Molecular tools from bacteriophages: A structural and functional characterisation of the BREX bacteriophage resistance system

WENT, SAMUEL,CHRISTOPHER

Molecular tools from bacteriophages: A structural and functional characterisation of the BREX bacteriophage resistance system

by Samuel Went

BSc (Hons), MSc

Thesis for the degree of Doctor of Philisophy in Biological Sciences,
Department of Chemistry, Durham University, 2023

Abstract

The interminable arms race between bacteriophages (phages) and their bacterial hosts has produced an abundance of phage defence system modalities. Phage-bacteria interactions have provided a great number of the molecular biology tools routinely applied in laboratories around the world. Additionally, the re-emergence of phage therapy provides a plausible solution to the rise of multi-drug resistant bacteria. Wide-scale use of phage therapy will require a thorough understanding of mechanisms by which bacteria resist phage infection. Recently, systematic approaches to defence system discovery have unearthed a plethora of new systems and attempts to characterise defence system functions and mechanisms have fallen behind. This study aimed to provide valuable insight into the mechanisms of BREX phage resistance systems through structural and functional characterisation of a type I BREX system from *Salmonella* Typhimurium strain D23580.

Through assaying the *Salmonella* D23580 system against the Durham Phage Collection alongside type I BREX systems from *Escherichia coli* and *Escherichia fergusonii*, it was shown that phage defence varies between systems against a given phage in a manner which does not correlate with the number of recognition motifs within respective phage genomes. Further, phages appear to encode mechanisms of inhibiting species specific BREX systems. Next, analysis of gene deletions demonstrated essential genes for host methylation and phage defence, again showing variation from similar studies in the literature. Unusually, deletion of *brxL* elicited an increase in phage defence by several orders of magnitude. To provide further insight into the function of individual components, the structure of the methyltransferase, PglX, was solved to a resolution of 3.4 Å. PglX displays distinct N and C-terminal domains joined by a central hinge, with conserved methyltransferase regions. To shed light on mechanisms of phage escape from BREX systems, the structure of PglX bound to the BREX inhibitor, Ocr, was also solved to 3.5 Å. Ocr binds along the C-terminal domain of PglX and provides insight on potential DNA binding positions. Finally, PglX was rationally mutated to alter the BREX recognition motif, both changing host methylation patterns and allowing defence against a previously resistant phage. As such, PglX is the sole specificity factor of BREX defence, despite other components encoding DNA binding functionalities. Mutations in PglX in nature would allow rapid retargeting of BREX defence against new phage threats. Together, these results will guide further studies into BREX systems towards understanding the molecular mechanisms of phage defence.

Declaration

This thesis is the outcome of my own work and is submitted solely for the degree of Doctor of Philosophy (PhD) in Biological Sciences, Durham University. Any results that arose from collaborative efforts are mentioned in-text. All research presented herein, unless otherwise stated, were performed in the laboratory of Professor Tim R. Blower, Durham University, during the period October 2019 to September 2023.

Statement of copyright

Contents

List of figures

List of tables

List of abbreviations

°C	Degrees celcius
2x YT	Yeast tryptone broth (growth media)
5mC	5-methylcytosine
Abi	Abortive infection
Acr	Anti-CRISPR
Ago	Argonaute
A-SEC	Analytical size exclusion chromatography
ATP	Adenosine triphosphate
ATPase	Catalyses hydrolysis of ATP to ADP
bp	Base paris
BREX	Bacteriophage Exclusion
CARF	CRISPR-associated rossmann fold
CBASS	Cyclic oligonucleotide-based antiphage signaling system
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
cryo-EM	Cryo electron microscopy
C-terminal	Carbon-terminal
CTP	Cytosine triphosphate
CTT	Coupled transcription-translation
CV	Column volume
dATP	Deoxy-adenosine triphosphate
DISARM	Defence island system associated with restriction-modification
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DTT	Dithiothreitol
dTTP	Thymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electromobility shift assay
EOP	Efficiency of plaquing
FPLC	Fast protein liquid chromatography
GA	Gibson assembly

GB	Gigabytes
GGA	Golden gate assembly
GOI	Gene of interest
GTP	Guanosine triphosphate
h	hours
ICTV	International committee for the taxonomy of viruses
IPTG	Isopropyl β- d-1-thiogalactopyranoside
kb	Kilobases
KSCN	Potassium thiocyanate
LIC	Ligation independent cloning
MGE	Mobile genetic element
MOI	Multiplicity of infection
mRNA	messenger RNA
N4mC	N4-methylcytosine
N6mA	N6-methyladenine
NAD	Nicotinamide adenine dinucleotide
NDP	Nucleotide diphosphate
NEB	New England Biolabs
N-termiinal	Amino-terminal
NTP	Nucleotide triphosphate
NTPase	Catalyses hydrolysis of NTPs to NDPs
OD_{600}	Optical density at 600nm
OLD	Overcoming lysogenisation defect
Oligo	Oligonucleotide
ONT	Oxford nanopore technologies
ORF	Open reading frame
PAM	Protospacer adjacent motif
PAPS	Phosphoadenosine phosphosulfate
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
Pgl	Phage growth limitation
PT	Posphothioation
RAM	Random access memory
RE	Restriction endonuclease

msDNA	multi-copy single stranded DNA (retron associated)
RM	Restriction modification
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-PCR	Reverse transcription PCR
SAH	S-Adenosyl-L-homocysteine
SAM	S-Adenosyl-L-methionine
SDS-PAGE	Sodium dodecylsulphafe polyacrylimide gel electrophoresis
SEC	Size exclusion chromatography
SMC	Structural modification of chromatin
ssDNA	single stranded DNA
TA	Toxin-antitoxin
TAE	Tris acetate EDTA
TRD	Target recognition domain
UK	United Kingdom
WGA	Whole genome amplification

Acknowledgments

As I sit here puzzling over what to write in this section, I realise that this thesis is not just a record of the last four years of research but the culmination of (almost) the last decade of my life. It goes without saying that this wouldn't have been possible without some wonderful and supportive people. A few paragraphs don't seem sufficient to give the thanks you deserve.

Firstly, Professor Tim Blower, thank you for inviting me to work in your lab, and for your guidance, support, enthusiasm, and friendship over the last four years. I look forward to using all that you have taught me as a Post Doc in your lab.

To the organisers of the MoSMed CDT, Ehmke, Mike, Emma and Selina, thank you for creating an enjoyable and inclusive environment. Thank you also to Ehmke for his help and guidance on all things crystallography over the years. And of course, thank you to the funding bodies which made this project possible.

Thank you to the members of Blower lab, firstly Dave, Ben and Izaak, then Abbie and later Tom, Jenny and Liam. You all contributed to making the lab and the office an enjoyable place to work (and those crosswords would have taken me ages on my own). I look forward to many more pints. Thank you also to the wider members of the office for many a chat over questionable cups of tea.

To my family, I genuinely would not be here without you. Thank you for putting up with me and for continuing to feed me. To Zoe, I do not have the words to describe what you support means to me. Thank you.

And finally, to Winston.

Publications

Work carried out over the course of this study has contributed to publication of the following articles:

Diverse Durham collection phages demonstrate complex BREX defense responses.

Applied and Environmental Microbiology. (2023). doi.org/10.1128/AEM.00623-23

*Kelly, A., *Went, S. C., Mariano, G., Shaw, L. P., Picton, D. M., Duffner, S. J., Coates, I., Herdman-Grant, R., Gordeeva, J., Drobiazko, A., Isaev, A., Lee, Y.-J., Luyten, Y., Morgan, R. D., Weigele, P., Severinov, K., Wenner, N., Hinton, J. C. D., & Blower, T. R.

(* indicates joint first authorship)

In addition, the following articles have been published through work not directly related to this study:

Toxin–antitoxin systems as mediators of phage defence and the implications for abortive infection.

Current Opinion in Microbiology. (2023). doi.org/10.1016/J.MIB.2023.102293

Kelly, A., Arrowsmith, T. J., Went, S. C., & Blower, T. R.

A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands.

Nucleic Acids Research. (2022). doi.org/10.1093/NAR/GKAC334

Picton, D. M., Harling-Lee, J. D., Duffner, S. J., Went, S. C., Morgan, R. D., Hinton, J. C. D., & Blower, T. R. *50*(9), 5191–5207.

Chapter 1 – Introduction

1.1 Bacteriophages

Bacteriophages (phages), viruses which infect bacteria, play a major role in shaping life on Earth [1–3]. Phages outnumber their bacterial hosts by an order of magnitude [4]; approximately 10^{30} are estimated to exist, occupying every niche of the biosphere in which their microbial counterparts can be found [1,5]. Phage-bacteria interactions have been the subject of intense study for over 100 years, leading to advances in molecular biology, biotechnology and antimicrobials to name but a few [6]. While study of phages has been intensive, the scale of the subject matter means that only a fraction of the whole picture is understood. Given the predicted number of phages in the biosphere, phage diversity is massively underrepresented in genomic databanks [7]. While new metagenomic approaches to discovery promise to reduce this deficit somewhat [8,9], it is vital that parallel efforts are made to increase the depth of our knowledge of phages. Indeed, phage-bacteria interactions have proven far more complex than previously imagined, with new modalities and mechanisms regularly being discovered. A full review of phages and their impact on biology is far beyond the scope of this study, though many excellent reviews exist [1,6,10–12]. Instead, this study focusses on phage-bacteria interactions in nature, methods bacteria use to resist phage infections, implications for applications of phages in phage therapy and the potential expansion of the molecular biology toolkit. In particular, this study aimed to shed new light on the recently discovered BREX (for BacteRiophage EXclusion) phage resistance system through the application of structural biology, microbiology and molecular biology techniques.

1.1.1 Phage therapeutics

The potential of phages as antimicrobial agents was recognised immediately on their independent discovery [13,14]. Phages were trialled as therapeutics over subsequent decades, but studies were often undermined by poor design and an inherent lack of knowledge of bacteriophages [4,15]. The advent of antibiotics and their large-scale manufacture during the second world war resulted in most western countries discontinuing pursuit of this goal [4]. In many eastern European countries however, research into phages as antimicrobials continued and application of phages to treat bacterial infections, known as phage therapy, was commonly utilised [15,16].

The recent emergence of multi-drug resistant bacterial strains has led to a global resurgence in research in this area [12,17]. A recent UK review estimated that antibiotic resistant infections accounted for around 700,000 deaths per year globally. An estimated 50 million deaths are estimated between the report publication and the year 2050, with additional economic cost estimates of 100 trillion US dollars [18]. Many of the deaths associated with antibiotic resistance are attributable to the nosocomial multi-drug resistant ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter* spp.) [19]. While primarily found to target immunocompromised hospital patients or those with weakened immune systems, increasing evidence is emerging of infection of otherwise healthy individuals and community infections [20,21].

With regards to resistance and discovery, phage therapy offers advantages to antibiotic treatments [21,22]. While bacteria may develop resistance against a given phage, phages in turn will actively evolve to counteract this. Theoretically, no bacteria can become resistant to all phages and phages should exist for all bacterial species. Additionally, phages demonstrate selective toxicity for their target and do not harm the patient or disturb the patient microbiome [23,24]. Many antibiotics suffer from poor access to infection sites through production of biofilms by pathogenic bacteria, or other physical barriers [25]. In contrast, many phages encode biofilm degrading enzymes, allowing better access to their target receptors [26,27]. As such, phage therapy and antibiotics have potential synergistic relationships when clearing persistent infections [28]. Biofilm degradation by phage therapy has also been shown to improve effectiveness of the immune response [29]. Additionally, phages replicate at the site of infection, ensuring high titres at target sites.

The antibacterial pipeline has proven largely unproductive in recent decades and large pharmaceutical companies are losing interest in funding research in this area [30], though recent progress has been made [31,32]. It is hoped that phages can provide an alternative or complementary solution to a significant threat. Effective lytic phages are commonly applied in a cocktail to minimise resistance events [24]. Recent successful applications of phage therapy where antibiotic treatments have proven unsuccessful have even reached mainstream media [33]. Conversely, a longer-term randomised phase 1/2 clinical trial of a 12 phage cocktail for the treatment of infected burn patients proved ineffective, with events of bacterial resistance present in failed treatments [34]. It should be noted here that phage application was four orders

of magnitude lower than initially planned, though this still resulted in a decrease in bacterial load. The application of phage therapy and the design of effective phage cocktails then, requires a detailed understanding of both phages and phage-bacteria interactions. Phages can also aid in the transfer of bacterial virulence factors, such as the stx phage of Escherichia coli O157 [35,36]. Finally, although phage therapy can be permitted on compassionate grounds in some western countries, much progress remains to be made on their licensing and regulation before phage therapy can become a widespread and accessible treatment [37]. In the UK, efforts to establish a national phage library and scale-up production processes to the internationally recognised good manufacturing practice (GMP) standard should aid availability [38].

1.1.2 Phages in molecular biology and biotechnology

In addition to utilisation as early anti microbials, phages are embedded in the history of molecular biology [6]. The simplistic nature of bacteriophages presented an excellent model for early studies on vital biological processes at a molecular level [11]: The first observations of random mutation in DNA were derived from the observation of differing resistance profiles of *E. coli* to phage T1 [39]; the relatively small genome size of phages allowed the phage ΦX174 to be the first fully sequenced DNA genome [40]; the simplistic life cycles of phages provided a model for the early study of genes and gene expression and the subsequent discovery of DNA structure [41–43]. Phage-bacteria interactions continue to produce a plethora of molecular biology tools [44–48]. More recently, biotechnological applications of filamentous phages have gained further industrial interest in the form of phage display [49]. Phage display involves displaying proteins of interest, usually antibody or other peptide drug candidates, as a fusion protein on the phage surface and selectivity for binding affinity to desired receptors [50,51]. This technique is used extensively in the pharmaceutical industry in target-based drug discovery for screening large libraries of candidates and producing mutants with improved binding affinity [49]. ϕ29 DNA polymerase has been utilised in DNA sequencing and forms the basis of PacBio SMRT sequencing technology [47]. Meanwhile, the discovery of CRISPR (clustered regularly interspaced short palindromic repeats)-Cas adaptive immunity in bacteria, discussed in detail in 1.3.5, has revolutionised molecular biology and genetic engineering [52–54] and provides an outstanding example of the potential applications that can be derived from phage-bacteria interactions.

1.1.3 Phage classifications and morphologies

The diversity of phages becomes immediately apparent from studying phage morphologies and genome formats. Historically, phages have been classified dependant on their capsid morphology and the format of their genetic material [55], though in the post genomic era it has become apparent that phage diversity is far greater than this simple system could describe [56,57]. As a result, the ICTV (International Committee on Taxonomy of Viruses) recently made broad changes to phage taxonomy and classification, including the abolition of the individual tailed morphological families, *Siphoviridae, Myoviridae* and *Podoviridae* and the order of *Caudovirales* and the new grouping of all tailed phages with double stranded DNA in a new class, *Caudoviricetes*, under a new binominal naming system (Figure 1.1).

Phages consist of a protein capsid containing phage genomic material in either single stranded or double stranded, DNA or RNA. A majority of known phages (96%) contain linear dsDNA and a majority of these are tailed [55], such as T4 (previously *Myoviridae*) and lambda (previously *Siphoviridae*). The polyhedral non-tailed *Microviridae* displays a circular single stranded DNA (ssDNA) genome. The *Plasmaviridae* family likewise possesses a circular genome but in the form of dsDNA, enclosed in a lipid-based membrane rather than a proteinaceous capsid. *Cystoviridae* on the other hand have a dsRNA with an outer membrane layer enclosing a double capsid layer. The *Inoviridae* family genome encodes just four ssRNA genes necessary for propagation in an icosahedral capsid. In contrast to these families, the length of the filamentous phages of the *Tubulavirales* order can measure in the region of micrometres rather than nanometres. Filamentous phages are unique in producing phage progeny without cell lysis, continuously producing phage virions which are secreted through the host cell membrane [58,59]. Filamentous phage infections have even been implicated in bacterial biofilm formation and maintenance, and persistence of infection in cystic fibrosis patients [60,61].

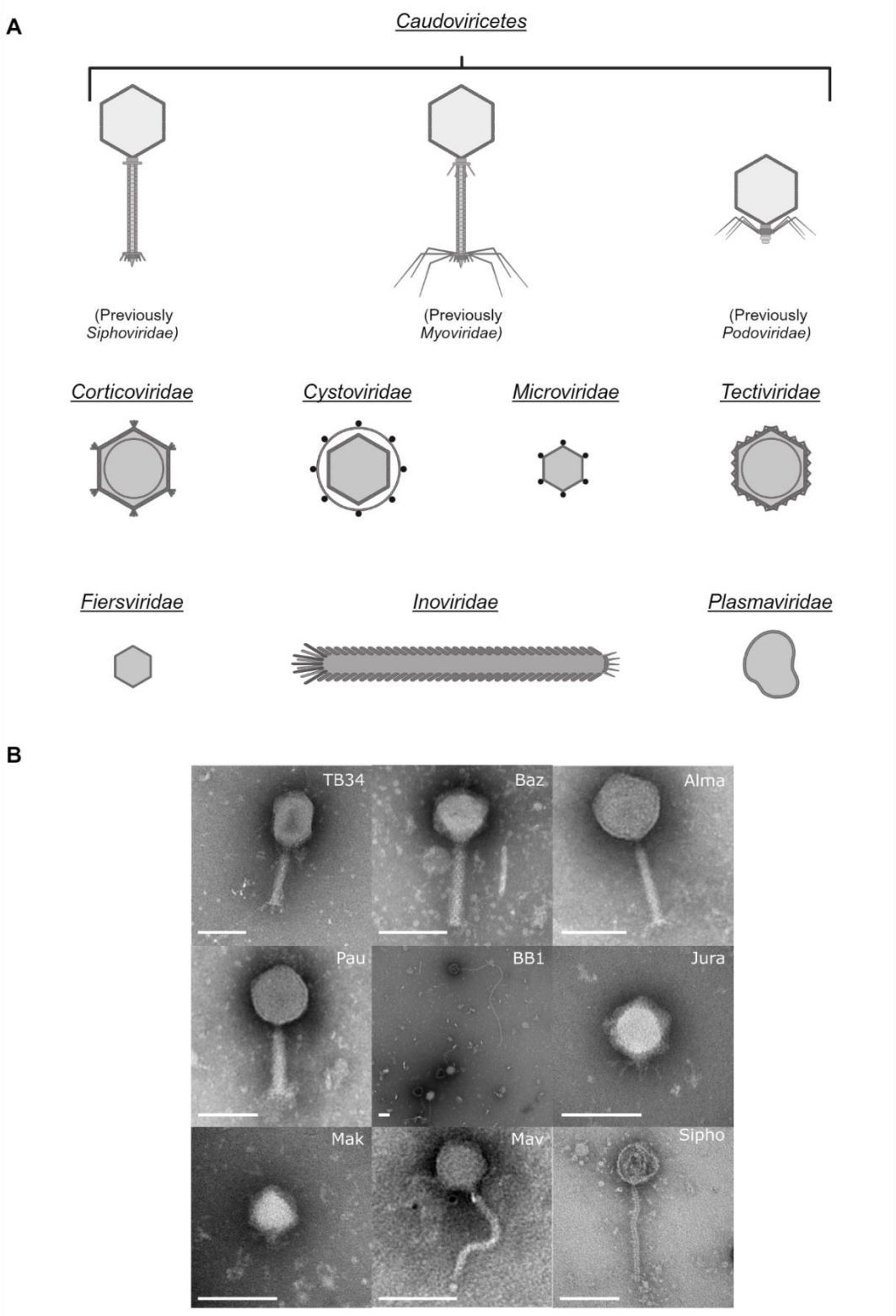

Figure 1.1 – Phage classifications. *A; Phage families and their general morphologies. B; Transmission electron microscopy of selected* Caudoviricetes *phages in the Durham Phage Collection* [62]*. Scale bar in whit represents 100 nm.*

1.2 Bacteriophage biology

Phages are obligate parasites, unable to replicate independently. Instead, phages inject their genomic material into the host cell and hijack host replication and expression machinery. As discussed in 1.1.3, phage diversity is enormous and a full discussion of the biology and lifestypes of phage classes is beyond the scope of this study. As this study utilises tailed phages (*Caudoviricetes*) isolated with *E. coli*, this section will predominantly focus on this group.

1.2.1 Adsorption and injection

Phages adsorb to host-specific ligands on the cell surface and inject their genetic material into the host cytoplasm [1]. Phage adsorption is highly specific and dictates phage host ranges. This specificity is a highly attractive property for antimicrobial agents and ensures that application of phage therapies does not overly disturb the microbiome of the patient [21,22,24]. Receptors are often essential membrane components such as membrane proteins (host cell signalling receptors, membrane channels, transport proteins) [63], lipopolysaccharides [64] and peptidoglycan [65]. Following irreversible adsorption, tail associated lysin proteins locally degrade the cell wall and a conformational change in the baseplate leads to contraction of the tail sheath from a high energy to low energy state [66,67]. This in turn causes a rotation of the tail tube which pushes the phage tail downwards, where the needle-like tip penetrates the bacterial membrane [66,67]. The extension of the tape measure protein forms a channel through the membrane for transfer of genetic material [68,69]. Genome injection is staggered, with early genes conditioning the host cell through functions including degradation of host DNA, inhibition of host gene transcription and translation, and repressing host defence systems [70].

1.2.2 Life cycles

From this point, phage infection cycles can be either lytic or lysogenic (figure 1.2). Lytic phages will replicate their genome and express phage capsule components using the host cell replication machinery, assemble into complete phage particles, lyse the cell and release phage progeny to infect new hosts. Temperate phages, such as phage λ, can enter either lytic or lysogenic cycles, depending on a range of environmental conditions and multiplicity of infection (MOI) [71]. Lysogenic phages will insert their genetic material into the host genome, with host replication also replicating phage DNA. Crucially, cytosolic repressor proteins expressed by lysogenised phages inhibit cellular reinfection. Thus, phages must balance rapid reproduction through lytic cycles with slower replication through lysogeny as available host cell numbers fluctuate to ensure host availability for progeny infections. In some cases, MOI can be determined via phage-phage quorum sensing-like arbitrium signalling systems and influences the lysis/lysogeny decision [72,73]. Phage λ genetics have been studied extensively and present the best model system to date of the lytic and lysogenic cycles [71].

1.2.2.1 Lysogenic cycle

In λ, early transcripts from the *pL* and *pR* promoters lead to expression of the regulatory genes *N* and *Cro*. Cro represses further transcription from *pL* and *pR* while N acts as an antitermination factor, inducing assembly of transcription complexes [74]. Delayed early genes encode lysogenic regulators CII and CIII as along with the late gene regulator Q. In lysogenic cycles, accumulation of the lysogenic regulator CII inhibits expression of *Q* and therefore expression of late genes by promoting transcription of the anti-sense RNA paQ. Q also promotes expression of the early gene repressor *CI*, preventing reactivation of the genetic pathway and also inhibiting phage reinfection [71]. A range of environmental cues, host cell stresses or other signalling events can cause the lysogenic phage to switch to the lytic cycle, form phage particles and lyse the cell. In most cases, prophage induction is instigated through activation of the host cell SOS response, promoting expression of RecA which cleaves the phage lytic repressor protein [71,75]. Prophage induction is heterogenous across the microbial community, ensuring phage lysis avoids host extinction [76].

1.2.2.2 Lytic cycle

In the lytic cycle, accumulation of Q results in expression of late genes through conformational changes in RNA polymerase causing insensitivity to terminator sequences [77]. Replication of phage DNA and expression of phage virion proteins proceeds utilising host machinery and resources. Phage heads assemble first [78], forming an icosahedral structure comprised of one or more phage capsid proteins with the aid of scaffold proteins [79]. Head assemblies also contain a portal protein, which acts as a strong molecular motor which packages DNA into an energetically unfavourable and highly condensed form [80,81]. Tail assembly then follows [78], with tail proteins either sequentially assembling from the capsid neck proteins [82] or assembling separately and attaching to assembled heads upon completion [78], and preassembled tail fibres are attached to the completed phage virion [78].

Host cell lysis follows in a process now known to be distinct and separately regulated from virion assembly. In Gram negative cells, lysis requires three stages for the degradation of the inner membrane, peptidoglycan and outer membrane, respectively [83]. In the classical model of phage λ, the inner membrane is targeted by holin proteins, allowing passage of endolysins into the inter-membrane space and disrupting the proton motive force [84–86]. Endolysins are then free to degrade the cell wall [83], exposing the outer membrane. Disruption of the outer membrane however is not achieved by holin enzymes but by spanins, which mediate the fusion of the inner and outer cell membranes, releasing phage progeny [87]. Timing of cell lysis is essential for the release of viable phage progeny, but temporal control of lysis is independent of phage replication and virion assembly pathways. Instead, timing of cell lysis is dependent on holin proteins themselves, through accumulation of holin proteins in the cytoplasmic membrane reaching an allele-specific critical threshold resulting in either mass aggregation and membrane lesion [88] or disruption of the proton motive force [89]. This releases phage progeny for subsequent infection cycles.

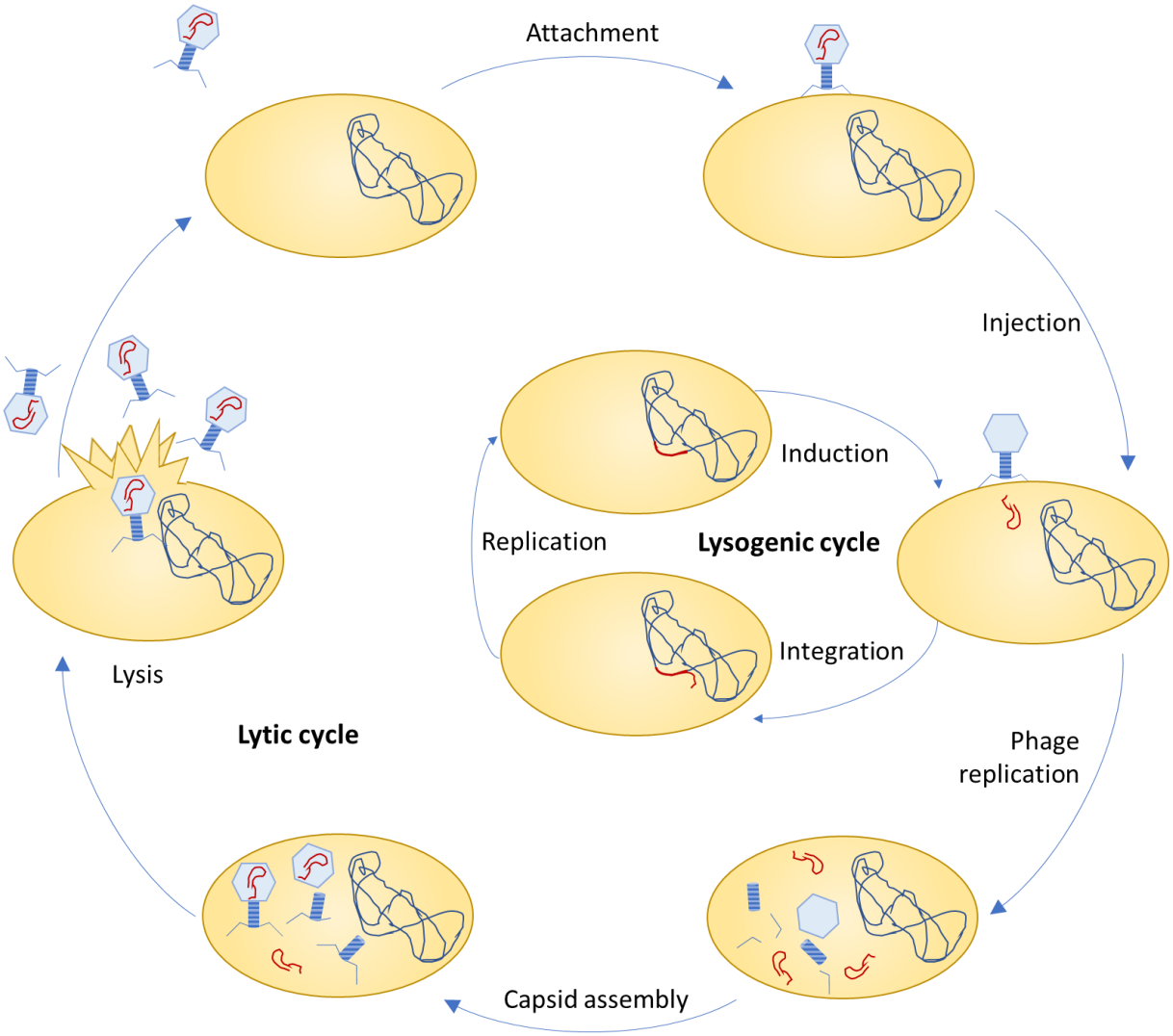

Figure 1.2 – Depiction of the lytic and lysogenic phage life cycles. *Phages adsorb to the host cell surface through interactions between the tail proteins and host specific ligands and inject their genetic material into the cell. Lysogenic phages then integrate their DNA into the chromosome and replicate with the host. Lytic phages begin replication using host machinery, form phage virions and lyse the cell. Progeny are then free to initiate new infections.*

1.2.3 Genomic organisation

In general, phage virion size is dictated by genome size and by capsid geometry. The latter can be described by the capsid triangulation number as a measure of the size and complexity of a viral capsid [90]. As most of the phage life cycle is dependent on hijacking host systems, phage genomes are comparatively small relative to their bacterial hosts [91]. Selectivity for reduced size has resulted in fewer non-coding DNA regions than other organisms [92,93]. Phage genomes are modular, with regions expressed at particular

points in the phage infection cycle [94,95]. This modularity allows a high degree of recombination between phages, both within and across species [96]. The high levels of horizontal gene transfer have in turn led to the enormous phage diversity. As a result, phages display a genetic mosaicism and do not follow traditional hierarchical phylogenetic patterns. Instead, a network-based phylogeny can be used to describe phage relationships utilising structural similarities of proteins rather than sequence similarities [57]. As lysogenic phages insert their genetic material into the host chromosome, phages are also drivers of bacterial evolution [97]. Horizontal gene transfer events instigated by lysogenic phages are often mutually beneficial, with phage genes offering antibiotic resistant capabilities and resistance to infection by related phages [98]. Lysogeny is also known to be the cause of changes in the pathogenic phenotype of bacterial species. Stx-phages for example are the source of the pathogenic phenotype of shiga toxin producing *E. coli*, a food borne pathogen responsible for may epidemics in Asia [35,36].

