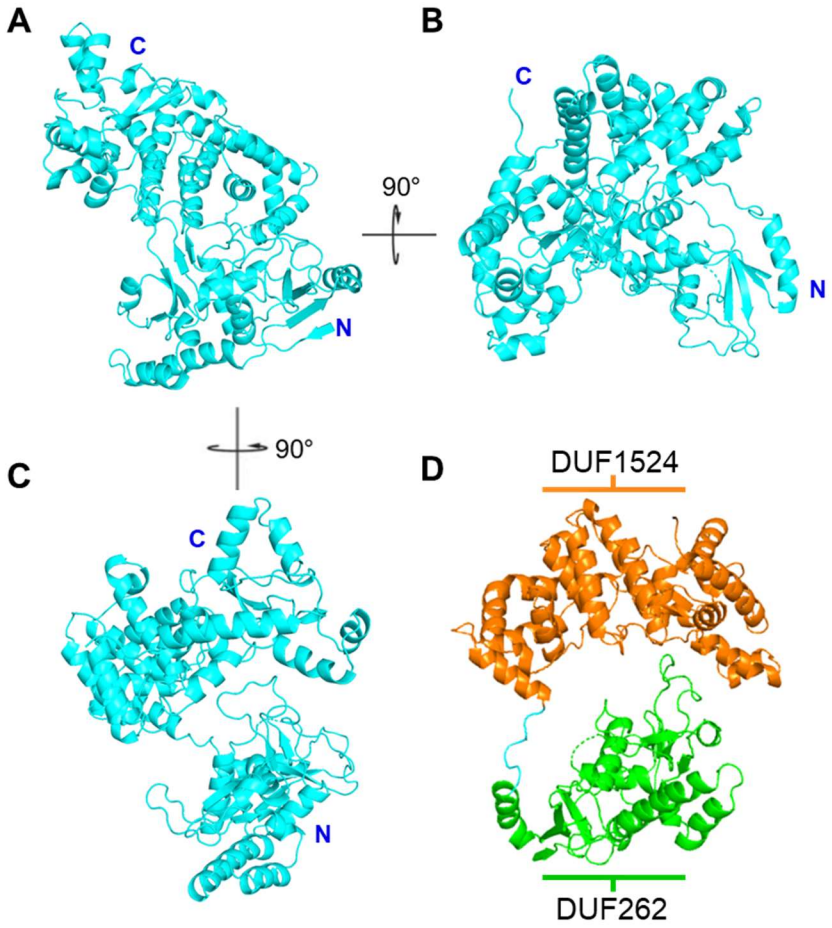


Figure 5.15: Protomeric structure of BrxU shown as cartoon. *A-C) Views of single BrxU subunits indicated by the rotation shown between each image. D) Domain architecture of BrxU is shown with N-terminal DUF262 (green) and C-terminal DUF1524 (orange) joined by a short linker sequence (cyan) forming a flexible loop region. D has been orientated to more clearly show the linker sequence.*

In panel D, BrxU has been orientated to show the distinct domain architecture of BrxU protomers consisting of two separate domains bridged by a short linker sequence. The N-terminal DUF262 domain consists of 8 alpha helices and 8 beta sheets, from residues M1 to M238. Within the DUF262 domain, residues Q25, R26 and E27 have not been built due to poor density. These residues form the most disordered section of an extended loop region extending from Q18 to R32. A short linker sequence extends from M238 to S246, connecting the DUF262 and DUF1524 domains (Figure 5.15D). The DUF1524 domain spans from Y247 to E587. It consists of mainly alpha helical regions with 2 short antiparallel beta sheets forming part of the DHIYP proposed nuclease site.

5.11 BrxU complex and dimeric interface

Two BrxU copies were observed in the asymmetric unit, showing each copy interlocking the other via the short linker sequence between each domain. Additional electron density surrounding each BrxU copy was attributed to crystallographic symmetry, and with the visualisation of the interlocking linker sequence, the biological assembly of BrxU has been determined to be dimeric (Figure 5.16). Rotation of the assembly reveals a channel of approximately the correct dimensions to indicate a double-stranded DNA binding cleft. Due to presence of Mg^{2+} and AMP-PnP in the crystallisation condition, and given that crystals do not form in their absence in this condition, it was expected that BrxU would be monomeric in this structure. However, metal ions and nucleotides could not be accounted for by unmodelled electron density during model building, and it appears as a dimer.

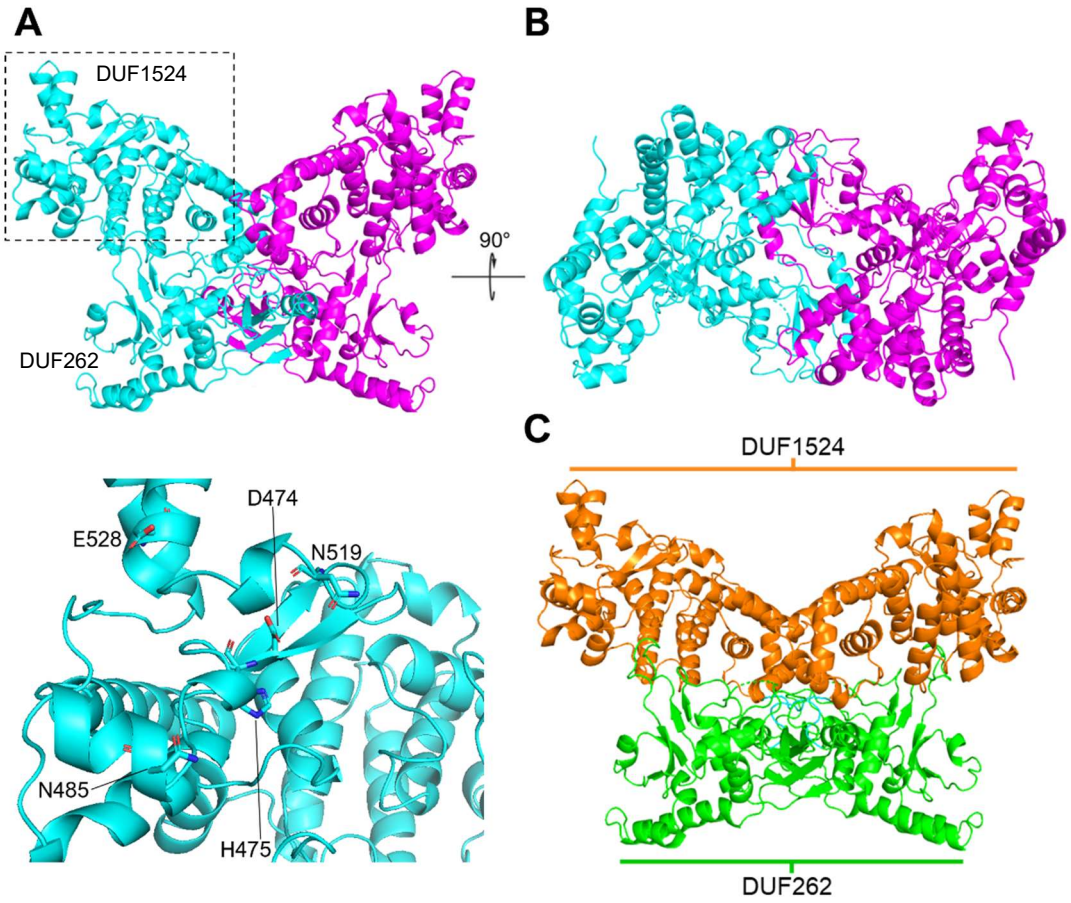


Figure 5.16: Dimeric BrxU shown as a cartoon. *A) Cartoon overview of the BrxU dimer Each protomer is coloured either cyan or magenta. The DHIYP motif is indicated by the box region, and a focused view of this region is shown with key residues. B) A 90° rotation of A is shown. C) Domain architecture of BrxU dimer shows assembly of DUF1524 (orange) and DUF262 domains (green) with linker region shown in cyan.*

5.12 Ligand identification

As mentioned in the previous section, BrxU was predicted to be in a monomeric form within these crystals due to the presence of Mg^{2+} and AMP-PnP. Size exclusion chromatography data showed a distinct shift from a calculated dimer mass to a monomer upon binding nucleotide cofactors (Figure 4.18), although AMP-PnP was not specifically tested by size exclusion. It was briefly proposed that these crystals may have formed irrespective of the Mg^{2+} and AMP-PnP in the drop, however control drops were found not to crystallise in the absence of either of these cofactors. Unmodelled electron density blobs observed during model building and structural refinement were identified and accounted for. Each subunit was found to bind 2 sulphate ions, a single chloride ion and a single molecule of glycerol. Ligand interactions are shown in Figure 5.17 for one subunit and are reflected in the biological assembly in the other subunit. Sulphate is present due to the 0.2 M $(NH_4)_2SO_4$ in the crystallisation condition, chloride is present due to the 0.15 M NaCl in the protein storage buffer. Glycerol is used in the purification process as a solubilising agent during chromatography and small quantities may be present following dialysis. The first sulphate ion can be seen modelled to bind within the DGQQR domain (Figure 5.17A). AMP-PnP could not account for the unmodelled density as it was significantly too large to fit. It is possible that the high concentration of sulphate in the crystallisation condition prevents stable binding of AMP-PnP. This would be accounted for by the strong ionic interactions with R102 and polar interactions with both Q100 and Q101. However, it is more likely that nucleotide binding is prevented due to the steric hindrance caused by protomer B (Figure 5.17A). The second sulphate ion is coordinated by the side chains of S392 and S397, likely due to hydrogen bonding between the oxygen atoms of sulphate and the hydrogens of serine (Figure 5.17B). The chloride ion is situated between S397 and N400, likely interacting with hydrogen in the asparagine side chain. Glycerol was modelled to bridge between the DUF262 and DUF1524 domains (Figure 5.17B).

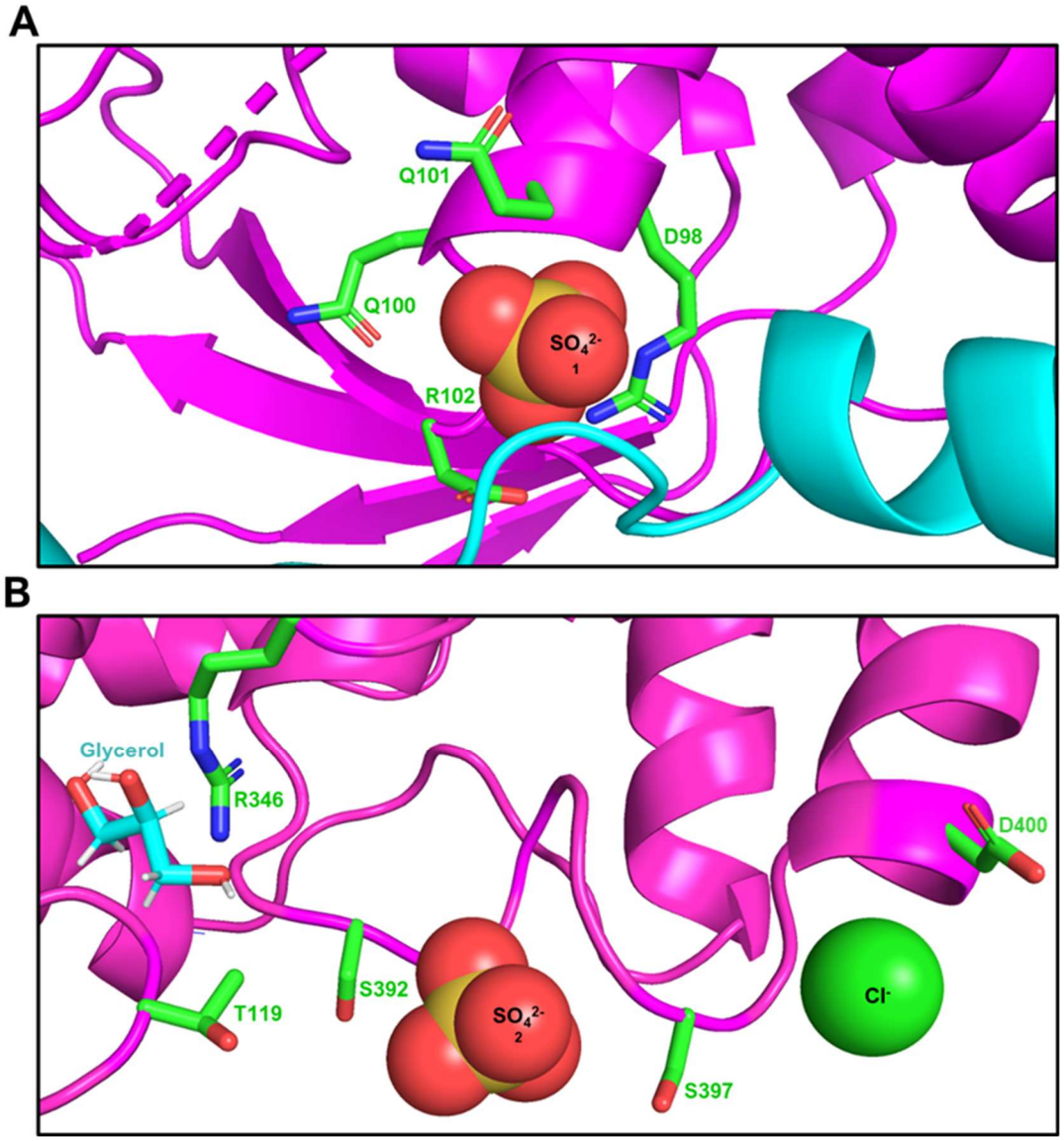


Figure 5.17: Ligand identification with key residues. *Protomer A shown in magenta, protomer B shown in cyan. Key residues are highlighted in green and side chains are shown as sticks. Oxygen shown as red, nitrogen as blue. A) A single sulphate ion is tightly coordinated within the DGQQR motif of DUF262. B) Sulphate, chloride and glycerol can be seen in the pocket formed between DUF262 and DUF1524.*

5.13 Linker domain and its role in dimerisation

As previously mentioned, there are two units in the biological assembly of this BrxU structure. The dimer interface is stabilised by the linker region of each subunit crossing over the other (Figure 5.18). This linker region separates the DUF262 and DUF1524 domains, allowing each domain to interact with its respective counterpart in the other unit. Flexibility in this linker would allow BrxU dimers to dissociate, as observed by size exclusion chromatography (Figure 4.18). The linker domains interlock the two subunits, with a loop from each N-terminal domain connecting with the corresponding C-terminal domain. This interaction is bridged by a glycerol molecule in Figure 5.17. In the orientation shown in Figure 5.18A, the linker region and C-terminal domain of chain A (magenta) is topologically in front of chain B. Note that only residues D151-S301 are shown in order to present the linker region clearly. The N-terminal domain of chain A can then be seen behind chain B. A wider perspective of the routing of each chain can be seen in Figure 5.18C. The interlocking of BrxU subunits suggests that in order for monomerization to occur, a conformational change must happen, allowing the domains of each subunit to move apart. This is likely driven by the binding of a nucleotide in the DGQQR motif. In Figure 5.18D, the electron density of the linker regions is shown to demonstrate the distinct routing of each protomer.

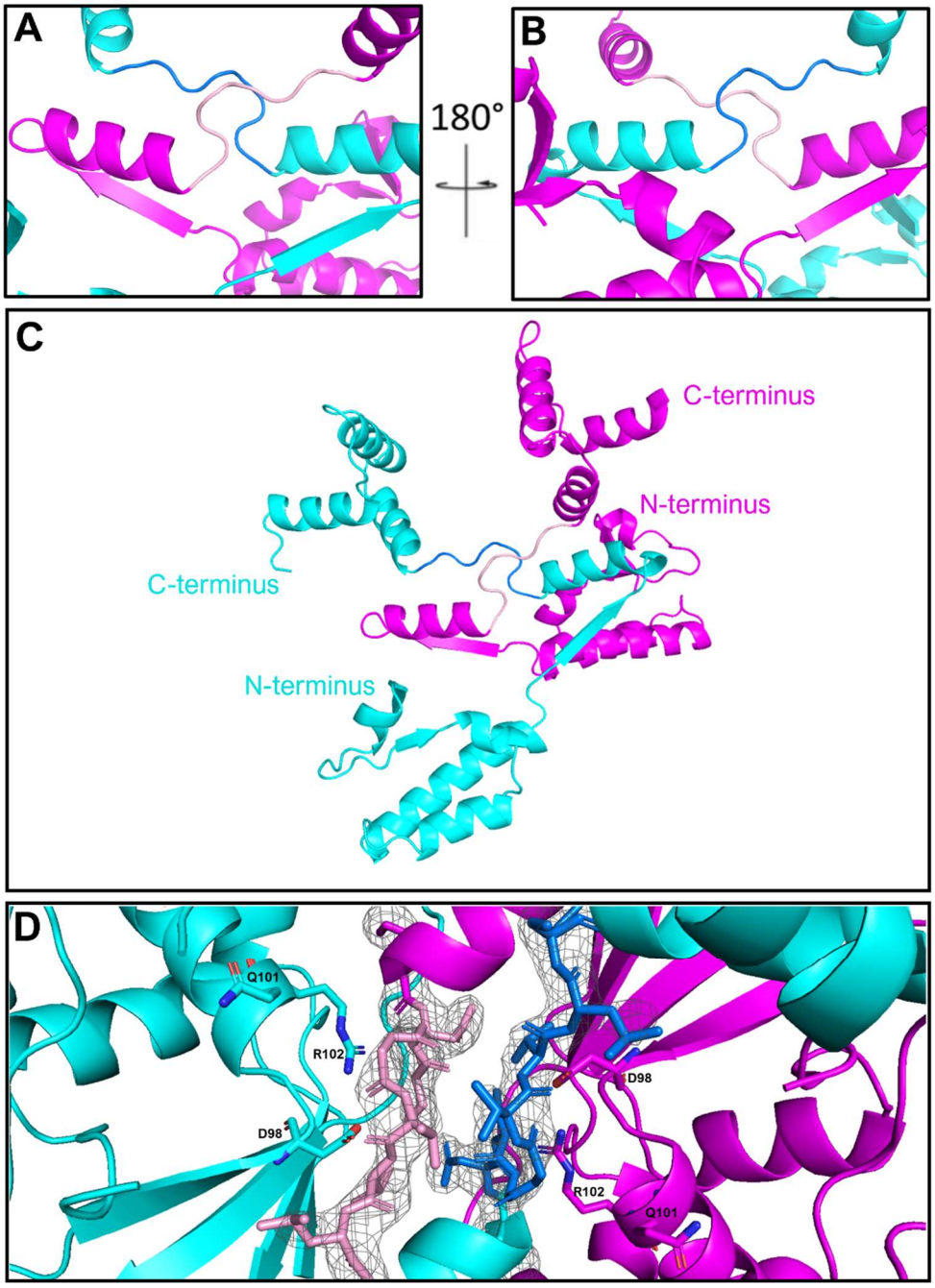


Figure 5.18: Cartoon of dimeric BrxU showing interlocked linker regions. *A and B) Views of the crosslinking of linker regions of each BrxU subunit. Protomer A (magenta) can be seen passing over and under protomer B (cyan). C) Wider perspective of interlocking linker regions. Perspective does not relate to panels A and B. Residues 151-301 are shown. D) Close up of the proposed nucleotide binding site and linker regions. A Fo-Fc map has been contoured to 2 σ to demonstrate the intertwining of the two protomers. Selected residues of the DGQQR motif are shown as sticks.*

5.14 Surface charge representation of BrxU

In order to visualise the DNA binding cleft, surface charges were calculated via APBS Electrostatics in PyMOL (Figure 5.19A-C). Mapping of the electrostatic surface potential of BrxU reveals a distribution of positive charges within the proposed DNA binding cleft that suggests a suitable binding surface for double phosphodiester backbones. In Figure 5.19A and B, the core of the BrxU can be seen to exhibit significant positive charges. In the conformation shown, the proposed binding cleft is 21 Å which is too narrow to allow DNA to dock. However, upon dissociation of the dimer, BrxU monomers would be able to bind DNA and then reassociate. In Figure 5.19C, the structure has been positioned to show how open pores form at each end of the complex, with highly positive electrostatic potential. Each pore is formed by the DUF262 and DUF1524 domains coming together, supported by an extended loop and a bridging glycerol at the tip of the loop. For reference, a cartoon representation of BrxU in the same perspective is shown beside it, showing the loop that closes the pore (Figure 5.19C). Due to the distribution of complementary charges, two models could be considered, wherein the dsDNA might thread through these pores, should they be able to open and close, or bind across the surface of BrxU, aligning past, but not through, the pore.

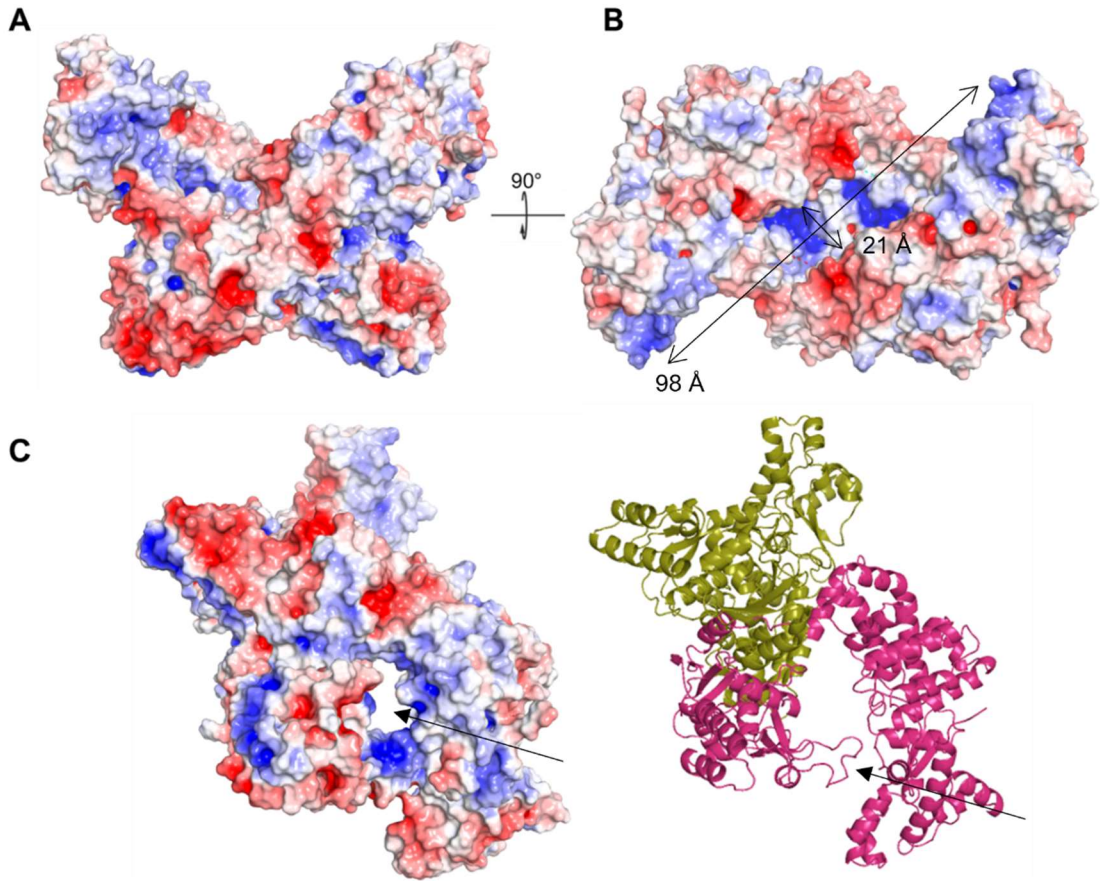


Figure 5.19: Surface charge representation of BrxU dimers. *Positive potential is shown in blue, negative potential is shown in red. Colour saturation corresponds to charge intensity. A-B) Two stereo views of BrxU showing the positive charge potential in the proposed DNA binding cleft. Dimension of the cleft are shown in B to demonstrate that in the current conformation, dsDNA cannot be docked. C) (Left) BrxU has been orientated to show the pore (arrow). (Right) Cartoon shown beside for reference, with the extended loop bridging the two domains (arrow).*

5.15 Key residue identification within BrxU

As detailed in Chapter 4, 12 BrxU mutants were generated based on alignments with GmrSD (Figure 4.11). Solving the structure of BrxU has allowed for close observation of these residues and identification of new residues as targets for mutagenesis studies. A key interaction between the protomers is shown below (Figure 5.20). In Figure 5.20A, a sulphate ion can be seen coordinated by the DGQQR motif. It was predicted that the DGQQR would bind the phosphates of nucleotide substrates (Machnicka et al. 2015). The linker region of protomer B extends directly past the DGQQR motif. In this conformation, this linker region causes steric hindrance to nucleotide binding and provides an explanation for the missing AMP-PnP. In close proximity to the DGQQR motif is an α-helix consisting of residues P33 to Q45. Closer observation of S42 (Figure 5.20), reveals an interaction with I234 of protomer B. The oxygen of S42 is shown 3.5 Å from I234 of protomer B. This interaction may be stabilising the proximity of these α-helices, closing the BrxU in a conformation that blocks the DGQQR from binding nucleotides. Mutants have been generated in PyMOL to show the change observed with a single residue mutation. The S42A mutant, which was shown to be permanently monomeric by size exclusion chromatography (Figure 4.26D) cannot form this interaction due to the increased distance to I234. The S42D mutant, which was shown to be permanently dimeric by size exclusion chromatography (Figure 4.26E), shows the oxygen of aspartate within 2.7 Å of I234, an interaction that is unfavourable and unlikely to occur (Figure 5.20D). This suggests that this region is destabilised in S42D and may cause perturbations in nucleotide binding and prevent BrxU dimers for dissociating. S42 is likely involved in the binding of the base of NTPs, and disturbing this interaction affects the dissociation of dimeric BrxU.