1.3 Bacterial defence against bacteriophages

While phages have become adept at manipulating bacterial genomes, these interactions are not one-directional. Over evolutionary history, bacteria have developed a number of defence systems against phages, many of which have proven to have biotechnological utility. Bacteria have developed mechanisms of inhibiting every step of the phage lifecycle [99]. Strategies range from simple mutations in phage receptors on the cell surface, to complex multi-component mechanisms to differentiate self from non-self, to adaptive strategies analogous to immune system memory in e*ukaryotes*, targeting previously encountered phage threats.

1.3.1 Adsorption and injection

While phage adsorption is highly specific and host receptor mutation may be sufficient to produce resistance, more active mechanisms of inhibiting absorption are employed. Many bacteria produce an extracellular matrix, often composed of polysaccharides, as a first line of defence against phages and as a protective layer against environmental hazards [27]. *E. coli* for example produces the exopolysaccharide

colanic acid, as well as protein-based fibre known as curli [100,101]. Colanic acid restricts phage penetration and curli blocks adsorption by covering cell receptors and binding phage particles [102]. Longer periods of biofilm maturation were shown to correlate with increased protection [102]. Naturally, phages have counter-evolved to overcome biofilm protection. Many phages encode exopolysaccharide degrading enzymes in the form of lyases and hydrolases, which can be either displayed on the phage surface or released from cells lysed in previous infections [103,104]. Phage adsorption can also be prevented by blocking phage receptors. The fertility plasmid of *E. coli* encodes an outer-membrane lipoprotein TraT, which masks OmpA, a common receptor for *E. coli*-targeting phages [105]. Conversely, phages have been found to mask their own receptors to prevent superinfection or inactivation of particles by way of binding to receptors on already lysed cells [106]. Competitive inhibitors of phage adsorption are employed by some bacteria: the antimicrobial molecule J25 can competitively inhibit phage T5 binding to the *E. coli* iron transporter FhuA [107]. Finally, release of outer membrane vesicles containing surface receptors by bacteria can function as decoy receptors for phages, sequestering phage particles and reducing productive infections [108]. Uptake of outer membrane vesicles can transiently expand phage host ranges, potentially providing a novel mechanism of inter-species horizontal gene transfer [109].

In contrast to other mechanisms of resistance, inhibition of DNA injection by bacteria is thought to be conferred by prophages as a mechanism of preventing superinfection [110,111]. In T-even coliphage infection, two phage encoded genes, *imm* and *sp*, are responsible for blocking DNA injection [112]. Imm is a transmembrane protein which prevents injection by altering the conformation of the membrane at the injection site [112,113]. Sp functions as an inhibitor of phage lysozyme, responsible for penetrating the peptidoglycan layer of the cell wall [112]. The superinfection exclusion protein Ltp is found in *Lactococcus lactis* genomes and is encoded by the prophage TP-J34 [114]. Ltp is an extracellularly secreted lipoprotein which blocks DNA injection through interfering with the phage tape measure protein [114,115]. The origin of superinfection exclusion proteins in phages again highlights the sometimes mutually beneficial relationship of bacteria and prophages by protecting host cells from infection by similar phages.

1.3.2 Abortive infection and toxin-antitoxin systems

In addition to individual cell defence, bacteria also display population level defence systems aimed at limiting effective phage progeny and reducing or minimising subsequent infections in the bacterial community. This is achieved in the form of a number of abortive infection (Abi) systems which induce altruistic cell death, prematurely lysing cells before phage virion assembly can be complete [99]. The mechanisms of Abi systems are diverse, and study is difficult due to the inherent toxicity of their components [116]. Though many of the known Abi systems function at late stages of phage infection – i.e. phage virion assembly: AbiC [117], phage DNA packaging: AbiE [118], AbiI [119], AbiQ [120], and premature cell lysis: AbiZ [121] – Abi systems functioning at all stages of the phage infection cycle have been observed [99].

1.3.2.1 Classic Abi systems

The Rex system of *E. coli* is perhaps the best characterised system and has provided a model for Abi system study which allowed early insights into phage biology [122]. The Rex system functions via formation of membrane channels, reducing membrane potential and consequently adenosine-triphosphate (ATP) production, halting the cellular replication machinery necessary for phage replication. RexA recognises and binds to phage replication complexes. Binding of two RexA molecules results in activation of the RexB ion channel. The Rex-like systems have been observed in evolutionary diverse bacterial strains [123]. In all cases, escape mutants arose through encoding mutations in proteins bound by the RexA component [123,124].

Several Abi systems have been shown to function as toxin-antitoxin (TA) systems, in which a toxic component is actively neutralised by its antitoxin. TA systems are divided into six types dependant on their mechanism of toxin and antitoxin interaction [125]: Type I involve mRNA based interference, preventing transcription; type II involve protein-protein interactions for toxin inactivation; type III antitoxin transcripts interact with the endoribonuclease toxins resulting in inactive TA complexes; type IV components have antagonistic effects on the same target but do not directly interact; type V antitoxins encode RNase domains which specifically degrade toxin mRNA transcripts; type VI antitoxins bind to toxins and promote degradation by a third party but do not independently inhibit toxin activity. In phage defence. Tight

regulation of TA systems is vital to avoid toxicity in favourable conditions, in this case, in the absence of phage threat.

Many of the known TA systems with phage defence properties originate in plasmid encoded lactococcal systems [126]. The function of AbiQ in phage resistance and its activity at the DNA packaging stage has been known for over two decades [120]. Around 15 years later, AbiQ was found to function through a type III TA mechanism [127,128]. The toxin, AbiQ, is an endoribonuclease which selectively degrades the antitoxin; a constitutively transcribed mRNA transcript consisting of 2.8 repeats of 35 nucleotides. In doing so, the antitoxic mRNA sequesters AbiQ, preventing cleavage of bacterial targets [127,128]. Though the exact mechanism of disruption of TA equilibrium during phage infection is unknown, it is proposed that a specific phage protein disrupts the regulatory TA complex, either through interfering with the antitoxin or from altering the activity of AbiQ [129]. In common with several other Abi systems, phage disruption is related to early-stage phage genes involved in nucleic acid pathways [127,128].

Unlike many other Abi systems, AbiE displays bacteriostatic activity rather than bactericidal [130]. AbiE is a type IV TA system consisting of a non-interacting TA pair in a bicistronic operon. Within the AbiE system, AbiEi has dual function both as both the antitoxin and in negative autoregulation of the AbiE operon [130]. The toxin, AbiEii, is a nucleotidyltransferase which transfers GTP to an unknown cellular target, resulting in growth inhibition [130,131]. Though the antitoxin mechanism of AbiEi is unknown, it has been demonstrated that only the C terminal domain is necessary for protection [130]. Meanwhile, the full protein is necessary for AbiE regulation [130,132]. This is somewhat analogous to type II system in which the N-terminal domain functions through transcriptional regulation and the C terminal domain interacts with the toxin component [133]. Though autoregulation is common in type II TA systems, the AbiE system is the only example in type IV systems. For AbiEi, the positively charged surface of the C terminal domain facilitates DNA binding of two monomers to two inverted repeats within the AbiE promoter region [132]. This produces a 72° bend in the DNA, with initial monomer binding promoting binding of the second, ultimately blocking access of DNA polymerase [132].

1.3.2.2 Current understanding of TA systems as mediators of abortive infection

Methods for the systematic discover of new phage defence systems (see 1.3.6) have resulted in the discovery of an abundance of TA systems in recent years [134–136]. This has revealed new and complex modalities and interactions of TA systems and phages and led to re-examining of the link between TA systems and Abi [137]. TA systems are widespread throughout nature, and many are not linked to phage defence or abortive infection. Strains of the human pathogen *Mycobacterium tuberculosis* for example have been found to encode as many as 88 TA systems providing traits including survivability of environmental stress, survival inside macrophages, resistance to antibiotics and pathogenicity, to name but a few [138]. Recent studies have also shed light on the activation and mechanisms of newly discovered TA systems (Figure 1.3), though the molecular details of many of these systems remains poorly understood [137]. Further insight has also been gained into mechanisms of TA system escape by phages, as detailed further in 1.3.9.

Recent evidence has demonstrated that activation of the type II TA systems, DarTG and CapRel^{SJ46}, proceeds without antitoxin proteolysis – in contrast to the classical type II TA system model – through direct activation by phage components [139,140]. Activated DarT ribosylates host and phage DNA, blocking replication [141]. CapRel ^{SJ46} phosphorylates tRNAs, blocking translation [140]. Activation of the type III TA system ToxIN has been confirmed to be the result of transcriptional shutdown by invading phages, causing destabilisation of the ToxN : ToxI ratio [142]. Once liberated, ToxN cleaves both host and phage mRNA transcripts, preventing formation of new phage virions [142]. Recent studies on the HokW-SokW system [143] and a tripartite Kinase-Kinase-Phosphatase (KKP) system [144] have shown that prophage encoded TA systems can function in aiding induction and regulation of prophages and prophage genes, respectively.

Bacterial retrons are composed of a reverse transcriptase, a non-coding multi-copy short DNA element (msDNA) and an accessory protein [145], and had previously been implicated in phage defence [146]. Only recently, however, retrons have been shown to act as tripartite TA systems functioning in phage defence [147–149]. Activation of retron toxins is poorly understood, through the retron Sen2 from *Salmonella enterica* was shown to be activated by methylation of, or degradation of, msDNA causing release of the RcaT toxin from the retron complex [149]. The *E. coli* retrons Ec48, Ec73 and Ec86 however were activated by recognition of specific phage components or by inhibition of the RecBCD complex by the msDNA within the retron complex [147,150]. The mechanisms of action of retron toxins remains elusive, though the RcaT

toxin of Sen2 is predicted to target nucleotides of nucleic acids [149]. As such, retrons appear to provide flexible activation mechanisms resulting in release of diverse toxins through a conserved mechanism.

In terms of the molecular mechanism of toxins within the cell, the link between toxin function and cell death is not always clear, with toxin activity related to inhibition of phage replication and cell death more likely linked to the invading phage shutting down host transcription/translation, resulting in growth arrest and ultimately, cell death [137]. ToxIN for example was previously hypothesised to cause growth arrest and cell death but has recently been shown to cleave phage mRNA transcripts, preventing phage virion formation [142]. The AbiEii system may provide similar modality, in which the nucleotidyltransferase activity may prevent tRNA charging and hence phage transcript translation, but resultant cell death is caused by phage inhibition of cellular machinery [130,151]. Thus, it has been suggested that modalities of new TA systems be considered in this same context, and that upon phage infection which would ultimately result in cell lysis regardless, toxin activity may result in impaired phage replication, preventing subsequent infections from phage progeny [137].

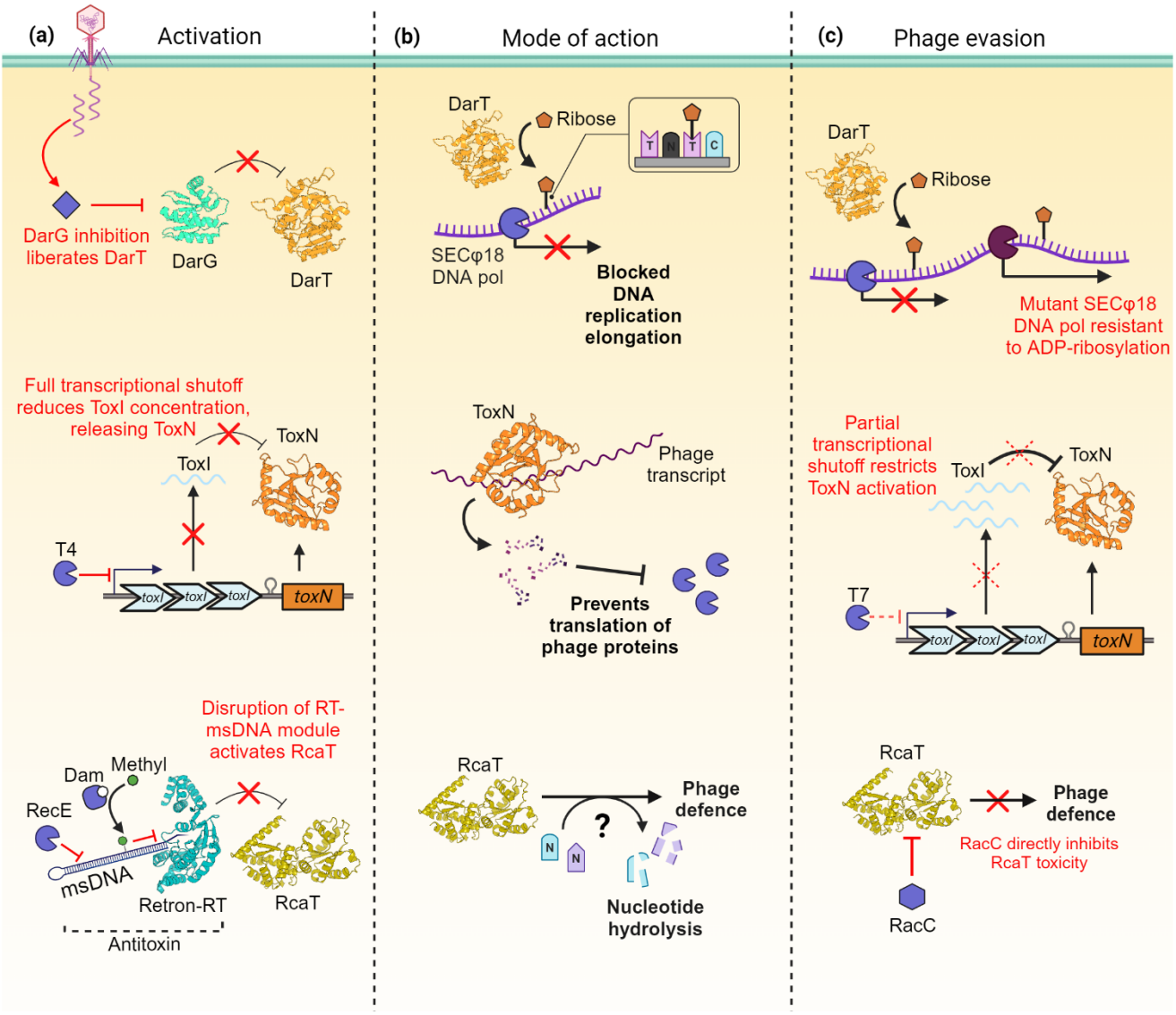

Figure 1.3 – Activation, mechanism of action and phage evasion for TA systems. *DarTG (top row) is activated by direct toxin release, whereupon DarT ribosylates DNA, blocking DNA replication. Phages can escape DarT activity by mutation DNA polymerase allowing replication elongation despite ribosylation. ToxIN is activated by phage-induced shutdown of host transcription, releasing ToxN which degrades phage transcripts. Phages can escape ToxN activity by only partially shutting down host transcription. For Retron-Sen2, disruption of the msDNA-Retron complex activates RcaT which hydrolyses available nucleotides. Phages can escape RcaT through direct inhibition mechanisms.*

1.3.3 Restriction-modification

A majority of known phage resistance mechanisms function after the point of DNA injection and prevent phage replication, often through degradation of phage DNA. Our earliest knowledge of phage resistance came in the form of restriction modification (RM) systems [152]. RM systems consist of a modification component which modifies the DNA of the bacterial host at a specific motif, and a restriction component, which cleaves un-modified DNA [153]. Methylation of host DNA provides recognition between self and non-self, preventing cleavage of the host genome. Thus, host DNA is protected from restriction, and foreign DNA from phages or other mobile genetic elements lacking modification at the same motif is degraded. RM systems are divided into four types and various subtypes based on their subunits, domain organisations, specificities and activities (Figure 1.4).

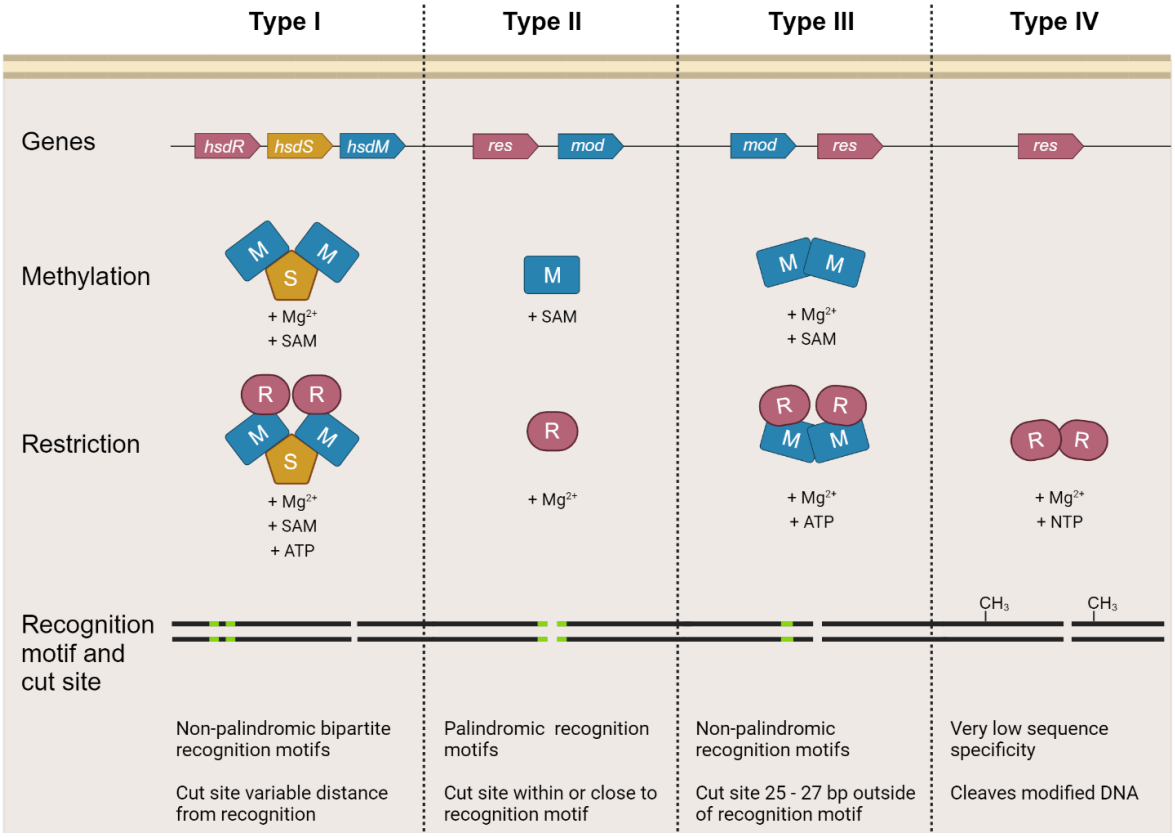

Figure 1.4 – Types of RM systems. *Genes, methylation complex and restriction complex components of type I, II, III and IV RM systems. Co-factors of methylation or restriction are shown under each complex, respectively. Positioning and cut site around recognition motifs are shown at the bottom. Type II systems show huge diversity and can be further separated into subtype (see 1.3.3.2), genes and complexes shown represent the most common organisations. M is methylase subunit; R is restriction subunit; S is specificity subunit.*

1.3.3.1 Type I

In type I systems, methyltransferase (HsdM), restriction (HsdR) and specificity domains (HsdS) are located on separate peptides [154]. Modification may occur independent of restriction components but requires presence of the specificity subunit for target recognition, with modification complexes formed of two HsdM components directed by one HsdS component [155]. A second complex of two HsdM subunits and two HsdR subunits alongside a HsdS subunit is capable of both methylation and restriction functionalities [156,157]. Type I RM complexes display bidirectional translocation along thousands of bases of DNA upon identification of unmodified recognition motifs [158,159]. Recognition motifs are bipartite, and cleavage is distant from recognition sites. Type I modification systems almost exclusively catalyse N6mA DNA modifications on each strand [154,160], though recently a novel type I RM system has been described which produces N6mA on one DNA strand and N4mC on the opposite strand through encoding two distinct HsdM subunits [161,162], opening the door to the possibility of alternative DNA modification by type I RMs. Despite over 50 years of study [163], structural data on type I RM systems is lacking, though some recent advances have been made in this area [155–157,159,162]. The specificity factor consists of two target recognition domains (TRDs) and two conserved domains which form helical bundles, separated by a long central helical region which produces the characteristic bipartite recognition/modification [155,156]. In the methylase complex, one HsdM subunit C-terminal interacts with one of each of the helical bundles and each target recognition domain of the HsdS subunit binds one of the bipartite recognition motifs through contacts between the positively charged HsdS surface and the negatively charged DNA phosphate backbone [155,156]. In the absence of recognition sequence, the methylase complex remains in an open conformation but forms a closed clamp-like conformation upon recognition motif binding, with additional interactions between the N-terminals of the HsdM domains [156]. The restriction complex can display either translocation or restriction conformations [157]. The $HsdS_1HsdM_2HsdR_2$ complex presents similarly to the closed methylase complex with a HsdR subunit bound to each of the HsdM subunits in a symmetrical manner through several interaction interfaces [157]. In the translocation conformation, the HsdM subunits are separated by around 22 Å, preventing full clamping around the DNA and thus, preventing methylase activity, and the bound DNA is positioned in proximity with the motor domain of each respective HsdR subunit [157]. To date, no structure of the restriction state of a $HsdS_1HsdM_2HsdR_2$ complex is available, though this state would likely require additional conformational changes to bring the DNA molecule into proximity of the nuclease domains.

1.3.3.2 Type II

Type II RMs have long been known as the workhorses of molecular biology and are by far the most well studied group [160,164]. The biotechnological utility of type II RMs is due to the cleavage by the endonuclease component at a fixed point within or close to their recognition motif, generating reproducible and predictable cleavage sites invaluable for DNA cloning [164,165]. Type II systems encode methyltransferase and endonuclease domains on separate peptides each containing a TRD, or on the same peptide under the control of a single TRD [164,165]. The sequence and structure of TRDs are distinct and methylase and endonuclease subunits bind recognition motifs via different molecular interactions [164]. Restriction and modification both occur independently, and recognition motifs are short palindromic sequences with cleavage within or adjacent to recognition sites. Most type II restriction enzymes contain catalytic domains of the PD-(D/E)-x-K superfamily, though sequence and structure around these domains is highly divergent [165]. The localised cleavage site allows type II RM systems to be ATP independent, unlike type I and III systems which require ATP for translocation. Nevertheless, type II RMs bind non-specifically to DNA and diffuse along it until a recognition motif is encountered [166,167]. The mechanism of this diffusion is still poorly understood, though it is thought to be a combination of DNA sliding facilitated by contacts with the DNA backbone and DNA hopping [164], whereby the RE dissociates from the DNA molecule and reattaches at another site.

Type II RM systems are highly diverse, and they are grouped into further sub-types dependant on their recognition motifs, methylation types, cleavage sites, domain organisations and cofactor requirements (Table1.1). With the creation of subtypes, 11 were initially defined, each with distinct but not necessarily unique properties [168]. As such, type II RMs can belong to more than one subtype (Table 1.1). The most commonly used RMs in molecular biology are the type IIP enzymes which recognise palindromic motifs and cut within the recognition sequence, or on the edge. Type IIC contain both endonuclease and methylase domains on a single peptide, along with a single TRD, and are capable of providing both restriction and methylation modalities. Type IIS RMs recognise asymmetric motifs and cut at a defined distance outside of their recognition sequences, making them useful for scarless cloning techniques [169,170]. Type IIT RMs contain two different catalytic sites, either on a single peptide or on two RMs forming heterodimers, which cleave within or at the edge of the recognition motif. Type IIT enzymes can easily be converted into nicking endonucleases by introducing mutations which inactivate one of the two active sites [171,172]. Recently,

a novel RM was described which contains both methylase and endonuclease domains on the same peptide but produced hemi-methylated DNA at non-palindromic recognition motifs [173]. The newly discovered RM, MmeI, was found to belong to a wider family of RMs with similar features and highly modular organisation and was proposed to form a new subtype, type IIL [174].

Table 1.1 – Type II RM subtypes and characteristics. *Type II RM systems can belong to multiple subtypes.*

RM type II subtype	Characteristics
IIP	Palindromic motif
IIA	Non-palindromic motif
IIC	Palindromic or non-palindromic motif, R and M functions on one protein
IIG	Palindromic or non-palindromic motif, cleavage requires SAM
IIS	Non-palindromic motif, cleavage outside of recognition site
IIM	Target methylated DNA
IIH	Palindromic or non-palindromic motif, methyltransferase has separate M and S subunits
IIT	Palindromic or non-palindromic motif, R subunit forms heterodimers with only one catalytic site each
IIB	Bipartite motifs, cleaves either side of target motif
IIE	Requires at least two pseudo-palindromic motifs for cleavage at one target site
IIF	Requires at least two pseudo-palindromic motifs for cleavage at two target sites

1.3.3.3 Type III

Type III systems also encode methyltransferase and endonuclease domains on separate peptides however, only the methyltransferase component possesses a TRD [175]. Thus, methylation occurs independently, requiring two methyltransferase subunits, but restriction requires presence of methyltransferase and restriction components for cleavage [175]. The composition of the restriction complex has been disputed and has been reported to exist as a heterotetramer of two methylase subunits and one restriction subunit [176] or as a heterotrimer with only one restriction subunit [177]. To date, no high-resolution structural model of a type III RM system exists, though X-ray scattering experiments suggest that the complex consists of two central methylase subunits with a restriction subunit located on each side, forming an elongated crescent shape [176]. Modification produces hemi-methylation at short non-palindromic sequences with cleavage 25 – 27 bp from recognition sites [175]. Currently, all known type III RMs produce N6mA modifications, though few systems have been biochemically characterised, [175]. Cleavage requires two restriction complexes each binding to a recognition motif in either a head-to-head or tail-to-tail orientation [178]. The requirement of two recognition sites for cleavage is considered a defining characteristic of type III RMs. Recognition sites may be separated by between dozens to thousands of bases and cleavage requires the translocation of one restriction complex to the site of another following motif recognition [179]. Translocation is ATP dependant and powered by a translocation domain, similar to type I systems [180].

1.3.3.4 Type IV

Type IV restriction systems independent restriction components with no associated methyltransferase which target and cleave methylated DNA and therefore, act inversely to most other RM systems [181]. The exception are the type IIM methylation dependant RM systems, though it is argued that this subtype should be merged into the type IV classification [181]. Type IV restriction systems in bacteria represent an innate immune system, degrading non-modified DNA non-discriminately. Type IV restriction systems recognise a variety of DNA modifications and are largely sequence-independent [160,181]. As such, the cleavage site is variable and can be dozens of bases from the recognised base modification. Type IV systems predominantly recognise 5-methylcytosine (5mC), N4-methylcytosine (N4mC) and N6-methyladenine

(N6mA) modifications, though recognition of hydroxymethylated and glycosylated bases has been observed *in vitro* [182,183]. Structural data on type IV restriction endonucleases (REs) is lacking, though a recent study provided the crystal structure of the type IV enzyme, BrxU, associated with a BREX phage defence system [183]. BrxU presents as an intertwined dimer in solution. N and C-terminal domains are connected by a linker region which loop each other and N and C-terminal regions within each monomer interact with each other. BrxU then undergoes nucleotide-dependant separation to produce the active monomeric form and cleaves DNA containing 5mC, N4mC and glc-N5hmC modifications [183]. Bacteriophages are known to produce exotic base modifications [184,185] and it would be expected that type IV restriction systems exist which target these mutations also.

1.3.3.5 Mechanisms of methylation and types of modification

Bacterial methylation invariably occurs on DNA bases, specifically cytosine and adenine, and in a majority of cases methyltransferase activity is dependent on a S-adenosyl-L-methionine (SAM) molecule as a methyl group donor (Figure 1.5). Methylation occurs on either the amino group of the fourth carbon of cytosine (N4mC), the fifth carbon of cytosine (5mC) or the amino group of the sixth carbon of adenine (6mA) [184,185]. Type II RM systems are the most well characterised both structurally and functionally due to their utility in molecular biology and several structures of RM components bound to DNA are available. The methylation component of type II systems transverse host DNA until recognition of target DNA motifs [164,165]. Binding of DNA elicits energetically unfavourable bending of the DNA molecule, the energy of which is in turn used to flip the target base out of the DNA helix [186]. Bacteriophages are known to encode more exotic modifications [185], including N5hmC (hydroxymethyl-cytosine) and derivatives with various sugar moieties attached, aminocarboxymethyl-adenine [185] and diverse 7-deazaguanine based modifications [187], to name but a few. These modifications shield phage genomes from degradation by traditional type IV restriction systems. In most cases, exotic DNA bases are produced through modification of free DNA bases before incorporation into DNA upon replication and then further hypermodified [185]. Modification of the DNA backbone is also possible [188–190]; systems encoded by bacteria can introduce sulphurs in place of the non-bridging oxygen and provide phages defence in an analogous manner to RM systems (see section 1.3.4). A single phage genome modification type can provide immunity to multiple diverse nuclease-containing phage defence systems [186].

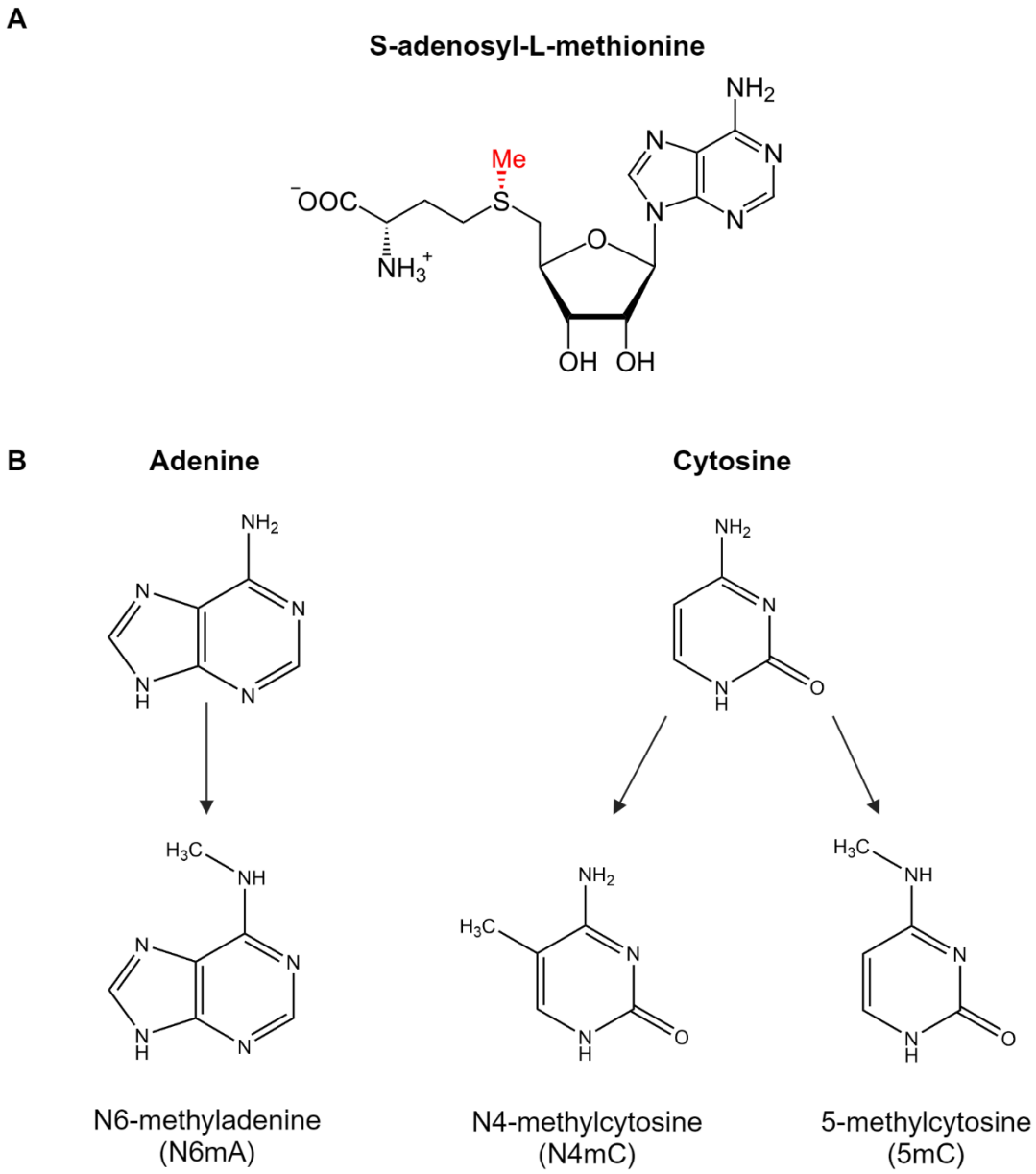

Figure 1.5 – Common bacterial DNA modifications. *A; RM-mediated methylation utilises S-adenosyl-L-methionine (SAM) as a methyl group donor, using the methyl group bound to the sulphur moiety. B; The three most common DNA base modifications in bacterial genomes. Adenine is modified at the N6 position to produce N6mA. Cytosine can be modified at the N4 or C5 position to produce N4mC or 5mC, respectively.*

1.3.3.6 Rational modification of motif recognition

As discussed in above sections, RMs, and particularly type II RMs, are invaluable molecular biology tools. A plethora of RMs are now commercially available and sequencing and bioinformatic advances have provided information on thousands of putative RM systems and recognition motifs [160]. Nevertheless, the sequence space covered by commercially available RMs is relatively small. The ability to rationally engineer novel recognition motifs would be invaluable to molecular biology, potentially allowing targeting of any given sequence on demand. Efforts towards this goal have so far been limited [191–194] and are hindered by the intrinsic link between both structure and sequence in motif recognition and binding. The type IIL RM, MmeI, contains both restriction and modification domains on a single peptide, guided by a single TRD and the MmeI family of enzymes are closely related and seemingly modular in design [174]. MmeI family enzymes recognise a non-palindromic 6 bp motif (TCCRAC) and produces N6mA modifications at the fifth, adenine residue on only one DNA strand [174]. Through multiple sequence alignments of TRDs of MmeI family members, residues implicated in motif recognition were inferred and rationally mutated [194]. This approach successfully generated MmeI mutants which recognised altered DNA bases at three of the non-adenine bases in the recognition motif [194]. Elucidation of the structure of MmeI in complex with DNA allowed further success through direct identification of residues involved in motif binding and allowed altered specificity at the remaining non-adenine bases [192]. In a similar study, the recognition motif of the type II restriction enzyme, R. MwoI, was altered based on structural homology to the DNA bound structure of the homologue, R. BglI, with mutation of residues implicated in DNA binding resulting in several novel recognition specificities [193].

1.3.4 Phosphorothioation

Analogous to RM systems, the Dnd and Ssp systems function by a phosphorothioation (PT) of the DNA backbone at specific motifs rather than modification of bases [188]. PT of DNA is a replacement of the non-bridging oxygen in the DNA backbone with a sulphur group [195,196]. PT has long been used as a biochemical method *in vitro* and is known to confer resistance to nuclease degradation [196,197]. Meanwhile, the discovery of a novel DNA modification in *Streptomyces lividans* appeared to promote DNA degradation and was found to depend on a gene cluster termed Dnd [198–200], though the exact

modification was unknown at the time. This modification was later revealed as a PT modification-based self/non-self recognition method for a phage defence system conferring resistance to a diverse range of phages [199,201] and found in a broad evolutionary range of bacteria [189]. Double stranded PT modification in achieved through the activities of DndABCDE. Briefly, DndB nicks DNA at specific motifs and stabilises structure and DndE tetramers provide affinity for negatively charged phosphorylated nicking sites. Meanwhile, DndA removes the sulphide group from a L-cysteine residue to produce a persulphide intermediate which is passed on to DndC. It is then proposed that DndC adenylates the target phosphate group, allowing attack and replacement by the persulphide [190]. Restriction of non-modified DNA is dependent on the DndHGF components, though the exact mechanism remains unclear. DndHGF form complexes which produce double strand cleavage of non-modified motifs [190,201].

Further investigations into PT modifications revealed a subset of bacteria in which PT modification occurred on a single DNA strand and was uneven throughout the chromosomes in strains lacking Dnd genes [188]. Further, study revealed the Ssp system which is analogous to the Dnd system in modification but distinct in its mechanism of phage resistance. Namely, unlike the Dnd system, the Ssp system did not degrade DNA. Instead, the Ssp system utilises a single protein effector, SspE, with dual nickase and NTPase domains. Upon recognition of unmodified motifs, SspE produces single strand nicking and NTPase mediated DNA translocation, resulting in inhibition of phage replication at an early stage of infection [188]. Thus, introduction of massive nicking in foreign DNA is a further mechanism of phage defence without requiring complete DNA degradation. In addition, the single stranded modification of the Ssp system provides protection against ssDNA phages.

1.3.5 CRISPR-Cas

Though RM systems are the most well-known and arguably most significant contributors to molecular biology to date, the diversity of phage resistance mechanisms is becoming apparent. The discovery of the CRISPR-Cas resistance system has already had a significant impact on the fields of molecular biology and gene editing [202,203]. More complex in its mechanism than RM systems, CRISPR-Cas represents an adaptive phage resistance system which imparts memory of past infections to its host, analogous to the mammalian adaptive immune response [48,204]. Like RM systems, CRISPR-Cas systems are diverse, with

two classes, six types and over 30 subtypes (Figure 1.6), each varying in their mechanisms and compositions [205]. This diversity reflects the requirement for rapid adaption to combat an ever-evolving phage threat. Common to all systems are three stages – adaptation, expression and interference (Figure 1.6) – which results in acquisition of immune memory of a phage in the shape of spacer sequences derived from foreign DNA [206]. Contrary to the expression and interference stages the adaptation stage is relatively conserved, consisting of the Cas1 endonuclease and the Cas2 structural protein. Together, Cas1 and Cas2 form complexes and incorporate foreign DNA fragments into CRISPR arrays, known as spacers, separated by short, type specific, protospacer adjacent motifs (PAM) [207].

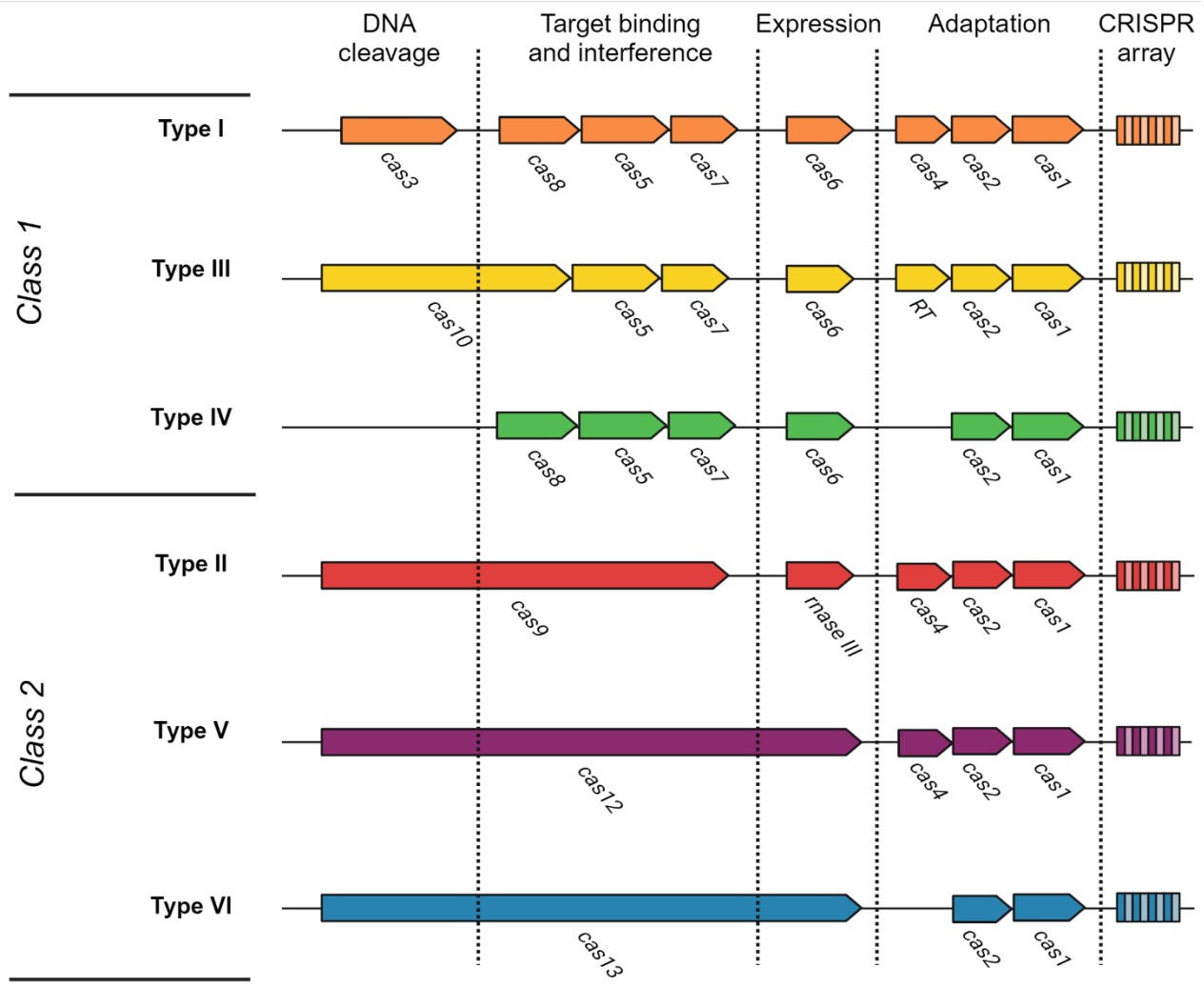

Figure 1.6 – CRISPR-Cas classes and types. *System compositions of the CRISPR-Cas types sorted by class. The roles each gene in CRISPR-Cas function are separated by dotted lines. Class 2 is distinguished by encoding a single component effector, in contrast to the multiple component effector complex of class 1.*

Spacers are invariably added to the 5' end of the array, which generates a stronger anti-phage response and therefore, prioritises immune response against the most immediate threats [208]. Spacer acquisition against new phages, termed naïve adaptation, typically leads to acquisition of a single spacer sequence, while acquisition of spacers from previously encountered phages, termed, primed adaptation, leads to multiple spacers. The low spacer acquisition of naïve adaptation is thought to be the result of low self/non-self discrimination, leading to high degrees of toxicity should spacers be incorporated from host DNA [209]. Thus, low spacer incorporation rates reduce the risk of cytotoxic autoimmunity. Nevertheless, discrimination between host DNA and mobile genetic elements is present, as demonstrated in host strains encoding adaptation components but deficient for interference [210,211]. Recent evidence suggests that self/non-self-recognition is the result of preferential spacer uptake from stalled replication forks at double strand breaks [212,213]. At double strand breaks, the repair complex RecBCD unwinds and degrades DNA until it reaches a Chi site, an asymmetric octamer motif [214]. RecBCD activity produces ssDNA fragments which are proposed to provide substrates for Cas1-Cas2 complexes [212,214]. The scarcity of Chi sites on mobile genetic elements relative to host chromosomes leads to a natural selectivity for foreign DNA due to the far greater quantity of ssDNA fragments produced by RecBCD. Interestingly, although CRISPR-Cas systems generally inhibit horizontal gene transfer events, transduction of bacterial genes by phages is enhanced [215]. In this way, some positive benefits of horizontal gene transfer are maintained, such as acquisition of antibiotic resistance and new virulence phenotypes. Conversely, phages are able to transduce spacers, and even entire CRISPR-Cas systems, which target other phages, allowing phage tolerance and reduced competition from other phages [215].

Primed adaptation is directly driven by complete or partial complementarity of pre-existing spacers to invading DNA [207,210]. This presents respectively both a positive feedback loop, reinforcing defence against persistent threats, and a mechanism for countering phage evasion by mutations withing the spacer recognition sequence. Indeed, mutations or deletions in the spacer and PAM sequences present the simplest mechanism for phage escape of CRISPR-Cas systems. Through the rapid acquisition of multiple new spacers, bacteria can maintain resistance. Single point mutations proximal to, or within the PAM sequence can be sufficient to allow phage escape, with more distal sites requiring higher degrees of mutation [216,217]. When considering the efficacy of CRISPR-Cas protection, the location of the spacer sequence in the phage genome is significant. CRISPR spacers targeting early genes are more effective than

those targeting late genes as early genes will have already begun to degrade host chromosome and inactivate CRISPR-Cas, and other phage defences, before second stage gene transfer [218].

In most type I and type III systems of class I, transcribed CRISPR arrays are processed by Cas6 [219,220]. Transcribed arrays (pre-crRNA) often display palindromic sequences allowing the formation of palindromic repeats. Cas6 recognises and binds stem loop structures and cleaves directly downstream of the hairpin to yield mature crRNA, composed of a single spacer sequence and repeat sequence [206,220]. In most cases Cas6 remains attached to the 3' repeat of the crRNA and is involved in the following CRISPR cascade. In type I, the CRISPR complex consists of Cas5, Cas6, six Cas7 subunits, Cas8 and two Cas11 subunits [221]. Cas8 facilitates unwinding of DNA for spacer sequence annealing and Cas11 binds the non-target strand, producing conformational changes in the complex which allow access of the Cas3 nuclease [222]. DNA nicking by Cas3 results in a further conformational change in Cas3, activating ATP-dependant helicase activity and allowing Cas3 translocation along the target DNA, where it degrades in a 3' to 5' direction [223]. Type III cascade complexes display high structural and compositional similarity to type I. The system differs in targeting RNA and DNA, with DNA degradation is dependent on transcription of target DNA into mRNA [224]. The complex assembles along the crRNA scaffold and binds to target mRNA transcripts. By targeting mRNA, no unwinding mechanism is required for spacer annealing to target sequences. Following binding, Cas10 cleaves both strands of DNA and Cas7 degrades mRNA [225]. Cas10 cleavage has also been demonstrated to promote host RNase activity through the production of cyclic oligoadenylate signalling molecules, upregulating activity of Csm6, an RNase implicit in degradation of invading RNA transcripts [226–228]. Thus, the system provides an additional layer of protection against foreign DNA, preventing the system from being overwhelmed or for when activity is lower due to target sequence mismatches. Indeed, the cyclic adenylate synthetase has been suggested as functioning in transcriptional regulation, possibly upregulating expression from other CRISPR loci or even other defence systems [228,229]. This communication highlights phage defence as a concerted effort by the entire arsenal of defence systems encoded by a bacterium rather than individual systems functioning independently. The function of bacterial quorum sensing in bacteriophage regulating phage defence at a population level is well documented [230], with upregulation of individual CRISPR-Cas systems demonstrated [231,232].

Class II systems are distinct in having single molecule effectors which also generate mature crRNA from transcribed arrays (Figure 1.7) [206]. In type II and type V-B systems, a trans-activating RNA molecule (tracrRNA) with similarity to the repeat regions is required for crRNA maturation [233]. Tracr and crRNA form a duplex, allowing binding of Cas9, which processes crRNA in association with host RNases to produce mature crRNA [233]. Cas9 initially probes for target DNA with similarity to the PAM sequence of the crRNA which causes a conformational change producing unwinding of adjacent DNA allowing subsequent binding of the remainder of the crRNA molecule should sequences match [234]. Further conformational changes result, leading to a DNA R-loop [235] and DNA cleavage by the dual nuclease domains [236].

Type V and VI systems do not require tracrRNA for crRNA maturation or for interference. Cas12 and Cas13 effector proteins respectively, recognise hairpin loops of crRNA and cleave to produce mature crRNA [237,238]. The interference stages of Cas12 and Cas13 differ. Similar to Cas9, Cas12 interrogates target sequences for PAM sequence complementarity and is then able to probe against adjacent sequence with the full crRNA molecule [237]. Unlike Cas9, Cas12 produces a staggered cut using a single nuclease domain resulting in a 5’ overhang [237,239], the mechanism of which is not yet known. In contrast to other class II types, Cas13 targets RNA rather than DNA and therefore protects against ssRNA phages [240]. crRNA binding to target ssRNA causes a conformational shift and activation of the HEPN catalytic domain, which cleaves target RNA and remains activated [241], allowing further indiscriminate ssRNA degradation. This feature has been shown to confer toxicity when Cas13 is highly expressed *in vitro* and must be tightly regulated [238,240].

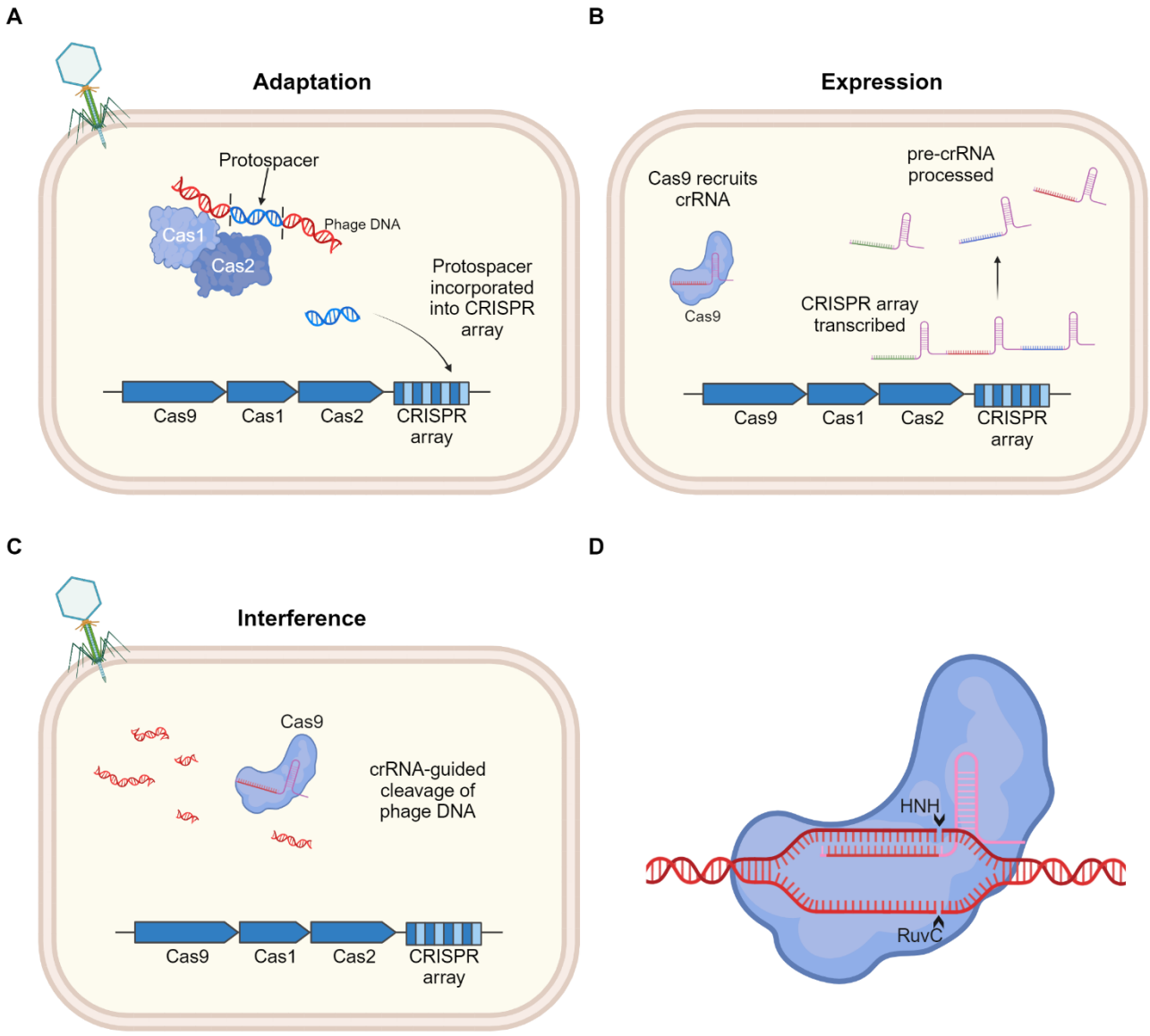

Figure 1.7 – CRISPR-Cas phage defence. *Depiction of the stages of phage defence mechanism of type II CRISPR-Cas systems. A; Fragments from invading phage genomes are processed and inserted into CRISPR arrays by Cas1 and Cas2. Fragments can originate from the activities of other phage defence systems, such as RM systems and the RecBCD complex. B; CRISPR-arrays are transcribed and pre-crRNA is processed into mature crRNA by RNaseIII in conjunction with Cas9, which then binds to crRNA. C; the Cas9-crRNA complex then actively searches for sequences with similarity to the crRNA sequence and produces targeted DNA cleavage. D; DNA restriction is guided by ennealing of crRNA molecules to unwound DNA, leading to cleavage of the DNA backbone by HNH and RuvC nuclease domains for top and bottom strands, respectively.*

1.3.6 Systematic discovery of new defence systems

The ever-expanding armoury of bioinformatic, genomic and proteomic tools available to researchers has allowed rapid expansion of our knowledge and understanding of the phage resistance landscape in recent decades. Resistance related genes are often clustered in loci termed 'defence islands', allowing systematic discovery of new and putative defence systems [134,136,150,242,243]. Initial attempts at "guilt-by-association" approaches to defence system discovery identified and then interrogated one system at a time and were successful in demonstrating new and varied defence mechanisms and systems [242]. More recently however, this process has been scaled up, with large genomic databases mined for defence islands and putative new defence systems by identifying genes and gene families that are enriched adjacent to known phage defence systems [136]. The first pass at this high-throughput methodology yielded nine new phage defence system families with diverse compositions and novel modalities [136]. Though extensive, this analysis is not exhaustive, and discovery of new defence systems allows these systems to be used for discovery of further new defence islands in subsequent rounds of analysis. Variations on the theme are also possible: a later study applied mining of large genomic datasets but this time with proteins clustered by sequence homology rather than using external domain annotations, producing 29 novel systems [244]. A following study using a similar methodology then yielded 21 new phage defence systems [134]. Though powerful, this method is still reliant on defence genes clustering on defence islands and is liable to miss any which do not follow this trend. An alternative high throughput method was developed to allow functional selection of genes and systems which provide phage defence in which ~40 kbp fragments of metagenomic *E. coli* DNA were subcloned onto fosmid vectors and individually screened for activity [245]. This methodology identified a further 21 new phage defence systems within *E. coli* strains alone which were not enriched in defence islands and therefore, were missed by previous bioinformatics approaches. However, although more comprehensive, this methodology cannot apply the same scale as bioinformatic approaches. Of particular interest, this study demonstrated that many phage defence systems are located in prophage and mobile genetic element regions [245]. Indeed, a study published earlier the same year had shown that prophages and their satellite regions were themselves enriched for phage defence genes, leading to the confirmation of phage defence from seven new systems [243]. Guilt-by-association based methods are continuing to unearth new phage defence systems [246], suggesting that the repertoire of defence systems in nature extends still further.

1.3.7 Newly discovered phage defence systems

The rapid discovery of phage defence systems has revealed a far more intricate and complex landscape of the phage bacteria arms race than had previously been envisioned in decades passed. With the glut of new defence systems identified, efforts to characterise these systems have quickly fallen behind. The PADLOC (for Prokaryotic Antiviral Defence LOCator) database contains over a hundred different defence systems which are further split into various types and subtypes, resulting in a database of over 240 systems at time of writing, many of which are uncharacterised [247]. Nevertheless, characterisation of phage defence systems is desirable as the study of phage-bacteria interactions have potentially valuable biotechnological applications (see 1.1.2). Several recently discovered systems with distinct mechanisms are discussed in this section, demonstrating the ever-increasing diversity of phage-bacteria interactions, though this accounts for only a fraction of discovered novel systems.

1.3.7.1 Argonaute

Discovery of new bacterial defence systems have revealed further similarities in mechanisms and functions to eukaryotic immune defences. Prokaryotic argonoaute (Ago) proteins have been found within defence islands [136,248] which suggested a role in host defence. In eukaryotes, Ago proteins are involved in RNA interference pathways responsible for gene regulation and antiviral activities [249]. In the prokaryotic system, Ago proteins facilitate DNA-guided DNA interference as a means of protection against foreign mobile genetic elements [250]. In most cases, Ago proteins associate with guide DNA molecules which direct degradation of complementary sequences [251] Though the mechanism of self/non-self recognition is not known, it is suggested that products of non-specific DNA cleavage can be loaded onto Ago proteins to significantly increase the rate of degradation [251]. The prokaryotic Ago system then may function through amplification of existing restriction-based defence system responses against currently active threats. This hypothesis is supported by the similar phylogenetic patterns of the Ago and CRISPR-Cas system, sometimes even encoded within the same operon [252]. This pattern seems logical as both systems rely on short DNA guide molecules for activity. As with many defence systems, further investigations reveal additional effector mechanisms. For example, a truncated Ago protein is associated with an effector which provides an Abi mechanism by depolarising the host cell membrane [253].

1.3.7.2 DISARM

The systematic discovery of defence systems through defence island association led to discovery of the DISARM system (Defence Island System Associated with Modification) [254]. Class I and II of the DISARM system encode adenine and cytosine methyltransferases respectively (DrmI and DrmII), alongside proteins encoding helicase (DrmA) and phospholipase D (DrmC) domains. Each class then encodes two additional genes with different putative functions suggesting different mechanisms of action [254]. Similar to RM systems, DISARM achieves self/non-self recognition through host methylation. Unlike RM systems, the DISARM system likely utilises an additional mechanism of phage DNA recognition as methylation of phage DNA and infection with phage lacking the recognition motif was not sufficient to provide defence in either case. Additionally, the putative nuclease DrmC was shown to be redundant in conferring defence against some phages and but others [254]. The close association of DISARM systems with RM systems therefore infers some functional overlap between the two and potential synergistic relationship, adding further complexity to the phage defence arsenal.

1.3.7.3 CBASS

CBASS (for Cyclic oligonucleotide-Based Antiphage Signalling System) phage defence systems are a widespread and diverse group which utilise cyclic nucleotide second messenger molecules and trigger Abi on sensing phage infection [229,255]. Thousands of CBASS systems have been identified in prokaryotes with diverse signalling and effector mechanisms [255]. The CBASS system is related to the cGAS-STING innate immune signalling system in animals and CD-NTase components show some structural homology to cGAS-like receptors [255,256]. All CBASS systems contain both a CD-NTase and CD-NTase-associated-protein (Cap) [229,255]. The CD-NTase senses phage infection and synthesises a specialised cyclic nucleotide [229,257]. Cap proteins encode receptor domains which specifically sense the signalling molecule and initiate an Abi response. The mechanisms of Cap function are diverse, and some require additional protein factors which are thought to regulate function [229].

CD-NTases encode domains for sensing phage infection and for synthesising cyclic nucleotides. The mechanisms by which CD-NTases sense phage infection are currently unknown, though as CBASS systems

respond to diverse phages, it is thought that triggers must be conserved phage features rather than phage specific factors. CD-NTases are highly diverse at the protein sequence level, reflecting the diversity of signalling molecules produced by CBASS systems [229]. Seven distinct forms have so far been identified as functioning in CBASS-mediated phage defence as both di and tri-nucleotides [256–261], though it has been calculated that recognition by Cap effectors of at least 180 different variants are possible [256]. The use of cyclic nucleotides in phage defence system signalling is not unique to CBASS; type III CRISPR-Cas [226], Pycsar [262] and Thoeris [136] systems utilise cyclic nucleotides to trigger downstream effectors. It is likely that systems encoding cyclic nucleotide second messengers have evolved and adapted from general nucleotide second messenger systems which regulate many functions and pathways in bacterial cells [263–266]. The diverse and complex nature of cyclic nucleotide species in CBASS ensures that potentially lethal Abi effector mechanisms are insulated from other common nucleotide messenger signals.