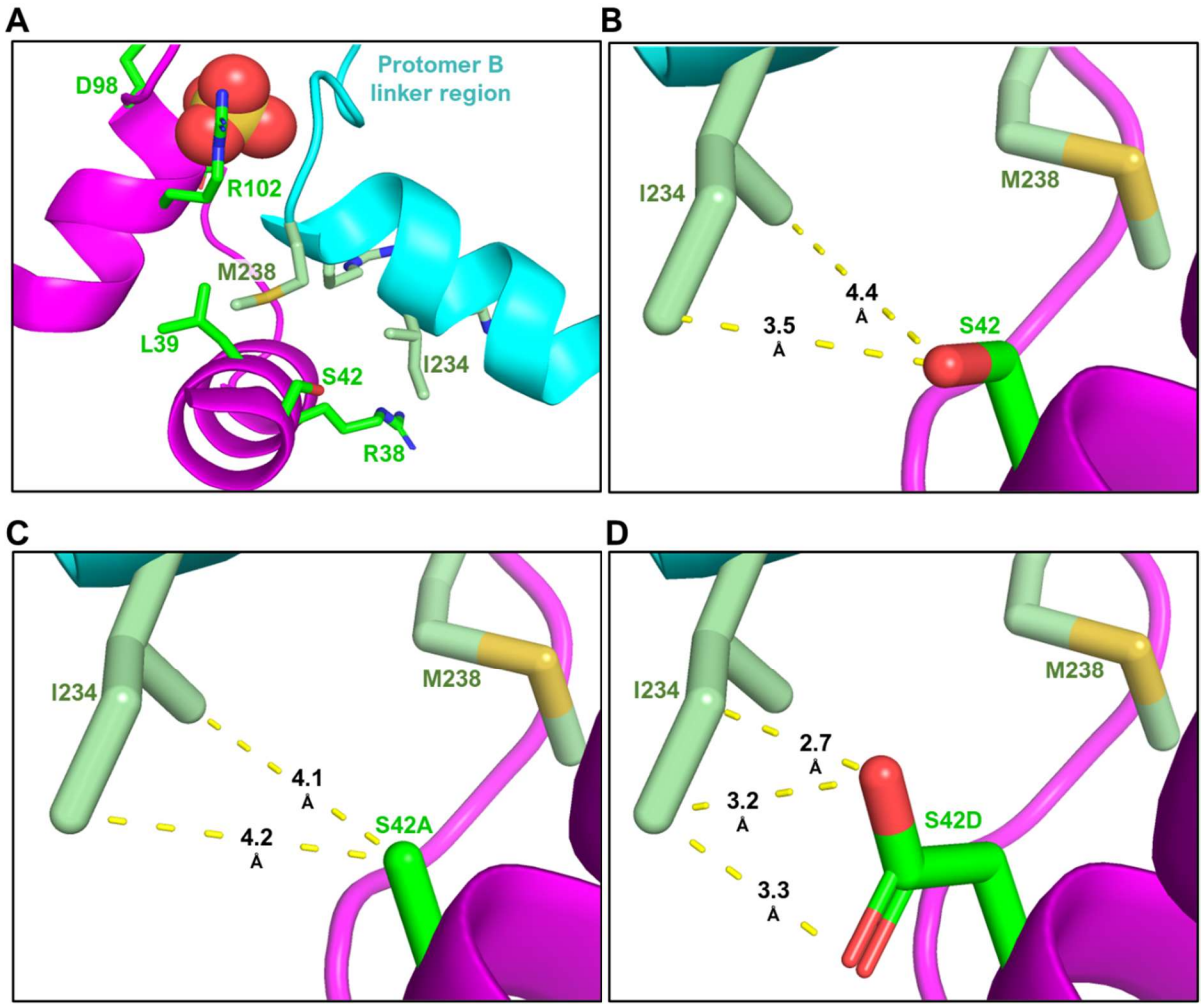


Figure 5.20: Interactions between BrxU protomers. *Protomer A in magenta, protomer B in cyan. Key side chains shown as green, oxygen as red, nitrogen as blue, sulphur as yellow. A) N-terminal DUF262 domain of protomer interaction with linker region of protomer B. Linker region of protomer B is blocking the DGQQR domain of protomer A from binding AMP-PnP. Sulphate appears to take the putative position of the gamma phosphate. B) Weak interactions shown between S42 of protomer A and I234 of protomer B. C and D) Residue mutagenesis for S42 to alanine (C) and aspartate (D). Weaker interactions are observed for S42A. S42D is shown in an unfavourable position due to the close proximity of its oxygens to I234.*

Within the C-terminal domain, key residues around the DHIYP nuclease motif are shown in Figure 5.21. Panels A and B show the same region in two orientations in order to show the architecture of the DHIYP motif as well as highlight the distance of N485 from the rest of the key residues (Figure 5.21). N485 and N519 are both located on the exterior of the protein, with sidechains protruding away from the nuclease motif. N505 and N509 have been identified as potentially key residues that form the predicted HNH nuclease motif within BrxU. In the HNH nuclease domain of *N. meningitidis* Cas9, the equivalent residue N509 forms a stack within the nuclease domain of BrxU (S. Taleb, unpublished MBiol thesis) (Harrington et al. 2017). N518 is shown stacking with D474, H475 and N509 (Figure 5.21B).

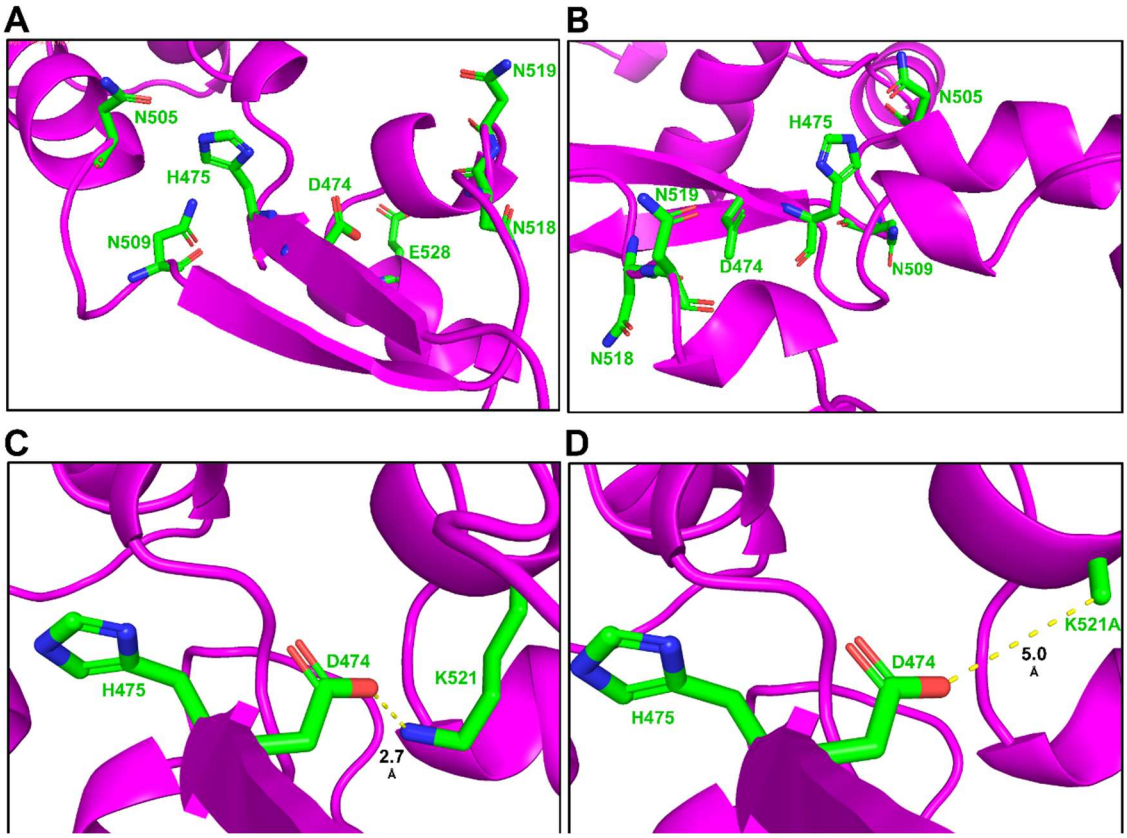


Figure 5.21: C-terminal DUF1524 key residue identification. *A and B) Views of C-terminal DHIYP domain and putative associated residues. Key side chains shown as green, oxygen as red, nitrogen as blue. C and D) Interaction of K521 with D474. C) D474 forms hydrogen bonds with K521. D) Mutagenesis of K521 to alanine disturbs this interaction, identifying a new key residue for mutational studies.*

Additionally, it appears that D474 is being coordinated by K521, forming hydrogen bonds. The oxygen of D474 is noted to be 2.7 Å from the nitrogen of K521 (Figure 5.21C). Mutation of this D474 to alanine would disrupt this interaction and may prevent D474 from functioning properly (Figure 5.21D). No direct interaction could be observed for E528, however its side chain can be seen oriented towards the DHIYP motif and therefore may be involved in hydrolysis or bonding with key residues. It was originally predicted that magnesium ions would be located near D474, however they could not be accounted for within this region. It is currently unclear as to why metal ions are absent despite being required for nuclease activity.

5.16 Alignment of BrxU with SspE

Phyre2 homology searches identified SspE (PDB: 6JIV) as a close structural homologue (Xiong et al. 2020). A sequence-structure alignment was produced via the PROMALS3D server (Pei, Kim, and Grishin 2008). The pre-deposit structure obtained for BrxU was aligned with 6JIV. The DHIYP motif of BrxU could not be aligned due to a lack of structural information in this region of 6JIV (data not shown). A sequence alignment of SspE and BrxU was produced (Figure 5.22). This sequence alignment allows for comparison within the two domains of key motifs. Within the DUF262 domain, the DGQQR domains can be seen as highly conserved between the two proteins. Immediately preceding this motif is a β-sheet of 5 hydrophobic residues. The first two positions before D98 are conserved as aliphatic, hydrophobic residues. A key observation in this alignment is the presence of the linker region. This linker region can be identified in Figure 5.23 which shows the superpositioning of BrxU over SspE. This alignment predicts N239 and S240 of BrxU to be part of an α-helix, however these residues form part of the flexible linker region. The DHIYP motif of BrxU aligns with the EHVAP motif of SspE, reflecting the consensus D/EHxxP of DUF1524 proteins.

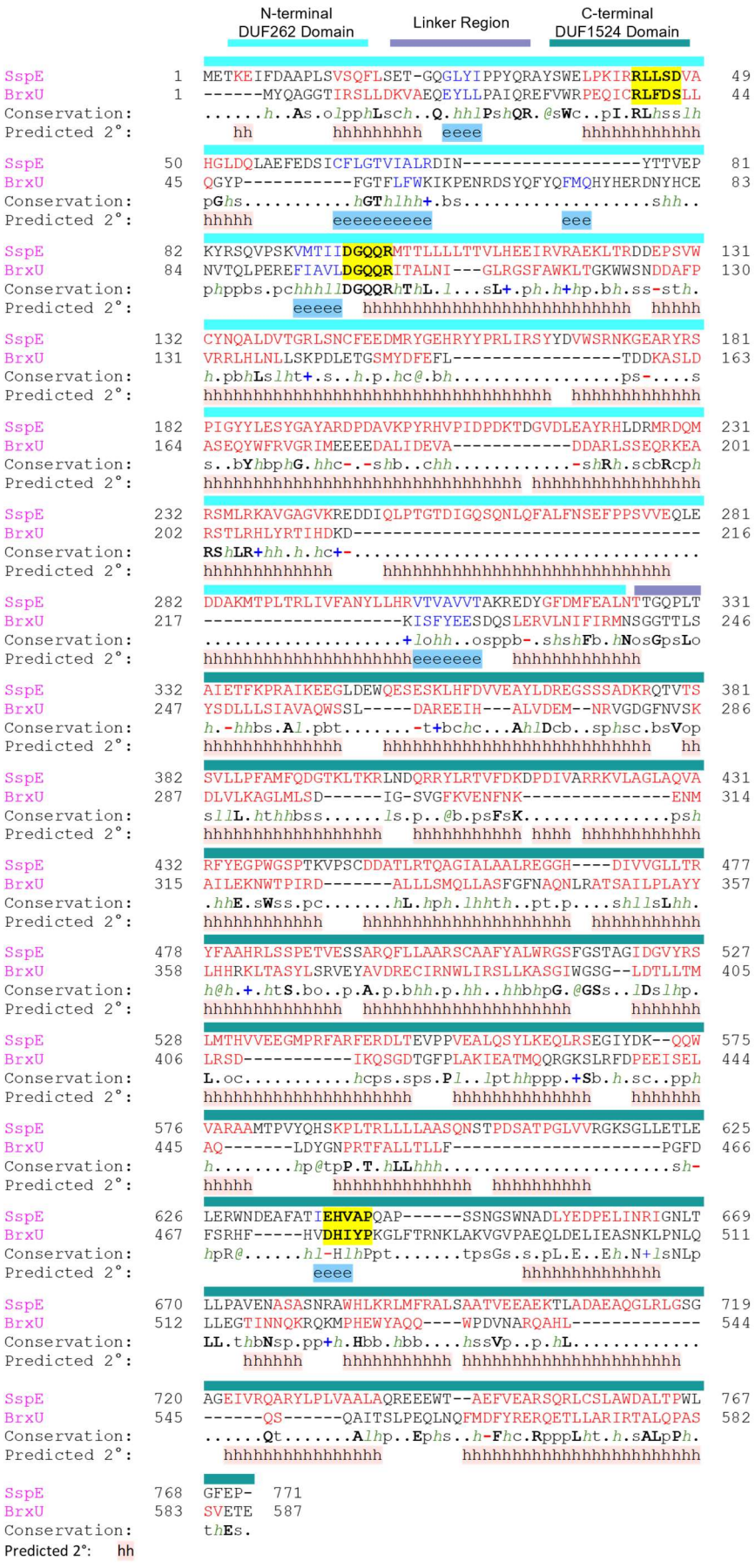

N-terminal
DUF262 Domain
Linker Region
C-terminal
DUF1524 Domain
SspE 1 METKEIFDAAPLSVSQFLSET-GQGLYIPPYQRAYSWELPKIRRLLSDVA 49
BrxU 1 ------MYQAGGTIRSLLDKVAEQEYLLPAIQREFVWRPEQICRLFDSLL 44
Conservation:
Predicted 2°:
SspE 50 HGLDQLAEFEDSICFLGTVIALRDIN----------YTTVEP 81
BrxU 45 QGYP----------FGTFLFWKIKPENRDSYQFYQFMQHYHERDNYHCE 83
Conservation:
Predicted 2°:
SspE 82 KYRSQVPSKVMTIIDGQQRMTTLLLLTTVLHEEIRVRAEKLTRDDEPSVW 131
BrxU 84 NVTQLPEREFIAVLDGQQRITALNI---GLRGSFAWKLTGKWWSNDDAFP 130
Conservation:
Predicted 2°:
SspE 132 CYNQALDVTGRLSNCFEEDMRYGEHRYYPRLIRSYYDVWSRNKGEARYRS 181
BrxU 131 VRRLHLNLLSKPDLETGSMYDFEFL----------TDDKASLD 163
Conservation:
Predicted 2°:
SspE 182 PIGYYLESYGAYARDPDAVKPYRHVPIDPDKTDGVDLEAYRHLDRMRDQM 231
BrxU 164 ASEQYWFRVGRIMEEEEDALIDEVA----------DDARLSSEQRKEA 201
Conservation:
Predicted 2°:
SspE 232 RSMLRKAVGAGVKREDDIQLPTGTDIGQSQNLQFALFNSEFPPSVVEQLE 281
BrxU 202 RSTLRHLYRTIHDKD---------- 216
Conservation:
Predicted 2°:
SspE 282 DDAKMTPLTRLIVFANYLLHRVTVAVVTAKREDYGFDMFEALNTTGQPLT 331
BrxU 217 ----------KISFYEESDQSLERVLNIFIRMNSGGTTLS 246
Conservation:
Predicted 2°:
SspE 332 AIETFKPRAIKEEGLDEWQESESKLHFDVVEAYLDREGSSSADKRQTVTS 381
BrxU 247 YSDLLLSIAVAQWSSL-----DAREEIH---ALVDEM--NRVGDGFNVSK 286
Conservation:
Predicted 2°:
SspE 382 SVLLPFAMFQDGTKLTKRLNDQRRYLRTVFDKDPDIVARRKVLAGLAQVA 431
BrxU 287 DLVLKAGLMLSD------IG-SVGFKVENFNK----------ENM 314
Conservation:
Predicted 2°:
SspE 432 RFYEGPWGSPTKVPSCDDATLRTQAGIALAALREGGH----DIVVGLLTR 477
BrxU 315 AILEKNWTPIRD-------ALLLSMQLLASFGFNAQNLRATSAILPLAYY 357
Conservation:
Predicted 2°:
SspE 478 YFAAHRLSSPETVESSARQFLLAARSCAAFYALWRGSFGSTAGIDGVYRS 527
BrxU 358 LHHRKLTASYLSRVEYAVDRECIRNWLIRSLLKASGIWGSG--LDTLLTM 405
Conservation:
Predicted 2°:
SspE 528 LMTHVVEEGMPRFARFERDLTEVPPVEALQSYLKEQLRSEGIYDK--QQW 575
BrxU 406 LRSD----------IKQSGDTGFPLAKIEATMQQRGKSLRFDPEEISEL 444
Conservation:
Predicted 2°:
SspE 576 VARAAMTPVYQHSKPLTRLLLLAASQNSTPDSATPGLVVRGKSGLLETLE 625
BrxU 445 AQ-------LDYGNPRTFALLTLLF---------PGFD 466
Conservation:
Predicted 2°:
SspE 626 LERWNDEAFATIEHVAPQAP------SSNGSWNADLYEDPELINRIGNLT 669
BrxU 467 FSRHF-----HVDHIYPKGLFTRNKLAKVGVPAEQLDELIEASNKLPNLQ 511
Conservation:
Predicted 2°:
SspE 670 LLPAVENASASNRAWHLKRLMFRALSAATVEEAEKTLADAEAQGLRLGSG 719
BrxU 512 LLEGTINNQKRQKMPHEWYAQQ----WPDVNARQAHL---------- 544
Conservation:
Predicted 2°:
SspE 720 AGEIVRQARYLPLVAALAQREEEWT--AEFVEARSQRLCSLAWDALTPWL 767
BrxU 545 ------QS------QAITSLPEQLNQFMDFYRERQETLLARIRTALQPAS 582
Conservation:
Predicted 2°:
SspE 768 GFEP- 771
BrxU 583 SVETE 587
Conservation:
Predicted 2°: hh

Figure 5.22: Sequence-structure alignment of SspE and BrxU. *SspE and BrxU were aligned using PROMALS3D. Predicted domain architectures are indicated. RLFDS, DGQQR and DHIYP motifs are highlighted in yellow and shown in bold. Conservation of residues is denoted using the following key: conserved amino acid residues are shown in bold and uppercase letters; conserved aliphatic residues (I, V, L), shown as l; conserved aromatic residues (Y, H, W, F), shown as @; conserved hydrophobic residues (W, F, Y, M, L, I, V, A, C, T, H), shown as h; conserved alcohol residues (S, T), shown as o; conserved polar residues (D, E, H, K, N, Q, R, S, T), shown as p; conserved "tiny" residues (A, G, C, S), shown as t; conserved small residues (A, G, C, S, V, N, D, T, P), shown as s; conserved bulky residues (E, F, I, K, L, M, Q, R, W, Y), shown as b; conserved positively charged residues (K, R, H), shown as +; conserved negatively charged residues (D, E), shown as -; conserved charged residues (D, E, K, R, H), shown as c. Secondary structure (2°) prediction denoted using the following key; α-helices, h, β-strands, s.*

In order to investigate the extent of the conserved regions, a structural alignment of chain D of SspE and chain A of BrxU was performed in PyMOL. The overall alignment is poor, with an RMSD of 8.548, but both proteins exhibit the same general architecture of an N-terminal DUF262 domain containing the DGQQR motif, which is preceded by a β-sheet. However, within the N-terminal domain there is little alignment of ordered components. In the C-terminal domain, the main area of interest is the DHIYP motif. The SspE residues that would align to this region are not present within 6JIV and so a comparison cannot be made. As a result, though SspE is the closest structural homologue available, detailed comparison provides little additional insight. Finally, whilst SspE was reported as a monomeric structure, the crystal form does show an intertwined dimer as per BrxU (Xiong et al. 2020).

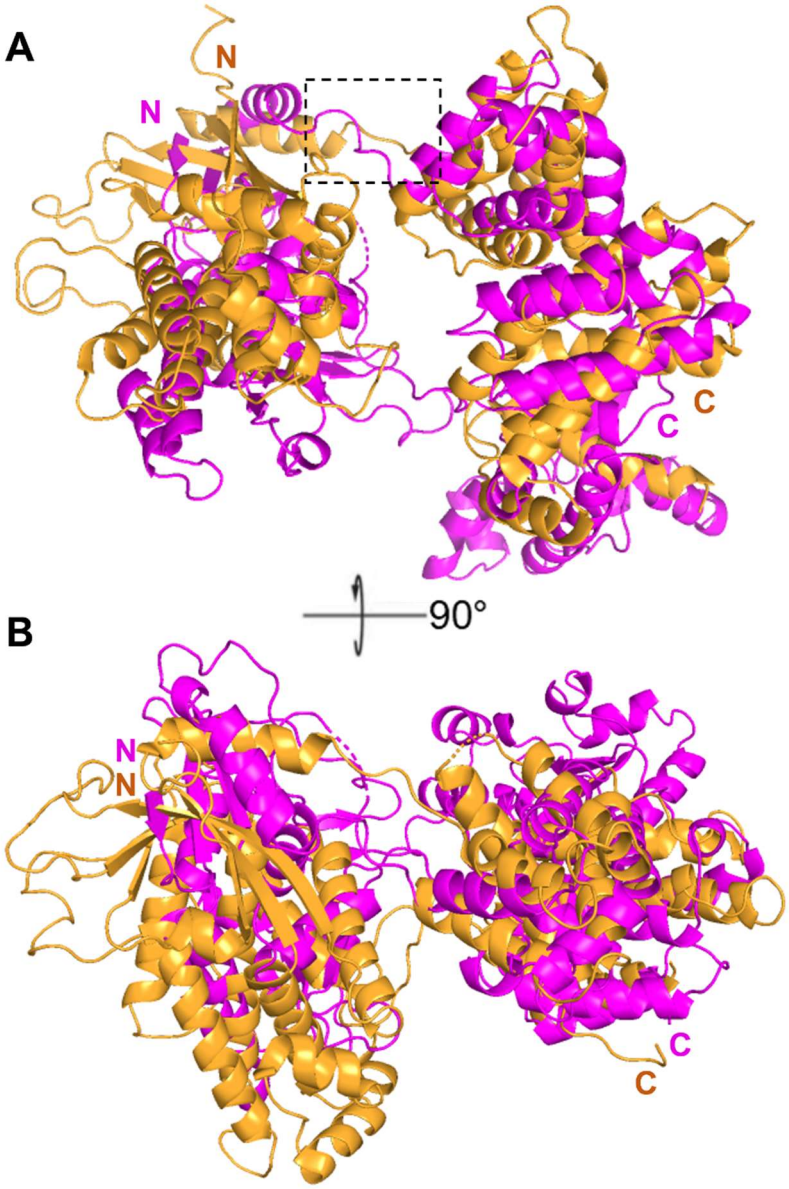


Figure 5.23: Superposition of BrxU and SspE monomers. *Chain A of BrxU in magenta is aligned with chain D of SspE (PDB 6JIV) in orange. A and B are shown related by a 90° rotation around the X axis. In both A and B, individual domain architecture can be observed, separated by a flexible linker region. The region is indicated in A by the boxed region. RMSD: 8.548*

5.17 Discussion

This chapter has detailed the advances in protein crystallisation of 4 BREX proteins and the crystallisation and structure solution for BrxU. Crystallisation conditions have been obtained for BrxA and PglZ. BrxA crystals have not yet been tested to see if they produce diffractions patterns and may present a relatively simple solution for obtaining a structure. The PDB deposition 3BHW of an uncharacterised protein from *Magnetospirillum magneticum* is predicted to be almost identical to BrxA, providing a strong molecular replacement model. PglZ crystals were obtained and with optimisation may become suitable for cryomounting and testing for diffraction. In contrast, BrxR and BrxL represent more attractive options for structural characterisation due to the progress already made. In order to fully characterise BrxU, other proteins had to be deprioritised and as a result, BrxR and BrxL remain a work in progress. As demonstrated by the phasing of the native BrxU dataset in Table 5.2, selenomethionine substitution and subsequent SAD phasing remains a viable option for BrxR and BrxL. A dataset has been collected from BrxL at a max resolution of 3.5 Å, which suggests with optimisation these crystals could be improved well below 3 Å.