Cap proteins encode a receptor domain which recognises specific cyclic nucleotide molecules and an effector domain which triggers Abi [229]. Several mechanisms of messenger molecule binding and recognition have been described but the most common is through a SAVED domain [256,267]. Elucidation of the structure of Cap4 demonstrated that SAVED domains are formed by a fusion of CARF domains, usually involved in signalling in association with CRISPR type III systems [256] but found to play a conserved regulatory role in phage defence systems [268]. This domain organisation which permits recognition of an incredibly diverse range of signalling nucleotides by allowing binding of asymmetric nucleotide molecules [256]. Many identified Cap proteins from CBASS systems encode no recognisable nucleotide binding domains, suggesting that there are unknown nucleotide binding modalities yet to be discovered [229,256]. The Abi mechanisms produced by Cap proteins are diverse, including disruption of the cell membrane [260], various nuclease functionalities [269], degradation of NAD^+ [259] and protease activity [256]. Activation of Cap proteins often results in multimerisation and the formation of larger effector complexes [229,269]. The modularity of CBASS systems allow interchange of domains within both CD-NTase and Cap proteins, potentially providing new combinations of sensing phage infection, signalling molecule synthesis and detection, and Abi mechanisms as required to overcome new or evolving phage threats.

1.3.7.4 Chemical mechanisms of community-level protection

Most interest in phage defence systems has centred around inhibition of phage infection at a DNA/protein level. It has long been known however, that bacteria can produce small molecule inhibitors of phage infection analogous to the production of antibiotics [270,271]. Design of successful phage therapeutics must take into account all mechanisms by which bacteria can resist phages. The increased interest in phage defence systems in recent decades however has not been paralleled in small molecule phage defence mechanisms. Many molecules with antibiotic properties also inhibit phage infection [272]. Intriguingly, secreted antimicrobials can even provide protection from phage infection to other bacteria in the local environment, presenting a possible mechanism of community-level protection [272,273]. Indeed, the concept of phage defence as a microbial community resource has previously been proposed [274]. In the context of phage therapeutics, functional relationships between phages and antibiotics used in clinical application would have been investigated. This would not take into account small molecules specific only to phage inhibition, however. Most early studies on small molecule inhibitors of phage infection focussed on *Actinomycetes* and *Streptomyces* species [272], the latter of which are the source of around two thirds of clinically utilised antibiotics [275]. A more recent study investigated small molecules produced by *Streptomyces* species for their ability to inhibit phage infection and found several which provided protection from phage infection through DNA intercalation [276]. All but one of the molecules discovered contained an anthracycline core, a class of molecule often used as an anti-tumour treatment [276]. As such, discovery of novel small molecule inhibitors of phage infection has potential therapeutic applications in cancer treatments. It would be interesting to determine whether small molecule phage inhibitors cooccur with phage defence islands, potentially allowing systematic discovery in the same manner as conventional phage defence systems (see 1.3.8).

1.3.8 Cooccurrence of phage defence systems

It is important to consider that bacterial genomes can encode multiple phage defence systems which can act in parallel to provide defence against invading phages (Figure 1.8). *E. coli* strains for example encode an average of six defence systems, both chromosomally and on mobile genetic elements [277]. These systems can interact both directly and indirectly, sometimes providing synergistic protection [277]. The

cooccurrence of RM systems and CRISPR-Cas systems has been documented, providing innate and adaptive immune functions [278]. Meanwhile, the type IV restriction system, BrxU, was shown to provide complementary defence to a type I BREX system in *Escherichia fergusonii*, in which BREX protects against methylated phage DNA and BrxU targets non-methylated phage DNA [183]. Phage defence systems have also been shown to utilise similar signalling strategies [226,227,255,262] and can be found under the control of universal defence system promoters [268,279–281]. The rapid increase in the rate of defence system discovery suggests that such interactions are likely to become more pertinent over the coming years. New synergies between cooccurring systems have recently been demonstrated [277]. Understanding these synergies will require further study on the molecular mechanisms of these systems. New bioinformatic tools for the identification of phage defence systems in bacterial genomes will expedite the study of interactions between systems [135]. The cooccurrence of phage defence systems was not found to correlate with compatibility of molecular mechanisms of said systems and were instead related to environmental factors and relevant phage threats [277]. Phage defence systems and defence islands are often encoded on mobile genetic elements [282], plasmids [183], prophages [243] and hypervariable regions [283], all of which are drivers of defence system turnover [284] leading to diverse combinations of systems across the microbiome.

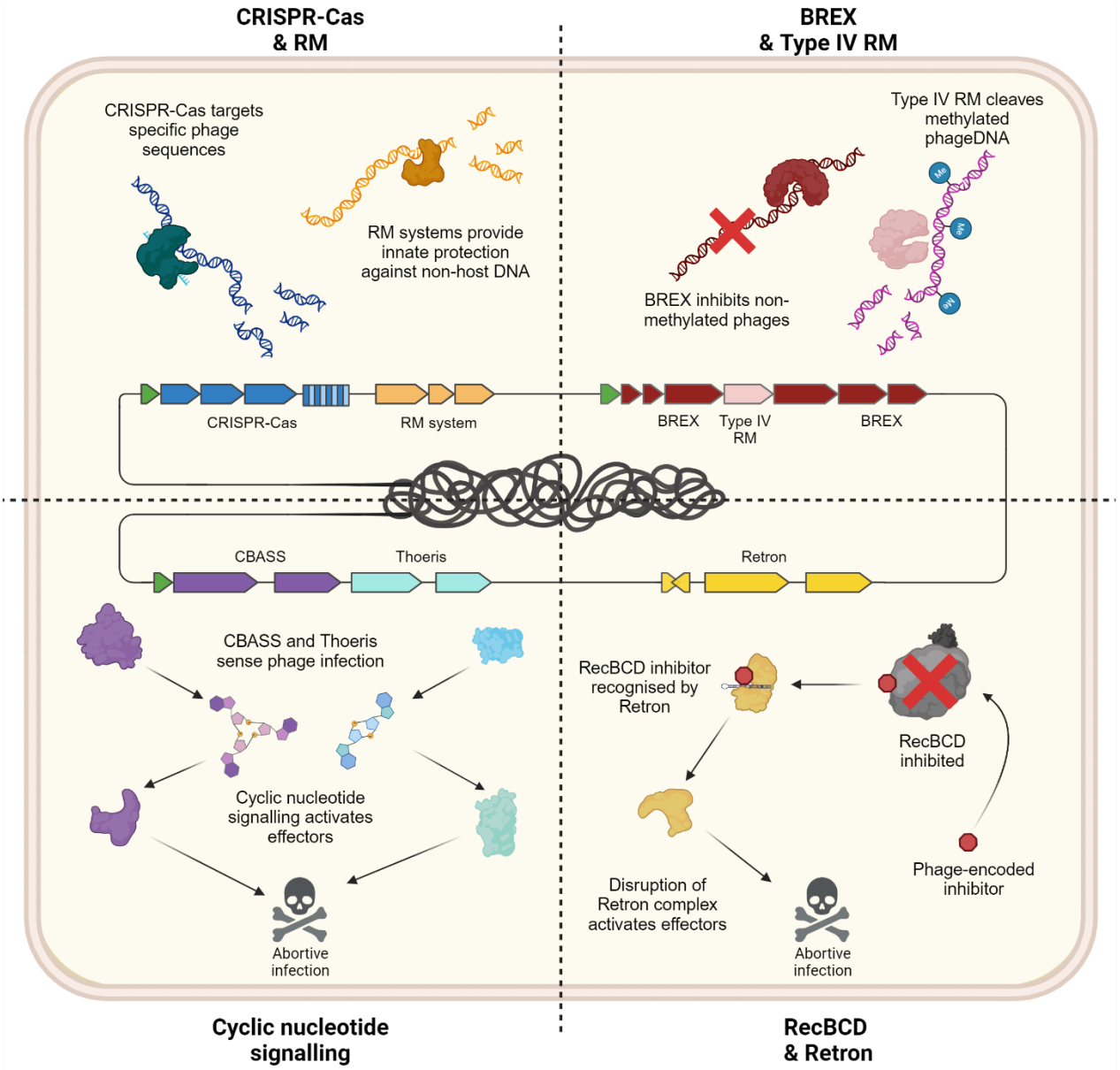

Figure 1.8 – Complementation and interaction of phage defence systems. *Top left; CRISPR-Cas and RM systems provide innate and specific immune responses, respectively. Top right; cooccurrence of BREX and type IV RMs inhibits replication of non-methylated and methylated DNA, respectively. Bottom left; defence systems encode common mechanisms. For example; CBASS, Thoeris and CRISPR-Cas type III systems produce cyclic nucleotide signalling molecules on sensing phage invasion which then activate downstream effectors. Bottom right; bacteria can encode contingencies for when phages inhibit one layer of defence, such as the activation of RcaT of Retron-Eco6 on inhibition of the RecBCD complex. Additionally, defence systems can be under the control of shared promoter mechanisms (depicted in green), allowing concerted activation of multiple defence systems upon sensing phage infection.*

1.3.9 Phage escape from defence systems

As bacteria have evolved new phage defence system modalities, phages have evolved to counteract or circumvent these defences. This relationship is the driving force behind the incredible complexity and multiplicity of phage defence systems across the microbiome. Though many studies omit defence system escape, others have utilised phage escape to provide valuable insight. The intense and interminable nature of the phage-bacteria arms race ensures that for any mechanisms that bacteria develop to inhibit phage infection, phages have or will develop mechanisms for evading or inhibiting them. Methods for circumventing bacterial alterations in receptor proteins, inhibition of adsorption and injection and the roles of prophages in preventing superinfection are discussed in 1.3.1. The remainder of this section will discuss phage encoded inhibition of phage defence systems following injection.

One of the earliest discovered and best understood mechanisms of phage defence system inhibition was the RM system inhibitor, Ocr [285]. Ocr is a DNA mimic protein, forming a dimer which mirrors around 20 – 24 bp of slightly bent B-form DNA, which competitively binds to type I RM components preventing their activity in phage DNA restriction [286–288]. The binding affinity of Ocr to type I RM systems is around 50x higher than that of DNA [287,289,290] and Ocr is highly expressed immediately following DNA injection [291,292], ensuring efficient inhibition of RM function. The basis of the high binding affinity of Ocr for type I RM systems is two-fold. First, Ocr has a highly negatively charged surface with 34 exposed glutamic acid and aspartic acid residues which mirror phosphate positions in the DNA backbone [288]. Second, binding of DNA to type I RM systems requires energetically unfavourable bending of target DNA [186,290]; the form of Ocr however is already in this bent conformation, allowing energetically preferential binding [287,288,290]. In addition to inhibition of host RM systems, Orc has also been shown to provide an additional role in inhibition of host transcription by binding to host RNA polymerase and preventing the recruitment of σ-factors through direct binding to RNA polymerase [291]. Universal inhibition of all DNA binding proteins in the host cell would be detrimental to phage replication however and Ocr proteins do show some specificity. Ocr does not inhibit type II or type III RM proteins, for example, likely highlighting key differences in the function of these systems [293]. Conversely, Ocr does inhibit type I BREX systems, implying similar DNA binding mechanisms [293,294].

A majority of discovered anti-CRISPR phage proteins (Acr) act at the interference stage, targeting the Cas components [295]. AcrIF2 of *Pseudomonas aeruginosa* for example is a DNA mimic, forming a pseudo-helical conformation which binds to Cas5-Cas8 components of the interference complex, preventing DNA binding by sterically inhibiting DNA access to the complex [296]. AcrIIC3 interacts with the HNH nuclease domain and REC lobe of Cas9, reducing DNA affinity, preventing cleavage and causing Cas9 dimerisation. Cas12a inhibition has been demonstrated by AcrVA1 through inducing cleavage of bound target recognition sequence RNA, leading to irreversible inactivation [297]. It is proposed that the diverse mechanisms of action of Acr proteins have in part driven the diversity of CRISPR-Cas compositions and mechanisms over evolutionary timescales. The use of Acr proteins has potential biotechnological applications. Acr proteins have been used as on/off switches *in vitro* to reduce off target effects in gene editing studies [298] and for the precise control of genetic circuits [299]. Control of CRISPR-Cas genome editing studies using Acr proteins and controllable inputs, such as light [300], may present vital routes to minimising off target effects, ultimately leading to the routine clinical application of gene editing treatments.

Phages have counter-evolved numerous mechanisms to escape resistance by Abi systems. This can include phage encoded homologues of antitoxin proteins. Most notably, the phage T4 encodes an ADP-ribosyltransferase Alt which impairs the function of the toxin components of the MazF TA system in *E. coli* [301]. Phages can also encode protease inhibitors which inhibit degradation of the antitoxin components of Type II and IV systems [302]. Finally, TA systems are commonly utilised by prophages to promote maintenance and are particularly prevalent in the persistence of cryptic prophages [116,303]. Prophages can even bootstrap host TA systems with their own TA modules, allowing additional layers of host control [149].

Understanding the mechanisms of phage escape can reveal key determinants of defence system function and can be particularly valuable in determining starting points for research on new defence system mechanisms [145] as well as identifying important trends in phage escape modalities [137]. An intensive study on phage escape determinants of 54 defence systems for example found that TA system escape mutants predominantly contained mutations in early phage genes; for ShosTA, ssDNA binding (SSB) proteins, an inhibitor of host RNA polymerase and a DNA primase/helicase; for Retron-Eco8, SSB family proteins; for Retron-Se72, a RecBCD inhibitor protein, a ssDNA annealing protein and an exonuclease; for

Retron-Ec67, the DNA transfer protein, A1 [150]. Such mutations are likely aimed at strategies of avoiding defence system activation rather than inhibition. Most escape mutants (83%) encoded only a single mutation and most mutations incurred a fitness cost. The results from this study demonstrated that defence system activation was usually the result of recognition of phage replication components, complexes and intermediates, recognition of phage structural proteins and host takeover mechanisms and defence system inhibition [150]. Activation of defence systems, predominantly Abi systems, upon inhibition of other defence systems supports the model of Abi systems operating as a second line of defence and acting as a contingency measure for when conventional defence systems are overcome. Conversely, the activation of many TA systems in the early stages on phage infection through recognition of early phage infection structures and components supports the notion that not all TA systems cause Abi [137,304,305].

Documentation of phage escape mutants in recent studies is sporadic but can be similarly insightful. Escape mutants from ToxIN have previously been described and encode a single mutation in a non-essential gene, preventing system activation [306]. The ToxIN system is activated through phage-mediated transcriptional shutoff [142,307]. A recent study demonstrated that phage T7 can circumvent ToxIN activity by only partially shutting off host transcription [142]. As with escape mutants produced in the study of phage determinants of defence system activation [150], overcoming ToxIN activity incurred a fitness cost but allowed infection of a bacteria which would otherwise have been resistant if transcription was fully inhibited [142]. The interplay between ToxIN and T7 implies that phage-bacteria interactions should perhaps be viewed as a gradient of phage infectivity rather than the black-and-white view of bacteria either being resistant or susceptible.

1.4 BREX phage defence systems

1.4.1 Phage growth limitation system

BREX (for BacteRiophage EXclusion) is a novel phage defence system present in approximately 10% of microbial genomes conferring resistance against a diverse range of phages [308]. The precursor to the

discovery of BREX was the Pgl (Phage Growth Limitation) system in *Streptomyces coelicolor*, which conferred protection against ϕC31 [309]. In the Pgl system, phages were able to replicate in the first round of infection, but infection was inhibited in subsequent cycles [309–312]. It was proposed that phage progeny from the first round of infection were modified, in contrast to traditional self from non-self recognition in phage defence systems [309]. The Pgl system encoded by *Streptomyces coelicolor* consists of four genes responsible for defence, PglW, PglY, PglZ and PglX [312]. Within the context of *Streptomyces coelicolor*, the Pgl system provides TA functionality, whereby the activity of the methyltransferase is toxic to the host in the absence of PglZ activity [313], though this activity is not thought to be the source of phage defence and the system does not function via an Abi mechanism.

1.4.2 BREX discovery, subtypes and system compositions

More recent bioinformatic analysis revealed that in more than half of the Pgl^+ genomes, PglZ is located in a 6 gene cluster [308]. A wider family of systems was subsequently discovered, renamed BREX and divided into 6 subtypes with varying compositions and organisations, of which the classical Pgl system represents type II [308]. All other BREX systems tested functionally so far represent type I systems and have demonstrated an alternative mechanism to the type II Pgl systems [183,281,308,314,315]; phages are restricted on the first round of infection and the host genome is modified rather than the target phage. It was postulated that the Pgl system may represent a specific BREX type which has adapted to an unmethylated host, in which introduction of methylated motifs is toxic. It is plausible that methylation of the phage genome leads to either restriction by type IV RMs or inhibition of phage replication by other BREX factors. The mechanism of phage defence is not currently understood, though it appears to operate at an early stage of phage infection and does not involve cleavage of phage DNA [308,314], making it distinct from RM system defence.

A majority of BREX systems (around 55%) are of type I [308]. Type I BREX systems have been the focus of most studies so far and consist of 6 genes (figure 1.2): BrxA has been shown to bind DNA [316]; BrxB has no known function (DUF1788 domain); BrxC is a putative ATPase; PglX is a SAM-dependant methyltransferase [314]; PglZ is a phosphodiesterase, as shown by biochemical investigation of the *pglZ* domain in PorX [317]; and BrxL is a DNA binding AAA^+ ATPase which forms multimeric complexes which

translocate along DNA [318]. The known and putative functions of BREX genes from type I systems will be probed and detailed much further in Chapter 3. Self from non-self recognition is achieved via methylation of a specific 6 bp DNA motif, producing N6mA modifications at the fifth base position [314]. Modification requires the presence of BrxB, BrxC and PglZ in conjunction with PglX, and methylation of phage DNA at these sites abolishes resistance [314]. The *Acinetobacter* BREX system also requires BrxA for methylation [281]. No methylation activity has been demonstrated *in vitro* from PglX in type I BREX systems [314], though the type II PglX component appeared to produce low levels of methylation activity in radio labelled SAM experiments [313]. Thus, type I BREX systems features epigenetic modification similar to RM systems however, the mechanism of phage inhibition remains unknown.

The differences in *in vitro* methylation activity from type II BREX PglX may relate to the inherent differences in phage defence modality. It has not yet been shown whether type III, IV, V or VI BREX systems provide phage defence via the type I of type II modalities or function by other distinct mechanisms. BREX subtypes encode an array of proteins, several of which are unique to a single subtype. Type II systems are the only subtype to encode a putative kinase (PglW), found to be essential for phage defence [312]. Although originally identified as containing four genes (*pglW*, *pglY*, *pglX*, and *pglZ*), 89% of Type II systems also encode putative ATP binding (BrxD) and helicase (BrxHI) components but all lack *brxL*, *brxA* or *brxB* genes of type I systems [308]. Interestingly, the helicase (either BrxHI or BrxHII) and BrxL components never cooccur within subtypes, suggesting analogous or redundant roles in BREX activity, or differing defence mechanisms. It is interesting that the type II system from *Streptomyces coelicolor* was able to function without either of these components, both in the native host and when transferred to *Streptomyces lividans* [312]. Type IV systems are the only subtype not to encode a methyltransferase component and are unique in containing a putative phosphoadenosine phosphosulfate (PAPS) reductase, though how these functions may be interchangeable is unclear [308]. Common themes in all subtypes appear to be an ATPase (either BrxC or PglY), a phosphodiesterase (PglZ), a methyltransferase (PglX or PglXI; swapped for the putative PAPS reductase, BrxP, in type IV) and either a helicase (BrxHI or BrxHII) or the AAA^+ ATPase, BrxL [308].

1.4.3 Regulation of BREX systems

Due to the increased metabolic load and/or inherent toxicity of phage defence systems, tight regulation is essential. Phage defence systems often encode conserved regulatory elements, such as the CARF (CRISPR Associated Rosman Fold) genes found adjacent to many CRISPR systems [268] and the WYL domain proteins found to regulate diverse defence systems [268,279–281]. The WYL domain protein BrxR has been shown to regulate BREX systems in both the pEFER plasmid encoded system from *E. fergusonii* [280] and the *Acinetobacter* species NEB394 [281]. BrxR forms a homodimer which binds to inverted repeats containing the -35 box motif, repressing transcription of the BREX operon [280,281]. BrxR is not required for phage defence and phage replication is inhibited when BrxR is mutated to abolish DNA binding activity, but deletion of BrxR caused a reduced level of phage defence and incurred some toxicity, implying that BrxR has a dual role in both transcriptional regulation and in regulation of BREX activity [281]. BrxR proteins contain a conserved helix-turn-helix (HTH)-WYL-WCX domain organisation, similar to that seen in PafBC which regulates DNA damage response pathways [280,281]. The type I BREX systems from *E. coli* [314] and *Salmonella* ER3625 [315] do not include BrxR components however, suggesting alternative methods of regulation. The control of phage defence systems through universal operators offers the substantial advantage of inducing a concerted response to phage invasion. As discussed in 1.3.8, bacteria can encode multiple defence systems with potentially synergistic or complementary activities. Thus, simultaneously deploying an expanded arsenal of systems maximises inhibition of phage infections.

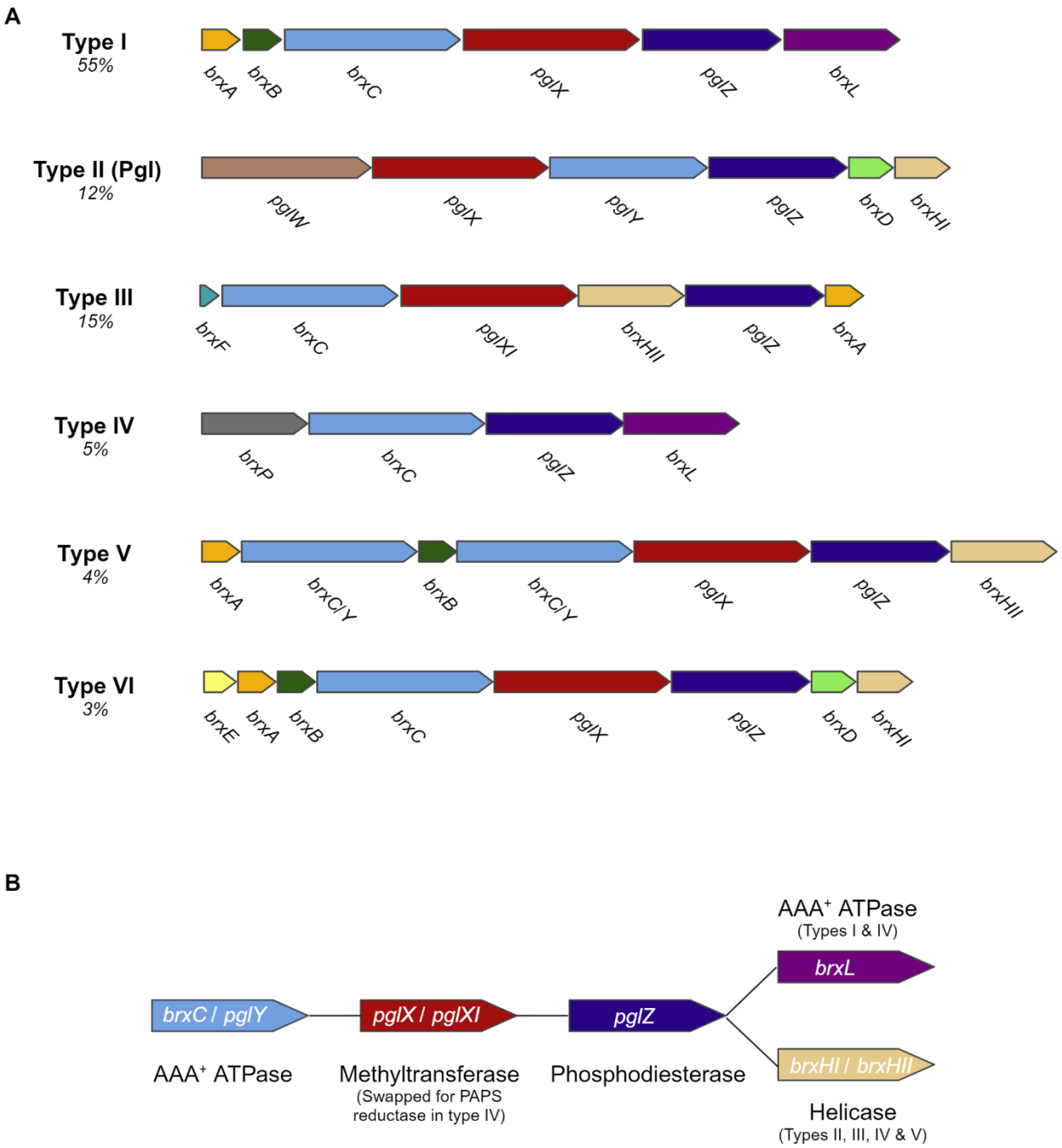

Figure 1.9 – BREX subtypes. *A; Common gene architectures of the six BREX system subtypes. Percentages indicate the proportion of BREX systems which are of the respective subtype. Type II was originally identified as the Phage Growth Limitation system (Pgl). B; BREX component functions common to all subtypes (unless stated otherwise).* brxL *and* brxHI/HII *never cooccur, but one of each is present in each subtype.*

1.5 Aims of this study

This study presents a structural and functional characterisation of the BREX system and its components in an attempt to add valuable insight into the mechanisms of phage defence. As previous studies by others in this lab have had difficulties expressing the PglX component of *E. fergusonii,* the BREX system from *Salmonella* Typhimurium strain D23580 (hereafter referred to as *Salmonella* D23580) from sub-Saharan Africa was selected for study. *Salmonella* D23580 is highly invasive and has been implicated in a rise in epidemic non-typhoidal *Salmonella* disease [319]. As such, findings have potential clinical relevance in the design of phage therapeutics to combat this multi drug resistant strain. Further, as many useful biotechnological tools have been derived from phage-resistance systems, investigating the mechanism of BREX also has the potential to provide useful reagents for research. In line with previous studies in our lab [183,280,316,320] and with collaborators [281,314], the *Salmonella* D23580 system represents a type I BREX system, allowing direct comparison. Beginning with initial bioinformatic analysis, this study first aimed to provide insight into the context of the BREX system in *Salmonella* D23580. Following, this study focussed on three objectives:

Functional characterisation – Through a combination of microbiological, molecular biology, and biochemistry techniques, this study aimed to characterise the *Salmonella* D23580 BREX system functionally, utilising the in-house Durham Phage Collection. It was hoped that by comparison to similar studies on type I BREX systems, new insight could be gained into common themes and differences which would further our understanding of BREX defence.

Structural Characterisation – Next, this study also aimed to structurally characterise two core components of BREX systems: the methyltransferase, PglX and the phosphodiesterase, PglZ, by application of X-ray crystallography techniques previously utilised in the lab. Elucidation of the molecular structures of these essential defence components would provide new detail on their molecular function as part of BREX activity.

Manipulation of BREX specificity – Finally, this study aimed to mutate the methyltransferase component, PglX, in an attempt to reprogramme the TRD to recognise an alternative DNA motif and potentially allow defence against a different subset of phages. The benefits of this are two-fold: documentation of rational modification of the methyltransferase component would add to a growing literature on rational modification of RM recognition motifs (see section 1.3.3.6) and potentially be a useful resource to guide similar studies; and, should modification of PglX to recognition of an alternative motif both shift host methylation sites and provide defence against a new subset of phages, PglX would be confirmed as the sole specificity factor of both host methylation and phage inhibition BREX functionalities.

Chapter 2 – Materials and Methods

2.1 Media and reagents

Details on media, antibiotics and supplements, and buffer solutions used throughout this study can be found in Tables 2.1, 2.2, and 2.3, respectively. Media, buffers and reagents were sterilised by autoclaving at 121 °C for 30 minutes, or filter sterilised using 0.22 µm syringe (Fisher Scientific) or vacuum filters (Sartorius), as appropriate.

2.2 Bacterial strains and growth conditions

Bacterial strains used in this study are detailed in Table 2.4. All *E. coli* strains were grown at 37 °C unless stated otherwise, with shaking at 180 rpm for liquid cultures. Growth was monitored by optical density at 600 nm (OD_{600}) measurement using a Biochrom WPA CO8000 spectrophotometer. Where necessary, growth media were supplemented with the appropriate antibiotics or protein expression inducing agents, as described later. Overnight cultures were grown for between 16 – 20 hours (h). For long term strain storage, 750 µl of overnight culture was added to 750 µl of 50% v/v glycerol and snap frozen in liquid nitrogen for storage at -80 °C.

Table 2.1 – Growth media used in this study.

Medium	**Ingredients per litre of MilliQ dH_2O**
Luria broth	10 g Casein Digest Peptone
	5 g Yeast Extract
	10 g NaCl
Luria broth agar	10 g Casein Digest Peptone
	5 g Yeast Extract
	10 g NaCl
	10 g agar for 1% w/v
	3.5 g agar for 0.35% w/v
	16 g Casein Digest Peptone
	10 g Yeast Extract

2x Yeast Tryptone (YT) broth	16 g Tryptone 10 g yeast extract 5 g NaCl

Table 2.2 – Antibiotics and supplements used in this study.

Chemical	**Stock concentration (mg/ml)** in MilliQ dH_2O, unless stated	**Working concentration (mg/ml)**
Antibiotics		
Ampicillin	100 mg/ml	100 µg/ml
Chloramphenicol	25 mg/ml, in ethanol	25 µg/ml, in ethanol
Kanomycin	50 mg/ml	50 µg/ml
Spectinomycin	50 mg/ml	50 µg/ml
Supplements		
L-arabinose	20% w/v	0.2% w/v
Isopropyl β – D – thiogalactopyranoside (IPTG)	1 M	0.5 mM

Table 2.3 – Buffers and solutions used in this study.