Biochemical characterisation of BrxU detailed in Chapter 4 provided insight into the requirements for BrxU activity and the key components it utilises. Identification of its requirement for not only nucleotides, but also divalent metal ions, aided protein crystallisation by the addition of AMP-PnP and ultimately led to a structural solution. However, one key question remains now that the BrxU structure has been obtained: where are the ligands? It was proposed that the earliest BrxU crystals tested (Figure 5.12A) diffracted poorly due to internal disorder, and crystal formation was highly irreproducible. Rescreening BrxU with AMP-PnP and $MgSO_4$ produced significantly more hits indicating the protein was more readily crystallisable in this form. Based on the FPLC data in Chapter 4 that showed a shift from a

dimer to a monomer upon nucleotide binding, it was hypothesised that the readily crystallisable form of BrxU was monomeric and bound to a nucleotide. However, analysis of the obtained structure reveals that the predicted nucleotide binding to the DGQQR motif within the DUF262 domain (Machnicka et al. 2015) is blocked by the linker region of the other BrxU subunit. This flexible linker region that joins the N-terminal DUF262 domain to the C-terminal DUF1524 domain runs directly across the nucleotide binding pocket, hindering the binding of a nucleotide. It was initially proposed that the added AMP-PnP and Mg^{2+} was non-essential for crystallisation, however BrxU failed to crystallise in the absence of these cofactors. Therefore, there is a direct role for these cofactors in the crystallisation of BrxU but in the structure shown, both are absent. To explain this discrepancy, it is important to consider the nature of the nucleotides used in each experiment. A shift from dimer to monomer was observed when using ATP and ATP-γ-S (Figure 4.19), but AMP-PnP was not specifically used in these experiments. AMP-PnP was used for crystallisation as it was considered more stable and therefore a better substrate for crystallisation studies. In retrospect, it would be important to confirm the size exclusion data using AMP-PnP. Furthermore, it has been shown that the nature of the nucleotide used can produce drastically different outcomes on structural studies, with different results obtained for ATP-γ-S and AMP-PnP (Thomsen and Berger 2012). To this end, it could be suggested that our structure represents an "*apo*" complex due to lack of nucleotide binding. Using ATP and ADP in future studies would then provide a post-hydrolysis ADP-bound complex (assuming ATP hydrolysis to ADP is rapid), and use of ADP-BeF_3 might provide the nucleotide bound pre-hydrolysis state. In support of this hypothesis, a distinctly different crystal form is observed when BrxU is bound to ATP or ADP (Figures 5.10 and 5.11), compared to AMP-PnP (Figure 5.9), indicative of a change in the crystal packing and likely representative of a nucleotide-bound form of BrxU with ATP and ADP. This matches previous examples where crystal forms differ greatly depending on choice of nucleotide (Thomsen and

Berger 2012). However, at this stage the changes observed for these crystals remains speculative. Irrespective of nucleotide binding, it is unclear why Mg^{2+} cannot be accounted for in the unmodelled density surrounding BrxU. Again, biochemical analysis has shown a direct requirement for divalent metal ions for nuclease and NTPase activity, which suggested at least two metal binding sites. Magnesium ions could not be detected in either domain. The absence of AMP-PnP has been accounted for by the steric hindrance induced by the linker region, however this would not block the binding of a singular, small ion such as magnesium. As observed by Thomsen and Berger, it is possible that magnesium interacts with the phosphate groups of ATP when bound by the NTPase domain (Thomsen and Berger 2012). Future studies will involve looking at co-crystallising BrxU with target DNA, and it is possible that in the presence of DNA, BrxU coordinates metal ions within the nuclease domain. All this of course does not yet explain why the crystals were obtained only in the presence of AMP-PnP and Mg, though this could potentially be attributed to altered conditions aiding nucleation and crystal growth.

Cross referencing with biochemical data in Chapter 4, a number of key residues were identified as required for protein activity, whether they contributed to NTPase or nuclease activity. In section 5.15, a role for S42 has been suggested. It was initially hypothesised that S42 was involved in the binding of the nucleotide base, and that mutagenesis to a negatively charged aspartate would alter the enzyme's nucleotide preference from ATP. This has yet to tested *in vitro*, however due to the absence of AMP-PnP in the structure solution, the role of S42 in nucleotide binding is still uncertain. However, it is shown that S42 is in close proximity to the α-helix immediately preceding the linker region of the other BrxU subunit (Figure 5.20). It is proposed that this side chain interacts with I234 of the other subunit to stabilise dimer formation. Mutagenesis of this residue results in a change in behaviour when analysed via FPLC. Alanine substitution at S42 results in monomer formation in the absence of nucleotides

which is unseen for the other mutants tested. Replacing serine with aspartate, originally designed to change the nucleotide preference, results in BrxU being locked as a dimer during size exclusion assays. Figure 5.21D shows the theoretical positioning of S42D in relation to I234, an arrangement that would be unfavourable to due to proximity of the aspartate's oxygen in relation to the hydrophobic I234. This suggests that S42D induces instability/disorder within this region, preventing nucleotide binding, so BrxU can no longer undergo a conformational change and dissociate into a monomer.

Additionally, N485 has been identified as redundant due to its outwards-facing side chain. This corresponds with EOP and hydrolysis assays which return similar values to WT. Replacement of N519 abolishes BrxU activity, however from the structure obtained it is unlikely that this residue forms part of the nuclease motif and is likely important in the structural integrity of this domain. How this is achieved given that no direct interactions can be seen for N519 remains unclear, as α-helix formation would still occur in this region when substituted for alanine. It is more likely that the HNH nuclease motif forms with N509, which can be observed on the base of the antiparallel β-sheet alongside H475. As a result, N509 has been identified as a candidate for mutagenesis and is predicted to be required for nuclease activity. D474 was predicted to interact with metal ions, however this was not observed as no metal ions could be accounted for. However, a potential interaction with K521 likely stabilises the proximity of these two α-helices. Experimentally, a number of issues were encountered during the process of obtaining phasing data for BrxU. For both native and SM substituted crystals, the rate of obtaining diffraction patterns that extended beyond 4 Å was roughly 1 in 20. As a result, for each synchrotron trip around 80 crystals would be harvested in the hope that 4 would diffract. What differentiated a well diffracting crystal from a non-diffracting crystal remained unclear for a significant portion of the BrxU project. The phasing data using SM-BrxU came from a fresh protein expression that had never been subjected to flash freezing in liquid nitrogen prior to

crystallisation. As a standard procedure, proteins would be aliquoted into 20 µl aliquots and flash frozen for storage at -80°C. From these results, it is foreseeable that BrxU becomes partially disordered during this freezing process and likely resulted in the poorer diffraction observed. This should be noted for future crystallisation experiments with BrxU.

SspE was identified as the closest structural homologue of BrxU (Table 4.1). Overall structural similarities can be observed between BrxU and SspE, notably the separation of DUF262 and DUF1524 domains by a flexible linker sequence. Despite observing intertwined protomers of SspE in the crystals, there is no comment on complex formation (Xiong et al. 2020). Beyond this architecture however, these enzymes align poorly (Figure 5.23) and so it is difficult to derive information from this alignment, and SspE is solved to only 3.3 Å. SspE is involved in the Ssp phosphorothioation (PT) sensing phage defence system, detecting PT modifications of phage genomes and introducing single stranded nicks (Xiong et al. 2020). Contrastingly, BrxU detects cytosine modifications with phage gDNA and introduces DSBs. Despite both enzymes contain a similar domain architecture and providing defence against phage infection, they exhibit different biochemical properties. This is reflected when comparing BrxU to more distant homologs such as *Homo sapien* Srx. In the N-terminal domain, H100 and R101 of *hs*Srx form hydrogen bonds with the β- and γ- phosphates of ATP, respectively, similar to Q101 and R102 of BrxU (Lee et al. 2006). BrxU was solved to 2.12 Å and currently represents the highest resolution structure for DUF262 and DUF1524 domains, as well as being the first structure of a GmrSD family type IV restriction enzyme.

Chapter 6: Final Discussion

Initial interest in *E. fergusonii* and its native plasmid, pEFER, stemmed from the idea that a plasmid-borne type I BREX system would allow ease of investigation over a chromosomally encoded system. As a member of the *Escherichia* genus, *E. fergusonii* components could be studied in lab strains of *E. coli* without having to perform experiments in a distantly related species. Upon further investigation, it became apparent that pEFER did not just encode a plasmid-borne BREX system, but a more complex phage defence island. Likewise, during the characterisation of the BREX system of *S. enterica* D23580, it was discovered that a secondary, uncharacterised phage defence system was encoded within the BREX locus. Whilst initial efforts were focused on the understanding of BREX in these organisms, this study has developed to examine the activities of multifaceted phage defence islands. By utilising several defence systems simultaneously, the host is protected from a wider variety of phages (Figure 6.1). Phages that are resistant to certain type II endonucleases due to the incorporation of non-canonical bases are subsequently restricted by type IV endonucleases such as BrxU. Phages that are resistant to multiple classes of restriction enzymes may be sensitive to BREX due to a lack of methylation at the recognised non-palindromic repeat (Figure 6.1). Phages such as T7 that encode anti-restriction proteins such as the DNA mimic Ocr will circumvent certain host BREX and REase defences, but in turn can activate Abi systems such as PARIS (Rousset et al. 2021). As phages have evolved mechanisms to evade defence systems, their prokaryotic hosts have in response developed additional counter measures, which have clustered in defence islands.

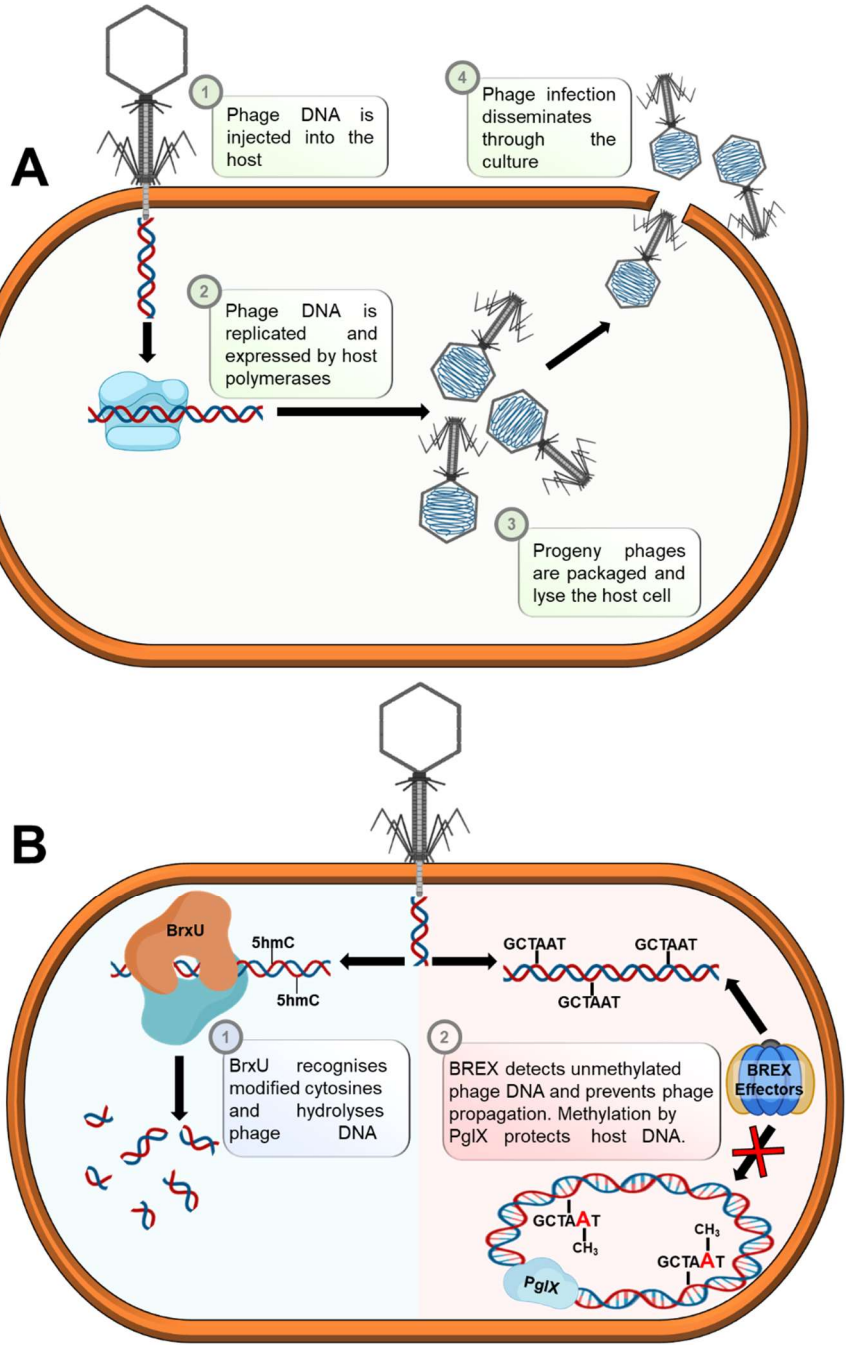


Figure 6.1: Encoding multiple phage defence systems manifests a bacterial immune system capable of suppressing multiple phage types*. A) Phage infection of a susceptible host cell results in the propagation of progeny phages, resulting in lysis of the hosts and dissemination of phage. B) Phage infection is prevented by a multi-strategy phage defence response. Left: Phage DNA is detected to contain modified cytosines and is subsequent hydrolysed by a BrxU. Right: BREX effector proteins recognise unmethylated non-palindromic motifs in phage DNA and prevent phage propagation. Methylation of this motif on host DNA by PglX protects the host from BREX effectors.*

6.1 Summary of Findings

BREX loci from *E. fergusonii* and *S. enterica* were observed to confer resistance to phages (Table 3.5 and 3.3). Generation of a D23580*ΔφΔBREX* was utilised to isolate 8 *S.* enterica phages from sewage effluent. Using these strains, it was demonstrated that the BREX locus encoded antiphage apparatus that affected 4 of the 8 phages isolated. It was assumed at this point that all resistance to these phages was attributable to BREX, however additional findings have introduced new variables. Bioinformatic analysis by another PhD student in the Blower group, Sam Went, identified *STM4493* and *STM4494* as a putative TA system. Furthermore, a pre-print publication in BioRxiv has since identified *STM4493* and *STM4494* homologues as novel Abi system, PARIS (Rousset et al. 2021). It is currently unclear if the reduced EOP values observed with D23580*Δφ* is resultant of BREX or PARIS. Whilst BREX functions by distinguishing self from non-self DNA by methylating non-palindromic motifs, PARIS detects the anti-restriction protein Ocr and triggers growth arrest. Fundamentally, the premise of this thesis is strengthened by this ambiguous result, as the BREX locus of D23580*Δφ* encompasses another phage defence island.

Preliminary work on pEFER performed at the beginning of this study, found that the introduction of a transposon into *brxA* abolished the resistance it once conferred to almost every phage tested (Table 3.2). Despite the inactivation of BREX, resistance was still conferred to Bam, Mak, Mav and Titus for which reduced EOP values were still observed when plated on pEFER-Km5. This became the first indication that BREX was not the only phage defence system encoded within pEFER. Whilst this additional system has not yet been characterised, it is predicted that it forms an additional layer of defence, as both Mak and Mav are resistant to both BREX and BrxU. Upon constructing a “BREX-only” plasmid construct via GGA, it became apparent that the inclusion of all genetic material from *brxR* (and its promoter regions)

to *brxL* was not sufficient to manifest an active BREX system. This was further clarified by PacBio sequencing, revealing that pEFER encoded an active methyltransferase, but pBREX did not. *EFER_p0017* and *EFER_p0019* were judged to be required components due to their conservation with other γ-proteobacterial BREX systems. Addition of these ORFs restored BREX function and the resulting "pBrxXL" was found to encode an active methyltransferase, and so the ORFs were denoted *brxS* and *brxT,* respectively. Deletion of *brxT* from pBrxXL abolished all antiphage activity of pBrxXL, suggesting that BrxT was a required component for BREX and BrxU activity (Section 3.6). However, if BrxT was a necessary component for BREX activity, then it is unlikely that would entirely neutralise BrxU activity. It is therefore possible that the deletion of *brxT* disrupted regulatory regions upstream of the BREX locus, resulting in suppression of transcription of the entire operon.

All BREX proteins with the exception of BrxS and BrxT, (which are yet to be confirmed as expressed proteins), have been overexpressed as soluble proteins and purified to homogeneity (Figure 4.5). BrxR was confirmed to be a transcriptional regulator as predicted from Phyre2 structural modelling (Section 4.2). The modelling of a wHTH domain indicated that BrxR interacted with double stranded DNA, and its location upstream of a type I BREX system strengthened this projection. Expression of BrxR was found to significantly reduce transcription initiation from Region S6, which contained the strongest promoter (Figure 4.9). Whilst a role has been determined for the N-terminal DNA binding domain of BrxR, it is currently unclear how the other protein domains function. As predicted, BrxR has been shown to form dimers in solution, characteristic of a transcriptional regulator (Figure 4.6). Potentially, the C-terminal region of BrxR could be involved in recognition of elements induced by phage infection. BrxR encodes a WYL-domain, which has been implicated in recognition of currently unidentified response ligands (Müller et al. 2019). There is potential for BrxR to bind stress

response molecules such as cyclic nucleotides that are upregulated upon phage infection (Høyland-Kroghsbo 2019).

Development of a biochemical assay for measuring inorganic phosphate production confirmed BrxC as an ATPase. No increase in inorganic phosphate was detected for PglZ and BrxL. PglZ and BrxL may exhibit phosphatase/NTPase activity, however in the conditions tested they did not. Additional components may be required and complex formation may be a necessity for hydrolysis. BrxC was predicted to have ATPase activity due to the presence of a Walker box motif in Phyre2 models (Section 4.2). Gordeeva et al. showed that deletion of BrxC abolished both methylation of host DNA by PglX and any resistance conferred to phage infection (Gordeeva et al. 2019). Gel filtration analysis shows that BrxC forms multimers, most likely pentamers or hexamers (Figure 4.7).

It is predicted from these results that BrxC provides the mechanical force for DNA translocation, and that it forms a complex with other BREX proteins, which use BrxC to move to their target regions. Deletion of PglX from pBrxXL resulted in the disabling of BREX, indicated by increased EOP values for BREX-sensitive phages (Table 3.5), aligning with findings by Gordeeva et al. PglX from pEFER was found to methylate the non-palindromic sequence GCTA**A**T at the 5th position. This differs from the target sequence of *E. coli* HS and *B. cereus* PglX enzymes, which target GGTA**A**G and TAGG**A**G, respectively. However, all 3 PglX homologues are adenine specific methyltransferases that methylate at the fifth position of their respective non-palindromic sequences. Whilst the biochemistry of PglZ and BrxL has not been prioritised during this study, both proteins represent novel proteins implicated in phage defence. Both have been demonstrated by Gordeeva et al. to be required for BREX to confer phage resistance, however only PglZ is required for PglX to methylate target DNA. Therefore, it is predicted that PglZ forms a complex with PglX and other BREX proteins, and

that BrxL is an effector protein that exhibits proteolytic activity resulting in a response against the invading phage DNA. Whilst the majority of structural work has focused on BrxU, advances have been in protein crystallisation for multiple BREX proteins. Crystals were obtained for BrxA and PglZ (Figure 5.4 and 5.5). BrxL has been crystallised and a native 3.5 Å dataset has been obtained. BrxL crystals shown in Figure 5.6 represent strong candidates for optimisation, and could be used for micro-seeding to identify additional crystallisation conditions. A BrxR native dataset was also collected at 2.8 Å, and optimisation of selenomethionine-BrxR crystals should allow the structure to be solved.

BrxU was originally predicted to have a non-essential role within BREX, as accessory genes were identified to associate within type I BREX systems (Goldfarb et al. 2015). Initial bioinformatic analysis identified an N-terminal DUF262 domain, indicative of a DNA binding protein. Upon consultation with Dr David Dryden, our interest in characterising BrxU was magnified and BrxU was no longer denoted as a BREX accessory component. Further bioinformatic characterisation of BrxU revealed the presence of a C-terminal DUF1524 domain, a member of the histidine-metal finger endonuclease superfamily, which include HNH nucleases. BrxU was observed to specifically hydrolyse gDNA substrates from phages that were affected by pBrxXL-*ΔpglX.* The phage resistance properties of BrxU were confirmed *in vivo*, as BrxU provided resistance against the same phages when expressed from pBAD30-*brxU* (Table 4.2). The activity of BrxU was dependent on the presence of Mg^{2+} and ATP. Not only is BrxU able to utilise Mn^{2+}, Co^{2+}, Ni^{2+}, Fe^{2+} and Fe^{3+} as a substitute for Mg^{2+}, it is able to use all canonical nucleotides in place of ATP (Figure 4.14 and 4.16). Whilst ATP was noted to be the most readily utilised nucleotide cofactor, dNTPs and rNTPs could be used by BrxU at lower efficiency. This promiscuity of cofactor requirements showcases the broad range of conditions BrxU in which exhibits activity. It was hypothesised that BrxU targeted modified phage genomes, as demonstrated by the homologue GmrSD (He et al. 2015). By substituting

dCTP for modified cytosines, synthetic DNA substrates were constructed (Figure 4.21). It was found that BrxU cleaved DNA substrates that incorporated 5mC-dCTP and 5hmC-dCTP, but not dCTP. This is in contrast with the activity of GmrSD which did not restrict DNA substrates containing 5mC. Furthermore, the glycosylation of target DNA by β-glucosyltransferase did not prevent BrxU from exhibiting nuclease activity. As a result, BrxU represents a single chain type IV restriction endonuclease that confers provides significant resistance against a wide range of susceptible phages, whilst being capable of utilising a extended array of cofactors.

A key finding within this study was the identification of a multimeric shift upon BrxU binding nucleotides. Gel filtration analysis showed that upon binding nucleotides, BrxU dimers dissociated to monomers. Whilst it is foreseeable that higher order structures are formed upon DNA binding, this result shows that nucleotide binding induces a significant change in enzyme morphology and phosphate hydrolysis is not specifically required to induce this change (Figure 4.19). This shift was aided by the addition of Mg^{2+}, which is likely due to its role in coordinating the terminal phosphate of the bound nucleotide, as demonstrated in other ATPase domain-containing proteins (Thomsen and Berger 2012). It was found that binding to ADP was also able to induce this shift, however AMP was not, also suggesting a direct role of the β-phosphate in inducing dimer dissociation. Figure 6.2 illustrates the nucleotide cycling of BrxU, and the role of nucleotides in allowing BrxU to bind target DNA. Although it remains unclear when ADP is ejected from BrxU subinits, and whether BrxU cuts once or twice per dimer, the role of the nucleotide in inducing monomer/ dimer of BrxU can be proposed.