Solution	**Components**
All stock solutions were made with MilliQ dH_2O.	
DNA work	
50x TAE buffer (1 L)	242 g Tris base 57.1 mL Glacial acetic acid (17.4 M) 100 mL EDTA pH 8.0 (0.5 M)
Agarose gel	0.5 - 1.5 % w/v agarose in 1x TAE buffer 500 ng/ml ethidium bromide
Solution A	9.9 mM $MnCl_2$

(For the creation of chemically competent *E. coli cells)*	49.5 mM $CaCl_2$ 9.9 mM MES
Solution B	9.9 mM $MnCl_2$
(For the creation of chemically competent *E. coli cells)*	49.5 mM $CaCl_2$ 9.9 mM MES 15% v/v glycerol
Elution buffer	10 mM Tris-HCl 0.1 mM EDTA pH 8.5

Bacteriophage work

Phage buffer	10 mM Tris pH 7.4 10 mM $MgSO_4$ 0.01% w/v sterile gelatin

Protein purification

A500 (Lysis buffer)	20 mM Tris pH 7.9 500 mM NaCl 5 mM - 150 mM imidazole, as required for purification 10% v/v glycerol
A100 (Low salt buffer)	20 mM Tris pH 7.9 100 mM NaCl 5 mM imidazole 10% v/v glycerol
B500 (Nickel elution buffer)	20 mM Tris pH 7.9 500 mM NaCl 250 mM imidazole 10% v/v glycerol

B100 (Nickel elution, low salt buffer)	20 mM Tris pH 7.9 100 mM NaCl 250 mM imidazole 10% v/v glycerol
C1000 (High salt buffer)	20 mM Tris pH 7.9 1 M NaCl 10% v/v glycerol
Strep elution buffer	20 mM Tris pH 7.9 500 mM NaCl 2.5 mM D-desthiobiotin 10% v/v glycerol
SEC (Size exclusion chromatography buffer)	50 mM Tris pH 7.9 500 mM KCl 10% v/v glycerol
Storage buffer (For -80 storage)	50 mM Tris pH 7.9 500 mM KCl 70% v/v glycerol
A-SEC (analytical SEC buffer)	50 mM Tris pH 7.9 150 mM NaCl

SDS-PAGE

10x running buffer	30.2 g Tris base 141 g glycine 10 g SDS pH 8.3 with concentrated HCl
3x protein loading dye	25 mM Tris pH 6.8 0.6 mg/ml bromophenol blue 2.5% w/v SDS 30% v/v glycerol

	125 mM dithiothreitol (DTT)
8% separating gel (for 24 ml)	12.7 ml MilliQ dH_2O
	4.8 ml 40% v/v acrylamide (Severn Biotech)
	6 ml 1.5 M Tris pH 8.8
	240 µl 10% SDS
	240 µl 10% ammonium persulphate
	24 µl TEMED
4% stacking gel (for 15 ml)	9.4 ml MilliQ dH_2O
	1.5 ml 40% v/v acrylamide (Severn Biotech)
	1.25 ml 1.5 M Tris pH 6.8
	150 µl 10% SDS
	150 µl 10% ammonium persulphate
	12 µl TEMED
Protein crystallisation	
Crystal buffer	20 mM Tris pH 7.9
	150 mM NaCl
	2.5 mM dithiothreitol (DTT)
2x Cryo buffer (For crystal storage)	25 mM Tris pH 7.9
	187.5 mM NaCl
	3.125 mM dithiothreitol (DTT)
	80% v/v glycerol
Methyltransferase assay	
4x reaction buffer	80 mM Tris pH 8.8
	200 mM NaCl
	4 mM EDTA
	12 mM $MgCl_2$
	4 mM dithiothreitol (DTT)

Table 2.4 – Bacterial strains used in this study.

Strain	Genotype	Source
Escherichia coli		
DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi1 gyrA96 relA1	Invitrogen
ER2796	K-12 F λ- fhuA2 Δ(lacZ)r1 glnV44 mcr-62 trp-31 dcm-6-zed-501::Tn10hisG1 argG6 rpsL104 dam-16::Kan xyl-7 mtlA2 metB1 (mcrB-hsd-mrr) 144::IS10	NEB
BL21 (DE3)	B; F^- *ompT* gal dcm lon $hsdS_B(r_B – m_B –)$ λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) $[malB^+]_{K-12}(\lambda^S)$	Invitrogen
ER2566	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10-- TetS)2 [dcm] R(zgb-210::Tn10-- TetS) endA1 Δ(mcrC-mrr)114::IS10	NEB

2.3 Bioinformatics

Bioinformatic tools utilised in this study are shown in Table 2.4. Full details of command line packages used for the analysis of methylation and sequencing of bacterial and phage genomes, respectively, is available in section 4.3. Use of AlphaFold for protein modelling is explained in detail in 2.3.1.

2.3.1 AlphaFold

AlphaFold was installed on a Linux operating system (Ubuntu 20.04) as per the installation instructions available on Github (github.com/deepmind/alphafold). Protein modelling was carried out using a NVIDIA Quadro RTX4000 graphics card (8 GB RAM) with a further 64 GB of system RAM available for use. Protein databases used for modelling were downloaded in February 2022.

Table 2.5 – Bioinformatics tools used in this study.

Bioinformatic tool	Function	Source
Blastn	Nucleotide sequence similarity search	blast.ncbi.nlm.nih.gov/Blast.cgi
Blastp	Protein sequence similarity search	blast.ncbi.nlm.nih.gov/Blast.cgi
Defence finder	Identification of phage defence genes and systems	github.com/mdmparis/defense-finder
Clustal omega	DNA and protein sequence alignment	www.ebi.ac.uk/Tools/msa/clustalo
SalComMac	Identification of transcription start sites in *Salmonella* Typhimurium	bioinf.gen.tcd.ie/cgi-bin/salcom.pl?db=salcom_mac_HL
InterPro	Identification of protein domains from protein sequence data	www.ebi.ac.uk/interpro/
DALI	Identification of protein structural homologues in the PDB	ekhidna2.biocenter.helsinki.fi/dali
PyMOL	Protein modelling	pymol.org
Expasy ProtParam	Estimation of protein biophysical and biochemical characteristics	web.expasy.org/protparam/
Benchling	Cloning design tool	benchling.com
AlphaFold	Protein structure prediction	github.com/deepmind/alphafold
ConSurf	Identification of conserved residues in protein structures	consurf.tau.ac.il

PDBePISA	Prediction of protein-protein interfaces	www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver
PoseView	Identification of protein residues involved in ligand binding	proteins.plus
ont_fast5_api	.fast5 file manipulation	github.com/nanoporetech/ont_fast5_api
samtools	Sorting and indexing of .bam and .sam file formats	github.com/samtools/samtools
Tombo	*De novo* detection of modified DNA motifs	github.com/nanoporetech/tombo
nanodisco	*De novo* detection of modified DNA motifs and specific modified base	github.com/fanglab/nanodisco
Megalodon	Per site methylation analysis	github.com/nanoporetech/megalodon

2.4 Molecular Biology

All molecular biology techniques applied in this study were performed using standard methods, unless stated otherwise. Primers and oligonucleotides (oligos) were synthesised by Integrated DNA Technologies (IDT). Plasmids and primers used in this study can be found in Tables 2.6 and 2.14, respectively. Where necessary, DNA was visualised on electrophoresis on 1% w/v agarose gels, with the appropriate Tris acetate EDTA (TAE) running buffer and loading dye, as detailed in Table 2.3. Gels were visualised and imaged using a Gel Doc XR+ System (Bio-Rad) running the Image Lab software package (Bio-Rad).

Table 2.6 – Plasmids used in this study.

Plasmid	Notes	Primers used	Reference
pBrxXL$_{Sty}$	Full *S. enterica* serovar Typhimurium coding region in a pGGA vector backbone, created by golden gate assembly from genomic *S. enterica* serovar Typhimurium D23580 DNA.	TRB1367-TRB1378	This study
pBrxXL$_{Sty}$-*ΔbrxA*	Cloned by Genscript from pBrxXL$_{Sty}$.	Genscript synthesis	This study (Genscript)
pBrxXL$_{Sty}$-*ΔbrxB*	Cloned by Genscript from pBrxXL$_{Sty.}$	Genscript synthesis	This study, (Genscript)
pBrxXL$_{Sty}$-*ΔbrxC*	Created by Gibson assembly from pBrxXL$_{Sty.}$	TRB1788, TRB1790-TRB1794	This study
pBrxXL$_{Sty}$-*ΔpglX*	Created by Gibson assembly from pBrxXL$_{Sty.}$	TRB1734-TRB1737	This study
pBrxXL$_{Sty}$-*ΔpglZ*	Cloned by Genscript from pBrxXL$_{Sty.}$	TRB904/907	This study, (Genscript)
pBrxXL$_{Sty}$-*ΔbrxL*	Cloned by Genscript from pBrxXL$_{Sty.}$	TRB904/958	This study
pBrxXL$_{Sty}$-*ΔariA*	Created by Gibson assembly from pBrxXL$_{Sty.}$	TRB1778, TRB1795-TRB1799	This study
pBrxXL$_{Sty}$-*ΔariB*	Cloned by Genscript from pBrxXL$_{Sty.}$	Genscript synthesis	This study, (Genscript)
pBrxXL$_{Sty}$-*ΔariAΔariB*	Created by Gibson assembly from pBrxXL$_{Sty.}$	TRB1748-TRB1743	This study
pBrxXL-AL	*E. coli* BREX		[314]
pBrxXL$_{Eferg}$	*E. fergusonii* BREX		[321]

pUC19	pUC19 with BsaI sites mutated out via “round-the-horn" mutagenesis	TRB1200, TRB1201	NEB
pTRB507	pGGA plasmid backbone containing 12400 – 14394 of the pEFER plasmid from *E. fergusonii,* used as a negative control.	TRB1146, TRB1147	[321]
pSAT1-LIC	pBAT4 derivative; pMB1 replicon		[322]
pSAT1-6xHis-SUMO-brxA	Created by LIC	TRB1579 TRB1580	This study
pSAT1-6xHis-SUMO-brxB	Created by LIC	TRB1581 TRB1582	This study
pSAT1-6xHis-SUMO-brxC	Created by LIC	TRB1585 TRB1586	This study
pSAT1-6xHis-SUMO-pglX	Created by LIC	TRB1587 TRB1588	This study
pSAT1-6xHis-SUMO-pglZ	Created by LIC	TRB1589 TRB1590	This study
pSAT1-6xHis-SUMO-brxL	Created by LIC	TRB1591 TRB1592	This study
pSAT1-6xHis-SUMO-ariA	Created by LIC	TRB1595 TRB1596	This study
pSAT1-ariB	Created by LIC	TRB1593 TRB1594	This study
pSAT1-Strep-6xHis-SUMO	Created by GA	TRB1856 TRB1857	This study
pSAT1-Strep-6xHis-SUMO-pglX	Created by LIC	TRB1587 TRB1588	This study
pBAD30-Strep-6xHis-SUMO-pglx	Created by LIC	TRB2025 TRB2026	This study

pBAD30-pglX	Created by GA	TRB2027 TRB2028	This study
pBAD30-ocr	Created by LIC	TRB1911 TRB2081	This study
pBAD30-*pglX*(mut.1–23)	Derived from pBAD30-*pglX.* Twenty-three plasmids each with different mutant *pglX* genes, described in 6.2.1.	Genscript synthesis	This study

2.4.1 DNA extraction and purification

2.4.1.1 Bacterial gDNA extraction

Bacterial DNA was extracted and purified using either Zymo Genomic DNA Extraction or NEB Monarch Genomic DNA Extraction kits. Extractions were carried out as per manufacturer's instruction for the NEB Monarch Genomic DNA Extraction kit. For the Zymo Genomic DNA Extraction kit, up to 10^9 bacterial cells were used as a starting material and the Tissue Buffer was used for cell lysis rather than Cell Buffer, with an incubation time of 2 – 4 h. Following this, double the recommended volume of genomic binding buffer was used to ensure DNA binding to column and an additional volume of gDNA wash buffer was used to remove impurities. This combination gave around ten times greater yields of gDNA of good purity, as was required for sequencing work. Genomic DNA was eluted from purification columns using nuclease free water (Invitrogen).

2.4.1.2 Phage gDNA extraction

Phage DNA was purified from high titre lysates, the production of which is described in 2.6.1. Briefly, 500 µl of phage lysate was first treated with 2 µl of DNase A (ThermoFisher) and incubated at 37 °C for 30 minutes to degrade contaminating bacterial DNA. DNase A was then deactivated by addition of 20 µl of 1 M EDTA and incubation at 75 °C for a further 30 minutes. Phage genomic DNA was then extracted by phenol-chloroform extraction. Five hundred microlitres of 25 : 24 : 1 phenol : chloroform : isoamyl alcohol

was added and samples were vortexed briefly to mix, and incubated for 10 minutes at 4 °C. Samples were the centrifuged for 10 minutes at >12000 x g and supernatant was removed to a clean tube. An equal volume of 24 : 1 chloroform : isoamyl alcohol was added and samples were centrifuged again at >12000 x g for 10 minutes. Supernatant was removed again, and the previous step was repeated. The final aqueous layer contained phage genomic material, and this was extracted and further purified by ethanol precipitation. The aqueous layer was transferred to a new tube and incubated with 45 µl 3 M sodium acetate pH 5.2 and 500 µl isopropanol for 15 minutes at room temperature. DNA was then pelleted by centrifugation at 4 °C for 20 minutes at 16,000 x g. The pellet was washed twice with 70% ethanol before being left to soak overnight at 4 °C in either elution buffer (10 mM Tris-HCl, 0.1 mM EDTA pH 8.5) or MilliQ nuclease free dH_2O to resuspend. Purity was determined by visualisation on a 1% w/v agarose gel and quantified using a Nanodrop One spectrophotometer (ThermoFisher). Phage genome extractions were carried out in collaboration with Dr Abigail Kelly.

2.4.1.3 Plasmid isolation and purification

Plasmid extraction and purification was carried out using a Monarch Plasmid Miniprep/Midiprep kit, as per manufacturer's instructions. Plasmid DNA was eluted using MilliQ nuclease free dH_2O. Purity was determined by visualisation on a 1% w/v agarose gel and quantified using a Nanodrop One spectrophotometer (ThermoFisher).

2.4.2 Preparation of heat shock competent cells

Chemically competent *E. coli* DH5α, ER2796, BL21 (DE3) and ER2796 cells were prepared to allow heat-shock transformation. Overnight cultures from glycerol stocks were used to seed 25 ml of LB broth or 2x yeast tryptone (YT) broth supplemented with 15 mM $MgCl_2$ at 1 : 100 v/v. Cultures were grown at 37 °C until an OD_{600} of 0.4 – 0.6 was reached then transferred to a pre-chilled falcon tube and incubated on ice for 1 hour. Cells were then centrifuged at 4200 x g for 10 minutes at 4 °C to pellet and the supernatant discarded. Cell pellets were resuspended in 10 ml of chilled Solution A, incubated on ice for a further 20 minutes, and centrifuged again, as previously. Supernatant was discarded and cell pellets were

resuspended in 2 ml of chilled solution B, dispensed into usable aliquots and snap frozen in liquid nitrogen for storage at -80 °C.

2.4.3 Heat shock transformation

All bacterial transformations carried out in this study used a standard heat-shock method. Briefly, 50 µl of previously prepared competent cells were thawed on ice, added to 20 – 100 ng of purified plasmid DNA, mixed thoroughly and incubated on ice for 30 minutes. Cells were then heat shocked at 42 °C for 30 seconds and returned to ice for another 5 minutes, after which 950 µl of LB broth or 2x YT broth were added for outgrowth at 37 °C for 1 hour. Outgrowth cultures were spun at 4200 x g for 3 minutes, supernatant was discarded, and cells were resuspended in 50 µl of relevant media. The entire 50 µl volume was then plated on agar plates containing the appropriate antibiotic for selection.

2.4.4 DNA manipulation and cloning

2.4.4.1 Polymerase chain reactions

Polymerase chain reactions (PCRs) were performed using Q5 high fidelity polymerase (NEB) and incubated on a Nexus GX2 thermocycler (Eppendorf). Annealing temperatures for primers were initially calculated using the NEB Tm Calculator tool or the Benchling cloning assembly wizard (benchling.com) and optimised by temperature gradients where necessary. PCR reactions and reaction conditions are shown in Tables 2.7 and 2.8, respectively. Genes were amplified either from purified genomic DNA or from purified and sequence verified plasmid DNA.

Target amplicons from PCR reactions were purified from contaminant amplicons and free Nucleotide triphosphates (NTPs) by gel extraction using a Monarch Gel extraction kit (NEB), as per manufacturer's instructions, and eluted in MilliQ nuclease free dH_2O.

Table 2.7 – PCR reaction composition used in this study.

Component	Volume (µl)
5x Q5 reaction buffer	5
2mM dNTPs	2.5
DMSO	1
10 mM Fwd primer	1.25
10 mM Rev primer	1.25
Template DNA (~100 ng/µl stock)	1
Q5 polymerase	0.1
MilliQ dH_2O	*to 25 µl*

Table 2.8 – PCR reaction conditions used in this study.

Reaction step	Temperature (°C)	Time
Initial denaturation	98	30 seconds
35 cycles		
Denaturation	98	10 seconds
Annealing	60 – 72	30 seconds
Extension	72	30 seconds per kb
Final extension	72	2 minutes
Hold	10	indefinitely

2.4.4.2 Restriction digests

Restriction digests were carried out according to the manufacturer's instructions (NEB or ThermoFisher), using appropriate buffers and incubation times and temperatures for each restriction enzyme.

2.4.4.3 LIC

Ligation independent cloning (LIC) was utilised to create protein overexpression plasmids from pSAT1-LIC and pBAD30-LIC. This allowed the expression of fusion proteins with cleavable tags for efficient purification of recombinant proteins. A full list of plasmids used in this study, including overexpression constructs produced by LIC, is shown in Table 2.6.

Table 2.9 – LIC reaction conditions used in this study.

Component	Volume (µl)
For LIC preparation of plasmid	
Purified, linearised vector	25
25 mM dTTP	5
10x NEB 2.1 buffer	5
100 mM DTT	2.5
T4 DNA polymerase (NEB)	1
Nuclease free water (Invitrogen)	11.5
For LIC preparation of insert	
Gel purified PCR amplicon	10
25 mM dATP	2
10x NEB 2.1 buffer	2
100 mM DTT	1
T4 DNA polymerase (NEB)	0.4
Nuclease free water (Invitrogen)	4.6

pSAT1-LIC was firs digested with StuI (NEB) as per manufacturer's instructions to produce a linearised blunt-ended plasmid and re-purified by agarose gel electrophoresis and gel extraction using a Monarch Gel Extraction kit (NEB). The gene of interest (GOI) to be inserted into pSAT1-LIC was amplified by PCR using primers which introduced complementary to the LIC site in pSAT1-LIC and purified by agarose gel electrophoresis and gel extraction. The LIC reaction involves the excision of DNA bases from the 3' ends of DNA by T4 DNA polymerase, creating complementary overhangs. The reaction composition is shown in Table 2.9. Reactions were carried out on a thermocycler at 22 °C for 30 minutes and then stopped by incubation at 75 °C for 20 minutes. The resulting reaction products are mixed directly at a 1 : 1 ratio, incubated for 5 minutes at room temperature and used to transform heat shock competent *E. coli*, as described above. Annealing of the complementary sequences in the plasmid and the PCR amplicon is sufficient to maintain their interaction and ligation is performed once inside the cells. Successful transformants are selected by plating on the appropriate antibiotic and plasmids from resulting colonies are then purified and sequenced to confirm correct assembly.

2.4.4.4 GGA

Golden gate assembly (GGA) was performed to construct the pBrxXL$_{Sty}$ plasmid containing the entire *Salmonella* D23580 BREX coding region, including the region 508 bp directly upstream of the *brxA* start codon to ensure that any promoters and transcription factors required for BREX expression and function were included. GGA allows the sequential seamless insertion of DNA fragments using type IIS restriction enzymes. This method was chosen as it was deemed preferable to divide the 15.7 kb BREX coding region into fragments rather than attempt to transfer the entire sequence at once. Transferring the entire *Salmonella* D23580 BREX coding sequence to a plasmid backbone would allow easy transfer to *E. coli* DH5α, allowing utilisation of the Durham Phage Collection of Coliphages and direct functional comparison to previous work on the pEFER BREX system. The Golden Gate Assembly kit (NEB) provides the pGGA destination plasmid and is the same plasmid backbone used for subcloning of the pEFER BREX system [183]. Six fragments of the BREX coding region were amplified from *Salmonella* D23580 genomic DNA [319] by PCR using primers from Table 2.14. Initial attempts to assemble the fragments on the pGGA plasmid backbone were unsuccessful. Instead, fragments were first subcloned onto pUC19 donor plasmids from which BsaI cut sites had been removed. Fragments were cloned using a standard heat shock method and plated on ampicillin plates containing X – gal for selection. Sequencing confirmed the successful cloning of

fragments into donor plasmids. Full GGA was then attempted following the NEB protocol. Donor plasmids were added at 2 : 1 molar ratio to pGGA with assembly mix and GGA reaction buffer. The reaction was cycled 30 times (37 °C, 5 minutes; 16 °C, 10 minutes) followed by a 55 °C, 10-minute inactivation step. Resulting reaction mixes were used to transform DH5α via heat shock, and the cells were plated on chloramphenicol for selection. Correct assembly was assessed first by digestion of the resulting pBrxXL$_{Sty}$ plasmid with EcoRI and visualisation of resulting banding patterns on an agarose gel (see Chapter 3), then by sequencing.

Table 2.10 – Golden gate assembly reaction composition.

Component	Volume (µl)
pGGA destination plasmid (75 ng/µl)	1
Insert/donor plasmid	75 ng of each plasmid, 2 : 1 molar ratio
10x T4 DNA ligase buffer	2
NEB Golden Gate Assembly Mix	1
Nuclease free water (Invitrogen)	*to 20 µl*

Table 2.11 – Golden gate assembly reaction conditions.

Reaction step	Temperature (°C)	Time (minutes)
30 cycles		
Restriction	37	5
Ligation	16	10
Inactivation	50	5

2.4.4.5 Gibson Assembly

The creation of individual gene knockouts utilised Gibson Assembly (GA). GA provides a sequence independent homology-based method for scarless assembly of DNA fragments in a single reaction. Individual gene knockouts were designed within the context of the pBrxXL$_{Sty}$ vector to allow direct comparison on the same plasmid backbone. PCR primers were designed to amplify the pBrxXL$_{Sty}$ plasmid sequence either side of the gene to be removed. Primers were designed with overlapping regions to allow ligation of the amplicons via GA. GA designs consisted of 2-3 fragments of pBrxXL$_{Sty}$ produced by PCR with primers containing 20 bp homologous overlaps from upstream and downstream of the gene to be removed (Table 2.14). Knockouts were designed for each of the eight genes in the BREX operon, alongside an additional double knockout of *ariA* and *ariB*. GA reaction buffers were made in-house (Table 2.12) based on the recipes available on the open access protocol sharing repository, OpenWetWare (openwetware.org/wiki/Gibson_Assembly), which were in turn derived from the initial article documenting its use [323]. Equimolar amounts of PCR-amplified, and gel-purified fragments were pooled in an equimolar ratio to a final volume of 5 µl and added to 15 µl of assembly master mix. Reaction mixtures were incubated at 50 °C for 1 hour, then visualised on and gel purified from agarose gels. Resulting products which displayed the correct size were used to transform DH5α and cells were plated on chloramphenicol agar plates. Plasmids from resulting colonies were extracted and sequenced to confirm correct assembly. Gene Kos for which GA was not successful were instead synthesised by Genscript.

Table 2.12 – Gibson reaction components used in Gibson assembly buffers.

5x Isothermal reaction mix

Component	Volume
1M Tris pH 7.5	3000
1M MgCl2	300
100 mM dGTP	60
100 mM dCTP	60
100 mM dATP	60
100 mM dTTP	60
1M DTT	300
PEG 8000	1.5 g
100 mM NAD	300
MilliQ dH_2O	*to 6 ml*

Assembly master mix

Component	Volume
5x Isothermal reaction mix	320
T5 exonuclease (10 U/ul)	0.64
Q5 DNA polymerase (2 U/ul)	20
Taq DNA ligase (40 U/ul)	160
MilliQ dH_2O	*to 1.2 ml*

2.5 Sequencing and methylation analysis

All cloning work was confirmed by sequencing via DBS Genomics (Durham University). Primers used for sequencing are available in Table 2.14. Phage genome sequencing was performed either by MinION Mk1C nanopore sequencing (Alma and TB34), or by MicrobesNG using either Illumina HiSeq (Baz, Pau and BB1) or Illumina Novaseq 6000 (remaining coliphages). For nanopore sequencing of Phage genomes, phage genomes were assembled *de novo* using the "Assembly tutorial" Jupiter notebooks workflow provided by EPI2ME Labs (labs.epi2me.io). Briefly, initial assembly was performed using Flye 2.8.1-b1676, with '–genome-size' set to 0.1m and '–asm-coverage' set to 100x to account for high coverage values (github.com/fenderglass/Flye), and final assemblies were polished with medaka 1.5.0 using medaka_consensus (github.com/nanoporetech/medaka). Optional settings were left as default, unless otherwise stated. For detection of BREX motifs in phage genomes a custom Python script created by Dr Liam Shaw was used to count motifs on the forward and reverse strands (available: github.com/liampshaw/BREX-phage-motifs). All bacterial genomic sequencing was performed by MinION Mk1C nanopore sequencing and is detailed below.

2.5.1 Library preparation and barcoding

All library preparation and barcoding were performed as per kit manufacturers instruction, as available on the Oxford Nanopore Technologies community pages (nanoporetech.com/community). Extraction and purification of bacterial genomic DNA is detailed in 2.4.1.1. Library preparation was performed using the Ligation Sequencing Kit (SQK-LSK109) with the NEBNext Companion Module (NEB) and barcoding was performed with the Native Barcoding expansion 1 – 12 (EXP-NBD104) and Native Barcoding Expansion 13 – 24 (EXP-NBD114) kits.

2.5.2 Sequencing and basecalling

Sequencing was carried out using a MinION Flow cell (R9.4.1) on a MinION Mk1C. Following generation of raw sequencing data, basecalling was performed by the Guppy basecalling package (github.com/nanoporetech/pyguppyclient) using the high accuracy (hac) basecalling model either during sequencing or post sequencing with data deconvoluted using the ont_fast5_api package (github.com/nanoporetech/ont_fast5_api).

2.5.3 Methylation analysis pipeline

Analysis of BREX motif methylation utilised three tools, Tombo (github.com/nanoporetech/tombo), nanodisco (github.com/fanglab/nanodisco) and Megalodon (github.com/nanoporetech/megalodon), providing complementary data and ensuring that results were robust. To provide a negative control sample containing no modified bases, a whole genome amplification (WGA) reaction was performed on pBrxXL$_{Sty}$ genomic DNA using a REPLI-g Kit (Qiagen), according to manufacturer's instructions.

2.5.3.1 Tombo

Tombo requires single sequence .fast5 input files, therefore reads were first converted to single read .fast5 files from the multi read .fast5 files output by Guppy using the ont_fast5_api package (github.com/nanoporetech/ont_fast5_api). Reads were then re-basecalled using the 'resquiggle' argument and N6mA signals were detected by comparison to the WGA sample negative control using the 'detect_modifications level_sample_compare' arguments. On the outputted statistics file, the 'text_output signif_sequence_context' arguments were applied to provide a .fasta file containing regions around detected differences in signal and MEME [324] was used to identify conserved motifs.

2.5.3.2 nanodisco

nanodisco was designed for *de novo* discovery of methylation in bacterial genomes and metagenomes. nanodisco was installed in a Singularity container, as per the installation instructions (github.com/fanglab/nanodisco), allowing the installation as a functional image, and analysis of methylation was carried out as described in the "Detailed Tutorial" tab in the nandisco documentation. The WGA sample provided a negative control for signal difference, allowing motif identification and refinement.

2.5.3.3 Megalodon

Megalodon performs per site methylation prediction, independent of control samples, and maps results to a reference genome. Megalodon was installed as per instructions (github.com/nanoporetech/megalodon). As Megalodon does not perform *de novo* identification of modified motifs, this package was run for detection of modification levels of motifs identified by Tombo and nanodisco. The rerio basecall model res_dna_r941_min_modbases-all-context_v001 (github.com/nanoporetech/rerio) was used for identification of modified bases and mapped .bam files containing modified basecalls were sorted and indexed using the samtools package (github.com/samtools/samtools). Individual bases were visualised using IGV (Broad Institute) and the .bam files could be interrogated as .csv files in standard text or

spreadsheet programs. Comparison of modification in sample was further visualised using the 'megalodon_extras validate aggregate_modified_bases' arguments.

2.6 Bacteriophage manipulation

All coliphages used in this study were isolated from fresh water sources around the city of Durham, UK. Environmental samples were collected and purified initially by undergraduate students at Durham University as part of a microbiology workshop module ran by Dr Tim Blower [62]. A list of bacteriophages used in this study can be found in Table 2.13.

2.6.1 Bacteriophage isolation and lysate production

Water samples were first passed through a 0.22 µm filter to remove microbial contaminants, added to 10 ml of LB broth and inoculated with overnight cultures of *E. coli* DH5α. Cultures were incubated overnight at 37 °C to allow propagation of bacteriophages. Bacteria were pelleted by centrifugation at 15000 x g for 5 minutes and supernatant was extracted and sterilised through the addition of 100 µl of chloroform. Supernatant was then serially diluted in 0.22 µm filter sterilised phage buffer, 10 µl of dilutions were added to 3 ml of molten 0.3% LB agar, inoculated with 200 µl of DH5α overnight culture and poured onto 1% LB agar plates. Plates were incubated at 37 °C overnight and the presence of coliphages could be visualised by the appearance of plaques in the bacterial lawn. Phages then underwent two rounds of plaque purification to ensure homogenous phage populations. Individual plaques were removed using a sterile pipette tip, resuspended in 200 µl of phage buffer and 20 µl of chloroform was added to kill off bacteria. Phage samples were then serially diluted and plated as previously and incubated at 37 °C overnight.

Table 2.13 – Bacteriophages used in this study.

Phage	Genome size (kb)	Sequenced via	ENA Accession	GenBank Accession	Taxonomic group
TB34	165.22	MinION Mk1C	ERS12074569	OX001802.1	*Myoviridae*
Trib	88.85	Illumina Novaseq 6000	ERS12037452	OX016465.1	*Myoviridae*
BHP	87.7	Illumina Novaseq 6000	N/A	N/A	*Myoviridae*
Baz	87.69	Illumina HiSeq	ERS4850617	LR880803.1	*Myoviridae*
Alma	136.98	MinION Mk1C	ERS6301234	OV101294.1	*Myoviridae*
Pau	146.55	Illumina HiSeq	ERS4860126	LR865361.1	*Myoviridae*
PATM	146.8	Illumina Novaseq 6000	ERS12117655	OX090893.1	*Myoviridae*
BB1	110.1	Illumina HiSeq	N/A1	MT843274.1	*Demerecviridae*
Jura	71.49	Illumina Novaseq 6000	ERS5597183	LR999871.1	Other (to be assigned T7 sub-group)
SPSP	70.98	Illumina Novaseq 6000	ERS8393721	OV049961.1	Other (to be assigned T7 sub-group)
AL25	70.98	Illumina Novaseq 6000	ERS8485899	OW991345.1	Other (to be assigned T7 sub-group)
Titus	40.01	Illumina Novaseq 6000	ERS12117656	OX090892.1	*Autographiviridae*
Mak	40.14	Illumina Novaseq 6000	ERS8486536	OX001577.1	*Autographiviridae*
Bam	40.14	Illumina Novaseq 6000	ERS12037537	OW991346.1	*Autographiviridae*
CS16	42.66	Illumina Novaseq 6000	ERS5596661	LR999870.1	*Siphoviridae*
Mav	42.62	Illumina Novaseq 6000	ERS5347931	LR990702.1	*Siphoviridae*
Sipho	50.81	Illumina Novaseq 6000	ERS6377148	OU734268.1	*Drexlerviridae*

1Was submitted via Geneious GenBank submission plug-in. Genome annotations and database submissions were carried out by Dr Abigail Kelly.

Once phage populations were producing single plaque morphologies, a phage lysate was produced. This process was also used to produce fresh high-titre lysate samples from already purified phages when required. Phage lysates were serially diluted and plated as previously and incubated at 37 °C overnight. Plates showing a syncytium of phage plaques were chosen for lysate production as this had previously been shown to provide the highest titre lysates. The top 0.3% LB agar layer was scraped off the plate using a sterile spreader into a sterile falcon tube. Three millilitres of phage buffer was added to the plate to wash off residual top agar and phage and added to the same falcon tube. Bacterial contaminants were eliminated by addition of 500 µl of chloroform and samples were vortexed for 2 minutes to break up the agar. Samples were then centrifuged at 4000 x g for 20 minutes at 4 °C to separate the chloroform, agar and aqueous layers and the later containing phage particles was extracted and transferred to a fresh sterile tube. A further 100 µl of chloroform was added to ensure sterility and lysates were stored at 4 °C for further use. This method typically produced titres in the range of 10^8 to 10^{12} plaque forming units per millilitre (pfu/ml).

2.6.2 Growth curves

Bacteriophage growth and infection curves were carried out to monitor phage resistance conferred by BREX in liquid culture. Growth and infection curves were performed in a SPECTROstar Nano plate reader (BMG Labtech) in a 96 well plate format at 37 °C with shaking, using 200 µl culture volumes and monitoring growth via OD_{600} readings at five-minute intervals. Initial screening of inoculation and infection conditions produced optimal results with initial inoculation from overnight culture to OD_{600} 0.1 and phage multiplicity of infection (MOI) of 10^{-6}. As well as infection with phage TB34, a negative control (phage T7) and a positive control (uninfected culture) were run in parallel. All strains other than DH5α were grown with 25 µg/ml chloramphenicol. All data represent the mean and the standard deviation of at least 3 biological and technical replicates.

2.6.3 Efficiency of plaquing analysis

Efficiency of plaquing (EOP) assays were carried out to assess plaquing ability of phages in the Durham Phage Collection against DH5α BREX and BREX knockout strains relative to control strains. Phage plaque plates were carried out using serial dilutions of high titre lysates in phage buffer, as described in 2.6.1, with

the exception of strains which required induction of expression from pBAD-30 expression vectors. Here, overnight cultures were induced with 0.2% w/v L-arabinose and incubated at 37 °C for 30 minutes prior to plating and both top and bottom agar layers included 0.2% w/v L-arabinose to ensure continuous expression over the course of lawn growth. For experiments using pBREX-AL or pBrxXL$_{Efer}$, 200 µl overnight of culture was used to inoculate molten top agar. DH5α pBrxXL$_{Sty}$ and derivative knockout strains required an additional overnight culture step and the addition of 400 µl of overnight culture to molten top agar to ensure high enough cell density for sufficient lawn growth for plaque enumeration. Plates were incubated overnight at 37 °C and pfu/ml were counted on each plate. The efficiency of plating was calculated by dividing the pfu of the test strain by the pfu of the control strain. Data shown are the mean and the standard deviation of at least 3 biological and technical replicates.

2.7 Protein expression and purification

Recombinant proteins were overexpressed from either pSAT1 or pBAD30 expression vectors created by LIC cloning, as described in 2.4.5.1, using variations of expression and purification conditions previously reported by our lab [183,280,316,322]. All purification steps were performed either on ice or at 4 °C. Fast protein liquid chromatography (FPLC) steps were carried out at 4 °C using an Akta Pure protein chromatography system (Cytiva). Buffers used for purification are available in Table 2.3. Protein concentration measurements were performed by Nanodrop One (Thermofisher), calibrated to the relative molecular weight and extinction coefficient, reading absorbance at 280 nm. Details on the specific protocols used for expression of PglX and PglZ in this study are described below.

2.7.1 Large scale expression and induction conditions

All large-scale protein expression was performed in 1 L volumes of 2x YT broth in 2 L flasks with shaking at 180 rpm. In all cases, colonies from fresh transformation plates were used to inoculate 5 ml of 2x YT broth and grown overnight at 37 °C. This culture was then used to seed a 65 ml volume of 2x YT broth at 1 : 100 v/v and grown overnight at 37 °C to produce a second overnight culture. This culture was then used to seed

1 L of 2x TY at a 1 : 200 ratio, cultures were grown at 37 °C until exponential growth phase (OD_{600} 0.3 – 0.7), induced, and protein was expressed at 18 °C overnight. Protein was induced from pSAT1 plasmid backbones by addition of 0.5 mM of 0.22 µm syringe filtered isopropyl β-D-thiogalactopyranoside (IPTG). Protein was induced from pBAD30 plasmid backbones by addition of 0.1% autoclave sterilised L-arabinose.

2.7.2 SDS-PAGE

Throughout the purification process, protein purity was monitored by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). Recipes for SDS-PAGE gels can be found in Table 2.3. Gels were prepared using BioRad 1 mm mini-gel casts and the final acrylamide percentage (v/v) was selected as appropriate for the desired separation range. Separating gel was prepared, 4.5 ml was added to each cast and an additional 200 µl volume of isopropanol was added to ensure an even interface between separating and stacking gels. Once the separating gel had set, isopropanol was poured off. Stacking gel was then prepared and added until the cast was overflowing and well combs were inserted. Gels were either used immediately or wrapped in damp tissue and stored in a sealed bag at 4 °C for up to one week for later use. Samples were mixed with 3x loading dye (Table 2.3) at a 2 : 1 ratio and incubated at 95 °C for 5 minutes to denature before loading onto a gel. Gels were run in BioRad cassettes and tanks in 1x running buffer (diluted from 10x running buffer with MilliQ dH_2O) at 180 – 200 V until the dye front reached the end of the gel, providing maximum separation. Gels were then removed from the cassettes and stained for 1 hour in Quick Coomassie protein stain (Protein Ark), then de-stained for between 1 hour and overnight in MilliQ dH_2O. Imaging was then performed using a BioRad ChemiDoc XRS+ with ImageLab software (BioRad) on the Coomassie setting.

2.7.3 Harvesting of cell material and extraction of soluble fraction

Following overnight protein expression, cells were harvested by centrifugation at 4000 x g for 25 minutes at 4 °C. Cells were then resuspended on ice in chilled A500 buffer then lysed by sonication using a Vibracell VCX500 ultrasonicator. The soluble fraction was separated from insoluble cell material by centrifugation at

20000 x g for 45 minutes at 4 °C and the supernatant was removed to a fresh, chilled falcon tube for purification.

2.7.4 Nickel affinity chromatography

Soluble cell lysate was applied to a 5 ml pre-packed Ni-NTA His-Trap HP column (Cytiva) using a benchtop peristaltic pump at around 1.5 ml/min to allow binding of the 6xHis tag to the nickel resin. Columns were then washed with between 5 – 10 column volumes (CVs) of A500 to remove residual unbound protein. From this point forward, purification protocols varied between PglX and PglZ.

For PglX, isocratic elution steps were performed using a peristaltic pump using A500 buffer with imidazole concentrations adjusted to 30 mM, 50 mM, 90 mM and 150 mM, followed by a final elution step using B500. Isocratic elution steps were determined by initial gradient elution trials using FPLC and protein purity and yield were consistently higher using this method. For PglZ, columns were washed into low salt with 5 CVs of A100 buffer and eluted directly onto a pre-equilibrated anion exchange column with B100 buffer using a peristaltic pump.

2.7.5 Anion exchange

Anion exchange chromatography was performed for PglZ and, initially, for PglX, allowing separation of proteins by charge. For PglX, 90 mM and 150 mM imidazole elution fractions were dialysed into A100 at 4 °C overnight then loaded onto a pre-equilibrated 5 ml HiTrap Q HP anion exchange column (Cytiva). For PglZ, protein had been eluted directly from Nickel affinity columns, as described in 2.7.4. In both cases, columns were then washed with 5 CVs of A100 using a peristaltic pump before being loaded onto the FPLC system. Gradient salt elutions were then performed with 10% – 100% C1000 buffer and elutions were separated into 2 ml fractions for analysis by SDS-PAGE.

2.7.6 Tag cleavage and second nickel affinity purification

Both the pSAT1 and pBAD30 expression vectors produce protein of interest fused with a cleavable SUMO tag, for cleavage of upstream purification tags. The SUMO tag is recognised and cleaved by the human sentrin/SUMO-specific protease 2 (hSENP2). Fractions containing protein of interest as identified by SDS-PAGE, were collected, and pooled and incubated overnight at 4 °C with hSENP2 to ensure homogenous tag cleavage. Samples were then applied to a second Ni-NTA His-Trap HP column, this time allowing the now untagged protein of interest to flow through and removing remaining nickel binding contaminants. Successful tag cleavage and subsequent protein purity was assessed by SDS-PAGE, with tag cleavage visible as a noticeable reduction in protein molecular weight relative to tagged protein.

2.7.7 Heparin affinity chromatography

The methyltransferase PglX is predicted to bind DNA. In an attempt to improve protein purity, PglX samples were also applied to a 5 ml HiTrap Heparin HP column (Cytiva), allowing separation of proteins with affinity for DNA. Elution from heparin resin is dependent on salt concentration, in a similar manner to anion exchange. As such, for PglX, the heparin column step was substituted for the anion exchange purification step following the initial nickel affinity purification step. The 90 mM and 150 mM isocratic elution fractions were dialysed into A100 at 4 °C overnight and applied to the heparin column using a peristaltic pump. The column was then washed with 5 CVs of A100 buffer and loaded onto the FPLC system for gradient salt elution. Elution was carried out with C1000 buffer, using the same process as described for elution from anion exchange columns. Elutions were separated into 2 ml fractions and purity was assessed by SDS-PAGE.

2.7.8 Size exclusion chromatography

Size exclusion chromatography (SEC) was used to separate proteins by size, using a HiPrep 16/60 Sephacryl S-200 SEC column (Cytiva) connected to the FPLC system. Protein samples were dialysed overnight at 4 °C into S500 buffer and concentrated to a 500 µl volume. The column was pre-equilibrated in S500, and the sample was loaded through a 500 µl volume capillary loop at 0.5 ml/min. Sample was eluted over 1.2 CVs at 0.5 ml/min and fractionated into 2 ml volumes for analysis by SDS-PAGE.

2.7.9 Storage conditions

Purified protein from SEC was concentrated to around 6 mg/ml and diluted in storage buffer at a 1 : 2 ratio of protein to buffer, respectively, giving a final concentration of around 2 mg/ml. Samples were split into appropriately sized aliquots, snap frozen in liquid nitrogen and stored at -80 for future use.

2.7.10 Trial expressions and purifications

Following the creation of the pSAT1 overexpression plasmids for each of the genes in the BREX operon, trial expressions were carried out to assess recombinant protein expression in this format. Due to the smaller scale cultures used, the format of protein expression and purification differed slightly from the large-scale protocols, though the principles remained the same. Fresh transformations were used to inoculate an initial 2xYT broth overnight culture and this culture was then used to seed 25 ml of 2x TY. Protein induction, expression, cell harvesting, and lysis then followed the same steps as large-scale expression as detailed in 2.7.1 and 2.7.3, with the exception of cell pellets being resuspended in 1 ml of chilled A500 buffer. Soluble supernatant was then incubated on a roller mixer with 200 µl of Ni-NTA resin for 4 h at 4 °C to allow binding of the 6xHis tag. The soluble fraction was then applied to a spin column (Proteus) for separation of unbound protein by centrifugation at 12000 x g, washed with 5 volumes of 400 µl of A500 buffer, then eluted into a fresh tube with 200 µl of B500 buffer. Assessment of protein expression and purity was then performed by SDS-PAGE.

2.8 X-ray crystallography

2.8.1 Crystallisation screening

Highly pure protein samples were used for crystallisation screening. Samples were either used immediately following purification or thawed on ice from -80 °C storage. Samples were buffer exchanged into crystal buffer (Table 2.3) using Vivaspin 20 ml concentrators (Sartorius) and further concentrated to 12 mg/ml. Protein concentration determination was performed using Nanodrop One (Thermofisher), calibrated to the relative molecular weight and extinction coefficient, reading absorbance at 280 nm. Crystal screens were set using the sitting drop vapour diffusion method either by hand or using a Mosquito Xtal3 liquid handling robot (SPT Labtech). Crystal screens were incubated at 18 °C. A full list of commercially available crystal screens and a detailed description of crystallisation optimisations carried out during this study can be found in Chapter 5. All commercially available crystal screens were produced by Molecular Dimensions. For PglX + S-adenosyl-L-methionine (SAM) samples, PglX was incubated with 1 mM SAM for 30 minutes on ice prior to addition to screens. For PglX + SAM + Ocr samples, PglX underwent the SAM incubation as above plus an additional 30 minute incubation on ice with 2.74 mg/ml of Ocr, providing a 1 : 2 molar ratio of PglX to Ocr for binding. Ocr was kindly provided in purified form by Dr David Dryden [291,325]. Ocr purity was confirmed by mass spectrometry, SDS-PAGE and analytical SEC. For PglX + SAM + DNA samples, PglX was incubated as above but with DNA oligos – at a 1 : 1.2 molar ratio of PglX to DNA – in place of Ocr. For PglZ samples set with $ZnCl_2$, samples were incubated with 0.5 mM of $ZnCl_2$ for 30 minutes on ice prior to addition to screens. $ZnCl_2$ concentrations >1 mM caused immediate precipitation of PglZ.

2.8.2 Crystal harvesting and shipment

Crystallisation was confirmed by microscopy, with larger crystals extracted for X-ray diffraction. To harvest, 20 µl of screen condition was mixed with 20 µl of cryo buffer (Table 2.3) and the solution was mixed thoroughly by vortexing. This solution was then added directly to the crystal drop at a 1 : 1 ratio. For some conditions for both PglX and PglZ, addition of cryo buffer caused crystal dissolution and this step was omitted when necessary. Prescence of cryo buffer appeared to have no impact on diffraction resolution.

Crystals were extracted using nylon cryo loops, submerged in liquid nitrogen, transferred into a unipuck and stored in liquid nitrogen until shipment.

2.8.3 Data collection

Data collection was carried out remotely at Diamond Light Source, Oxford, UK on beamlines I04 and I24. Crystals were initially screened using the screening utility and the grid scan tool to allow prioritisation of diffracting crystals. Full data collections comprised of full 360° datasets of 3600 images. Data collection parameters and conditions are available in Chapter 5.

2.8.4 Data processing

Initial data processing was performed by automated processes on iSpyB (Diamond Light Source) using the Xia2-DIALS X-ray data processing and integration tool [326]. The same program was used to merge multiple datasets and provide initial data on the space groups and unit cell sizes. Further data reduction and production of dataset statistics was carried out using AIMLESS within CCP4i2 [327].

2.8.5 Structure determination

Merged datasets were first processed in CCP4i2 and then built and refined in Coot [328] and Phenix [329], respectively. Details of the final datasets after refinement are available in Tables 5.4 and 5.5 for PglX and PglX + Ocr, respectively. Quality of the final model was assessed using a combination of CCP4i2, Phenix, Coot and the wwPDB validation server. Visualisation and structural figure generation was performed in PyMol (Schrödinger).

2.8.5.1 PglX

The crystal structure of PglX was solved by molecular replacement in Phaser [330] using the PglX predicted model produced by AlphaFold. Details on the setup and parameters of AlphaFold used to generate PglX can be found in section 2.3.1. The resulting model then underwent iterative rounds of building and refinement in Coot [328] and Phenix [329], respectively. The SAM molecule was downloaded from the PDB ligand repository and placed manually in coot and similarly iteratively built and refined. Building and refinement of this structure was carried out by Dr Tim Blower.

2.8.5.2 PglX + Ocr

The structure of the PglX + Ocr heterodimer complex was solved by molecular replacement in Phaser [330] using the PglX + SAM structure solved previously and the PDB structure of Ocr (1S7Z). The resulting model then underwent iterative rounds of building and refinement in Coot [328] and Phenix [329], respectively. The SAM molecule was downloaded from the PDB ligand repository and placed manually in coot and similarly iteratively built and refined.

2.9 Biochemical and biophysical experiments

Purified PglX and PglZ were used for biochemical and biophysical characterisation. BrxB was expressed, purified and kindly provided by Dr Abigail Kelly. Ocr was kindly provided in purified form by Dr David Dryden.

2.9.1 Methyltransferase assay

SAM-dependant N6mA DNA methylation activity of PglX was probed *in vitro* using a MTase-Glo Methytransferase Assay kit (Promega). The kit allows indirect measurement of SAM dependent methyltransferase activity via production of the S-Adenosyl-L-homocysteine (SAH) reaction product. Through a proprietary two step reaction, SAH is used to produce ADP then ATP, which in turn is used by a luciferase reporter enzyme to generate a measurable luminescence signal. Signal can then be correlated to that produced by a SAH standard curve. The methyltransferase assay was carried out as per manufacturer's instructions in a 96-well plate format. PglX, PglZ and BrxB were buffer exchanged into the methyltransferase assay reaction buffer (Table 2.3) and concentrated to 1 µM. Protein samples were then either added to the plate separately or combined in equimolar quantities to wells of the assay plate, where required, and diluted with nuclease free water (Invitrogen). As a substrate, 100 ng of DH5α genomic DNA was used per reaction as this should provide ample *Salmonella* BREX recognition motifs for methylation. DNA substrate and SAM stock was added to the 4x Methyltransferase reaction buffer preparation and diluted with Nuclease free water to create a final 2x Methyltransferase reaction buffer stock, containing the 100 ng of DNA and 20 µM SAM. The 2x reaction mix was then combined with the protein samples at a 1 : 1 ratio and the reaction was incubated at room temperature for 30 minutes. The SAH standard curve was prepared by two-fold serial dilutions of a 1 µM SAH stock in 1x Methyltransferase reaction buffer The MTase-Glo reagent was added to samples and SAH standards, and the reaction was incubated for a further 30 minutes. MTase-Glo Detection Solution was then added to samples and SAH standards and incubated at room temperature for a further 30 minutes. Luminescence was then measured on a Biotek Synergy 2 plate reader.

2.9.2 Analytical SEC

Analytical SEC (A-SEC) was performed on a Superose 6 10/300 GL SEC column (Cytiva, discontinued) connected to an Akta Pure protein chromatography system (Cytiva). The column, system and loading loop were washed between each run and equilibrated with 1.2 CVs of A-SEC buffer (Table 2.3). Protein samples were buffer exchanged into A-SEC buffer and concentrated. Final concentration ranged between 1 µM and 5 µM, as required to give a distinct measurable elution peak. Protein was loaded onto the system via a 100

µl capillary loop loaded using a 100 µl Hamilton syringe. For PglX + SAM + Ocr samples, PglX was incubated with each on ice in the same process as that used for crystallisation screening (see 2.8.1). Protein in capillary loops was injected onto the column with 1.5 ml of A-SEC buffer and eluted over 1.2 CVs with A-SEC buffer at 0.5 ml/min. For estimation of protein molecular weight, relative to elution volume (V_e), a calibration curve was produced from commercially available high and low molecular weight protein calibration kits (Cytiva). Peaks were identified using the Unicorn 7 software package (Cytiva).

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

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

The molecular weight calibration curve is then plotted as K_{av} against $Log_{10}(M_r, kDa)$. The stokes radius calibration curve plotted as $Log_{10}(R_{st}, Å)$ against K_{av}, allowing calculation of sample stokes radius measurements. Estimated stokes radius calculations were carried out using the HullRad stokes radius estimation server [331], as described in Chapter 5.

Table 2.14 – Primers used in this study.

Primer	Sequence	Notes
Golden-gate assembly of *S. enterica* serovar Typhimurium BREX into pGGA		
TRB1367	tttTCTAGAGGTCTCCGGAGAATGGTTCATCTGGCGCT	*S. enterica* serovar Typhimurium BREX frag1 FWD
TRB1368	tttTCTAGAGGTCTCCTAACGAGGATTCAATGTCGC	*S. enterica* serovar Typhimurium BREX frag1 REV
TRB1369	tttTCTAGAGGTCTCGGTTAAAGAGCACAGCAATGAATATTG	*S. enterica* serovar Typhimurium BREX frag2 FWD
TRB1370	tttTCTAGAGGTCTCGAGgCCAAAAATATTATTTTCCAGAattag	*S. enterica* serovar Typhimurium BREX frag2 REV
TRB1371	tttTCTAGAGGTCTCGGcCTCGACATTGACGACCGTGCT	*S. enterica* serovar Typhimurium BREX frag3 FWD
TRB1372	tttTCTAGAGGTCTCGTTCATTAAATAATCTCCGGTGCATTGCCG	*S. enterica* serovar Typhimurium BREX frag3 REV
TRB1373	tttTCTAGAGGTCTCGTGAAAATAGCGGCGATTTATAC	*S. enterica* serovar Typhimurium BREX frag4 FWD
TRB1374	tttTCTAGAGGTCTCGGGTCACAATCCCATTTCATTCCA	*S. enterica* serovar Typhimurium BREX frag4 REV
TRB1375	tttTCTAGAGGTCTCGGACCTTGCAAAATCAGGAATTTATT	*S. enterica* serovar Typhimurium BREX frag5 FWD
TRB1376	tttTCTAGAGGTCTCGCACTTAAAAGAAATCATCCTGAAATGC	*S. enterica* serovar Typhimurium BREX frag5 REV
TRB1377	tttTCTAGAGGTCTCCAGTGAGGCGCTATGCAAAC	*S. enterica* serovar Typhimurium BREX frag6 FWD
TRB1378	tttTCTAGAGGTCTCCATGGAGGGAAACCAGGGGTTAC	*S. enterica* serovar Typhimurium BREX frag6 REV

TRB1200	aCGCGGTATCATTGCAGCACTGG	“Round-the-horn" mutagenesis FWD to remove BsaI site in Amp gene of pUC19
TRB1201	GACCCACGCTCACCGGCTCCAG	“Round-the-horn" mutagenesis FWD to remove BsaI site in Amp gene of pUC19
Gibson assembly primers for creation of pBrxXL*Sty* knockouts (KO)		
TRB1734	CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAA	pBrxXL*Sty* *pglX* Gibson assembly KO FWD (1)
TRB1735	CGCAGGCGATCGTCGAAGCTTTACTGAAGACGAATCCGGT	pBrxXL*Sty* *pglX* Gibson assembly KO REV (1)
TRB1736	ACCGGATTCGTCTTCAGTAAAGCTTCGACGATCGCCTGCG	pBrxXL*Sty* *pglX* Gibson assembly KO FWD (2)
TRB1737	TTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGA	pBrxXL*Sty* *pglX* Gibson assembly REV (2)
TRB1738	ATGCACCGGAGATTATTTAAATCTGGAATGAAATGGGATT	pBrxXL*Sty* PARIS Gibson assembly FWD (1)
TRB1739	TTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCG	pBrxXL*Sty* PARIS Gibson assembly REV (1)
TRB1740	CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAA	pBrxXL*Sty* PARIS Gibson assembly FWD (2)
TRB1741	TTCTTCGTTGGTCAGAAACACGTATTTGTCTTCAACACGT	pBrxXL*Sty* PARIS Gibson assembly REV (2)
TRB1742	ACGTGTTGAAGACAAATACGTGTTTCTGACCAACGAAGAA	pBrxXL*Sty* PARIS Gibson assembly FWD (3)
TRB1743	AATCCCATTTCATTCCAGATTTAAATAATCTCCGGTGCAT	pBrxXL*Sty* PARIS Gibson assembly KO REV (3)
TRB1788	GATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACA	pBrxXL*Sty* *brxC* Gibson assembly KO FWD (1)

TRB1790	GAATCCGGTCACCTGCCTCAGCTGTGCTCTTTAACGAGGA	pBrxXL*Sty* *brxC* Gibson assembly KO REV (1)
TRB1791	TCCTCGTTAAAGAGCACAGCTGAGGCAGGTGACCGGATTC	pBrxXL*Sty* *brxC* Gibson assembly KO FWD (2)
TRB1792	CCGGCGCGATCAGATAACTTTCAATACAAAAACGCGGAAGAACC	pBrxXL*Sty* *brxC* Gibson assembly KO REV (2)
TRB1793	CTTCCGCGTTTTTGTATTGAAAGTTATCTGATCGCGCCGG	pBrxXL*Sty* *brxC* Gibson assembly KO FWD (3)
TRB1794	GTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCT	pBrxXL*Sty* *brxC* Gibson assembly KO REV (3)
TRB1778	GATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAAC	pBrxXL*Sty* *ariA* Gibson assembly KO FWD (1)
TRB1795	AGTGTTTTCAGTAAGTTGCCATCGAACTCATCAAGAATGC	pBrxXL*Sty* *ariA* Gibson assembly KO REV (1)
TRB1796	GCATTCTTGATGAGTTCGATGGCAACTTACTGAAAACACT	pBrxXL*Sty* *ariA* Gibson assembly KO FWD (2)
TRB1797	TTGCCATCGGTTAAATCAATTTAAATAATCTCCGGTGCAT	pBrxXL*Sty* *ariA* Gibson assembly KO REV (2)
TRB1798	ATGCACCGGAGATTATTTAAATTGATTTAACCGATGGCAA	pBrxXL*Sty* *ariA* Gibson assembly KO FWD (3)
TRB1799	GTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATC	pBrxXL*Sty* *ariA* Gibson assembly KO REV (3)
LIC cloning primers for creation of overexpression plasmids		
TRB1579	caacagcagacgggaggtAAAAACGACAAGGCATGGAT	FWD LIC BrxA Salmonella D23580
TRB1580	gcgagaaccaaggaaaggttattaCCGTTGTCCCTCCAGAATAG	REV LIC BrxA Salmonella D23580
TRB1581	caacagcagacgggaggtATCGATCCCGTGCTTGAATA	FWD LIC BrxB Salmonella D23580

TRB1582	gcgagaaccaaggaaaggttattaACGAGGATTCAATGTCGCCG	REV LIC BrxB Salmonella D23580
TRB1585	caacagcagacgggaggtAATATTGAACAGATCTTTGA	FWD LIC BrxC Salmonella D23580
TRB1586	gcgagaaccaaggaaaggttattaCTGAAGACGAATCCGGTCAC	REV LIC BrxC Salmonella D23580
TRB1587	caacagcagacgggaggtAATACCAATAACATCAAAAA	FWD LIC PglX Salmonella D23580
TRB1588	gcgagaaccaaggaaaggttattaAATAATCTCCGGTGCATTGC	REV LIC PglX Salmonella D23580
TRB1589	caacagcagacgggaggtACCGACCAGTCGCAGCTGGC	FWD LIC PglZ Salmonella D23580
TRB1590	gcgagaaccaaggaaaggttattaAAAGAAATCATCCTGAAATG	REV LIC PglZ Salmonella D23580
TRB1591	caacagcagacgggaggtCAAACCCATCATGACTTACC	FWD LIC BrxL Salmonella D23580
TRB1592	gcgagaaccaaggaaaggttattaGTTTACTCCCAACGCCTTAT	REV LIC BrxL Salmonella D23580
TRB1593	caacagcagacgggaggtCGGAAAATGGCCAGCGTTAA	FWD LIC ariB Salmonella D23580
TRB1594	gcgagaaccaaggaaaggttattaCAATCCCATTTCATTCCAGA	REV LIC ariB Salmonella D23580
TRB1595	caacagcagacgggaggtAAAATAGCGGCGATTTATAC	FWD LIC ariA Salmonella D23580
TRB1596	gcgagaaccaaggaaaggttattaACGCTGGCCATTTTCCGCAT	REV LIC ariA Salmonella D23580
TRB1856	tggagccacccgcagttcgaaaaaTCAGGAGTCAAGACTGAG	FWD for cloning Strep site into pSAT1-LIC
TRB1857	tttttcgaactgcgggtggctccaCGGATGATGATGATGATGATG	REV for cloning Strep site into pSAT1-LIC

TRB2025	CAACAGCAGACGGGAGGTCCTCTAGAAATAATTTTGTTTAAC	FWD LIC pBAD30-RBS-His-Strep-PglX
TRB2026	GCGAGAACCAAGGAAAGGTTATTAGTTATTAAATAATCTCCGGTG	REV LIC pBAD30-RBS-His-Strep-PglX
TRB2027	CAACAGCAGACGGGAGGT GAAGGAGATATATCCATG AATACCAATAACATCAAAAAATAC	FWD LIC pBAD30-RBS-PglX
TRB2028	GCGAGAACCAAGGAAAGGTTATTA TTATTAAATAATCTCCGGTGC	REV LIC pBAD30-RBS-PglX
Sequencing primers (ranges indicate position of primer in *Salmonella* BREX coding region)		
TRB710	CGTTACCTGGAACCATTCGT	152-171
TRB711	CCCTATGGATAGCTGGGATG	865-884
TRB712	GCAGGACGTGATGGGTTTTA	1551-1570
TRB713	GCCAATACGACGCGTTTAAG	2266-2285
TRB714	GTCTATCCGGACCAAAGGTG	2957-2976
TRB715	CGGCTGCATTTTAATTCGTT	3659-3678
TRB716	GCACAAACTATGGCGGAAAT	4362-4381
TRB717	ACGGATGCCGAGAAGAAGAT	5058-5077
TRB718	GATAACCCGACAGGCTTTGA	5751-5770
TRB719	GTCTCGACATTGACGACCG	6463-6481
TRB720	ATTACGATGGCACATTTGGG	7155-7174
TRB721	GGATTAATGTGCACTCCGGT	7863-7882
TRB722	AAATCTCGAATTTATCGCCG	8567-8586
TRB723	ATTGGCTGGGCACGGGTA	9261-9278
TRB724	GGCGGTGTTGTACTCATTGAT	9968-9988
TRB725	CCGGTTTTAATACTGCGTTTC	10651-10671
TRB726	CTTGAAAGGCCTGGTCACTG	11368-11387
TRB727	TTGCTGAATCTGCGTAATCG	12056-12075
TRB728	GCATAACACCATTGATGCCA	12751-12770

TRB729	CGAATTTCAATCGCCGTAAT	13461-13480
TRB730	GTATTTACCGCACGTACGCA	14154-14173
TRB731	AACCAGCGCGACGTTATC	14835-14870
TRB732	CGCGGTAGATATTCCGACTG	15566-15585
TRB733	CAACCTCATCCTCTTCACCTG	FOR_24
DNA oligos used for PglX crystallisation trials		
TRB2129	TAGATCAGACAAC	PglX oligo 1 FWD
TRB2130	GTTGTCTGATCTA	PglX oligo 1 REV
TRB2131	TTCGATCAGTCCCA	PglX oligo 2 FWD
TRB2132	ATGGGACTGATCGA	PglX oligo 2 REV

Chapter 3 – Bioinformatic analysis and cloning of *Salmonella* BREX

3.1 Introduction

This study aimed to further structurally and functionally characterise BREX bacteriophage resistance. Though several papers on BREX systems have been published [183,280,281,294,308,314–316,318,321], understanding of the mechanistic functions of both individual components and BREX complexes is lacking. A chromosomal BREX system encoded by the clinically relevant *Salmonella enterica* subspecies *enterica* serovar Typhimurium strain D23580 (hereafter referred to as *Salmonella* D23580) was chosen for study. Phage defence activity from the *Salmonella* D23580 BREX system (hereafter referred to as the *Salmonella* BREX system) in a non-pathogenic *Salmonella* D23580 derivative strain has been shown against *Salmonella* phages [321]. Previous work undertaken by members of this lab had focussed on a BREX system from the multidrug resistant pEFER plasmid from *E. fergusonii* [183,280,316,321] and included functional profiling against the Durham Phage collection; a repository of coliphages [62,321]. It was desirable to utilise the Durham Phage Collection in the same manner with the *Salmonella* BREX system as this would provide data directly comparable to the pEFER BREX system. To achieve this, subcloning of the *Salmonella* BREX system onto a transferable plasmid backbone would be required. To further characterise individual components both structurally and functionally, individual protein expression vectors would be required. The design, creation and testing of these constructs is detailed in this chapter. The *Salmonella* BREX system was assayed for phage defence activity against the Durham Phage Collection and compared to phage defence profiles from both the pEFER BREX system and a BREX system from *E. coli* [314] against the same phages.