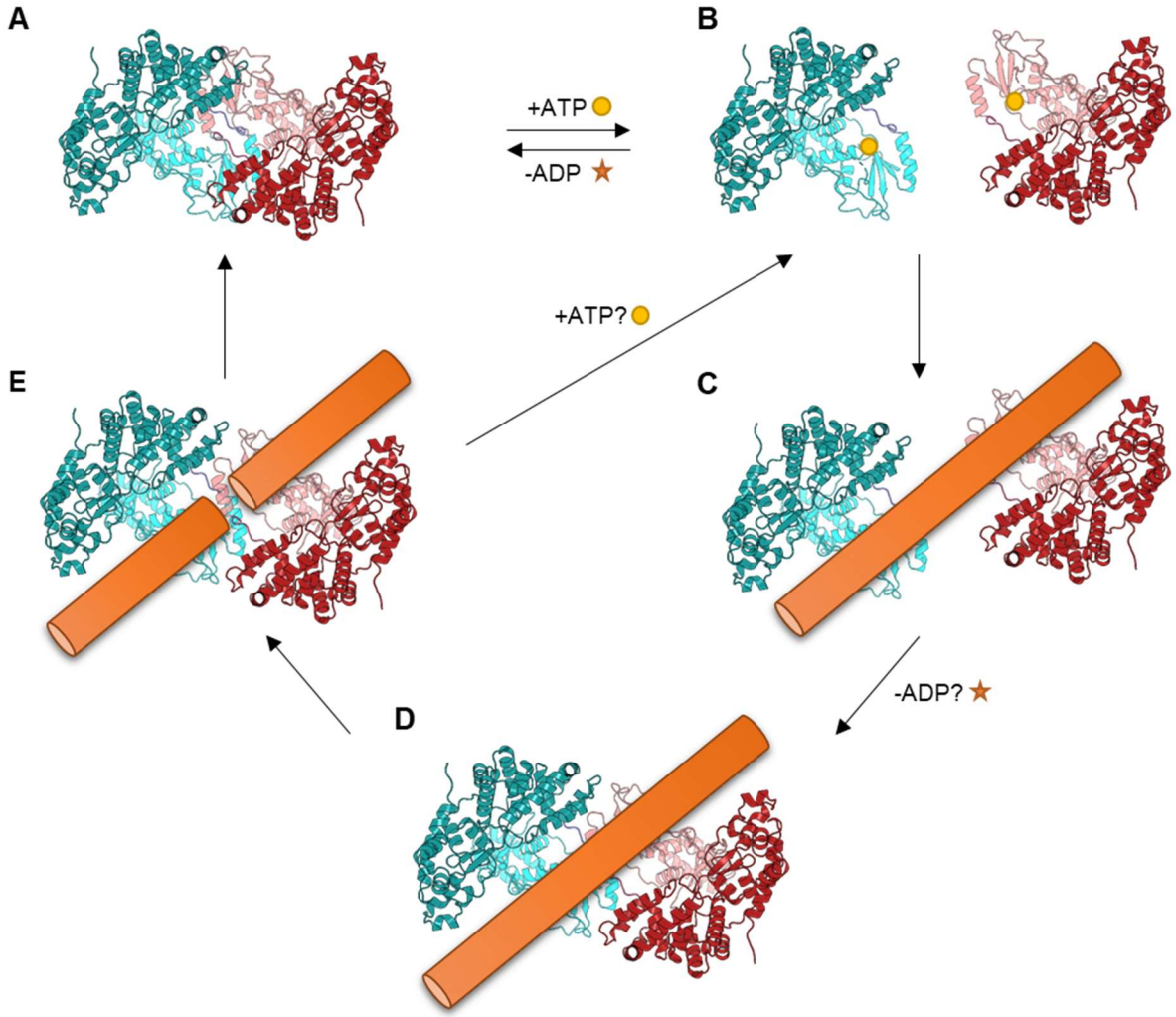


Figure 6.2: Proposed multi-step reaction cycle of BrxU. *A) When not bound to a nucleotide, BrxU exists as a dimer. B) Upon binding a nucleotide, dimeric BrxU dissociates into monomers. C) BrxU monomers can associate and dock with dsDNA (orange). D) BrxU recognises its target sequence which contains modified cytosine(s). It is currently unclear when ADP is released by BrxU subunits. E) dsDNA is cleaved. Binding of ATP to the dimeric BrxU can initiate the next cleavage cycle. This may be required for the release of the cleaved dsDNA.*

Crystallisation trials were repeated for BrxU, this time with the addition of metal and nucleotide cofactors. AMP-PnP was substituted in place of ATP, as it was predicted that AMP-PnP would remain bound and induce a stable monomeric shift, as its γ-phosphate is only very slowly hydrolysed. The addition of AMP-PnP produced a much greater number of crystal hits, suggesting AMP-PnP was bound. However, upon structure solution via phasing a native 2.12 Å dataset using SM-SAD, no nucleotide ligand could be accounted for in the unmodelled electron density. Furthermore, BrxU was found to be a dimer in its crystal form, contrasting gel filtration analysis. Dimers were observed to be 'locked' together, with linker regions between DUF262 and DUF1524 domain interlocking the protomers. Initially, this was particularly perplexing as it had been assumed that AMP-PnP was bound in BrxU crystals, as the absence of AMP-PnP in these conditions resulted in no crystal formation. AMP-PnP perhaps therefore aids crystal formation, likely due to a weak interaction at the nucleotide binding site of DUF262, however, BrxU does not interact strongly enough with it to bind. The non-hydrolysable analogues ATP-γ-S and GTP-γ-S were shown to induce monomer formation by analytical gel filtration (Figure 4.19). However, as an oversight, it was assumed that AMP-PnP would interact with BrxU in a similar manner and gel filtration analysis with this nucleotide was not performed. Analysis of ATP analogue binding by Thomsen *et al* reveals that AMP-PnP does not bind any of the six ATPase domains of Rho, whereas ADP and ADP-BeF_3 do bind. As there is precedent, it is possible that AMP-PnP does not induce monomer formation, and this will be tested in future. In hindsight, it is strikingly apparent that AMP-PnP could behave differently to other ATP analogues such as ATP-γ-S, however it was chosen due to its superior stability. This is further corroborated by the fact that the bridging oxygen between the β and γ-phosphates of ATP is replaced by nitrogen in AMP-PnP. In ATP-γ-S, a mercapto group replaces a hydroxyl group on the γ-phosphate. As a direct role of the β-phosphate has been implicated for binding BrxU and inducing dimer dissociation, it is foreseen

that the nitrogen in AMP-PnP's phosphate groups prevents interaction with the DUF262 domain.

A number of key residues within BrxU were identified by a combination of comparative alignments, point mutagenesis and structural analysis. The DGQQR motif of DUF262 is conserved between GmrSD and BrxU, and alanine substitution of Q101 and R102 inactivates BrxU *in vivo* and *in vitro*. As shown in Figure 5.17, the linker region of the adjacent protomer passes directly across the DGQQR motif. Binding of a nucleotide to this region likely induces a conformational change resulting in the multimeric shift observed in gel filtration analyses. Interestingly, the S42A mutant was found to exist as a monomer in the absence of nucleotides (Figure 4.26), suggesting that this serine side chain is involved in stabilising dimers. Conversely, S42D did not form monomers when incubated with ATP, potentially due to the disruption caused at the dimer interface by introducing a highly negatively charged side in close proximity to the hydrophobic I234 on the other protomer. It is possible that this induces a conformation change that prevents successful nucleotide binding in the DGQQR motif, and stops the shift to a monomeric state, however this has yet to be tested.

6.2 Future Work on BREX and BrxU

Progress has been made on the characterisation of BREX and BrxU, however certain questions remain to be answered. In particular for BREX, how do the components interact, and what are the effector mechanisms? PglX is documented to distinguish foreign DNA from host DNA by methylating host DNA at non-palindromic motifs, but how does this translate to preventing viral infection? All BREX proteins, other than *BrxS* and *BrxT*, have been expressed and purified, and their multimeric states in solution as individual proteins are known. Protein-protein interactions can now be investigated by gel filtration analysis, looking for shifts in

elution volumes caused by complex formation. This can also be assessed by expressing 6His-tagged variants of BREX proteins and binding to Ni-NTA. Incubation of a tagged variant with the other untagged BREX proteins followed by washing and elution of the tagged variant would allow for identification of interactions by SDS-PAGE. A 2.8 Å native dataset has been obtained for BrxR, and phasing with SM remains a viable option. Previously, SM-labelled BrxR did not crystallise well enough to be mounted in cryoloops, however, due to limited yield additional purification steps could not be performed. Subsequent attempts will scale up expression of SM-labelled BrxR to allow for more efficient purification. Additionally, BrxL crystals will be used in micro-seeding experiments in order obtain new crystal conditions for optimisation to improve upon the 3.5 Å dataset obtained. BrxA represents a fairly simple structural target due to the identification of highly similar homologue, however this is dependent on the obtained crystals diffracting.

Whilst the observation that BrxU was dimeric in the obtained structure was unexpected at first, it fortuitously provided the *apo* structure, and provides an opportunity to obtain structures of both the ATP and ADP-bound forms. In the absence of AMP-PnP, BrxU crystals were poorly ordered and diffracted to a max resolution of 5.05 Å (Figure 5.12). As shown in Figure 5.10 and 5.11, a distinctly different crystal morphology is observed when BrxU is crystallised with ATP and ADP, respectively. These crystals are to be optimised and harvested for data collection and will likely be the ADP-bound form (assuming ATP is rapidly hydrolysed). Using ADP-BeF_3 could also potentially trap the ATP bound state. A single protomer of the BrxU model can then be used to phase these monomeric forms of BrxU via molecular replacement. This would allow for direct comparison of the nucleotide bound and unbound forms of BrxU. Additionally, tests against type IV restriction enzyme inhibitors such as IPI from T4, and potential co-crystallisation with BrxU, could reveal even more insight into the mechanics of this novel type IV restriction enzyme. In reflection, a study which was conceived with the

notion of purely investigating the mechanism of the BREX system in *E. fergusonii* has led to the identification of new phage defence systems that cluster to form defence islands. In this "molecular arms race", the acquisition of multiple defence systems provides the host with a stronger, more robust barricade to the omnipresent phages that seek to destroy them.

Bibliography

Abreu, Afonso G., and Angela S. Barbosa. 2017. 'How Escherichia coli Circumvent Complement-Mediated Killing', *Frontiers in Immunology*, 8.

Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. 2010. 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta crystallographica. Section D, Biological crystallography*, 66: 213-21.

Adesina, Tomilola, Obinna Nwinyi, Nandita De, Olayemi Akinnola, and Emmanuel Omonigbehin. 2019. 'First Detection of Carbapenem-Resistant Escherichia fergusonii Strains Harbouring Beta-Lactamase Genes from Clinical Samples', *Pathogens*, 8: 164.

Adhikari, Satish, and Patrick D. Curtis. 2016. 'DNA methyltransferases and epigenetic regulation in bacteria', *FEMS Microbiology Reviews*, 40: 575-91.

Agnes Ullmann. 2009. 'Escherichia coli Lactose Operon', *eLS*.

Ali, Ibraheem, Ryan J. Conrad, Eric Verdin, and Melanie Ott. 2018. 'Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics', *Chemical Reviews*, 118: 1216-52.

Allet, Bernard, and Ahmad I. Bukhari. 1975. 'Analysis of bacteriophage Mu and λ-Mu hybrid DNAs by specific endonucleases', *Journal of Molecular Biology*, 92: 529-40.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. 'Basic local alignment search tool', *J Mol Biol*, 215: 403-10.

Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. 1981. 'Structure of the cro repressor from bacteriophage lambda and its interaction with DNA', *Nature*, 290: 754-8.

Anton, Brian P., Emmanuel F. Mongodin, Sonia Agrawal, Alexey Fomenkov, Devon R. Byrd, Richard J. Roberts, and Elisabeth A. Raleigh. 2015. 'Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of Escherichia coli K-12', *PLOS ONE*, 10: e0127446.

Aravind, L., K. S. Makarova, and E. V. Koonin. 2000. 'SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories', *Nucleic acids research*, 28: 3417-32.

Aslanidis, C., and P. J. de Jong. 1990. 'Ligation-independent cloning of PCR products (LIC-PCR)', *Nucleic acids research*, 18: 6069-74.

Ballandras-Colas, Allison, Monica Brown, Nicola J. Cook, Tamaria G. Dewdney, Borries Demeler, Peter Cherepanov, Dmitry Lyumkis, and Alan N. Engelman. 2016. 'Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function', *Nature*, 530: 358-61.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. 2007. 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315: 1709-12.

Barrangou, Rodolphe, and John van der Oost. 2015. 'Bacteriophage exclusion, a new defense system', *The EMBO journal*, 34: 134-35.

Beaulaurier, John, Eric E. Schadt, and Gang Fang. 2019. 'Deciphering bacterial epigenomes using modern sequencing technologies', *Nature reviews. Genetics*, 20: 157-72.

Bernheim, Aude, and Rotem Sorek. 2020. 'The pan-immune system of bacteria: antiviral defence as a community resource', *Nature Reviews Microbiology*, 18: 113-19.

Bertani, G., and J. J. Weigle. 1953. 'Host controlled variation in bacterial viruses', *Journal of Bacteriology*, 65: 113-21.

Bervoets, Indra, and Daniel Charlier. 2019. 'Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology', *FEMS Microbiology Reviews*, 43: 304-39.

Bickle, Thomas A. 2004. 'Restricting restriction', *Molecular Microbiology*, 51: 3-5.

Blank, Kathrin, Michael Hensel, and Roman G. Gerlach. 2011. 'Rapid and Highly Efficient Method for Scarless Mutagenesis within the Salmonella enterica Chromosome', *PLOS ONE*, 6: e15763.

Blower, Tim R., Terry J. Evans, Rita Przybilski, Peter C. Fineran, and George P. C. Salmond. 2012. 'Viral Evasion of a Bacterial Suicide System by RNA–Based Molecular Mimicry Enables Infectious Altruism', *PLOS Genetics*, 8: e1003023.

Blower, Tim R., Francesca L. Short, Peter C. Fineran, and George P. C. Salmond. 2012. 'Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication', *Bacteriophage*, 2: 234-38.

Bourniquel, A. A., and T. A. Bickle. 2002. 'Complex restriction enzymes: NTP-driven molecular motors', *Biochimie*, 84: 1047-59.

Boyer, H. W. 1971. 'DNA restriction and modification mechanisms in bacteria', *Annu Rev Microbiol*, 25: 153-76.

Braga, Lucas P. P., Aymé Spor, Witold Kot, Marie-Christine Breuil, Lars H. Hansen, João C. Setubal, and Laurent Philippot. 2020. 'Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios', *Microbiome*, 8: 52.

Bragg, Robert, Wouter van der Westhuizen, Ji-Yun Lee, Elke Coetsee, and Charlotte Boucher. 2014. 'Bacteriophages as Potential Treatment Option for Antibiotic Resistant Bacteria.' in Rameshwar Adhikari and Santosh Thapa (eds.), *Infectious Diseases and Nanomedicine I: First International Conference (ICIDN – 2012), Dec. 15-18, 2012, Kathmandu, Nepal* (Springer India: New Delhi).

Brickner, M., and J. Chmielewski. 1998. 'Inhibiting the dimeric restriction endonuclease EcoRI using interfacial helical peptides', *Chem Biol*, 5: 339-43.

Bryson, Alexandra L., Young Hwang, Scott Sherrill-Mix, Gary D. Wu, James D. Lewis, Lindsay Black, Tyson A. Clark, and Frederic D. Bushman. 2015. 'Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9', *mBio*, 6: e00648-15.

Burch, Lauranell H., Leilei Zhang, Frank G. Chao, Hong Xu, and John W. Drake. 2011. 'The Bacteriophage T4 Rapid-Lysis Genes and Their Mutational Proclivities', *Journal of Bacteriology*, 193: 3537-45.

Butterer, Annika, Christian Pernstich, Rachel M. Smith, Frank Sobott, Mark D. Szczelkun, and Júlia Tóth. 2014. 'Type III restriction endonucleases are heterotrimeric: comprising one helicase–nuclease subunit and a dimeric methyltransferase that binds only one specific DNA', *Nucleic acids research*, 42: 5139-50.

Callahan, Scott J., Yvette A. Luyten, Yogesh K. Gupta, Geoffrey G. Wilson, Richard J. Roberts, Richard D. Morgan, and Aneel K. Aggarwal. 2016. 'Structure of Type IIL Restriction-Modification Enzyme MmeI in Complex with DNA Has Implications for Engineering New Specificities', *PLOS Biology*, 14: e1002442.

Casadesus, J., and D. Low. 2006. 'Epigenetic gene regulation in the bacterial world', *Microbiology and molecular biology reviews : MMBR*, 70: 830-56.

Chatterjee, Sujoy, and Eli Rothenberg. 2012. 'Interaction of bacteriophage l with its E. coli receptor, LamB', *Viruses*, 4: 3162-78.

Chen, Bo-Wei, Ming-Hsing Lin, Chen-Hsi Chu, Chia-En Hsu, and Yuh-Ju Sun. 2015. 'Insights into ParB spreading from the complex structure of Spo0J and parS', *Proceedings of the National Academy of Sciences*, 112: 6613-18.

Cheng, Xiaodong, and Richard J. Roberts. 2001. 'AdoMet-dependent methylation, DNA methyltransferases and base flipping', *Nucleic acids research*, 29: 3784-95.

Chopin, M. C., A. Chopin, and E. Bidnenko. 2005. 'Phage abortive infection in lactococci: variations on a theme', *Curr Opin Microbiol*, 8: 473-9.

Coffey, A., and R. P. Ross. 2002. 'Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application', *Antonie Van Leeuwenhoek*, 82: 303-21.

Cohen, D., S. Melamed, A. Millman, G. Shulman, Y. Oppenheimer-Shaanan, A. Kacen, S. Doron, G. Amitai, and R. Sorek. 2019. 'Cyclic GMP-AMP signalling protects bacteria against viral infection', *Nature*, 574: 691-95.

Cohen, Stanley N., Annie C. Y. Chang, Herbert W. Boyer, and Robert B. Helling. 1973. 'Construction of Biologically Functional Bacterial Plasmids In Vitro', *Proceedings of the National Academy of Sciences*, 70: 3240-44.

Cooper, Laurie P., Gareth A. Roberts, John H. White, Yvette A. Luyten, Edward K. M. Bower, Richard D. Morgan, Richard J. Roberts, Jodi A. Lindsay, and David T. F. Dryden. 2017. 'DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus', *Nucleic acids research*, 45: 3395-406.

Cowling, Victoria H. 2009. 'Regulation of mRNA cap methylation', *The Biochemical journal*, 425: 295-302.

Cowtan, K. 2006. 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta crystallographica. Section D, Biological crystallography*, 62: 1002-11.

Creixell, Pau, Erwin M. Schoof, Chris Soon Heng Tan, and Rune Linding. 2012. 'Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367: 2584-93.

Cymerman, I. A., A. Obarska, K. J. Skowronek, A. Lubys, and J. M. Bujnicki. 2006. 'Identification of a new subfamily of HNH nucleases and experimental characterization of a representative member, HphI restriction endonuclease', *Proteins*, 65: 867-76.

Datsenko, K. A., and B. L. Wanner. 2000. 'One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97: 6640-5.

Derous, Veerle, Francine Deboeck, Jean-Pierre Hernalsteens, and Henri De Greve. 2011. 'Reproducible gene targeting in recalcitrant Escherichia coli isolates', *BMC Research Notes*, 4: 213.

Dryden, D. T., N. E. Murray, and D. N. Rao. 2001. 'Nucleoside triphosphate-dependent restriction enzymes', *Nucleic acids research*, 29: 3728-41.

Dussoix, D., and W. Arber. 1962. 'Host specificity of DNA produced by Escherichia coli. II. Control over acceptance of DNA from infecting phage lambda', *J Mol Biol*, 5: 37-49.

Dy, R. L., L. A. Rigano, and P. C. Fineran. 2018. 'Phage-based biocontrol strategies and their application in agriculture and aquaculture', *Biochem Soc Trans*, 46: 1605-13.

Ellis, E. L., and M. Delbrück. 1939. 'The growth of bacteriophage', *The Journal of general physiology*, 22: 365-84.

Emsley, P., and K. Cowtan. 2004. 'Coot: model-building tools for molecular graphics', *Acta crystallographica. Section D, Biological crystallography*, 60: 2126-32.

Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. 2008. 'A One Pot, One Step, Precision Cloning Method with High Throughput Capability', *PLOS ONE*, 3: e3647.

Erni, Bernhard. 2006. 'The Mannose Transporter Complex: an Open Door for the Macromolecular Invasion of Bacteria', *Journal of Bacteriology*, 188: 7036-38.

Feder, M., and J. M. Bujnicki. 2005. 'Identification of a new family of putative PD-(D/E)XK nucleases with unusual phylogenomic distribution and a new type of the active site', *BMC Genomics*, 6: 21.

Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. 1976. 'Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene', *Nature*, 260: 500-7.

Fineran, Peter C., Tim R. Blower, Ian J. Foulds, David P. Humphreys, Kathryn S. Lilley, and George P. C. Salmond. 2009. 'The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair', *Proceedings of the National Academy of Sciences of the United States of America*, 106: 894-99.

Flick, K. E., M. S. Jurica, R. J. Monnat, Jr., and B. L. Stoddard. 1998. 'DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI', *Nature*, 394: 96-101.

Fuller-Pace, F. V., and N. E. Murray. 1986. 'Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, 83: 9368-72.

Gajiwala, K. S., and S. K. Burley. 2000. 'Winged helix proteins', *Current opinion in structural biology*, 10: 110-16.

Gasiunas, Giedrius, Giedrius Sasnauskas, Gintautas Tamulaitis, Claus Urbanke, Dalia Razaniene, and Virginijus Siksnys. 2008. 'Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family', *Nucleic acids research*, 36: 938-49.

Goldfarb, T., H. Sberro, E. Weinstock, O. Cohen, S. Doron, Y. Charpak-Amikam, S. Afik, G. Ofir, and R. Sorek. 2015. 'BREX is a novel phage resistance system widespread in microbial genomes', *The EMBO journal*, 34: 169-83.

Golovenko, Dmitrij, Elena Manakova, Giedre Tamulaitiene, Saulius Grazulis, and Virginijus Siksnys. 2009. 'Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII', *Nucleic acids research*, 37: 6613-24.

Gong, T., M. Lu, X. Zhou, A. Zhang, B. Tang, J. Chen, M. Jing, and Y. Li. 2019. 'CRISPR-Cas Systems in Streptococci', *Curr Issues Mol Biol*, 32: 1-38.

Gordeeva, Julia, Natalya Morozova, Nicolas Sierro, Artem Isaev, Tomas Sinkunas, Ksenia Tsvetkova, Mikhail Matlashov, Lidija Truncaite, Richard D. Morgan, Nikolai V. Ivanov, Virgis Siksnys, Lanying Zeng, and Konstantin Severinov. 2019. 'BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site', *Nucleic acids research*, 47: 253-65.

Griffiths AJF, Miller JH, Suzuki DT, et al.,. 2000. *Introduction to Genetic Analysis.* (W. H. Freeman: New York).

Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. 1995. 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter', *Journal of Bacteriology*, 177: 4121-30.

Harrington, Lucas B., Kevin W. Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J. Knott, Alireza Edraki, Bianca Garcia, Nadia Amrani, Janice S. Chen, Joshua C. Cofsky, Philip J. Kranzusch, Erik J. Sontheimer, Alan R. Davidson, Karen L. Maxwell, and Jennifer A. Doudna. 2017. 'A Broad-Spectrum Inhibitor of CRISPR-Cas9', *Cell*, 170: 1224-33.e15.

Hayes, F. 2003. 'Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest', *Science*, 301: 1496-9.

He, Xinyi, Victoria Hull, Julie A. Thomas, Xiaoqing Fu, Sonal Gidwani, Yogesh K. Gupta, Lindsay W. Black, and Shuang-yong Xu. 2015. 'Expression and purification of a single-chain Type IV restriction enzyme Eco94GmrSD and determination of its substrate preference', *Scientific reports*, 5: 9747.

Hein, Stephanie, Ingeborg Scholz, Björn Voß, and Wolfgang R. Hess. 2013. 'Adaptation and modification of three CRISPR loci in two closely related cyanobacteria', *RNA biology*, 10: 852-64.

Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. 'The origins and ongoing evolution of viruses', *Trends Microbiol*, 8: 504-8.

Hernday, Aaron D., Bruce A. Braaten, and David A. Low. 2003. 'The Mechanism by which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch', *Molecular Cell*, 12: 947-57.

Hershey, A. D., and M. Chase. 1952. 'Independent functions of viral protein and nucleic acid in growth of bacteriophage', *The Journal of general physiology*, 36: 39-56.

Hilbert, Brendan J., Janelle A. Hayes, Nicholas P. Stone, Caroline M. Duffy, Banumathi Sankaran, and Brian A. Kelch. 2015. 'Structure and mechanism of the ATPase that powers viral genome packaging', *Proceedings of the National Academy of Sciences of the United States of America*, 112: E3792-E99.

Hinton, D. M. 1991. 'Transcription from a bacteriophage T4 middle promoter using T4 motA protein and phage-modified RNA polymerase', *J Biol Chem*, 266: 18034-44.

Hinton, Deborah M. 2010. 'Transcriptional control in the prereplicative phase of T4 development', *Virology Journal*, 7: 289.

Høyland-Kroghsbo, Nina Molin. 2019. 'Cyclic Nucleotide Signaling: A Second Messenger of Death', *Cell Host & Microbe*, 26: 567-68.

Hui, Wenyan, Wenyi Zhang, Lai-Yu Kwok, Heping Zhang, Jian Kong, and Tiansong Sun. 2019. 'A Novel Bacteriophage Exclusion (BREX) System Encoded by the pglX Gene in Lactobacillus casei Zhang', *Applied and environmental microbiology*, 85: e01001-19.

Iida, Shigeru, Markus B. Streiff, Thomas A. Bickle, and Werner Arber. 1987. 'Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA− phages', *Virology*, 157: 156-66.

Interthal, H., J. J. Pouliot, and J. J. Champoux. 2001. 'The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily', *Proceedings of the National Academy of Sciences of the United States of America*, 98: 12009-14.

Irizarry, R. A., C. Ladd-Acosta, B. Carvalho, H. Wu, S. A. Brandenburg, J. A. Jeddeloh, B. Wen, and A. P. Feinberg. 2008. 'Comprehensive high-throughput arrays for relative methylation (CHARM)', *Genome Res*, 18: 780-90.

Isaev, Artem, Alena Drobiazko, Nicolas Sierro, Julia Gordeeva, Ido Yosef, Udi Qimron, Nikolai V Ivanov, and Konstantin Severinov. 2020. 'Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence', *Nucleic acids research*.

Ishino, Yoshizumi, Mart Krupovic, and Patrick Forterre. 2018. 'History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology', *Journal of Bacteriology*, 200: e00580-17.

Jablonska, Jagoda, Dorota Matelska, Kamil Steczkiewicz, and Krzysztof Ginalski. 2017. 'Systematic classification of the His-Me finger superfamily', *Nucleic acids research*, 45: 11479-94.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. 'Individual-specific ‘fingerprints’ of human DNA', *Nature*, 316: 76-79.

Jiang, Fuguo, and Jennifer A. Doudna. 2017. 'CRISPR–Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, 46: 505-29.

Jin, Zelin, and Yun Liu. 2018. 'DNA methylation in human diseases', *Genes & diseases*, 5: 1-8.

Jindrova, Eva, Stefanie Schmid-Nuoffer, Fabienne Hamburger, Pavel Janscak, and Thomas A. Bickle. 2005. 'On the DNA cleavage mechanism of Type I restriction enzymes', *Nucleic acids research*, 33: 1760-66.

Jurkowski, Tomasz P., Nils Anspach, Liliya Kulishova, Wolfgang Nellen, and Albert Jeltsch. 2007. 'The M.EcoRV DNA-(Adenine N6)-methyltransferase Uses DNA Bending for Recognition of an Expanded EcoDam Recognition Site', *Journal of Biological Chemistry*, 282: 36942-52.

Kaminska, Katarzyna, and Janusz Bujnicki. 2008. 'Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acyltransferase superfamily', *Cell cycle (Georgetown, Tex.)*, 7: 120-1.

Kazmierczak, Mark J., Martin Wiedmann, and Kathryn J. Boor. 2005. 'Alternative Sigma Factors and Their Roles in Bacterial Virulence', *Microbiology and Molecular Biology Reviews*, 69: 527-43.

Kchromov, I. S., V. V. Sorotchkina, T. G. Nigmatullin, and T. I. Tikchonenko. 1980. 'A new nitrogen base 5-hydroxycytosine in phage N-17 DNA', *FEBS Letters*, 118: 51-54.

Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg. 2015. 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature protocols*, 10: 845-58.

Kennaway, Christopher K., Agnieszka Obarska-Kosinska, John H. White, Irina Tuszynska, Laurie P. Cooper, Janusz M. Bujnicki, John Trinick, and David T. F. Dryden. 2009. 'The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein', *Nucleic acids research*, 37: 762-70.

Khudyakov, I. Y., M. D. Kirnos, N. I. Alexandrushkina, and B. F. Vanyushin. 1978. 'Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine', *Virology*, 88: 8-18.

Kobayashi, Ichizo. 2001. 'Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution', *Nucleic acids research*, 29: 3742-56.

Kong, H., R. D. Morgan, R. E. Maunus, and I. Schildkraut. 1993. 'A unique restriction endonuclease, BcgI, from Bacillus coagulans', *Nucleic acids research*, 21: 987-91.

Koonin, Eugene, Valerian Dolja, Mart Krupovic, Arvind Varsani, Yuri Wolf, Natalya Yutin, Francisco Zerbini, and Jens Kuhn. 2019. *Create a megataxonomic framework, filling all principal taxonomic ranks, for DNA viruses encoding vertical jelly roll-type major capsid proteins*.

Kozdon, Jennifer B., Michael D. Melfi, Khai Luong, Tyson A. Clark, Matthew Boitano, Susana Wang, Bo Zhou, Diego Gonzalez, Justine Collier, Stephen W. Turner, Jonas Korlach, Lucy Shapiro, and Harley H. McAdams. 2013. 'Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle', *Proceedings of the National Academy of Sciences*, 110: E4658-E67.

Krishnamurthy, Vishnu Vardhan, John S. Khamo, Ellen Cho, Cara Schornak, and Kai Zhang. 2015. 'Polymerase chain reaction-based gene removal from plasmids', *Data in brief*, 4: 75-82.

Kruger, D. H., G. J. Barcak, M. Reuter, and H. O. Smith. 1988. 'EcoRII can be activated to cleave refractory DNA recognition sites', *Nucleic acids research*, 16: 3997-4008.

Kruger, D. H., and T. A. Bickle. 1983. 'Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts', *Microbiol Rev*, 47: 345-60.

Kumar, Suresh, Viswanathan Chinnusamy, and Trilochan Mohapatra. 2018. 'Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond', *Frontiers in Genetics*, 9.

Kutter, E. M., and J. S. Wiberg. 1969. 'Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4', *Journal of virology*, 4: 439-53.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. 'Bacteriophage resistance mechanisms', *Nat Rev Micro*, 8: 317-27.

Lepikhov, K., A. Tchernov, L. Zheleznaja, N. Matvienko, J. Walter, and T. A. Trautner. 2001. 'Characterization of the type IV restriction modification system BspLU11III from Bacillus sp. LU11', *Nucleic acids research*, 29: 4691-8.

Lewis, Dale, Phuoc Le, Chiara Zurla, Laura Finzi, and Sankar Adhya. 2011. 'Multilevel autoregulation of λ repressor protein CI by DNA looping in vitro', *Proceedings of the National Academy of Sciences*, 108: 14807-12.

Lewis, M. 2005. 'The lac repressor', *C R Biol*, 328: 521-48.

Lim, Derek H K, and Eamonn R Maher. 2010. 'DNA methylation: a form of epigenetic control of gene expression', *The Obstetrician & Gynaecologist*, 12: 37-42.

Little, J. W. 1984. 'Autodigestion of lexA and phage lambda repressors', *Proceedings of the National Academy of Sciences of the United States of America*, 81: 1375-79.

Liu, K., Y. Zhang, K. Severinov, A. Das, and M. M. Hanna. 1996. 'Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA', *The EMBO journal*, 15: 150-61.

Loenen, Wil A. M., David T. F. Dryden, Elisabeth A. Raleigh, and Geoffrey G. Wilson. 2014. 'Type I restriction enzymes and their relatives', *Nucleic acids research*, 42: 20-44.

Loenen, Wil A. M., David T. F. Dryden, Elisabeth A. Raleigh, Geoffrey G. Wilson, and Noreen E. Murray. 2013. 'Highlights of the DNA cutters: a short history of the restriction enzymes', *Nucleic acids research*, 42: 3-19.

Loenen, Wil A. M., and Elisabeth A. Raleigh. 2014. 'The other face of restriction: modification-dependent enzymes', *Nucleic acids research*, 42: 56-69.

Lopatina, A., N. Tal, and R. Sorek. 2020. 'Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy', *Annu Rev Virol*, 7: 371-84.

Luria, S. E., and M. L. Human. 1952. 'A nonhereditary, host-induced variation of bacterial viruses', *Journal of Bacteriology*, 64: 557-69.

Machnicka, Magdalena A., Katarzyna H. Kaminska, Stanislaw Dunin-Horkawicz, and Janusz M. Bujnicki. 2015. 'Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes', *BMC Bioinformatics*, 16: 336.

MacWilliams, M. P., and T. A. Bickle. 1996. 'Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene', *The EMBO journal*, 15: 4775-83.

Madeira, Fábio, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R. N. Tivey, Simon C. Potter, Robert D. Finn, and Rodrigo Lopez. 2019. 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic acids research*, 47: W636-W41.

Maghsoodi, Ameneh, Anupam Chatterjee, Ioan Andricioaei, and Noel C. Perkins. 2019. 'How the phage T4 injection machinery works including energetics, forces, and dynamic pathway', *Proceedings of the National Academy of Sciences*, 116: 25097-105.

Makarova, K. S., V. Anantharaman, N. V. Grishin, E. V. Koonin, and L. Aravind. 2014. 'CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems', *Front Genet*, 5: 102.

Makarova, Kira S., and Eugene V. Koonin. 2015. 'Annotation and Classification of CRISPR-Cas Systems', *Methods in molecular biology (Clifton, N.J.)*, 1311: 47-75.

Marino, Nicole D., Rafael Pinilla-Redondo, Bálint Csörgő, and Joseph Bondy-Denomy. 2020. 'Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies', *Nature Methods*, 17: 471-79.

Marinus, Martin G., and Josep Casadesus. 2009. 'Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more', *FEMS Microbiology Reviews*, 33: 488-503.

Marks, Phil, John McGeehan, Geoff Wilson, Neil Errington, and Geoff Kneale. 2003. 'Purification and characterisation of a novel DNA methyltransferase, M.AhdI', *Nucleic acids research*, 31: 2803-10.

Marshall, Jacqueline J. T., Darren M. Gowers, and Stephen E. Halford. 2007. 'Restriction Endonucleases that Bridge and Excise Two Recognition Sites from DNA', *Journal of Molecular Biology*, 367: 419-31.

Mathew, R., and D. Chatterji. 2006. 'The evolving story of the omega subunit of bacterial RNA polymerase', *Trends Microbiol*, 14: 450-5.

Matthews, B. W., D. H. Ohlendorf, W. F. Anderson, and Y. Takeda. 1982. 'Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor', *Proceedings of the National Academy of Sciences of the United States of America*, 79: 1428-32.

McClelland, S. E., and M. D. Szczelkun. 2004. 'The Type I and III Restriction Endonucleases: Structural Elements in Molecular Motors that Process DNA.' in Alfred M. Pingoud (ed.), *Restriction Endonucleases* (Springer Berlin Heidelberg: Berlin, Heidelberg).

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. 'Phaser crystallographic software', *J Appl Crystallogr*, 40: 658-74.

McKay, David B., and Thomas A. Steitz. 1981. 'Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA', *Nature*, 290: 744-49.

McKenzie, Joanna L., Jennifer Robson, Michael Berney, Tony C. Smith, Alaine Ruthe, Paul P. Gardner, Vickery L. Arcus, and Gregory M. Cook. 2012. 'A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria', *Journal of Bacteriology*, 194: 2189-204.

Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe. 1999. 'Food-related illness and death in the United States', *Emerg Infect Dis*, 5: 607-25.

Mellén, Marian, Pinar Ayata, Scott Dewell, Skirmantas Kriaucionis, and Nathaniel Heintz. 2012. 'MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system', *Cell*, 151: 1417-30.

Miller, Eric S., Elizabeth Kutter, Gisela Mosig, Fumio Arisaka, Takashi Kunisawa, and Wolfgang Rüger. 2003. 'Bacteriophage T4 genome', *Microbiology and molecular biology reviews : MMBR*, 67: 86-156.

Miller, J. H. 1972. *Experiments in Molecular Genetics* (Cold Spring Harbor: New York).

Mojica, Francisco J. M., Chc sar Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. 2005. 'Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements', *Journal of Molecular Evolution*, 60: 174-82.

Moore, Lisa D., Thuc Le, and Guoping Fan. 2013. 'DNA methylation and its basic function', *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 38: 23-38.

Morgan, Richard D., Tanya K. Bhatia, Lindsay Lovasco, and Theodore B. Davis. 2008. 'MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection', *Nucleic acids research*, 36: 6558-70.

Morgan, Richard D., Elizabeth A. Dwinell, Tanya K. Bhatia, Elizabeth M. Lang, and Yvette A. Luyten. 2009. 'The MmeI family: type II restriction–modification enzymes that employ single-strand modification for host protection', *Nucleic acids research*, 37: 5208-21.

Morgan, Richard D., Yvette A. Luyten, Samuel A. Johnson, Emily M. Clough, Tyson A. Clark, and Richard J. Roberts. 2016. 'Novel m4C modification in type I restriction-modification systems', *Nucleic acids research*, 44: 9413-25.

Morrow, John F., and Paul Berg. 1972. 'Cleavage of Simian Virus 40 DNA at a Unique Site by a Bacterial Restriction Enzyme', *Proceedings of the National Academy of Sciences*, 69: 3365-69.

Müller, Andreas U., Marc Leibundgut, Nenad Ban, and Eilika Weber-Ban. 2019. 'Structure and functional implications of WYL domain-containing bacterial DNA damage response regulator PafBC', *Nature Communications*, 10: 4653.

Murray, N. E. 2000. 'Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle)', *Microbiology and molecular biology reviews : MMBR*, 64: 412-34.

Nakagawa, H., F. Arisaka, and S. Ishii. 1985. 'Isolation and characterization of the bacteriophage T4 tail-associated lysozyme', *Journal of virology*, 54: 460-66.

Narberhaus, Franz, and Sylvia Balsiger. 2003. 'Structure-function studies of Escherichia coli RpoH (sigma32) by in vitro linker insertion mutagenesis', *Journal of Bacteriology*, 185: 2731-38.

NCBI. 2020. 'Monodnaviria'. https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2731342.

NEB. 2020. 'Golden Gate Assembly', New England Biolabs. https://international.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/golden-gate-assembly.

NEB. 2020. 'Types of Restriction Endonucleases,'. https://international.neb.com/products/restriction-endonucleases/restriction-endonucleases/types-of-restriction-endonucleases.

Nechaev, S., and K. Severinov. 2008. 'The elusive object of desire--interactions of bacteriophages and their hosts', *Curr Opin Microbiol*, 11: 186-93.

Nikolskaya, II, M. I. Tediashvili, N. G. Lopatina, T. G. Chanishvili, and S. S. Debov. 1979. 'Specificity and functions of guanine methylase of Shigella sonnei DDVI phage', *Biochim Biophys Acta*, 561: 232-9.

Nikolskaya, I. I., N. G. Lopatina, and S. S. Debov. 1976. 'Methylated guanine derivative as a minor base in the DNA of phage DDVI Shigella disenteriae', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 435: 206-10.

Niv, M. Y., D. R. Ripoll, J. A. Vila, A. Liwo, E. S. Vanamee, A. K. Aggarwal, H. Weinstein, and H. A. Scheraga. 2007. 'Topology of Type II REases revisited; structural classes and the common conserved core', *Nucleic acids research*, 35: 2227-37.

Ofir, Gal, Sarah Melamed, Hila Sberro, Zohar Mukamel, Shahar Silverman, Gilad Yaakov, Shany Doron, and Rotem Sorek. 2018. 'DISARM is a widespread bacterial defence system with broad anti-phage activities', *Nature microbiology*, 3: 90-98.

Orhanović, S., and M. Pavela-Vrancic. 2003. 'Dimer asymmetry and the catalytic cycle of alkaline phosphatase from Escherichia coli', *Eur J Biochem*, 270: 4356-64.

Orhanović, Stjepan, and Maja Pavela-Vrančič. 2003. 'Dimer asymmetry and the catalytic cycle of alkaline phosphatase from Escherichia coli', *European Journal of Biochemistry*, 270: 4356-64.

Owen, Siân V., Nicolas Wenner, Rocío Canals, Angela Makumi, Disa L. Hammarlöf, Melita A. Gordon, Abram Aertsen, Nicholas A. Feasey, and Jay C. D. Hinton. 2017. 'Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1', *Frontiers in microbiology*, 8: 235-35.

Pandey, Deo Prakash, and Kenn Gerdes. 2005. 'Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes', *Nucleic acids research*, 33: 966-76.

Panne, D., S. A. Müller, S. Wirtz, A. Engel, and T. A. Bickle. 2001. 'The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides', *The EMBO journal*, 20: 3210-17.

Parsons, R. J., M. Breitbart, M. W. Lomas, and C. A. Carlson. 2012. 'Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea', *Isme j*, 6: 273-84.

Payelleville, Amaury, Ludovic Legrand, Jean-Claude Ogier, Céline Roques, Alain Roulet, Olivier Bouchez, Annabelle Mouammine, Alain Givaudan, and Julien Brillard. 2018. 'The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated Adenines', *Scientific reports*, 8: 12091.

Pedruzzi, I., J. P. Rosenbusch, and K. P. Locher. 1998. 'Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein', *FEMS Microbiol Lett*, 168: 119-25.

Pei, Jimin, Bong-Hyun Kim, and Nick V. Grishin. 2008. 'PROMALS3D: a tool for multiple protein sequence and structure alignments', *Nucleic acids research*, 36: 2295-300.

Pfeifer, Gerd P., Wenying Xiong, Maria A. Hahn, and Seung-Gi Jin. 2014. 'The role of 5-hydroxymethylcytosine in human cancer', *Cell and tissue research*, 356: 631-41.

Piekarowicz, A., M. Golaszewska, A. O. Sunday, M. Siwinska, and D. C. Stein. 1999. 'The HaeIV restriction modification system of Haemophilus aegyptius is encoded by a single polypeptide', *J Mol Biol*, 293: 1055-65.

Pingoud, A., M. Fuxreiter, V. Pingoud, and W. Wende. 2005. 'Type II restriction endonucleases: structure and mechanism', *Cell Mol Life Sci*, 62: 685-707.

Pingoud, Alfred, Geoffrey G. Wilson, and Wolfgang Wende. 2014. 'Type II restriction endonucleases--a historical perspective and more', *Nucleic acids research*, 42: 7489-527.

Price, A. R. 1974. 'Bacteriophage PBS2-induced deoxycytidine triphosphate deaminase in Bacillus subtilis', *Journal of virology*, 14: 1314-7.

Qimron, U., B. Marintcheva, S. Tabor, and C. C. Richardson. 2006. 'Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 19039-44.

Rajamanickam, Karthic, and Sidney Hayes. 2018. 'The Bacteriophage Lambda CII Phenotypes for Complementation, Cellular Toxicity and Replication Inhibition Are Suppressed in cII-oop Constructs Expressing the Small RNA OOP', *Viruses*, 10: 115.

Rao, Desirazu N., David T. F. Dryden, and Shivakumara Bheemanaik. 2014. 'Type III restriction-modification enzymes: a historical perspective', *Nucleic acids research*, 42: 45-55.

Rao, Venigalla B., and Lindsay W. Black. 2005. 'DNA Packaging in Bacteriophage T4.' in, *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism* (Springer US: Boston, MA).

Rauch, B. J., M. R. Silvis, J. F. Hultquist, C. S. Waters, M. J. McGregor, N. J. Krogan, and J. Bondy-Denomy. 2017. 'Inhibition of CRISPR-Cas9 with Bacteriophage Proteins', *Cell*, 168: 150-58.e10.

Rauhut, R., and G. Klug. 1999. 'mRNA degradation in bacteria', *FEMS Microbiol Rev*, 23: 353-70.

Revel, Helen R., and C. P. Georgopoulos. 1969. 'Restriction of nonglucosylated T-even bacteriophages by prophage P1', *Virology*, 39: 1-17.

Reyes-Robles, T., R. S. Dillard, L. S. Cairns, C. A. Silva-Valenzuela, M. Housman, A. Ali, E. R. Wright, and A. Camilli. 2018. 'Vibrio cholerae Outer Membrane Vesicles Inhibit Bacteriophage Infection', *J Bacteriol*, 200.

Rifat, D., N. T. Wright, K. M. Varney, D. J. Weber, and L. W. Black. 2008a. 'Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target', *J Mol Biol*, 375: 720-34.

Rifat, Dalin, Nathan Wright, Kristen Varney, David Weber, and Lindsay Black. 2008b. 'Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target', *Journal of Molecular Biology*, 375: 720-34.

Roa, M. 1979. 'Interaction of bacteriophage K10 with its receptor, the lamB protein of Escherichia coli', *J Bacteriol*, 140: 680-6.

Roberts, Gareth A., Augoustinos S. Stephanou, Nisha Kanwar, Angela Dawson, Laurie P. Cooper, Kai Chen, Margaret Nutley, Alan Cooper, Garry W. Blakely, and David T. F. Dryden. 2012. 'Exploring the DNA mimicry of the Ocr protein of phage T7', *Nucleic acids research*, 40: 8129-43.

Roberts, Richard J., Marlene Belfort, Timothy Bestor, Ashok S. Bhagwat, Thomas A. Bickle, Jurate Bitinaite, Robert M. Blumenthal, Sergey Kh Degtyarev, David T. F. Dryden, Kevin Dybvig, Keith Firman, Elizaveta S. Gromova, Richard I. Gumport, Stephen E. Halford, Stanley Hattman, Joseph Heitman, David P. Hornby, Arvydas Janulaitis, Albert Jeltsch, Jytte Josephsen, Antal Kiss, Todd R. Klaenhammer, Ichizo Kobayashi, Huimin Kong, Detlev H. Krüger, Sanford Lacks, Martin G. Marinus, Michiko Miyahara, Richard D. Morgan, Noreen E. Murray, Valakunja Nagaraja, Andrzej Piekarowicz, Alfred Pingoud, Elisabeth Raleigh, Desirazu N. Rao, Norbert Reich, Vladimir E. Repin, Eric U. Selker, Pang-Chui Shaw, Daniel C. Stein, Barry L. Stoddard, Waclaw Szybalski, Thomas A. Trautner, James L. Van Etten, Jorge M. B. Vitor, Geoffrey G. Wilson, and Shuang-yong Xu. 2003. 'A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes', *Nucleic acids research*, 31: 1805-12.

Roberts, Richard J., Tamas Vincze, Janos Posfai, and Dana Macelis. 2015. 'REBASE--a database for DNA restriction and modification: enzymes, genes and genomes', *Nucleic acids research*, 43: D298-D99.

Rodríguez-Rodríguez, Diana Raquel, Ramiro Ramírez-Solís, Mario Alberto Garza-Elizondo, María De Lourdes Garza-Rodríguez, and Hugo Alberto Barrera-Saldaña. 2019. 'Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review)', *International journal of molecular medicine*, 43: 1559-74.

Rodwell, Ella V., Nicolas Wenner, Caisey V. Pulford, Yueyi Cai, Arthur Bowers-Barnard, Alison Beckett, Jonathan Rigby, David M. Picton, Tim R. Blower, Nicholas A. Feasey, Jay C. D. Hinton, and Blanca M. Perez-Sepulveda. 2021. 'Isolation and characterisation of bacteriophages with activity against invasive non-typhoidal *Salmonella* causing bloodstream infection in Malawi', *Viruses*, 13: 478.

Rohwer, Forest, and Anca M. Segall. 2015. 'A century of phage lessons', *Nature*, 528: 46-47.

Rojo, F. 1999. 'Repression of transcription initiation in bacteria', *Journal of Bacteriology*, 181: 2987-91.

Rousset, François, Julien Dowding, Aude Bernheim, Eduardo P.C. Rocha, and David Bikard. 2021. 'Prophage-encoded hotspots of bacterial immune systems', *bioRxiv*: 2021.01.21.427644.

Safa, Ashrafus, G. Balakrish Nair, and Richard Y. C. Kong. 2010. 'Evolution of new variants of Vibrio cholerae O1', *Trends in Microbiology*, 18: 46-54.

Salmond, George P. C., and Peter C. Fineran. 2015. 'A century of the phage: past, present and future', *Nat Rev Micro*, 13: 777-86.

Sambrook, Joseph. 2001. *Molecular cloning : a laboratory manual* (Cold Spring Harbor, N.Y).

Samson, Julie E., Alfonso H. Magadan, Mourad Sabri, and Sylvain Moineau. 2013. 'Revenge of the phages: defeating bacterial defences', *Nat Rev Micro*, 11: 675-87.

Sander, Jeffry D., and J. Keith Joung. 2014. 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nat Biotech*, 32: 347-55.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, and M. Smith. 1977. 'Nucleotide sequence of bacteriophage φX174 DNA', *Nature*, 265: 687-95.

Sarkar-Banerjee, Suparna, Sachin Goyal, Ning Gao, John Mack, Benito Thompson, David Dunlap, Krishnananda Chattopadhyay, and Laura Finzi. 2018. 'Specifically bound lambda repressor dimers promote adjacent non-specific binding', *PLOS ONE*, 13: e0194930.

Sausset, R., M. A. Petit, V. Gaboriau-Routhiau, and M. De Paepe. 2020. 'New insights into intestinal phages', *Mucosal Immunology*, 13: 205-15.

Schaefer, Jorrit, Goran Jovanovic, Ioly Kotta-Loizou, and Martin Buck. 2016. 'Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader', *Analytical biochemistry*, 503: 56-57.

Schierling, Benno, Nadine Dannemann, Lilia Gabsalilow, Wolfgang Wende, Toni Cathomen, and Alfred Pingoud. 2012. 'A novel zinc-finger nuclease platform with a sequence-specific cleavage module', *Nucleic acids research*, 40: 2623-38.

Seed, Kimberley D. 2015. 'Battling Phages: How Bacteria Defend against Viral Attack', *PLoS Pathogens*, 11: e1004847.