This chapter begins however with bioinformatic characterisation of the *Salmonella* BREX system, allowing initial understanding of the architecture of the system and putative functions of system components. Further, it is becoming more apparent that phage defence systems function as part of a wider defence arsenal [134,135,332,333], both at a genomic and a population level [274], and that interplay and interactions between phage defence systems are important determinants of the outcome of phage infection [334]. Thus, understanding the phage defence landscape of *Salmonella* D23580 will ensure that functional and structural characterisation is anchored in this context.

3.2 *Salmonella* D23580 BREX defence island contains multiple phage defence systems.

This study of the *Salmonella* BREX defence island began by carrying out bioinformatic characterisation, beginning at the DNA level. The BREX defence island encoded on a 15,163 bp region of the *Salmonella* D23580 chromosome was identified (Figure 3.1), comprising eight open reading frames (ORFs). As previously described, the *Salmonella* D23580 BREX system represents a type I BREX system [308], containing six core BREX genes. In addition, two ORFs sit between *pglX* and *pglZ*. These genes were later shown to be homologues of the recently discovered PARIS type II phage defence system [243]. PARIS type II phage defence was reported to be triggered by anti-restriction factors, such as the T7-encoded Ocr protein. The significance of this association will be discussed later in this chapter, and further in later chapters. Association of BREX systems with additional phage defence systems has previously been demonstrated [183] and clustering of phage defence components in genomic islands is common in nature [136,335].

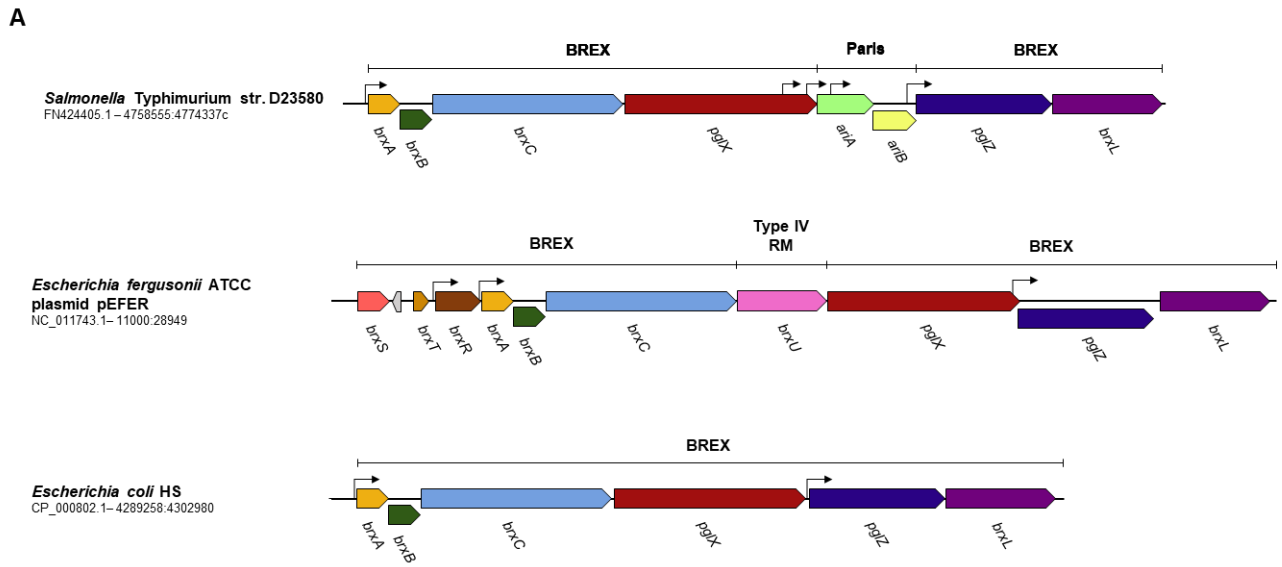

Figure 3.1 – Type I BREX systems display diverse architectures. *Scaled linear representation of the BREX systems utilised in this study. Putative promoter sequences are indicated with arrows.*

In the context of the *Salmonella* D23580 genome (GenBank: FN424405.1), the BREX operon is located near the 3' end on the reverse strand, between a putative endonuclease at the 3' end, and a hypothetical protein of unknown function and a putative methylase at the 5' end. In order to provide further understanding of the phage defence arsenal available to *Salmonella* D23580, known phage

defence systems within the genome were searched for using the python based "DefenseFinder" package (Table 3.1) [135,336]. DefenseFinder takes the proteome of an organism, in the form of a protein fasta file, and searches against a library of model profiles created for 151 types and subtypes of known phage defence systems. In total, DefenseFinder identified thirteen different phage defence systems. Interestingly, there appeared to be two sub-types of PARIS systems present, alongside three restriction enzymes of types I, III and IV. The system did not identify the endonuclease or methylase on either side of the *Salmonella* BREX system as phage defence components but did identify the putative helicase downstream of the BREX system (CBG27443.1_4413) as a form of the newly identified Mokosh phage defence system [134]. The BREX defence island then could potentially be extended to include three, and possibly four, putative defence systems. Elsewhere, another cluster of three systems – Lamassu, PARIS type I and PD-T4-1 – present another distinct phage defence island. Elsewhere, a type I and a second type IV system cluster together, while BstA, type III RM, Cas class I and retron type II systems appear alone.

Recent mapping of *Salmonella* D23580 promoters via reverse transcription PCR (RT-PCR) has allowed genome-wide analysis of transcriptionally active regions and has been made available as a public resource online [337]. Using this resource, several promoter regions within the *Salmonella* BREX operon were identified: One upstream of *brxA*; two within the C-terminal region of *pglX*; one at the beginning of *ariA*; one at the beginning of *ariB*. Seemingly, the defence island is split into three discrete transcribed regions, with *brxA* to *pglX*, transcribed together, then the PARIS system duet transcribed together, and finally, the *pglZ* and *brxL* genes transcribed together (Figure 3.1). The *Salmonella* BREX system also seems to lack a universal WYL-domain transcriptional regulator, often found upstream of BREX systems [279–281]. The hypothetical protein upstream of the operon could provide this function. Though this protein is over 800 bp away, it bears a structural similarity to two YebC family proteins (highest similarity: 1kon-A; Z-score 24.4, %id 36), known to be transcriptional regulators sensitive to a variety of stimuli [338–340]. Further upstream again is another putative methylase. It is unknown whether this methylase is related to the downstream endonuclease, but this would fit the trend of defence-related proteins clustering and it is possible that the BREX system was inserted between these two components.

Table 3.1* – Salmonella *D23580 encodes multiple defence systems. *Putative defence systems in the* Salmonella *D23580 genome, as found by the DefenseFinder python package.*

Type	Subtype	System start	System end	No. of genes
BstA	BstA	CBG23363.1_345	CBG23363.1_345	1
RM	RM_Type_III	CBG23436.1_418	CBG23437.1_419	2
Lamassu-Fam	Lamassu-Fam	CBG25716.1_2698	CBG25720.1_2702	2
Rst_PARIS	PARIS_I	CBG25720.1_2702	CBG25721.1_2703	2
PD-T4-1	PD-T4-1	CBG25728.1_2710	CBG25728.1_2710	1
Cas	CAS_Class1-Subtype-I-E	CBG25907.1_2889	CBG25914.1_2896	8
Retron	Retron_II	CBG26843.1_3813	CBG26844.1_3814	2
Mokosh	Mokosh_TypeII	CBG27443.1_4413	CBG27443.1_4413	1
*RM	*RM_Type_IV	CBG27444.1_4414	CBG27443.1_4414	1
BREX	BREX_I	CBG27445.1_4415	CBG27452.1_4422	6
Rst_PARIS	PARIS_II	CBG27447.1_4417	CBG27448.1_4418	2
RM	RM_Type_I	CBG27479.1_4449	CBG27481.1_4451	3
RM	RM_Type_IV	CBG27482.1_4452	CBG27482.1_4452	1

* System was identified manually.

The function of individual proteins encoded by the BREX operon was then analysed (Figure 3.2). For proteins of unknown function, putative functions were predicted using the InterPro protein domain prediction server from protein sequence input [341]. Structural homologues were then searched for via the DALI server [342], using either known structures or models generated using AlphaFold [343] to further infer function.

Recent work in our lab has identified BrxA as a dsDNA binding protein and the protein structure has been solved (Figure 3.2B) [316]. BrxA presents a small globular protein composed of α-helices and contains two helix-turn-helix domains allowing DNA binding (Figure 3.2). Though low scoring, BrxA shows structural similarity to nicking endonucleases SspB and FokI, and to type IIS Restriction endonuclease (RE), BpuJI (Table 3.2). Interestingly, SspB is an essential component in phosphorothioation of host DNA in the SspABCD phage defence system [344]. It is likely, however, that BrxA plays some role in regulation of the BREX complex, possibly through altering transcription. Intriguingly, it has previously been shown to be dispensable for restriction and methylation in *E. coli* [314] but is essential for both in *Actinobacter* [281].

BrxB is identical in sequence length to BrxA and returns no recognisable domains, families, or superfamilies from sequence level comparisons (Table 3.2). AlphaFold modelling produces a small, globular protein with four β-sheets at its core and a potential helix-turn-helix motif suggestive of a DNA binding modality (Figure 3.2C). Structural comparisons with DALI suggest structural similarity to domain III chromosomal replication factor, DnaA (Table 3.2), required for ATP binding and interaction with DnaB [345]. BrxB appears to lack a functional ATP binding motif however and, as with BrxA, it is likely that BrxB plays some ancillary role in BREX defence, either through regulation or complex formation.

BrxC is a large protein containing a AAA+ domain (Table 3.2). Alphafold modelling shows a long C-terminal "tail" domain composed of several elongated helices which loop back around towards the N-terminus (Figure 3.2D). The closest structural matches are to Origin of Replication Complex formation proteins and, interestingly, a AAA+ domain protein (Table 3.2) which regulates complex states in HORMA protein phage defence [258]. AAA+ domain proteins are often utilised as multimeric scaffold proteins and facilitate movement along DNA strands [346]. Thus, BrxC likely plays a role in complex formation and translocation in conjunction with one or more other BREX components.

PglX has previously been characterised as an N6mA methyltransferase and was demonstrated to be responsible for methylation of host DNA at the non-palindromic BREX motif sites [314]. The closest structural homologues (Table 3.2) are the MmeI type IIL restriction system [192] and an endonuclease and methylase, LlaGI [347]. PglX has been shown to be essential for both modification and restriction, but alone is insufficient to methylate host DNA [281,314]. A previous study reported that PglX from a BREX system in the *Salmonella* ER3625 lab strain methylated host DNA at the fifth adenine base of a

GATC**A**G motif [315]. The modelling and structure of PglX will be discussed in much greater detail in later chapters.

PglZ (Figure 3.2H) has a predicted alkaline phosphatase domain (Table 3.2). This domain is a structural homologue of the PorX orphan response regulator PglZ domain [317] and has been shown by other members of our lab to have phosphodiesterase activity (unpublished data). PglZ catalyses the Zn^{2+} dependent cleavage of poly- and cyclic- oligoadenylates to single nucleotides. Cyclic oligoadenylates are produced by CRISPR-Cas type III, CBASS, Pycsar and Thoeris phage defence as signalling molecules [136,227,255,262]. The function of phosphodiesterase activity in BREX phage defence is currently unknown, though it has been shown to be essential for defence [281,314].

BrxL shows Lon-like protease domains (Table 3.2), though no protease activity has previously been shown by this protein. Closer investigation showed that the Lon-like protease regions of BrxL correspond to the structural, ring forming region of the Lon-like protease. Thus, protease activity from BrxL seems unlikely and the similarity likely corresponds to similar structural assemblies. This hypothesis was validated by the recent publication of a cryo-electron microscopy (cryo-EM) structure of BrxL (Figure 3.32I) [318]. BrxL forms a barrel-like structure as either a homo-hexamer or heptamer and is able to bind DNA through its hollow centre. BrxL was shown to be required for phage defence, but not methylation in both the *Actinobacter* BREX [281] and *E. coli* BREX systems [314].

PARIS systems have demonstrated phage defence in response to phage encoded Ocr, the mechanisms involved are not yet understood. The larger AriA protein shows both AAA+ ATPase and Rad50/SsbC type AAA domain signatures from InterPro. Modelling produces a protein with a globular fold, containing three distinct β-sheet regions (Figure 3.2F). The closest structural homologues include chromosomal partitioning proteins (or Structural Modification of Chromosome – SMC - proteins) and SbbCD, a double strand break repair protein and a bacterial homologue of eukaryotic Rad50. The SMC protein domain was found in the recently discovered Lamassu phage defence system [136]. AriB shows no recognisable domains beyond a conserved DUF4435 domain. Structural modelling and comparison (Figure 3.2G) return several low-scoring OLD family nucleases and Ago proteins.

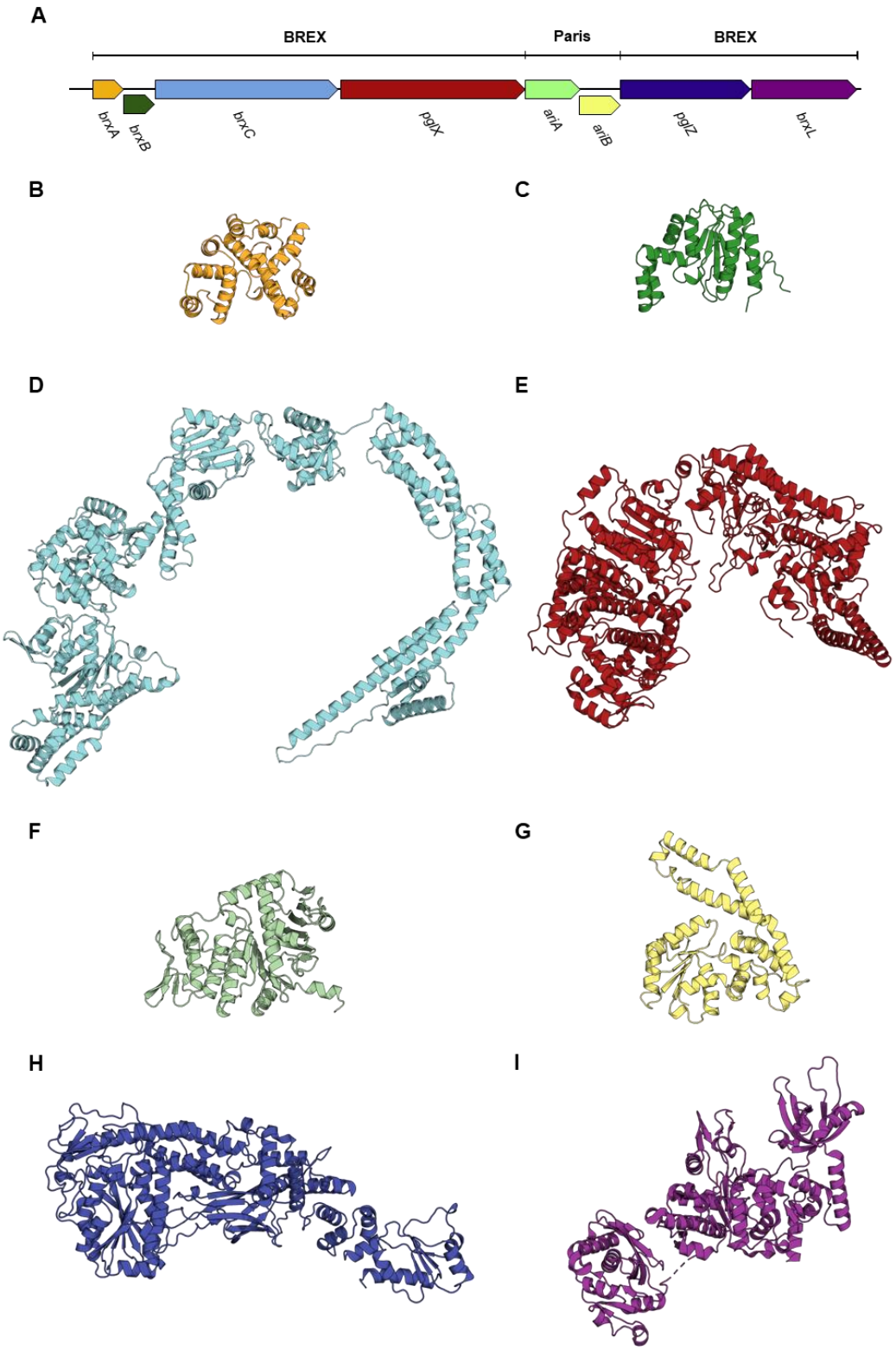

Figure 3.2 *–* ***Alphafold models and known structures of Sal BREX proteins****. The Salmonella BREX operon is shown again in panel A. Cartoon depictions of the protein structure of each of the individual components of the BREX operon are shown in panels B to I. BrxA (panel B) and BrxL (panel I) are from solved structures deposited in the PDB. BrxB (panel C), BrxC (panel D), PglX (panel E), AriA (panel F), AriB (panel G) and PglZ (panel H) are predicted structures produced by AlphaFold.*

Table 3.2 – BREX defence components possess diverse biochemical functions. *Sequence and structural level analysis of BREX component function. Protein biophysical characteristics were calculated with ProtParam. Sequence and structural based homology were inferred using InterPro and DALI, respectively. Proteins with no solved structure were modelled using Alphafold. Bold text indicates experimentally proven function.*

Gene name	Genomic locus	Protein length (AAs)	Theoretical protein PI	Protein molecular weight (kDa)	Predicted domains (InterPro)	Structural homologues (DALI)
brxA	4763227-4763829	200	7.86	22683	**DNA binding**	SspB, FokI, restriction endonuclease BpuJI
brxB	4762628-4763230	200	6.97	22970	*None*	Chromosomal replication initiation protein DnaA,
brxC	4768975-4762616	1213	5.57	139218	*AAA+ ATPase*	Origin of replication complex subunit, AAA+ ATPase
pglX	4765252-4768929	1225	5.65	141337	**Adenine specific DNA methylase,** MmeI-like DNA methyltransferase	Endonuclease and methylase LlaGI, MmeI type IIL restriction - modification system
aria	4764161-4765252	363	5.33	41201	*AAA+ ATPase, Rad50/SbcC-type AAA*	SbbCD, Rad50, Chromosomal partitioning protein
ariB	4763362-4764183	273	6.66	31447	*DUF4435*	OLD family nucease
pglZ	4760762-4763365	867	5.69	100353	Alkaline phosphatase	**Ectonucleotide phosphodiesterase**
brxL	4768667-4760751	694	6.25	77462	*Peptidase S16 Lon proteolytic domain, AAA^{+} ATPase*	Lon-like protease, **DNA binding**

3.3 All six core *Salmonella* BREX genes can be expressed in *E. coli.*

To facilitate further structural and biochemical study of individual genes in the *Salmonella* BREX system, each gene was subcloned into pSAT1-LIC; an IPTG-inducible, T7-driven expression vector. The use of ligation independent cloning (LIC) cloning (Figure 3.3A) allows homology-based insertion of sequences assisted by homologous overhangs which are exposed by the exonuclease activity of T4 polymerase [348]. LIC into pSAT1-LIC also allows addition of both 6x His and SUMO tags to the protein of interest for subsequent purification steps (Figure 3.3B). Individual genes were first amplified from genomic *Salmonella* D23580 DNA using primers which incorporated sequence homologous to those either side of a StuI digest site in pSAT1-LIC (Figure 3.3C). Both *pglX* and *brxC* required temperature gradient optimisation and gel extraction to ensure purity (Figure 3.3C). All other genes underwent PCR cleanup procedures to remove residual Nucleotide triphosphates (NTPs). Purified amplicons treated with DNA polymerase and dTTP were then mixed with StuI digested vector backbone treated with T4 polymerase and dATP. The mixture was used directly to transform chemically competent DH5α and plated on ampicillin plates for selection. Colonies were present for all strains after 18 h and successful cloning was confirmed by sequencing.

Initial small scale (25 ml) trial expressions were undertaken in *E. coli* BL21 to assess for protein expression from pSAT1-LIC plasmids. Protein purification was limited to initial spin column based Ni-NTA resin separation with isocratic 250 mM imidazole elution. BrxB, BrxC, BrxL and AriA all expressed well, while BrxA expressed poorly (Figure 3.3D). In this initial trial, no expression was evident from any of PglX, PglZ or AriB. A second, larger scale attempt (6 L culture volume) produced better expression for PglX and PglZ (Figure 3.3D). There remained no evidence of protein expression from the pSAT1-*ariB* plasmid. As PglX was of particular interest to this study, additional purification steps were undertaken, including anion exchange, tag cleavage and second Ni-NTA column, and size exclusion chromatography (SEC), producing a protein prep of reasonable purity (Figure 3.3D) and presenting a starting point for further optimisation (as discussed in Chapter 5). Protein expression from all six BREX genes in this format provides a solid foundation for study of the BREX protein products, both for this study (PglX, PglZ) and for future studies (BrxA, BrxB, BrxC, BrxL).

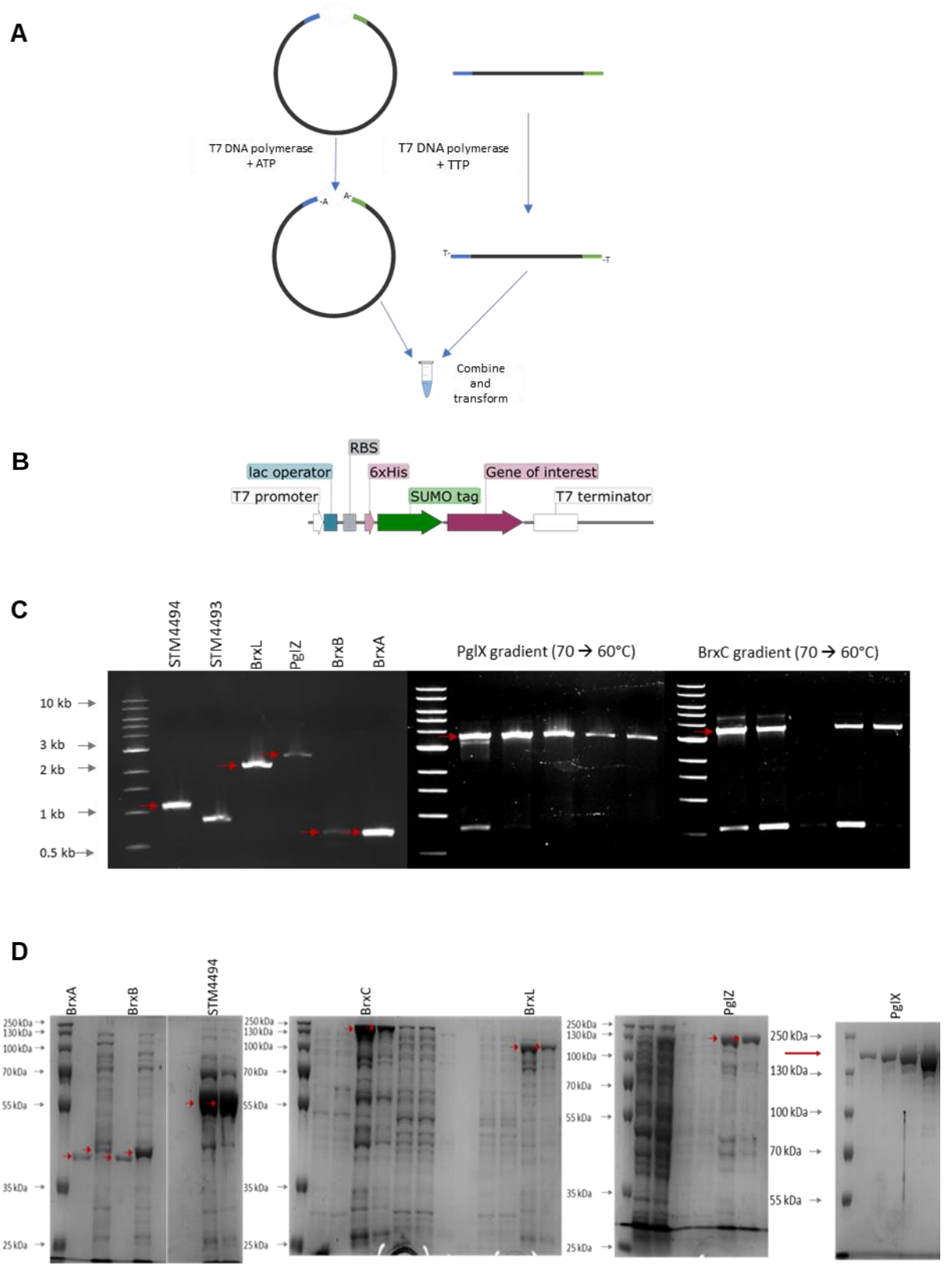

Figure 3.3 – All six core BREX genes were cloned and could be overexpressed. *A; Schematic of the ligation independent cloning (LIC) method used for subcloning each of the genes in the* Salmonella *D23580 BREX operon into the pSAT1-LIC plasmid. B; architecture of the pSAT1 overexpression system. C; PCR amplification of each gene in the* Salmonella *BREX operon. Correctly amplified band is indicated by a red arrow. D; Over expression of seven of the genes in the operon. Correct protein band is indicated by a red arrow.*

3.4 The *Salmonella* BREX locus was subcloned onto a transferable plasmid backbone.

Alongside the work described in 3.3, a method was also required for characterising the activity of the *Salmonella* BREX system *in vivo*. Limited characterisation of *Salmonella* BREX activity has already been carried out by our lab, demonstrating phage defence activity against several *Salmonella* phages [321]. It was decided that it would be useful to employ our in-house library of coliphages for further characterisation of *Salmonella* BREX, and this would also allow direct comparison to similar characterisation of the pEFER BREX system. It was therefore decided to clone the *Salmonella* BREX system onto a plasmid for use in an *E. coli* host. Transferring the entire system onto a plasmid would also provide the advantage of being able to use comparatively simple cloning techniques for generation of mutants. Nevertheless, the challenge of cloning over 15 kilobases (kb) of DNA from the *Salmonella* D232580 genome and assembling it on a plasmid was not insignificant. GGA is commonly used for the sequential assembly of multiple DNA fragments (Figure 3.4A and B) and allows simultaneous assembly with no residual scar sequence [169,170]. It was decided to split the BREX locus into 6 fragments, with the first fragment containing 508 bp of non-coding DNA upstream of the *brxA* start site to account for possible promoter regions. Each of the six regions of the BREX locus were amplified by PCR using primers designed to introduce appropriate type II RE sites for sequential reassembly on the pGGA plasmid backbone (NEB). Initial attempts were unsuccessful, however. To attempt to improve assembly efficiency, donor vector intermediates were utilised. Each of the BREX locus fragments were subcloned onto six donor vectors, digested and re-assembled onto pGGA, as previously. This time, the cloning appeared to be successful, and colonies were present on chloramphenicol transformation plates after selection. Colonies were subcultured and plasmids were extracted for further analysis by restriction digest with EcoRI (Figure 3.4 C). Having identified putative positive colonies, subsequent sequencing confirmed correct assembly of the full *Salmonella* BREX locus on pGGA (Figure 3.4 D). The resulting plasmid will be referred to as pBrxXL$_{Sty}$ for the remainder of this study, in keeping with the naming convention used for the pBrxXL$_{Efer}$ plasmid created by the same method when subcloning the pEFER BREX system [183].