Seidel, Ralf, Joost G. P. Bloom, Cees Dekker, and Mark D. Szczelkun. 2008. 'Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I', *The EMBO journal*, 27: 1388-98.

Severinov, K., M. Kashlev, E. Severinova, I. Bass, K. McWilliams, E. Kutter, V. Nikiforov, L. Snyder, and A. Goldfarb. 1994. 'A non-essential domain of Escherichia coli RNA polymerase required for the action of the termination factor Alc', *J Biol Chem*, 269: 14254-9.

Sharp, Phillip A., Bill Sugden, and Joe Sambrook. 1973. 'Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose-ethidium bromide electrophoresis', *Biochemistry*, 12: 3055-63.

Sheldrick, George. 2008. 'A Short History of ShelX', *Acta crystallographica. Section A, Foundations of crystallography*, 64: 112-22.

Shen, Betty W, Lindsey Doyle, Phil Bradley, Daniel F Heiter, Keith D Lunnen, Geoffrey G Wilson, and Barry L Stoddard. 2018. 'Structure, subunit organization and behavior of the asymmetric Type IIT restriction endonuclease BbvCI', *Nucleic acids research*, 47: 450-67.

Shen, Betty W., Derrick Xu, Siu-Hong Chan, Yu Zheng, Zhenyu Zhu, Shuang-yong Xu, and Barry L. Stoddard. 2011. 'Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI', *Nucleic acids research*, 39: 8223-36.

Short, Francesca L., Chidiebere Akusobi, William R. Broadhurst, and George P. C. Salmond. 2018. 'The bacterial Type III toxin-antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent antiviral abortive infection activity', *Scientific reports*, 8: 1013-13.

Sistla, S., and D. N. Rao. 2004. 'S-Adenosyl-L-methionine-dependent restriction enzymes', *Crit Rev Biochem Mol Biol*, 39: 1-19.

Snyder, Larry. 1995. 'Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents?', *Molecular Microbiology*, 15: 415-20.

Sommer, N., R. Depping, M. Piotrowski, and W. Ruger. 2004. 'Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism', *Biochem Biophys Res Commun*, 323: 809-15.

Sommer, N., R. Depping, M. Piotrowski, and W. Rüger. 2004. 'Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism', *Biochem Biophys Res Commun*, 323: 809-15.

Song, Chun-Xiao, Senlin Yin, Li Ma, Amanda Wheeler, Yu Chen, Yan Zhang, Bin Liu, Junjie Xiong, Weihan Zhang, Jiankun Hu, Zongguang Zhou, Biao Dong, Zhiqi Tian, Stefanie S. Jeffrey, Mei-Sze Chua, Samuel So, Weimin Li, Yuquan Wei, Jiajie Diao, Dan Xie, and Stephen R. Quake. 2017. '5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages', *Cell research*, 27: 1231-42.

Song, Sooyeon, and Thomas K. Wood. 2020. 'A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition', *Frontiers in microbiology*, 11.

Souther, A., R. Bruner, and J. Elliott. 1972. 'Degradation of Escherichia coli chromosome after infection by bacteriophage T4: role of bacteriophage gene D2a', *Journal of virology*, 10: 979-84.

Srikumar, Shabarinath, Carsten Kröger, Magali Hébrard, Aoife Colgan, Siân V. Owen, Sathesh K. Sivasankaran, Andrew D. S. Cameron, Karsten Hokamp, and Jay C. D. Hinton. 2015. 'RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium', *PLoS Pathogens*, 11: e1005262.

Stanaway, Jeffrey D., Andrea Parisi, Kaushik Sarkar, Brigette F. Blacker, Robert C. Reiner, Simon I. Hay, Molly R. Nixon, Christiane Dolecek, Spencer L. James, Ali H. Mokdad, Getaneh Abebe, Elham Ahmadian, Fares Alahdab, Birhan Tamene T. Alemnew, Vahid Alipour, Fatemeh Allah Bakeshei, Megbaru Debalkie Animut, Fereshteh Ansari, Jalal Arabloo, Ephrem Tsegay Asfaw, Mojtaba Bagherzadeh, Quique Bassat, Yaschilal Muche Muche Belayneh, Félix Carvalho, Ahmad Daryani, Feleke Mekonnen Demeke, Asmamaw Bizuneh Bizuneh Demis, Manisha Dubey, Eyasu Ejeta Duken, Susanna J. Dunachie, Aziz Eftekhari, Eduarda Fernandes, Reza Fouladi Fard, Getnet Azeze Gedefaw, Birhanu Geta, Katherine B. Gibney, Amir Hasanzadeh, Chi Linh Hoang, Amir Kasaeian, Amir Khater, Zelalem Teklemariam Kidanemariam, Ayenew Molla Lakew, Reza Malekzadeh, Addisu Melese, Desalegn Tadese Mengistu, Tomislav Mestrovic, Bartosz Miazgowski, Karzan Abdulmuhsin Mohammad, Mahdi Mohammadian, Abdollah Mohammadian-Hafshejani, Cuong Tat Nguyen, Long Hoang Nguyen, Son Hoang Nguyen, Yirga Legesse Nirayo, Andrew T. Olagunju, Tinuke O. Olagunju, Hadi Pourjafar, Mostafa Qorbani, Mohammad Rabiee, Navid Rabiee, Anwar Rafay, Aziz Rezapour, Abdallah M. Samy, Sadaf G. Sepanlou, Masood Ali Shaikh, Mehdi Sharif, Mika Shigematsu, Belay Tessema, Bach Xuan Tran, Irfan Ullah, Ebrahim M. Yimer, Zoubida Zaidi, Christopher J. L. Murray, and John A. Crump. 2019. 'The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet Infectious Diseases*, 19: 1312-24.

Stanley, Sabrina Y., and Karen L. Maxwell. 2018. 'Phage-Encoded Anti-CRISPR Defenses', *Annual Review of Genetics*, 52: 445-64.

Steczkiewicz, Kamil, Anna Muszewska, Lukasz Knizewski, Leszek Rychlewski, and Krzysztof Ginalski. 2012. 'Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily', *Nucleic acids research*, 40: 7016-45.

Stephenson, Stacy Ann-Marie, and Paul D. Brown. 2016. 'Epigenetic Influence of Dam Methylation on Gene Expression and Attachment in Uropathogenic Escherichia coli', *Frontiers in Public Health*, 4.

Stern, A., and R. Sorek. 2011a. 'The phage-host arms race: shaping the evolution of microbes', *Bioessays*, 33: 43-51.

Stern, Adi, and Rotem Sorek. 2011b. 'The phage-host arms-race: Shaping the evolution of microbes', *Bioessays*, 33: 43-51.

Stewart, F. J., D. Panne, T. A. Bickle, and E. A. Raleigh. 2000a. 'Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme', *J Mol Biol*, 298: 611-22.

Stewart, F. J., and E. A. Raleigh. 1998. 'Dependence of McrBC cleavage on distance between recognition elements', *Biological chemistry*, 379: 611-16.

Stewart, Fiona J., Daniel Panne, Thomas A. Bickle, and Elisabeth A. Raleigh. 2000b. 'Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme 11Edited by J. Karn', *Journal of Molecular Biology*, 298: 611-22.

Sulakvelidze, Alexander. 2011. 'Bacteriophage: A new journal for the most ubiquitous organisms on Earth', *Bacteriophage*, 1: 1-2.

Sumby, Paul, and Margaret C. M. Smith. 2002. 'Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2)', *Molecular Microbiology*, 44: 489-500.

Summers, William C. 2001. 'Bacteriophage Therapy', *Annual Review of Microbiology*, 55: 437-51.

Suttle, Curtis A. 2007. 'Marine viruses — major players in the global ecosystem', *Nature Reviews Microbiology*, 5: 801-12.

Takahashi, Noriko, Yasuhiro Naito, Naofumi Handa, and Ichizo Kobayashi. 2002. 'A DNA Methyltransferase Can Protect the Genome from Postdisturbance Attack by a Restriction-Modification Gene Complex', *Journal of Bacteriology*, 184: 6100-08.

Tamulaitis, G., G. Sasnauskas, M. Mucke, and V. Siksnys. 2006. 'Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease', *J Mol Biol*, 358: 406-19.

Thomsen, Nathan D, and James M. Berger. 2008. 'Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases', *Molecular Microbiology*, 69: 1071-90.

Thomsen, Nathan D., and James M. Berger. 2012. 'Crystallization and X-ray structure determination of an RNA-dependent hexameric helicase', *Methods in enzymology*, 511: 171-90.

Thursby, Elizabeth, and Nathalie Juge. 2017. 'Introduction to the human gut microbiota', *The Biochemical journal*, 474: 1823-36.

Tripathi, Lakshmi, Yan Zhang, and Zhanglin Lin. 2014. 'Bacterial Sigma Factors as Targets for Engineered or Synthetic Transcriptional Control', *Frontiers in Bioengineering and Biotechnology*, 2.

Usón, Isabel, and George M. Sheldrick. 2018. 'An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features', *Acta crystallographica. Section D, Structural biology*, 74: 106-16.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, and G. N. Murshudov. 2004. 'REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta crystallographica. Section D, Biological crystallography*, 60: 2184-95.

van Miert, A. S. 1994. 'The sulfonamide-diaminopyrimidine story', *J Vet Pharmacol Ther*, 17: 309-16.

van Noort, John, Thijn van der Heijden, Christina F. Dutta, Keith Firman, and Cees Dekker. 2004. 'Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion', *Nucleic acids research*, 32: 6540-47.

Vanamee, Éva Scheuring, Sandro Santagata, and Aneel K. Aggarwal. 2001. 'FokI requires two specific DNA sites for cleavage11Edited by T. Richmond', *Journal of Molecular Biology*, 309: 69-78.

Vasu, Kommireddy, and Valakunja Nagaraja. 2013. 'Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense', *Microbiology and molecular biology reviews : MMBR*, 77: 53-72.

Venkataraman, Sangita, Burra V. L. S. Prasad, and Ramasamy Selvarajan. 2018. 'RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution', *Viruses*, 10: 76.

Vidakovic, L., P. K. Singh, R. Hartmann, C. D. Nadell, and K. Drescher. 2018. 'Dynamic biofilm architecture confers individual and collective mechanisms of viral protection', *Nature microbiology*, 3: 26-31.

Vieux, Ellen F., Matthew L. Wohlever, James Z. Chen, Robert T. Sauer, and Tania A. Baker. 2013. 'Distinct quaternary structures of the AAA+ Lon protease control substrate degradation', *Proceedings of the National Academy of Sciences*, 110: E2002-E08.

Virus Metadata Resource. 2019. 'Virus Taxonomy: 2019 Release'. https://talk.ictvonline.org/taxonomy/.

Vitkute, J., Z. Maneliene, M. Petrusyte, and A. Janulaitis. 1997. 'BplI, a new BcgI-like restriction endonuclease, which recognizes a symmetric sequence', *Nucleic acids research*, 25: 4444-6.

Vlot, Marnix, Joep Houkes, Silke J. A. Lochs, Daan C. Swarts, Peiyuan Zheng, Tim Kunne, Prarthana Mohanraju, Carolin Anders, Martin Jinek, John van der Oost, Mark J. Dickman, and Stan J. J. Brouns. 2018. 'Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes', *Nucleic acids research*, 46: 873-85.

Waite-Rees, P. A., C. J. Keating, L. S. Moran, B. E. Slatko, L. J. Hornstra, and J. S. Benner. 1991. 'Characterization and expression of the Escherichia coli Mrr restriction system', *Journal of Bacteriology*, 173: 5207-19.

Walling, Lauren R., and J. Scott Butler. 2018. 'Homologous VapC Toxins Inhibit Translation and Cell Growth by Sequence-Specific Cleavage of tRNA^{fMet}', *Journal of Bacteriology*, 200: e00582-17.

Wang, N., S. Gottesman, M. C. Willingham, M. M. Gottesman, and M. R. Maurizi. 1993. 'A human mitochondrial ATP-dependent protease that is highly homologous to bacterial Lon protease', *Proceedings of the National Academy of Sciences of the United States of America*, 90: 11247-51.

Wang, Z., and D. W. Mosbaugh. 1989. 'Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase', *J Biol Chem*, 264: 1163-71.

Wei, Hua, Guojie Zhao, Tianyu Hu, Suming Tang, Jiquan Jiang, Bo Hu, and Yifu Guan. 2016. 'Mapping the nicking efficiencies of nickase R.BbvCI for side-specific LNA-substituted substrates using rolling circle amplification', *Scientific reports*, 6: 32560-60.

Weigele, Peter, and Elisabeth A. Raleigh. 2016. 'Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses', *Chemical Reviews*, 116: 12655-87.

Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. 'Prokaryotes: the unseen majority', *Proceedings of the National Academy of Sciences of the United States of America*, 95: 6578-83.

Williams, S. A., and S. E. Halford. 2001. 'SfiI endonuclease activity is strongly influenced by the non-specific sequence in the middle of its recognition site', *Nucleic acids research*, 29: 1476-83.

Winn, Martyn D., Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei Vagin, and Keith S. Wilson. 2011. 'Overview of the CCP4 suite and current developments', *Acta crystallographica. Section D, Biological crystallography*, 67: 235-42.

Wyatt, G. R., and S. S. Cohen. 1953. 'The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine', *The Biochemical journal*, 55: 774-82.

Xiong, Xiaolin, Geng Wu, Yue Wei, Liqiong Liu, Yubing Zhang, Rui Su, Xianyue Jiang, Mengxue Li, Haiyan Gao, Xihao Tian, Yizhou Zhang, Li Hu, Si Chen, You Tang, Susu Jiang, Ruolin Huang, Zhiqiang Li, Yunfu Wang, Zixin Deng, Jiawei Wang, Peter C. Dedon, Shi Chen, and Lianrong Wang. 2020. 'SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities', *Nature microbiology*, 5: 917-28.

Xu, S. Y., A. R. Corvaglia, S. H. Chan, Y. Zheng, and P. Linder. 2011. 'A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300', *Nucleic acids research*, 39: 5597-610.

Yamaichi, Y., and H. Niki. 2000. 'Active segregation by the Bacillus subtilis partitioning system in Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 97: 14656-61.

Yap, Moh Lan, and Michael G. Rossmann. 2014. 'Structure and function of bacteriophage T4', *Future microbiology*, 9: 1319-27.

Yousuf, Shehla, Joyce E. Karlinsey, Stephanie L. Neville, Christopher A. McDevitt, Stephen J. Libby, Ferric C. Fang, and Elaine R. Frawley. 2020. 'Manganese import protects Salmonella enterica serovar Typhimurium against nitrosative stress', *Metallomics*, 12: 1791-801.

Yuan, Z., A. Riera, L. Bai, J. Sun, S. Nandi, C. Spanos, Z. A. Chen, M. Barbon, J. Rappsilber, B. Stillman, C. Speck, and H. Li. 2017. 'Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1', *Nat Struct Mol Biol*, 24: 316-24.

Zabet, Nicolae Radu, Marco Catoni, Filippo Prischi, and Jerzy Paszkowski. 2017. 'Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies', *Nucleic acids research*, 45: 3777-84.

Zahran, Mai, Tomasz Berezniak, Petra Imhof, and Jeremy C. Smith. 2011. 'Role of magnesium ions in DNA recognition by the EcoRV restriction endonuclease', *FEBS Letters*, 585: 2739-43.

Zhang, Yadong, Zhewen Zhang, Hao Zhang, Yongbing Zhao, Zaichao Zhang, and Jingfa Xiao. 2019. 'PADS Arsenal: a database of prokaryotic defense systems related genes', *Nucleic acids research*, 48: D590-D98.

Zhao, Guojie, Jun Li, Zhaoxue Tong, Bin Zhao, Runqing Mu, and Yifu Guan. 2013. 'Enzymatic Cleavage of Type II Restriction Endonucleases on the 2′-O-Methyl Nucleotide and Phosphorothioate Substituted DNA', *PLOS ONE*, 8: e79415.

Zheng, Yu, Devora Cohen-Karni, Derrick Xu, Hang Gyeong Chin, Geoffrey Wilson, Sriharsa Pradhan, and Richard J. Roberts. 2010. 'A unique family of Mrr-like modification-dependent restriction endonucleases', *Nucleic acids research*, 38: 5527-34.

Zhou, Yan Ning, Lucyna Lubkowska, Monica Hui, Carolyn Court, Shuo Chen, Donald L. Court, Jeffrey Strathern, Ding Jun Jin, and Mikhail Kashlev. 2013. 'Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in Escherichia coli', *The Journal of biological chemistry*, 288: 2700-10.

Zou, N., S. Ditty, B. Li, and S. C. Lo. 2003. 'Random priming PCR strategy to amplify and clone trace amounts of DNA', *Biotechniques*, 35: 758-60, 62-5.

Abreu, Afonso G., and Angela S. Barbosa. 2017. 'How Escherichia coli Circumvent Complement-Mediated Killing', *Frontiers in Immunology*, 8.

Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. 2010. 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta crystallographica. Section D, Biological crystallography*, 66: 213-21.

Adesina, Tomilola, Obinna Nwinyi, Nandita De, Olayemi Akinnola, and Emmanuel Omonigbehin. 2019. 'First Detection of Carbapenem-Resistant Escherichia fergusonii Strains Harbouring Beta-Lactamase Genes from Clinical Samples', *Pathogens*, 8: 164.

Adhikari, Satish, and Patrick D. Curtis. 2016. 'DNA methyltransferases and epigenetic regulation in bacteria', *FEMS Microbiology Reviews*, 40: 575-91.

Agnes Ullmann. 2009. 'Escherichia coli Lactose Operon', *eLS*.

Ali, Ibraheem, Ryan J. Conrad, Eric Verdin, and Melanie Ott. 2018. 'Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics', *Chemical Reviews*, 118: 1216-52.

Allet, Bernard, and Ahmad I. Bukhari. 1975. 'Analysis of bacteriophage Mu and λ-Mu hybrid DNAs by specific endonucleases', *Journal of Molecular Biology*, 92: 529-40.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. 'Basic local alignment search tool', *J Mol Biol*, 215: 403-10.

Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. 1981. 'Structure of the cro repressor from bacteriophage lambda and its interaction with DNA', *Nature*, 290: 754-8.

Anton, Brian P., Emmanuel F. Mongodin, Sonia Agrawal, Alexey Fomenkov, Devon R. Byrd, Richard J. Roberts, and Elisabeth A. Raleigh. 2015. 'Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of Escherichia coli K-12', *PLOS ONE*, 10: e0127446.

Aravind, L., K. S. Makarova, and E. V. Koonin. 2000. 'SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories', *Nucleic acids research*, 28: 3417-32.

Aslanidis, C., and P. J. de Jong. 1990. 'Ligation-independent cloning of PCR products (LIC-PCR)', *Nucleic acids research*, 18: 6069-74.

Ballandras-Colas, Allison, Monica Brown, Nicola J. Cook, Tamaria G. Dewdney, Borries Demeler, Peter Cherepanov, Dmitry Lyumkis, and Alan N. Engelman. 2016. 'Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function', *Nature*, 530: 358-61.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. 2007. 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315: 1709-12.

Barrangou, Rodolphe, and John van der Oost. 2015. 'Bacteriophage exclusion, a new defense system', *The EMBO journal*, 34: 134-35.

Beaulaurier, John, Eric E. Schadt, and Gang Fang. 2019. 'Deciphering bacterial epigenomes using modern sequencing technologies', *Nature reviews. Genetics*, 20: 157-72.

Bernheim, Aude, and Rotem Sorek. 2020. 'The pan-immune system of bacteria: antiviral defence as a community resource', *Nature Reviews Microbiology*, 18: 113-19.

Bertani, G., and J. J. Weigle. 1953. 'Host controlled variation in bacterial viruses', *Journal of Bacteriology*, 65: 113-21.

Bervoets, Indra, and Daniel Charlier. 2019. 'Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology', *FEMS Microbiology Reviews*, 43: 304-39.

Bickle, Thomas A. 2004. 'Restricting restriction', *Molecular Microbiology*, 51: 3-5.

Blank, Kathrin, Michael Hensel, and Roman G. Gerlach. 2011. 'Rapid and Highly Efficient Method for Scarless Mutagenesis within the Salmonella enterica Chromosome', *PLOS ONE*, 6: e15763.

Blower, Tim R., Terry J. Evans, Rita Przybilski, Peter C. Fineran, and George P. C. Salmond. 2012. 'Viral Evasion of a Bacterial Suicide System by RNA–Based Molecular Mimicry Enables Infectious Altruism', *PLOS Genetics*, 8: e1003023.

Blower, Tim R., Francesca L. Short, Peter C. Fineran, and George P. C. Salmond. 2012. 'Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication', *Bacteriophage*, 2: 234-38.

Bourniquel, A. A., and T. A. Bickle. 2002. 'Complex restriction enzymes: NTP-driven molecular motors', *Biochimie*, 84: 1047-59.

Boyer, H. W. 1971. 'DNA restriction and modification mechanisms in bacteria', *Annu Rev Microbiol*, 25: 153-76.

Braga, Lucas P. P., Aymé Spor, Witold Kot, Marie-Christine Breuil, Lars H. Hansen, João C. Setubal, and Laurent Philippot. 2020. 'Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios', *Microbiome*, 8: 52.

Bragg, Robert, Wouter van der Westhuizen, Ji-Yun Lee, Elke Coetsee, and Charlotte Boucher. 2014. 'Bacteriophages as Potential Treatment Option for Antibiotic Resistant Bacteria.' in Rameshwar Adhikari and Santosh Thapa (eds.), *Infectious Diseases and Nanomedicine I: First International Conference (ICIDN – 2012), Dec. 15-18, 2012, Kathmandu, Nepal* (Springer India: New Delhi).

Brickner, M., and J. Chmielewski. 1998. 'Inhibiting the dimeric restriction endonuclease EcoRI using interfacial helical peptides', *Chem Biol*, 5: 339-43.

Bryson, Alexandra L., Young Hwang, Scott Sherrill-Mix, Gary D. Wu, James D. Lewis, Lindsay Black, Tyson A. Clark, and Frederic D. Bushman. 2015. 'Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9', *mBio*, 6: e00648-15.

Burch, Lauranell H., Leilei Zhang, Frank G. Chao, Hong Xu, and John W. Drake. 2011. 'The Bacteriophage T4 Rapid-Lysis Genes and Their Mutational Proclivities', *Journal of Bacteriology*, 193: 3537-45.

Butterer, Annika, Christian Pernstich, Rachel M. Smith, Frank Sobott, Mark D. Szczelkun, and Júlia Tóth. 2014. 'Type III restriction endonucleases are heterotrimeric: comprising one helicase–nuclease subunit and a dimeric methyltransferase that binds only one specific DNA', *Nucleic acids research*, 42: 5139-50.

Callahan, Scott J., Yvette A. Luyten, Yogesh K. Gupta, Geoffrey G. Wilson, Richard J. Roberts, Richard D. Morgan, and Aneel K. Aggarwal. 2016. 'Structure of Type IIL

Restriction-Modification Enzyme MmeI in Complex with DNA Has Implications for Engineering New Specificities', *PLOS Biology*, 14: e1002442.

Casadesus, J., and D. Low. 2006. 'Epigenetic gene regulation in the bacterial world', *Microbiology and molecular biology reviews : MMBR*, 70: 830-56.

Chatterjee, Sujoy, and Eli Rothenberg. 2012. 'Interaction of bacteriophage l with its E. coli receptor, LamB', *Viruses*, 4: 3162-78.

Chen, Bo-Wei, Ming-Hsing Lin, Chen-Hsi Chu, Chia-En Hsu, and Yuh-Ju Sun. 2015. 'Insights into ParB spreading from the complex structure of Spo0J and parS', *Proceedings of the National Academy of Sciences*, 112: 6613-18.

Cheng, Xiaodong, and Richard J. Roberts. 2001. 'AdoMet-dependent methylation, DNA methyltransferases and base flipping', *Nucleic acids research*, 29: 3784-95.

Chopin, M. C., A. Chopin, and E. Bidnenko. 2005. 'Phage abortive infection in lactococci: variations on a theme', *Curr Opin Microbiol*, 8: 473-9.

Coffey, A., and R. P. Ross. 2002. 'Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application', *Antonie Van Leeuwenhoek*, 82: 303-21.

Cohen, D., S. Melamed, A. Millman, G. Shulman, Y. Oppenheimer-Shaanan, A. Kacen, S. Doron, G. Amitai, and R. Sorek. 2019. 'Cyclic GMP-AMP signalling protects bacteria against viral infection', *Nature*, 574: 691-95.

Cohen, Stanley N., Annie C. Y. Chang, Herbert W. Boyer, and Robert B. Helling. 1973. 'Construction of Biologically Functional Bacterial Plasmids In Vitro', *Proceedings of the National Academy of Sciences*, 70: 3240-44.

Cooper, Laurie P., Gareth A. Roberts, John H. White, Yvette A. Luyten, Edward K. M. Bower, Richard D. Morgan, Richard J. Roberts, Jodi A. Lindsay, and David T. F. Dryden. 2017. 'DNA target recognition domains in the Type I restriction and modification systems of Staphylococcus aureus', *Nucleic acids research*, 45: 3395-406.

Cowling, Victoria H. 2009. 'Regulation of mRNA cap methylation', *The Biochemical journal*, 425: 295-302.

Cowtan, K. 2006. 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta crystallographica. Section D, Biological crystallography*, 62: 1002-11.

Creixell, Pau, Erwin M. Schoof, Chris Soon Heng Tan, and Rune Linding. 2012. 'Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367: 2584-93.

Cymerman, I. A., A. Obarska, K. J. Skowronek, A. Lubys, and J. M. Bujnicki. 2006. 'Identification of a new subfamily of HNH nucleases and experimental characterization of a representative member, HphI restriction endonuclease', *Proteins*, 65: 867-76.

Datsenko, K. A., and B. L. Wanner. 2000. 'One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97: 6640-5.

Derous, Veerle, Francine Deboeck, Jean-Pierre Hernalsteens, and Henri De Greve. 2011. 'Reproducible gene targeting in recalcitrant Escherichia coli isolates', *BMC Research Notes*, 4: 213.

Dryden, D. T., N. E. Murray, and D. N. Rao. 2001. 'Nucleoside triphosphate-dependent restriction enzymes', *Nucleic acids research*, 29: 3728-41.

Dussoix, D., and W. Arber. 1962. 'Host specificity of DNA produced by Escherichia coli. II. Control over acceptance of DNA from infecting phage lambda', *J Mol Biol*, 5: 37-49.

Dy, R. L., L. A. Rigano, and P. C. Fineran. 2018. 'Phage-based biocontrol strategies and their application in agriculture and aquaculture', *Biochem Soc Trans*, 46: 1605-13.

Ellis, E. L., and M. Delbrück. 1939. 'The growth of bacteriophage', *The Journal of general physiology*, 22: 365-84.

Emsley, P., and K. Cowtan. 2004. 'Coot: model-building tools for molecular graphics', *Acta crystallographica. Section D, Biological crystallography*, 60: 2126-32.

Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. 2008. 'A One Pot, One Step, Precision Cloning Method with High Throughput Capability', *PLOS ONE*, 3: e3647.

Erni, Bernhard. 2006. 'The Mannose Transporter Complex: an Open Door for the Macromolecular Invasion of Bacteria', *Journal of Bacteriology*, 188: 7036-38.

Feder, M., and J. M. Bujnicki. 2005. 'Identification of a new family of putative PD-(D/E)XK nucleases with unusual phylogenomic distribution and a new type of the active site', *BMC Genomics*, 6: 21.

Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. 1976. 'Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene', *Nature*, 260: 500-7.

Fineran, Peter C., Tim R. Blower, Ian J. Foulds, David P. Humphreys, Kathryn S. Lilley, and George P. C. Salmond. 2009. 'The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair', *Proceedings of the National Academy of Sciences of the United States of America*, 106: 894-99.

Flick, K. E., M. S. Jurica, R. J. Monnat, Jr., and B. L. Stoddard. 1998. 'DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI', *Nature*, 394: 96-101.

Fuller-Pace, F. V., and N. E. Murray. 1986. 'Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, 83: 9368-72.

Gajiwala, K. S., and S. K. Burley. 2000. 'Winged helix proteins', *Current opinion in structural biology*, 10: 110-16.

Gasiunas, Giedrius, Giedrius Sasnauskas, Gintautas Tamulaitis, Claus Urbanke, Dalia Razaniene, and Virginijus Siksnys. 2008. 'Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family', *Nucleic acids research*, 36: 938-49.

Goldfarb, T., H. Sberro, E. Weinstock, O. Cohen, S. Doron, Y. Charpak-Amikam, S. Afik, G. Ofir, and R. Sorek. 2015. 'BREX is a novel phage resistance system widespread in microbial genomes', *The EMBO journal*, 34: 169-83.

Golovenko, Dmitrij, Elena Manakova, Giedre Tamulaitiene, Saulius Grazulis, and Virginijus Siksnys. 2009. 'Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII', *Nucleic acids research*, 37: 6613-24.

Gong, T., M. Lu, X. Zhou, A. Zhang, B. Tang, J. Chen, M. Jing, and Y. Li. 2019. 'CRISPR-Cas Systems in Streptococci', *Curr Issues Mol Biol*, 32: 1-38.

Gordeeva, Julia, Natalya Morozova, Nicolas Sierro, Artem Isaev, Tomas Sinkunas, Ksenia Tsvetkova, Mikhail Matlashov, Lidija Truncaite, Richard D. Morgan, Nikolai V. Ivanov, Virgis Siksnys, Lanying Zeng, and Konstantin Severinov. 2019. 'BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site', *Nucleic acids research*, 47: 253-65.

Griffiths AJF, Miller JH, Suzuki DT, et al.,. 2000. *Introduction to Genetic Analysis.* (W. H. Freeman: New York).

Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. 1995. 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter', *Journal of Bacteriology*, 177: 4121-30.

Harrington, Lucas B., Kevin W. Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J. Knott, Alireza Edraki, Bianca Garcia, Nadia Amrani, Janice S. Chen, Joshua C. Cofsky, Philip J. Kranzusch, Erik J. Sontheimer, Alan R. Davidson, Karen L. Maxwell, and Jennifer A. Doudna. 2017. 'A Broad-Spectrum Inhibitor of CRISPR-Cas9', *Cell*, 170: 1224-33.e15.

Hayes, F. 2003. 'Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest', *Science*, 301: 1496-9.

He, Xinyi, Victoria Hull, Julie A. Thomas, Xiaoqing Fu, Sonal Gidwani, Yogesh K. Gupta, Lindsay W. Black, and Shuang-yong Xu. 2015. 'Expression and purification of a single-chain Type IV restriction enzyme Eco94GmrSD and determination of its substrate preference', *Scientific reports*, 5: 9747.

Hein, Stephanie, Ingeborg Scholz, Björn Voß, and Wolfgang R. Hess. 2013. 'Adaptation and modification of three CRISPR loci in two closely related cyanobacteria', *RNA biology*, 10: 852-64.

Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. 'The origins and ongoing evolution of viruses', *Trends Microbiol*, 8: 504-8.

Hernday, Aaron D., Bruce A. Braaten, and David A. Low. 2003. 'The Mechanism by which DNA Adenine Methylase and PapI Activate the Pap Epigenetic Switch', *Molecular Cell*, 12: 947-57.

Hershey, A. D., and M. Chase. 1952. 'Independent functions of viral protein and nucleic acid in growth of bacteriophage', *The Journal of general physiology*, 36: 39-56.

Hilbert, Brendan J., Janelle A. Hayes, Nicholas P. Stone, Caroline M. Duffy, Banumathi Sankaran, and Brian A. Kelch. 2015. 'Structure and mechanism of the ATPase that powers viral genome packaging', *Proceedings of the National Academy of Sciences of the United States of America*, 112: E3792-E99.

Hinton, D. M. 1991. 'Transcription from a bacteriophage T4 middle promoter using T4 motA protein and phage-modified RNA polymerase', *J Biol Chem*, 266: 18034-44.

Hinton, Deborah M. 2010. 'Transcriptional control in the prereplicative phase of T4 development', *Virology Journal*, 7: 289.

Høyland-Kroghsbo, Nina Molin. 2019. 'Cyclic Nucleotide Signaling: A Second Messenger of Death', *Cell Host & Microbe*, 26: 567-68.

Hui, Wenyan, Wenyi Zhang, Lai-Yu Kwok, Heping Zhang, Jian Kong, and Tiansong Sun. 2019. 'A Novel Bacteriophage Exclusion (BREX) System Encoded by the pglX Gene in Lactobacillus casei Zhang', *Applied and environmental microbiology*, 85: e01001-19.

Iida, Shigeru, Markus B. Streiff, Thomas A. Bickle, and Werner Arber. 1987. 'Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA− phages', *Virology*, 157: 156-66.

Interthal, H., J. J. Pouliot, and J. J. Champoux. 2001. 'The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily', *Proceedings of the National Academy of Sciences of the United States of America*, 98: 12009-14.

Irizarry, R. A., C. Ladd-Acosta, B. Carvalho, H. Wu, S. A. Brandenburg, J. A. Jeddeloh, B. Wen, and A. P. Feinberg. 2008. 'Comprehensive high-throughput arrays for relative methylation (CHARM)', *Genome Res*, 18: 780-90.

Isaev, Artem, Alena Drobiazko, Nicolas Sierro, Julia Gordeeva, Ido Yosef, Udi Qimron, Nikolai V Ivanov, and Konstantin Severinov. 2020. 'Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence', *Nucleic acids research*.

Ishino, Yoshizumi, Mart Krupovic, and Patrick Forterre. 2018. 'History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology', *Journal of Bacteriology*, 200: e00580-17.

Jablonska, Jagoda, Dorota Matelska, Kamil Steczkiewicz, and Krzysztof Ginalski. 2017. 'Systematic classification of the His-Me finger superfamily', *Nucleic acids research*, 45: 11479-94.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. 'Individual-specific ‘fingerprints’ of human DNA', *Nature*, 316: 76-79.

Jiang, Fuguo, and Jennifer A. Doudna. 2017. 'CRISPR–Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, 46: 505-29.

Jin, Zelin, and Yun Liu. 2018. 'DNA methylation in human diseases', *Genes & diseases*, 5: 1-8.

Jindrova, Eva, Stefanie Schmid-Nuoffer, Fabienne Hamburger, Pavel Janscak, and Thomas A. Bickle. 2005. 'On the DNA cleavage mechanism of Type I restriction enzymes', *Nucleic acids research*, 33: 1760-66.

Jurkowski, Tomasz P., Nils Anspach, Liliya Kulishova, Wolfgang Nellen, and Albert Jeltsch. 2007. 'The M.EcoRV DNA-(Adenine N6)-methyltransferase Uses DNA Bending for Recognition of an Expanded EcoDam Recognition Site', *Journal of Biological Chemistry*, 282: 36942-52.

Kaminska, Katarzyna, and Janusz Bujnicki. 2008. 'Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acyltransferase superfamily', *Cell cycle (Georgetown, Tex.)*, 7: 120-1.

Kazmierczak, Mark J., Martin Wiedmann, and Kathryn J. Boor. 2005. 'Alternative Sigma Factors and Their Roles in Bacterial Virulence', *Microbiology and Molecular Biology Reviews*, 69: 527-43.

Kchromov, I. S., V. V. Sorotchkina, T. G. Nigmatullin, and T. I. Tikchonenko. 1980. 'A new nitrogen base 5-hydroxycytosine in phage N-17 DNA', *FEBS Letters*, 118: 51-54.

Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg. 2015. 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature protocols*, 10: 845-58.

Kennaway, Christopher K., Agnieszka Obarska-Kosinska, John H. White, Irina Tuszynska, Laurie P. Cooper, Janusz M. Bujnicki, John Trinick, and David T. F. Dryden. 2009. 'The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein', *Nucleic acids research*, 37: 762-70.

Khudyakov, I. Y., M. D. Kirnos, N. I. Alexandrushkina, and B. F. Vanyushin. 1978. 'Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine', *Virology*, 88: 8-18.

Kobayashi, Ichizo. 2001. 'Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution', *Nucleic acids research*, 29: 3742-56.

Kong, H., R. D. Morgan, R. E. Maunus, and I. Schildkraut. 1993. 'A unique restriction endonuclease, BcgI, from Bacillus coagulans', *Nucleic acids research*, 21: 987-91.

Koonin, Eugene, Valerian Dolja, Mart Krupovic, Arvind Varsani, Yuri Wolf, Natalya Yutin, Francisco Zerbini, and Jens Kuhn. 2019. *Create a megataxonomic framework, filling all principal taxonomic ranks, for DNA viruses encoding vertical jelly roll-type major capsid proteins*.

Kozdon, Jennifer B., Michael D. Melfi, Khai Luong, Tyson A. Clark, Matthew Boitano, Susana Wang, Bo Zhou, Diego Gonzalez, Justine Collier, Stephen W. Turner, Jonas Korlach,

Lucy Shapiro, and Harley H. McAdams. 2013. 'Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle', *Proceedings of the National Academy of Sciences*, 110: E4658-E67.

Krishnamurthy, Vishnu Vardhan, John S. Khamo, Ellen Cho, Cara Schornak, and Kai Zhang. 2015. 'Polymerase chain reaction-based gene removal from plasmids', *Data in brief*, 4: 75-82.

Kruger, D. H., G. J. Barcak, M. Reuter, and H. O. Smith. 1988. 'EcoRII can be activated to cleave refractory DNA recognition sites', *Nucleic acids research*, 16: 3997-4008.

Kruger, D. H., and T. A. Bickle. 1983. 'Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts', *Microbiol Rev*, 47: 345-60.

Kumar, Suresh, Viswanathan Chinnusamy, and Trilochan Mohapatra. 2018. 'Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond', *Frontiers in Genetics*, 9.

Kutter, E. M., and J. S. Wiberg. 1969. 'Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4', *Journal of virology*, 4: 439-53.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. 'Bacteriophage resistance mechanisms', *Nat Rev Micro*, 8: 317-27.

Lee, Duck-Yeon, Sung Jun Park, Woojin Jeong, Ho Jin Sung, Taena Oho, Xiongwu Wu, Sue Goo Rhee, and James M. Gruschus. 2006. 'Mutagenesis and Modeling of the Peroxiredoxin (Prx) Complex with the NMR Structure of ATP-Bound Human Sulfiredoxin Implicate Aspartate 187 of Prx I as the Catalytic Residue in ATP Hydrolysis', *Biochemistry*, 45: 15301-09.

Lepikhov, K., A. Tchernov, L. Zheleznaja, N. Matvienko, J. Walter, and T. A. Trautner. 2001. 'Characterization of the type IV restriction modification system BspLU11III from Bacillus sp. LU11', *Nucleic acids research*, 29: 4691-8.

Lewis, Dale, Phuoc Le, Chiara Zurla, Laura Finzi, and Sankar Adhya. 2011. 'Multilevel autoregulation of λ repressor protein CI by DNA looping in vitro', *Proceedings of the National Academy of Sciences*, 108: 14807-12.

Lewis, M. 2005. 'The lac repressor', *C R Biol*, 328: 521-48.

Lim, Derek H K, and Eamonn R Maher. 2010. 'DNA methylation: a form of epigenetic control of gene expression', *The Obstetrician & Gynaecologist*, 12: 37-42.

Little, J. W. 1984. 'Autodigestion of lexA and phage lambda repressors', *Proceedings of the National Academy of Sciences of the United States of America*, 81: 1375-79.

Liu, K., Y. Zhang, K. Severinov, A. Das, and M. M. Hanna. 1996. 'Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA', *The EMBO journal*, 15: 150-61.

Loenen, Wil A. M., David T. F. Dryden, Elisabeth A. Raleigh, and Geoffrey G. Wilson. 2014. 'Type I restriction enzymes and their relatives', *Nucleic acids research*, 42: 20-44.

Loenen, Wil A. M., David T. F. Dryden, Elisabeth A. Raleigh, Geoffrey G. Wilson, and Noreen E. Murray. 2013. 'Highlights of the DNA cutters: a short history of the restriction enzymes', *Nucleic acids research*, 42: 3-19.

Loenen, Wil A. M., and Elisabeth A. Raleigh. 2014. 'The other face of restriction: modification-dependent enzymes', *Nucleic acids research*, 42: 56-69.

Lopatina, A., N. Tal, and R. Sorek. 2020. 'Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy', *Annu Rev Virol*, 7: 371-84.

Luria, S. E., and M. L. Human. 1952. 'A nonhereditary, host-induced variation of bacterial viruses', *Journal of Bacteriology*, 64: 557-69.

Machnicka, Magdalena A., Katarzyna H. Kaminska, Stanislaw Dunin-Horkawicz, and Janusz M. Bujnicki. 2015. 'Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes', *BMC Bioinformatics*, 16: 336.

MacWilliams, M. P., and T. A. Bickle. 1996. 'Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene', *The EMBO journal*, 15: 4775-83.

Madeira, Fábio, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R. N. Tivey, Simon C. Potter, Robert D. Finn, and Rodrigo Lopez. 2019. 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic acids research*, 47: W636-W41.

Maghsoodi, Ameneh, Anupam Chatterjee, Ioan Andricioaei, and Noel C. Perkins. 2019. 'How the phage T4 injection machinery works including energetics, forces, and dynamic pathway', *Proceedings of the National Academy of Sciences*, 116: 25097-105.

Makarova, K. S., V. Anantharaman, N. V. Grishin, E. V. Koonin, and L. Aravind. 2014. 'CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems', *Front Genet*, 5: 102.

Makarova, Kira S., and Eugene V. Koonin. 2015. 'Annotation and Classification of CRISPR-Cas Systems', *Methods in molecular biology (Clifton, N.J.)*, 1311: 47-75.

Marino, Nicole D., Rafael Pinilla-Redondo, Bálint Csörgő, and Joseph Bondy-Denomy. 2020. 'Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies', *Nature Methods*, 17: 471-79.

Marinus, Martin G., and Josep Casadesus. 2009. 'Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more', *FEMS Microbiology Reviews*, 33: 488-503.

Marks, Phil, John McGeehan, Geoff Wilson, Neil Errington, and Geoff Kneale. 2003. 'Purification and characterisation of a novel DNA methyltransferase, M.AhdI', *Nucleic acids research*, 31: 2803-10.

Marshall, Jacqueline J. T., Darren M. Gowers, and Stephen E. Halford. 2007. 'Restriction Endonucleases that Bridge and Excise Two Recognition Sites from DNA', *Journal of Molecular Biology*, 367: 419-31.

Mathew, R., and D. Chatterji. 2006. 'The evolving story of the omega subunit of bacterial RNA polymerase', *Trends Microbiol*, 14: 450-5.

Matthews, B. W., D. H. Ohlendorf, W. F. Anderson, and Y. Takeda. 1982. 'Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor', *Proceedings of the National Academy of Sciences of the United States of America*, 79: 1428-32.

McClelland, S. E., and M. D. Szczelkun. 2004. 'The Type I and III Restriction Endonucleases: Structural Elements in Molecular Motors that Process DNA.' in Alfred M. Pingoud (ed.), *Restriction Endonucleases* (Springer Berlin Heidelberg: Berlin, Heidelberg).

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. 'Phaser crystallographic software', *J Appl Crystallogr*, 40: 658-74.

McKay, David B., and Thomas A. Steitz. 1981. 'Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA', *Nature*, 290: 744-49.

McKenzie, Joanna L., Jennifer Robson, Michael Berney, Tony C. Smith, Alaine Ruthe, Paul P. Gardner, Vickery L. Arcus, and Gregory M. Cook. 2012. 'A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria', *Journal of Bacteriology*, 194: 2189-204.

Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe. 1999. 'Food-related illness and death in the United States', *Emerg Infect Dis*, 5: 607-25.

Mellén, Marian, Pinar Ayata, Scott Dewell, Skirmantas Kriaucionis, and Nathaniel Heintz. 2012. 'MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system', *Cell*, 151: 1417-30.

Miller, Eric S., Elizabeth Kutter, Gisela Mosig, Fumio Arisaka, Takashi Kunisawa, and Wolfgang Rüger. 2003. 'Bacteriophage T4 genome', *Microbiology and molecular biology reviews : MMBR*, 67: 86-156.

Miller, J. H. 1972. *Experiments in Molecular Genetics* (Cold Spring Harbor: New York).

Mojica, Francisco J. M., Chc sar Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. 2005. 'Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements', *Journal of Molecular Evolution*, 60: 174-82.

Moore, Lisa D., Thuc Le, and Guoping Fan. 2013. 'DNA methylation and its basic function', *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 38: 23-38.

Morgan, Richard D., Tanya K. Bhatia, Lindsay Lovasco, and Theodore B. Davis. 2008. 'MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection', *Nucleic acids research*, 36: 6558-70.

Morgan, Richard D., Elizabeth A. Dwinell, Tanya K. Bhatia, Elizabeth M. Lang, and Yvette A. Luyten. 2009. 'The MmeI family: type II restriction–modification enzymes that employ single-strand modification for host protection', *Nucleic acids research*, 37: 5208-21.

Morgan, Richard D., Yvette A. Luyten, Samuel A. Johnson, Emily M. Clough, Tyson A. Clark, and Richard J. Roberts. 2016. 'Novel m4C modification in type I restriction-modification systems', *Nucleic acids research*, 44: 9413-25.

Morrow, John F., and Paul Berg. 1972. 'Cleavage of Simian Virus 40 DNA at a Unique Site by a Bacterial Restriction Enzyme', *Proceedings of the National Academy of Sciences*, 69: 3365-69.

Müller, Andreas U., Marc Leibundgut, Nenad Ban, and Eilika Weber-Ban. 2019. 'Structure and functional implications of WYL domain-containing bacterial DNA damage response regulator PafBC', *Nature Communications*, 10: 4653.

Murray, N. E. 2000. 'Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle)', *Microbiology and molecular biology reviews : MMBR*, 64: 412-34.

Nakagawa, H., F. Arisaka, and S. Ishii. 1985. 'Isolation and characterization of the bacteriophage T4 tail-associated lysozyme', *Journal of virology*, 54: 460-66.

Narberhaus, Franz, and Sylvia Balsiger. 2003. 'Structure-function studies of Escherichia coli RpoH (sigma32) by in vitro linker insertion mutagenesis', *Journal of Bacteriology*, 185: 2731-38.

NCBI. 2020. 'Monodnaviria'. https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2731342.

NEB. 2020. 'Golden Gate Assembly', New England Biolabs. https://international.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/golden-gate-assembly.

NEB. 2020. 'Types of Restriction Endonucleases,'. https://international.neb.com/products/restriction-endonucleases/restriction-endonucleases/types-of-restriction-endonucleases.

Nechaev, S., and K. Severinov. 2008. 'The elusive object of desire--interactions of bacteriophages and their hosts', *Curr Opin Microbiol*, 11: 186-93.

Nikolskaya, II, M. I. Tediashvili, N. G. Lopatina, T. G. Chanishvili, and S. S. Debov. 1979. 'Specificity and functions of guanine methylase of Shigella sonnei DDVI phage', *Biochim Biophys Acta*, 561: 232-9.

Nikolskaya, I. I., N. G. Lopatina, and S. S. Debov. 1976. 'Methylated guanine derivative as a minor base in the DNA of phage DDVI Shigella disenteriae', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 435: 206-10.

Niv, M. Y., D. R. Ripoll, J. A. Vila, A. Liwo, E. S. Vanamee, A. K. Aggarwal, H. Weinstein, and H. A. Scheraga. 2007. 'Topology of Type II REases revisited; structural classes and the common conserved core', *Nucleic acids research*, 35: 2227-37.

Ofir, Gal, Sarah Melamed, Hila Sberro, Zohar Mukamel, Shahar Silverman, Gilad Yaakov, Shany Doron, and Rotem Sorek. 2018. 'DISARM is a widespread bacterial defence system with broad anti-phage activities', *Nature microbiology*, 3: 90-98.

Orhanović, S., and M. Pavela-Vrancic. 2003. 'Dimer asymmetry and the catalytic cycle of alkaline phosphatase from Escherichia coli', *Eur J Biochem*, 270: 4356-64.

Orhanović, Stjepan, and Maja Pavela-Vrančič. 2003. 'Dimer asymmetry and the catalytic cycle of alkaline phosphatase from Escherichia coli', *European Journal of Biochemistry*, 270: 4356-64.

Owen, Siân V., Nicolas Wenner, Rocío Canals, Angela Makumi, Disa L. Hammarlöf, Melita A. Gordon, Abram Aertsen, Nicholas A. Feasey, and Jay C. D. Hinton. 2017. 'Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1', *Frontiers in microbiology*, 8: 235-35.

Pandey, Deo Prakash, and Kenn Gerdes. 2005. 'Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes', *Nucleic acids research*, 33: 966-76.

Panne, D., S. A. Müller, S. Wirtz, A. Engel, and T. A. Bickle. 2001. 'The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides', *The EMBO journal*, 20: 3210-17.

Parsons, R. J., M. Breitbart, M. W. Lomas, and C. A. Carlson. 2012. 'Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea', *Isme j*, 6: 273-84.

Payelleville, Amaury, Ludovic Legrand, Jean-Claude Ogier, Céline Roques, Alain Roulet, Olivier Bouchez, Annabelle Mouammine, Alain Givaudan, and Julien Brillard. 2018. 'The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated Adenines', *Scientific reports*, 8: 12091.

Pedruzzi, I., J. P. Rosenbusch, and K. P. Locher. 1998. 'Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein', *FEMS Microbiol Lett*, 168: 119-25.

Pei, Jimin, Bong-Hyun Kim, and Nick V. Grishin. 2008. 'PROMALS3D: a tool for multiple protein sequence and structure alignments', *Nucleic acids research*, 36: 2295-300.

Pfeifer, Gerd P., Wenying Xiong, Maria A. Hahn, and Seung-Gi Jin. 2014. 'The role of 5-hydroxymethylcytosine in human cancer', *Cell and tissue research*, 356: 631-41.

Piekarowicz, A., M. Golaszewska, A. O. Sunday, M. Siwinska, and D. C. Stein. 1999. 'The HaeIV restriction modification system of Haemophilus aegyptius is encoded by a single polypeptide', *J Mol Biol*, 293: 1055-65.

Pingoud, A., M. Fuxreiter, V. Pingoud, and W. Wende. 2005. 'Type II restriction endonucleases: structure and mechanism', *Cell Mol Life Sci*, 62: 685-707.

Pingoud, Alfred, Geoffrey G. Wilson, and Wolfgang Wende. 2014. 'Type II restriction endonucleases--a historical perspective and more', *Nucleic acids research*, 42: 7489-527.

Price, A. R. 1974. 'Bacteriophage PBS2-induced deoxycytidine triphosphate deaminase in Bacillus subtilis', *Journal of virology*, 14: 1314-7.

Qimron, U., B. Marintcheva, S. Tabor, and C. C. Richardson. 2006. 'Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 19039-44.

Rajamanickam, Karthic, and Sidney Hayes. 2018. 'The Bacteriophage Lambda CII Phenotypes for Complementation, Cellular Toxicity and Replication Inhibition Are Suppressed in cII-oop Constructs Expressing the Small RNA OOP', *Viruses*, 10: 115.

Rao, Desirazu N., David T. F. Dryden, and Shivakumara Bheemanaik. 2014. 'Type III restriction-modification enzymes: a historical perspective', *Nucleic acids research*, 42: 45-55.

Rao, Venigalla B., and Lindsay W. Black. 2005. 'DNA Packaging in Bacteriophage T4.' in, *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism* (Springer US: Boston, MA).

Rauch, B. J., M. R. Silvis, J. F. Hultquist, C. S. Waters, M. J. McGregor, N. J. Krogan, and J. Bondy-Denomy. 2017. 'Inhibition of CRISPR-Cas9 with Bacteriophage Proteins', *Cell*, 168: 150-58.e10.

Rauhut, R., and G. Klug. 1999. 'mRNA degradation in bacteria', *FEMS Microbiol Rev*, 23: 353-70.

Revel, Helen R., and C. P. Georgopoulos. 1969. 'Restriction of nonglucosylated T-even bacteriophages by prophage P1', *Virology*, 39: 1-17.

Reyes-Robles, T., R. S. Dillard, L. S. Cairns, C. A. Silva-Valenzuela, M. Housman, A. Ali, E. R. Wright, and A. Camilli. 2018. 'Vibrio cholerae Outer Membrane Vesicles Inhibit Bacteriophage Infection', *J Bacteriol*, 200.

Rifat, D., N. T. Wright, K. M. Varney, D. J. Weber, and L. W. Black. 2008a. 'Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target', *J Mol Biol*, 375: 720-34.

Rifat, Dalin, Nathan Wright, Kristen Varney, David Weber, and Lindsay Black. 2008b. 'Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target', *Journal of Molecular Biology*, 375: 720-34.

Roa, M. 1979. 'Interaction of bacteriophage K10 with its receptor, the lamB protein of Escherichia coli', *J Bacteriol*, 140: 680-6.

Roberts, Gareth A., Augoustinos S. Stephanou, Nisha Kanwar, Angela Dawson, Laurie P. Cooper, Kai Chen, Margaret Nutley, Alan Cooper, Garry W. Blakely, and David T. F. Dryden. 2012. 'Exploring the DNA mimicry of the Ocr protein of phage T7', *Nucleic acids research*, 40: 8129-43.

Roberts, Richard J., Marlene Belfort, Timothy Bestor, Ashok S. Bhagwat, Thomas A. Bickle, Jurate Bitinaite, Robert M. Blumenthal, Sergey Kh Degtyarev, David T. F. Dryden, Kevin Dybvig, Keith Firman, Elizaveta S. Gromova, Richard I. Gumport, Stephen E. Halford, Stanley Hattman, Joseph Heitman, David P. Hornby, Arvydas Janulaitis, Albert Jeltsch, Jytte Josephsen, Antal Kiss, Todd R. Klaenhammer, Ichizo Kobayashi, Huimin Kong, Detlev H. Krüger, Sanford Lacks, Martin G. Marinus, Michiko Miyahara, Richard D. Morgan, Noreen E. Murray, Valakunja Nagaraja, Andrzej Piekarowicz, Alfred Pingoud, Elisabeth Raleigh, Desirazu N. Rao, Norbert Reich, Vladimir E. Repin, Eric U. Selker, Pang-Chui Shaw, Daniel C. Stein, Barry L. Stoddard, Waclaw Szybalski, Thomas A. Trautner, James L. Van Etten, Jorge M. B. Vitor, Geoffrey G. Wilson, and Shuang-yong Xu. 2003. 'A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes', *Nucleic acids research*, 31: 1805-12.

Roberts, Richard J., Tamas Vincze, Janos Posfai, and Dana Macelis. 2015. 'REBASE--a database for DNA restriction and modification: enzymes, genes and genomes', *Nucleic acids research*, 43: D298-D99.

Rodríguez-Rodríguez, Diana Raquel, Ramiro Ramírez-Solís, Mario Alberto Garza-Elizondo, María De Lourdes Garza-Rodríguez, and Hugo Alberto Barrera-Saldaña. 2019. 'Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review)', *International journal of molecular medicine*, 43: 1559-74.

Rodwell, Ella V., Nicolas Wenner, Caisey V. Pulford, Yueyi Cai, Arthur Bowers-Barnard, Alison Beckett, Jonathan Rigby, David M. Picton, Tim R. Blower, Nicholas A. Feasey, Jay C. D. Hinton, and Blanca M. Perez-Sepulveda. 2021. 'Isolation and characterisation of bacteriophages with activity against invasive non-typhoidal *Salmonella* causing bloodstream infection in Malawi', *Viruses*, 13: 478.

Rohwer, Forest, and Anca M. Segall. 2015. 'A century of phage lessons', *Nature*, 528: 46-47.

Rojo, F. 1999. 'Repression of transcription initiation in bacteria', *Journal of Bacteriology*, 181: 2987-91.

Rousset, François, Julien Dowding, Aude Bernheim, Eduardo P.C. Rocha, and David Bikard. 2021. 'Prophage-encoded hotspots of bacterial immune systems', *bioRxiv*: 2021.01.21.427644.

Safa, Ashrafus, G. Balakrish Nair, and Richard Y. C. Kong. 2010. 'Evolution of new variants of Vibrio cholerae O1', *Trends in Microbiology*, 18: 46-54.

Salmond, George P. C., and Peter C. Fineran. 2015. 'A century of the phage: past, present and future', *Nat Rev Micro*, 13: 777-86.

Sambrook, Joseph. 2001. *Molecular cloning : a laboratory manual* (Cold Spring Harbor, N.Y).

Samson, Julie E., Alfonso H. Magadan, Mourad Sabri, and Sylvain Moineau. 2013. 'Revenge of the phages: defeating bacterial defences', *Nat Rev Micro*, 11: 675-87.

Sander, Jeffry D., and J. Keith Joung. 2014. 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nat Biotech*, 32: 347-55.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, and M. Smith. 1977. 'Nucleotide sequence of bacteriophage φX174 DNA', *Nature*, 265: 687-95.

Sarkar-Banerjee, Suparna, Sachin Goyal, Ning Gao, John Mack, Benito Thompson, David Dunlap, Krishnananda Chattopadhyay, and Laura Finzi. 2018. 'Specifically bound lambda repressor dimers promote adjacent non-specific binding', *PLOS ONE*, 13: e0194930.

Sausset, R., M. A. Petit, V. Gaboriau-Routhiau, and M. De Paepe. 2020. 'New insights into intestinal phages', *Mucosal Immunology*, 13: 205-15.

Schaefer, Jorrit, Goran Jovanovic, Ioly Kotta-Loizou, and Martin Buck. 2016. 'Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader', *Analytical biochemistry*, 503: 56-57.

Schierling, Benno, Nadine Dannemann, Lilia Gabsalilow, Wolfgang Wende, Toni Cathomen, and Alfred Pingoud. 2012. 'A novel zinc-finger nuclease platform with a sequence-specific cleavage module', *Nucleic acids research*, 40: 2623-38.

Seed, Kimberley D. 2015. 'Battling Phages: How Bacteria Defend against Viral Attack', *PLoS Pathogens*, 11: e1004847.

Seidel, Ralf, Joost G. P. Bloom, Cees Dekker, and Mark D. Szczelkun. 2008. 'Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I', *The EMBO journal*, 27: 1388-98.

Severinov, K., M. Kashlev, E. Severinova, I. Bass, K. McWilliams, E. Kutter, V. Nikiforov, L. Snyder, and A. Goldfarb. 1994. 'A non-essential domain of Escherichia coli RNA polymerase required for the action of the termination factor Alc', *J Biol Chem*, 269: 14254-9.

Sharp, Phillip A., Bill Sugden, and Joe Sambrook. 1973. 'Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose-ethidium bromide electrophoresis', *Biochemistry*, 12: 3055-63.

Sheldrick, George. 2008. 'A Short History of ShelX', *Acta crystallographica. Section A, Foundations of crystallography*, 64: 112-22.

Shen, Betty W, Lindsey Doyle, Phil Bradley, Daniel F Heiter, Keith D Lunnen, Geoffrey G Wilson, and Barry L Stoddard. 2018. 'Structure, subunit organization and behavior of the asymmetric Type IIT restriction endonuclease BbvCI', *Nucleic acids research*, 47: 450-67.

Shen, Betty W., Derrick Xu, Siu-Hong Chan, Yu Zheng, Zhenyu Zhu, Shuang-yong Xu, and Barry L. Stoddard. 2011. 'Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI', *Nucleic acids research*, 39: 8223-36.

Short, Francesca L., Chidiebere Akusobi, William R. Broadhurst, and George P. C. Salmond. 2018. 'The bacterial Type III toxin-antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent antiviral abortive infection activity', *Scientific reports*, 8: 1013-13.

Sistla, S., and D. N. Rao. 2004. 'S-Adenosyl-L-methionine-dependent restriction enzymes', *Crit Rev Biochem Mol Biol*, 39: 1-19.

Snyder, Larry. 1995. 'Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents?', *Molecular Microbiology*, 15: 415-20.

Sommer, N., R. Depping, M. Piotrowski, and W. Ruger. 2004. 'Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism', *Biochem Biophys Res Commun*, 323: 809-15.

Sommer, N., R. Depping, M. Piotrowski, and W. Rüger. 2004. 'Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism', *Biochem Biophys Res Commun*, 323: 809-15.

Song, Chun-Xiao, Senlin Yin, Li Ma, Amanda Wheeler, Yu Chen, Yan Zhang, Bin Liu, Junjie Xiong, Weihan Zhang, Jiankun Hu, Zongguang Zhou, Biao Dong, Zhiqi Tian, Stefanie S. Jeffrey, Mei-Sze Chua, Samuel So, Weimin Li, Yuquan Wei, Jiajie Diao, Dan Xie, and Stephen R. Quake. 2017. '5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages', *Cell research*, 27: 1231-42.

Song, Sooyeon, and Thomas K. Wood. 2020. 'A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition', *Frontiers in microbiology*, 11.

Souther, A., R. Bruner, and J. Elliott. 1972. 'Degradation of Escherichia coli chromosome after infection by bacteriophage T4: role of bacteriophage gene D2a', *Journal of virology*, 10: 979-84.

Srikumar, Shabarinath, Carsten Kröger, Magali Hébrard, Aoife Colgan, Siân V. Owen, Sathesh K. Sivasankaran, Andrew D. S. Cameron, Karsten Hokamp, and Jay C. D. Hinton. 2015. 'RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium', *PLoS Pathogens*, 11: e1005262.

Stanaway, Jeffrey D., Andrea Parisi, Kaushik Sarkar, Brigette F. Blacker, Robert C. Reiner, Simon I. Hay, Molly R. Nixon, Christiane Dolecek, Spencer L. James, Ali H. Mokdad, Getaneh Abebe, Elham Ahmadian, Fares Alahdab, Birhan Tamene T. Alemnew, Vahid Alipour, Fatemeh Allah Bakeshei, Megbaru Debalkie Animut, Fereshteh Ansari, Jalal Arabloo, Ephrem Tsegay Asfaw, Mojtaba Bagherzadeh, Quique Bassat, Yaschilal Muche Muche Belayneh, Félix Carvalho, Ahmad Daryani, Feleke Mekonnen Demeke, Asmamaw Bizuneh Bizuneh Demis, Manisha Dubey, Eyasu Ejeta Duken, Susanna J. Dunachie, Aziz Eftekhari, Eduarda Fernandes, Reza Fouladi Fard, Getnet Azeze Gedefaw, Birhanu Geta, Katherine B. Gibney, Amir Hasanzadeh, Chi Linh Hoang, Amir Kasaeian, Amir Khater, Zelalem Teklemariam Kidanemariam, Ayenew Molla

Lakew, Reza Malekzadeh, Addisu Melese, Desalegn Tadese Mengistu, Tomislav Mestrovic, Bartosz Miazgowski, Karzan Abdulmuhsin Mohammad, Mahdi Mohammadian, Abdollah Mohammadian-Hafshejani, Cuong Tat Nguyen, Long Hoang Nguyen, Son Hoang Nguyen, Yirga Legesse Nirayo, Andrew T. Olagunju, Tinuke O. Olagunju, Hadi Pourjafar, Mostafa Qorbani, Mohammad Rabiee, Navid Rabiee, Anwar Rafay, Aziz Rezapour, Abdallah M. Samy, Sadaf G. Sepanlou, Masood Ali Shaikh, Mehdi Sharif, Mika Shigematsu, Belay Tessema, Bach Xuan Tran, Irfan Ullah, Ebrahim M. Yimer, Zoubida Zaidi, Christopher J. L. Murray, and John A. Crump. 2019. 'The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet Infectious Diseases*, 19: 1312-24.

Stanley, Sabrina Y., and Karen L. Maxwell. 2018. 'Phage-Encoded Anti-CRISPR Defenses', *Annual Review of Genetics*, 52: 445-64.

Steczkiewicz, Kamil, Anna Muszewska, Lukasz Knizewski, Leszek Rychlewski, and Krzysztof Ginalski. 2012. 'Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily', *Nucleic acids research*, 40: 7016-45.

Stephenson, Stacy Ann-Marie, and Paul D. Brown. 2016. 'Epigenetic Influence of Dam Methylation on Gene Expression and Attachment in Uropathogenic Escherichia coli', *Frontiers in Public Health*, 4.

Stern, A., and R. Sorek. 2011a. 'The phage-host arms race: shaping the evolution of microbes', *Bioessays*, 33: 43-51.

Stern, Adi, and Rotem Sorek. 2011b. 'The phage-host arms-race: Shaping the evolution of microbes', *Bioessays*, 33: 43-51.

Stewart, F. J., D. Panne, T. A. Bickle, and E. A. Raleigh. 2000a. 'Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme', *J Mol Biol*, 298: 611-22.

Stewart, F. J., and E. A. Raleigh. 1998. 'Dependence of McrBC cleavage on distance between recognition elements', *Biological chemistry*, 379: 611-16.

Stewart, Fiona J., Daniel Panne, Thomas A. Bickle, and Elisabeth A. Raleigh. 2000b. 'Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme 11Edited by J. Karn', *Journal of Molecular Biology*, 298: 611-22.

Sulakvelidze, Alexander. 2011. 'Bacteriophage: A new journal for the most ubiquitous organisms on Earth', *Bacteriophage*, 1: 1-2.

Sumby, Paul, and Margaret C. M. Smith. 2002. 'Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2)', *Molecular Microbiology*, 44: 489-500.

Summers, William C. 2001. 'Bacteriophage Therapy', *Annual Review of Microbiology*, 55: 437-51.

Suttle, Curtis A. 2007. 'Marine viruses — major players in the global ecosystem', *Nature Reviews Microbiology*, 5: 801-12.

Takahashi, Noriko, Yasuhiro Naito, Naofumi Handa, and Ichizo Kobayashi. 2002. 'A DNA Methyltransferase Can Protect the Genome from Postdisturbance Attack by a Restriction-Modification Gene Complex', *Journal of Bacteriology*, 184: 6100-08.

Tamulaitis, G., G. Sasnauskas, M. Mucke, and V. Siksnys. 2006. 'Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease', *J Mol Biol*, 358: 406-19.

Thomsen, Nathan D, and James M. Berger. 2008. 'Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases', *Molecular Microbiology*, 69: 1071-90.

Thomsen, Nathan D., and James M. Berger. 2012. 'Crystallization and X-ray structure determination of an RNA-dependent hexameric helicase', *Methods in enzymology*, 511: 171-90.

Thursby, Elizabeth, and Nathalie Juge. 2017. 'Introduction to the human gut microbiota', *The Biochemical journal*, 474: 1823-36.

Tripathi, Lakshmi, Yan Zhang, and Zhanglin Lin. 2014. 'Bacterial Sigma Factors as Targets for Engineered or Synthetic Transcriptional Control', *Frontiers in Bioengineering and Biotechnology*, 2.

Usón, Isabel, and George M. Sheldrick. 2018. 'An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features', *Acta crystallographica. Section D, Structural biology*, 74: 106-16.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, and G. N. Murshudov. 2004. 'REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta crystallographica. Section D, Biological crystallography*, 60: 2184-95.

van Miert, A. S. 1994. 'The sulfonamide-diaminopyrimidine story', *J Vet Pharmacol Ther*, 17: 309-16.

van Noort, John, Thijn van der Heijden, Christina F. Dutta, Keith Firman, and Cees Dekker. 2004. 'Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion', *Nucleic acids research*, 32: 6540-47.

Vanamee, Éva Scheuring, Sandro Santagata, and Aneel K. Aggarwal. 2001. 'FokI requires two specific DNA sites for cleavage11Edited by T. Richmond', *Journal of Molecular Biology*, 309: 69-78.

Vasu, Kommireddy, and Valakunja Nagaraja. 2013. 'Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense', *Microbiology and molecular biology reviews : MMBR*, 77: 53-72.

Venkataraman, Sangita, Burra V. L. S. Prasad, and Ramasamy Selvarajan. 2018. 'RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution', *Viruses*, 10: 76.

Vidakovic, L., P. K. Singh, R. Hartmann, C. D. Nadell, and K. Drescher. 2018. 'Dynamic biofilm architecture confers individual and collective mechanisms of viral protection', *Nature microbiology*, 3: 26-31.

Vieux, Ellen F., Matthew L. Wohlever, James Z. Chen, Robert T. Sauer, and Tania A. Baker. 2013. 'Distinct quaternary structures of the AAA+ Lon protease control substrate degradation', *Proceedings of the National Academy of Sciences*, 110: E2002-E08.

Virus Metadata Resource. 2019. 'Virus Taxonomy: 2019 Release'. https://talk.ictvonline.org/taxonomy/.

Vitkute, J., Z. Maneliene, M. Petrusyte, and A. Janulaitis. 1997. 'BplI, a new BcgI-like restriction endonuclease, which recognizes a symmetric sequence', *Nucleic acids research*, 25: 4444-6.

Vlot, Marnix, Joep Houkes, Silke J. A. Lochs, Daan C. Swarts, Peiyuan Zheng, Tim Kunne, Prarthana Mohanraju, Carolin Anders, Martin Jinek, John van der Oost, Mark J. Dickman, and Stan J. J. Brouns. 2018. 'Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes', *Nucleic acids research*, 46: 873-85.

Waite-Rees, P. A., C. J. Keating, L. S. Moran, B. E. Slatko, L. J. Hornstra, and J. S. Benner. 1991. 'Characterization and expression of the Escherichia coli Mrr restriction system', *Journal of Bacteriology*, 173: 5207-19.

Walling, Lauren R., and J. Scott Butler. 2018. 'Homologous VapC Toxins Inhibit Translation and Cell Growth by Sequence-Specific Cleavage of tRNA^{fMet}', *Journal of Bacteriology*, 200: e00582-17.

Wang, N., S. Gottesman, M. C. Willingham, M. M. Gottesman, and M. R. Maurizi. 1993. 'A human mitochondrial ATP-dependent protease that is highly homologous to bacterial Lon protease', *Proceedings of the National Academy of Sciences of the United States of America*, 90: 11247-51.

Wang, Z., and D. W. Mosbaugh. 1989. 'Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase', *J Biol Chem*, 264: 1163-71.

Wei, Hua, Guojie Zhao, Tianyu Hu, Suming Tang, Jiquan Jiang, Bo Hu, and Yifu Guan. 2016. 'Mapping the nicking efficiencies of nickase R.BbvCI for side-specific LNA-substituted substrates using rolling circle amplification', *Scientific reports*, 6: 32560-60.

Weigele, Peter, and Elisabeth A. Raleigh. 2016. 'Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses', *Chemical Reviews*, 116: 12655-87.

Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. 'Prokaryotes: the unseen majority', *Proceedings of the National Academy of Sciences of the United States of America*, 95: 6578-83.

Williams, S. A., and S. E. Halford. 2001. 'SfiI endonuclease activity is strongly influenced by the non-specific sequence in the middle of its recognition site', *Nucleic acids research*, 29: 1476-83.

Winn, Martyn D., Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei Vagin, and Keith S. Wilson. 2011. 'Overview of the CCP4 suite and current developments', *Acta crystallographica. Section D, Biological crystallography*, 67: 235-42.

Wyatt, G. R., and S. S. Cohen. 1953. 'The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine', *The Biochemical journal*, 55: 774-82.

Xiong, Xiaolin, Geng Wu, Yue Wei, Liqiong Liu, Yubing Zhang, Rui Su, Xianyue Jiang, Mengxue Li, Haiyan Gao, Xihao Tian, Yizhou Zhang, Li Hu, Si Chen, You Tang, Susu Jiang, Ruolin Huang, Zhiqiang Li, Yunfu Wang, Zixin Deng, Jiawei Wang, Peter C. Dedon, Shi Chen, and Lianrong Wang. 2020. 'SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities', *Nature microbiology*, 5: 917-28.

Xu, S. Y., A. R. Corvaglia, S. H. Chan, Y. Zheng, and P. Linder. 2011. 'A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300', *Nucleic acids research*, 39: 5597-610.

Yamaichi, Y., and H. Niki. 2000. 'Active segregation by the Bacillus subtilis partitioning system in Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 97: 14656-61.

Yap, Moh Lan, and Michael G. Rossmann. 2014. 'Structure and function of bacteriophage T4', *Future microbiology*, 9: 1319-27.

Yousuf, Shehla, Joyce E. Karlinsey, Stephanie L. Neville, Christopher A. McDevitt, Stephen J. Libby, Ferric C. Fang, and Elaine R. Frawley. 2020. 'Manganese import protects Salmonella enterica serovar Typhimurium against nitrosative stress', *Metallomics*, 12: 1791-801.

Yuan, Z., A. Riera, L. Bai, J. Sun, S. Nandi, C. Spanos, Z. A. Chen, M. Barbon, J. Rappsilber, B. Stillman, C. Speck, and H. Li. 2017. 'Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1', *Nat Struct Mol Biol*, 24: 316-24.

Zabet, Nicolae Radu, Marco Catoni, Filippo Prischi, and Jerzy Paszkowski. 2017. 'Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies', *Nucleic acids research*, 45: 3777-84.

Zahran, Mai, Tomasz Berezniak, Petra Imhof, and Jeremy C. Smith. 2011. 'Role of magnesium ions in DNA recognition by the EcoRV restriction endonuclease', *FEBS Letters*, 585: 2739-43.

Zhang, Yadong, Zhewen Zhang, Hao Zhang, Yongbing Zhao, Zaichao Zhang, and Jingfa Xiao. 2019. 'PADS Arsenal: a database of prokaryotic defense systems related genes', *Nucleic acids research*, 48: D590-D98.

Zhao, Guojie, Jun Li, Zhaoxue Tong, Bin Zhao, Runqing Mu, and Yifu Guan. 2013. 'Enzymatic Cleavage of Type II Restriction Endonucleases on the 2′-O-Methyl Nucleotide and Phosphorothioate Substituted DNA', *PLOS ONE*, 8: e79415.

Zheng, Yu, Devora Cohen-Karni, Derrick Xu, Hang Gyeong Chin, Geoffrey Wilson, Sriharsa Pradhan, and Richard J. Roberts. 2010. 'A unique family of Mrr-like modification-dependent restriction endonucleases', *Nucleic acids research*, 38: 5527-34.

Zhou, Yan Ning, Lucyna Lubkowska, Monica Hui, Carolyn Court, Shuo Chen, Donald L. Court, Jeffrey Strathern, Ding Jun Jin, and Mikhail Kashlev. 2013. 'Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in Escherichia coli', *The Journal of biological chemistry*, 288: 2700-10.

Zou, N., S. Ditty, B. Li, and S. C. Lo. 2003. 'Random priming PCR strategy to amplify and clone trace amounts of DNA', *Biotechniques*, 35: 758-60, 62-5.