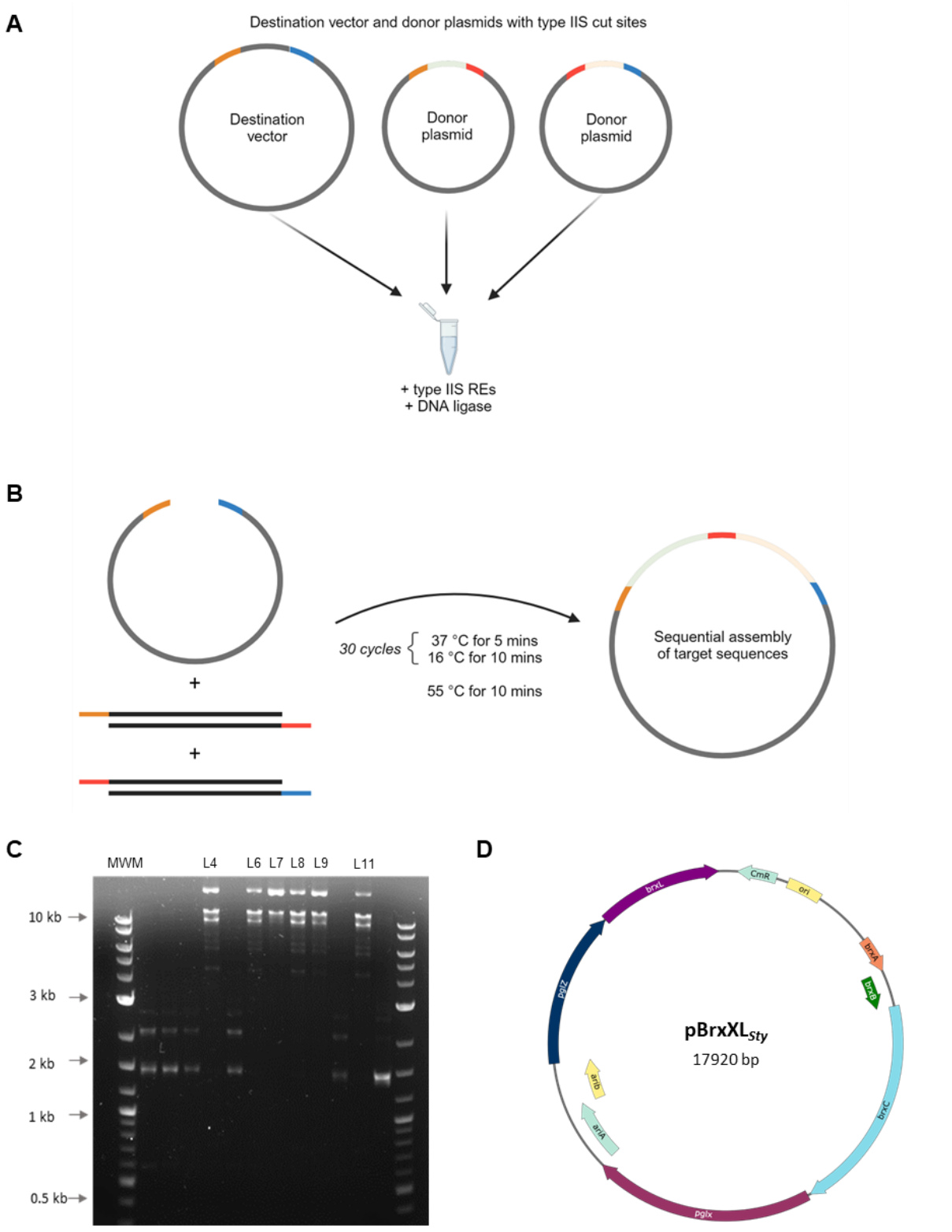

Figure 3.4 – Subcloning the* Salmonella *D23580 BREX operon onto a plasmid backbone. *A; Golden gate assembly utilises the cleavage of DNA by type IIS restriction endonucleases (REs) outside of their recognition motifs (shown as orange, blue and red) to assemble multiple target fragments (shown in pale green and pale orange) in a single reaction. B; Digested fragments are incubated in a thermal cycler and sequentially reassemble onto the target vector. Homologous overhangs produced by restriction anneal and fragments are ligated by a DNA ligase. C; restriction digests of assembled pBrxXL$_{Sty}$ constructs, showing expected digestion patterns in lanes 4, 6, 7, 8, 9, and 11. MWM is Generuler 1 kb Plus molecular weight marker D; the fully assembled pBrxXL$_{Sty}$ produced by this method.*

3.5 *Salmonella* D23580 BREX provides defence against coliphages.

Having subcloned *Salmonella* BREX, it needed testing for phage defence activity against coliphages in its new plasmid context within an *E. coli* host. Testing began using liquid cultures and examining bacterial growth during phage infections. Phage TB34 was selected as a model test phage as it was shown to be resistant to restriction by the type IV RE, BrxU, and was therefore predicted to have an unmodified genome and could therefore be susceptible to *Salmonella* BREX [183]. As a negative phage control, the model phage T7 was chosen, as this phage did not have any of the previously reported *Salmonella* ER3625 BREX motifs in its genome [315]. Additionally, phage T7 encodes the anti-restriction protein, Ocr, which has recently been demonstrated to inhibit phage defence by the *E. coli* BREX system [294]. Plasmid pTRB507 was used as a negative plasmid control; a pGGA plasmid backbone containing fragment 1 of the *E. fergusonii* BREX locus used in GGA, encoding the upstream non-coding region and some of *brxR* [183,321]. As expected, both TB34 and T7 were able to infect *E. coli* DH5α without plasmid (Figure 3.5, left panel), with both infected cultures displaying a rapid decrease in OD_{600} after around 2 h of growth post-infection at 37 °C. For the TB34-infected culture, growth was marginally restored after around 10 h, likely due to the rise of spontaneously phage-resistant mutants. For phage T7, there was no evidence of phage-resistant mutants and OD_{600} readings remained close to zero. DH5α pTRB507 cultures were infected by both phages and showed similar growth dynamics to DH5α without plasmid (Figure 3.5, middle panel), though Late stage OD_{600} readings were slightly lower than DH5α cultures, likely due to the inclusion of chloramphenicol in the growth media for plasmid maintenance. DH5α pBrxXL$_{Sty}$ cultures grew in the presence of TB34 but not phage T7 (Figure 3.5, right panel), suggesting that the *Salmonella* BREX system is active in *E. coli* against TB34. Growth of non-infected DH5α pBrxXL$_{Sty}$ was lower again than with uninfected DH5α pTRB507 strains (Figure 3.5, right panel). This trend was observed throughout this study for all DH5α pBrxXL$_{Sty}$ strains and mutant derivatives and is likely due to the increased fitness cost imposed by harbouring a large BREX plasmid.

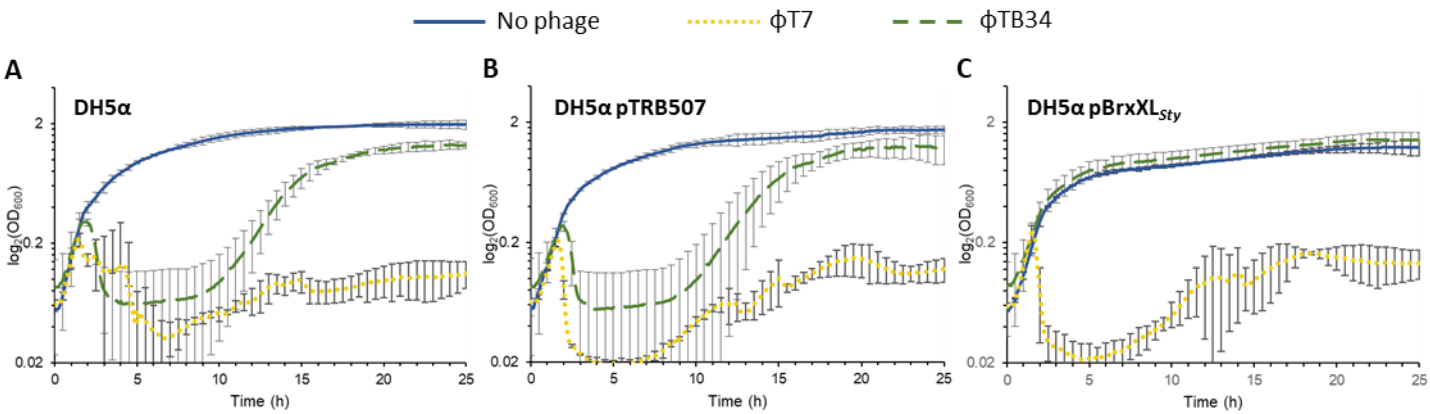

***Figure 3.5* – Salmonella *BREX is active in* E. coli.** *Growth and infection curves for Dh5α (left), pTRB507 (middle) and pBrxXL$_{Sty}$ strains, alone or infected with either phage TB34 or phage T7. Cultures were infected at an MOI of 10^{-6}. Error bars represent standard deviation.*

Production of the previous growth and infection data required a very low starting MOI of infection (MOI) of 10^{-6}, suggesting a weak phage defence phenotype. It was deemed valuable to gain some further insight into the level of protection provided by *Salmonella* BREX in comparison to previous studies in the literature and to that the pEFER BREX system done previously in our lab. Therefore, plaque assays for DH5α, pTRB507 and pBrxXL$_{Sty}$ strains against TB34 and T7 were undertaken. Plaque assays utilise serial dilutions and bacterial lawns to allow enumeration of individual phage infections, visualised as distinct phage plaques. From this, efficiency of plaquing (EOP) values can be calculated as a ratio of number of plaques in a test phage or strain relative to a control. EOP assays are considered the gold standard for quantification and comparison of phage tire calculation and comparisons of phage defence activity [349]. Phage defence conferred by pBrxXL$_{Sty}$ against TB34 produced a relatively modest ~100-fold reduction in phage plaquing efficiency. This is in line with efficacy observed previously for the *E. coli* and *Actinobacter* BREX system [281,314], and with that previously shown against *Salmonella* phages in the native *Salmonella* host [321]. Interestingly, TB34 was found to be resistant to the pEFER BREX system despite having numerous pEFER BREX recognition motif sites, suggesting strain specific inhibition of pEFER BREX activity. In summary, the *Salmonella* D23580 BREX system was shown to be active in *E. coli* against coliphage TB34 and further study of the system in this context was validated.

Table 3.3* – Salmonella *BREX provides modest protection against coliphages. Efficiency Of Plating (EOP) results for DH5α pBrxXL$_{Sty}$ against phages TB34 and T7. EOPs were calculated against the DH5α pTRB507 plasmid control strain.

Construct	TB34 EOP		T7 EOP	
	AV	SD	AV	SD
pBrxXL$_{Sty}$	2.86×10^{-2}	1.54×10^{-2}	1.36	3.03×10^{-1}

3.6 Testing against full coliphage panel

To provide a more complete comparison of phage defence phenotypes the plaque assay analysis was extended to our complete phage library. The Durham Phage Collection is composed of environmental phage isolates collected from waterways around Durham, UK [62]. Phages are collected and isolated by undergraduate students as part of an undergraduate microbiology workshop (in which I participate as a Demonstrator) and in this way, the collection is expanded each year. The phage collection has previously been used to characterise the phage defence activity of the pEFER BREX associated type IV RE, BrxU, a promiscuous enzyme which recognises and cleaves multiple DNA modification [183]. As such, there is a good indication of which of the phages in our library have modified genomes. More comprehensive identification of DNA modification would be complicated as many phages produce more complex hyper-modification, which can be recalcitrant to traditional sequencing-based modification detection methods (such as nanopore sequencing) and is otherwise beyond the scope of this study. As such, the BrxU-resistant phages were selected for pBrxXL$_{Sty}$ characterisation as these phages are likely unmodified and potentially susceptible. The resulting phage panel presented sixteen additional, phylogenetically diverse, environmentally isolated phages [62]. In order to extend this comparison, EOP assays were also performed for a third BREX system, from *E. coli* [314] allowing us to compare phage defence profiles of three BREX systems from different species against the same phage panel (Figure 3.6). The *E. coli* BREX system had previously been subcloned onto onto a pBTB-2 plasmid backbone – producing the pBREXAL plasmid [314] – and the empty pBTB-2 negative control [350]. All three systems represent archetypal type I BREX systems [308]. Collection of the EOP data for the three systems was a concerted effort between Sam Went, Dr David Picton [183], Dr Abbie Kelly, and Sam Duffner. Sam Went collected EOP data for all DH5α pBrxXL$_{Sty}$ strains and derivatives hereafter, and for several of the phages against *E. coli* BREX strains. Figure 3.6 was generated with the aid of Dr Liam Shaw.

In all, four phages (including TB34) were susceptible to defence conveyed by pBrxXL$_{Sty}$ (Fiigure 3.6A), categorised by a >10-fold reduction in plaquing efficiency (TB34, Alma, CS16, Sipho). It is also worth noting that phages Mav, PATM and Paula consistently showed reduced EOP values, but did not exceed a 10-fold reduction or greater. BREX systems from pEFER and *E. coli* produced different phage defence responses. The *E. coli* BREX system provided defence against phages TB34, Trib, Titus, Mak, Bam, CS16, Mav and Sipho, while the pEFER BREX system provided defence against Trib, Paula, PATM, CS16 and Sipho (Figure 3.6A). Where a phage was susceptible to multiple defence systems, the level of reduction

in EOP would often vary. For example, phage CS16 EOP was reduced >10-fold, >100-fold and by >10^6 by pBrxXL$_{Sty}$, pEFER and *E. coli* systems, respectively (Figure 3.6A). In addition to reductions in EOP, several, but not all, susceptible phages displayed smaller plaques (figure 3.6C). Smaller burst size indicates that even in a successful infection, production of new phage virions is inhibited.

Some phages were susceptible to BREX defence by one species but not another, suggesting species specific mechanisms of BREX inhibition. In order to identify whether this inhibition was the result of absence of recognition motifs within the phage genomes, the phages were sequenced using either nanopore or Illumina sequencing. Further, to ascertain whether the level of defence conveyed correlates with the number of recognition motifs within the phage genome, the number and location of motifs were mapped for each system in each of the phage genomes (figure 3.6B) using a custom python code (github.com/liampshaw/BREX-phage-motifs). CS16 shows some correlation in that there are ten more pEFER motif sites than pBrxXL$_{Sty}$ and ten more *E. coli* sites than pEFER, with slightly increasing levels of defence respectively. There is no correlation evident for any of the other phages tested, however. TB34 for example has twice as many pBrxXL$_{Sty}$ motifs as *E. coli* motifs but each produces similar reductions in EOP, while the EOP of Sipho is reduced by >100-fold by the *E. coli* system and only >10-fold by the pBrxXL$_{Sty}$ and pEFER systems, despite having eighteen fewer and 6 more *E. coli* motifs, respectively. As would be expected, lack of BREX motifs results in no phage defence. Seemingly, the presence of a single BREX motif is insufficient to illicit phage defence, as seen for phages Titus, Mak and Bam against pBrxXL$_{Sty}$, though it is possible that this is the result of some other additional mechanism of inhibition. Thus, beyond the complete absence of recognition motifs, the defence phenotype conveyed by a BREX system cannot be predicted simply by the number of target motifs in the genome.

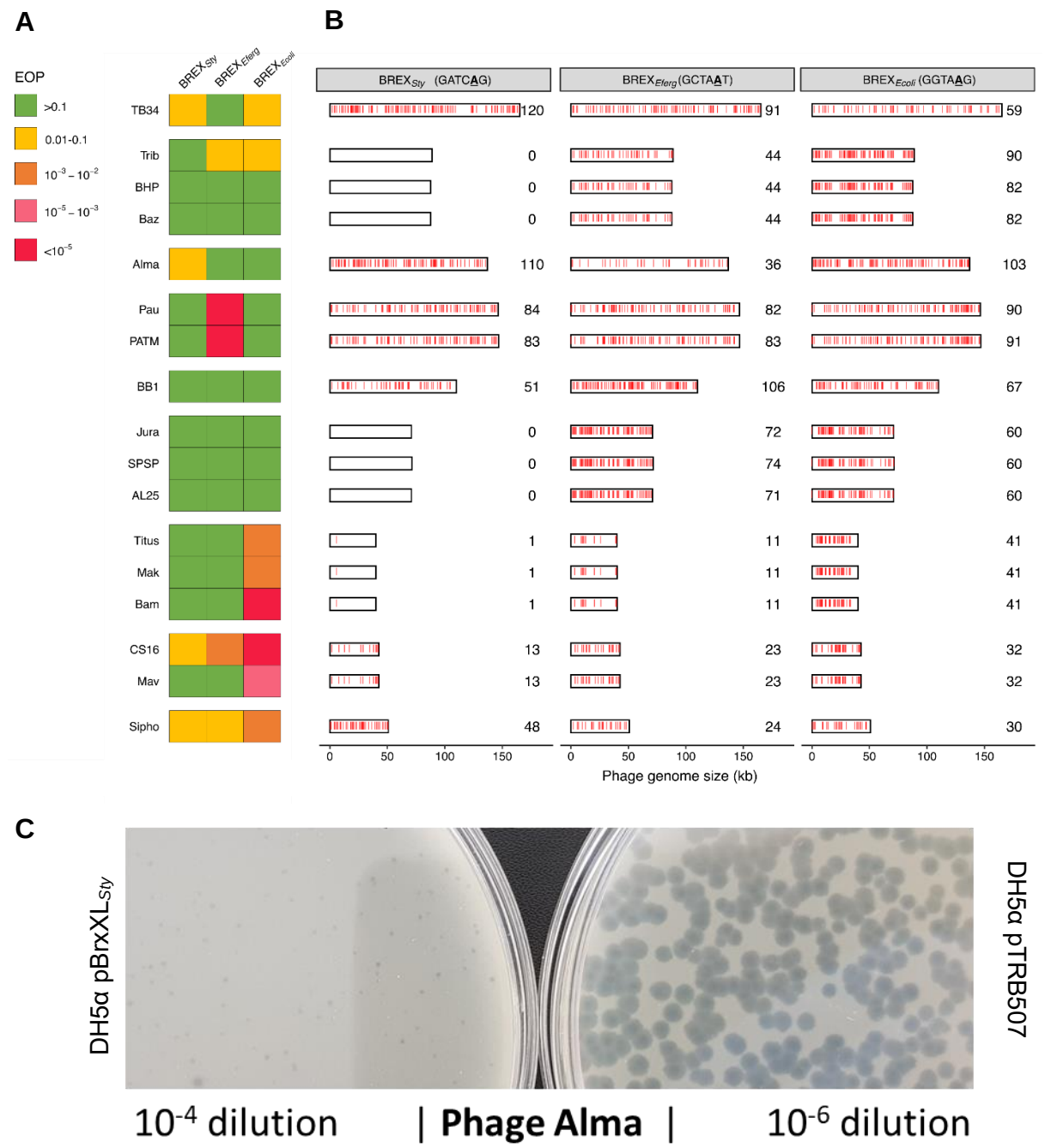

Figure 3.6 – BREX systems produce a diverse response to phages. *A: EOP values of three type I BREX systems against the Durham phage collection. B: Recognition motif positions of the different BREX systems in each of the phage genomes. A motif is indicated by a red vertical line on the to-scale bar representation of the phage genome. Total number of motifs in each genome is indicated to the right. C: Representative plaque assay result showing discrepancy in plaque sizes between pBrxXL$_{Sty}$ strain and plasmid control against phage Alma. Panels A and B were produced by Dr Liam Shaw, further details are available Chapter 2, section 2.5.*

3.7 Discussion

3.7.1 Bioinformatics

Initial bioinformatic analysis was carried out to allow understanding of the *Salmonella* D23580 BREX system composition and provide some genomic context, both in terms of local coding regions and for the complete phage defence landscape of *Salmonella* D23580. As this study is focussed on BREX, only the BREX operon was excised for characterisation. It is important, however, that findings are anchored in this context; it is becoming increasingly evident that bacteria often possess multiple phage defence systems [136,335], thanks to the development to new systematic approaches to defence system discovery. These defence systems can be additive or even synergistic [277], such as the recently described protection provided by the pEFER BREX system in which BrxU protects against modified DNA and the type I BREX system protects against unmodified DNA [183]. The PARIS system embedded within the *Salmonella* BREX coding region suggests a similar complementary relationship. PARIS has been shown to provide Abi mediated phage defence on encountering the phage encoded anti-restriction protein, Ocr, which has in turn been shown to inhibit BREX defence in *E. coli* [243,294]. This potential cooccurrence will be investigated and discussed further in later chapters. Nevertheless, this again highlights the need to consider phage defence systems within the context of the wider phage defence landscape.

Analysis of the *Salmonella* D23580 genome with DefenceFinder revealed that the defence island on which the BREX-PARIS operon sits may be larger than just two systems. The Mokosh type II system adjacent to the 3’ end of BREX consists of an RNA helicase domain and has been postulated to recognise phage RNA or RNA/DNA intermediates to provide phage protection via an as yet unknown mechanism [134]. Additionally, a putative type IV RE is sandwiched between BREX and Mokosh. This relationship between BREX and a type IV endonuclease could be homologous to that seen for pEFER BREX and BrxU [183]. Further investigation into the type IV nuclease using InterPro suggested a Mrr family domain, a type IV family often involved in DNA SOS response and attributed with recognising and cutting both methylated adenine and cytosine residues [351]. Structural modelling suggested the hypothetical endonuclease was structural homologue of the type IV endonuclease, MspJI, a eukaryotic enzyme attributed with recognising and cutting methylated and hydroxy-methylated cytosines [352]. Recognition of modified cytosines is the most likely scenario as BREX confers adenine methylation. In addition to this, another type IV Mrr homologue sits upstream of a type I RM system, which is in turn

predicted by the RM database REBASE to produce N6mA modifications. It is interesting that, in a different format, there is again a pairing of type IV Mrr RE with a defence system producing adenine methylation, though it remains to be seen whether these putative systems are active.

Elsewhere in the genome is a second apparent defence island composed of Lamassu, PARIS type I and PD-T4-1 systems. The significance of PARIS will be discussed further in later chapters – mechanistic understanding of Lamassu and PD-T4-1 remains lacking. A better understanding of these functions of these systems would feed into our understanding of the phage defence landscape in *Salmonella* D23580 and could ultimately allow better informed design of phage-based therapeutics. Current models for *in silico* detection of phage defence systems, such as tools like DefenseFinder, can detect only previously known and characterised phage defence systems. As such, the *Salmonella* D23580 genome may contain as hitherto unknow phage defence systems. Guilt by association approaches to system discovery could be performed, but even these are not exhaustive. Recently, a functional selection-based method was derived allowing detection of defence systems irrespective of genomic context [353]. Repeating such work with *Salmonella* strains could well provide a similar abundance of undiscovered systems.

The *Salmonella* D23580 BREX operon appears to lack any universal transcriptional regulator, such as the widespread WYL-domain regulator, BrxR [183,279–281]. The operon itself is split into three coding regions (*brxA* to *pglX*, *ariA* to *ariB*, *pglZ* to *brxL*). The two-operon format of *Salmonella* BREX has previously been documented in *Bacillus cereus* [308], though the *Acinetobacter* sp. NEB394 BREX system (hereafter referred to as the *Acinetobacter* BREX system) was reportedly transcribed as a single large transcript [281]. It could be postulated that the presence of the promoter sequence in the 3' region of *pglX*, intended for the transcription of *pglZ* and *brxL*, allowed the integration of the PARIS system within the BREX operon, inadvertently providing transcription for the PARIS genes. In turn, *ariB* contains a promoter region in its 3' region, ensuring that its integration does not inhibit surrounding genes and systems. It is possible then that phage defence genes harbour 3' promoter sequences to aid in horizontal gene transfer, providing a modularity that promotes successful integration. In addition to universal transcriptional regulator proteins such as BrxR, it may be of interest to identify common promoter sequences within phage defence genes which may act universal regulators at a DNA sequence level. In fitting with current trends of guilt by association approaches to identification of phage defence genes [135,136], this may in turn provide a novel mechanism of identifying new phage defence systems. Phage defence genes are often subject to horizontal gene transfer and are commonly

found clustered on mobile genetic elements (MGEs) [238,349,350]. Just upstream of the Mokosh gene is a *tRNA*^{LeuX} gene, which marks the beginning of a hypervariable region of the *Salmonella* genome [283]. This *tRNA*^{LeuX} gene is followed by a truncated integrase gene. It is likely that this integrase is inactive in its truncated form and is the remains of a degraded MGE. A recent study on phage defence systems in a related strain of *Salmonella* Typhimurium showed that may of the defence systems were situated proximal to integrases [333]. MGEs are considered one of the major driving forces for defence system and defence island turnaround in bacterial communities [354,355] and is likely the mechanism by which *Salmonella* D23580 acquired the BREX system.

Returning to the BREX operon, individual functions and putative functions of operon components were examined, to help understand what biochemical and functional activities are involved in BREX phage defence. Overall, the functions are diverse and complex. Several components have either been shown to, or predicted to, bind DNA, including BrxA, PglX and BrxL. Additionally, BrxC possesses a AAA+ domain which commonly function as ATP-driven molecular motors which provide translocate along DNA strands [356]. AAA+ proteins commonly form multimeric ring assemblies and pass DNA through a central pore [346]. A structural study of BrxL meanwhile demonstrated that BrxL forms a barrel-like structure around DNA, composed of a dimer of hexamers [318]. As such, BREX activity seems to require two DNA-binding complexes of BrxC and BrxL, respectively, each translocating along the DNA strand presumably in conjunction with other BREX components. Large multimeric complexes are not uncommon in phage defence; the *dndABCDE–dndFGH* phage defence system requires multimeric complexes of DndBCDE for host modification and complexes of DndFGH for restriction [357], while the recently discovered RADAR system forms multimeric complexes of RdrA and RdrB which come together to form complexes up to 10 MDa [358]. While BrxB does not show any recognisable domains, work elsewhere in our lab has suggested that BrxB co-purifies with BrxC, PglX, PglZ and BrxL (unpublished data), suggesting that it is an important intermediate in recruitment and interaction between these BREX complexes. Recently, the structural elucidation of PglZ-domain containing protein PorX allowed identification of PglZ as a phosphodiesterase. The PorX phosphodiesterase cleaves cyclic and linear oligonucleotides (oligos) [317]. Such oligos are often implicated in anti-phage based signalling pathways, such as that of type III CRISPR and the recently discovered cBASS systems [359–361], though what function this may play in BREX phage defence is currently unknown. PglX is a methyltransferase and responsible for host genome methylation and has been shown to also be essential to restriction [281,314]. Although sequence and structural homology searches label it as a type IIL RM, no nuclease activity has ever been reported, either by PglX alone or by BREX as a whole. As it is a focus of this study, the function of PglX will be discussed in greater detail in later chapters.

3.7.2 Cloning and Trial expressions

Following bioinformatic analysis, cloning was undertaken to allow further characterisation of both individual proteins within the BREX operon and of the BREX operon as a whole. First, overexpression constructs for production and purification of individual BREX proteins were produced. The pSAT1-LIC plasmid was chosen due to the ease of cloning into this vector and the cleavable amino-terminal (N-terminal) fusion tag it generates. The T7 promoter drives high expression levels and has been shown to provide high yields of target protein for us in the past and the addition of a cleavable tag allows for efficient two-step purification. Expression of each of the six core BREX genes within the BREX operon provides a solid foundation for further studies on these proteins. Of some note were point mutations which arose in some clones of PglX. The occurrence of point mutations in PglX is concurrent with previous studies, showing a high degree phase variation in PglX genes [357–359]. It is thought that such mechanisms may be employed to protect host bacteria from the toxic activity of these genes in the absence of phage infection. Indeed, strains harbouring pSAT1-PglX displayed slower growth than other BREX genes and overexpression of pEFER PglX has previously been found to be toxic [321]. Induction of expression of AriB also appeared to be toxic and no protein expression was evident. This is perhaps to be expected since PARIS provides defence through an abortive defence mechanism [243]. Expression under a more tightly controlled promoter, such as an araBAD arabinose inducible promoter, or co-expression alongside AriA may prove more successful. Constructs have been designed and produced to co-express AriB alongside AriA, hopefully mitigating toxicity, for future studies. Optimisation and purification of both PglZ and PglX will be detailed in later chapters.

3.7.3 Functional characterisation against coliphages

Alongside sub-cloning of individual genes, the complete BREX operon was successfully sub-cloned onto a pGGA plasmid backbone, allowing directly comparative characterisation in *E. coli*. The inclusion of around 500 bp of the non-coding region upstream of the *brxA* start site ensured that any promoter sequences or other promoter elements would be maintained in the resulting construct and that transcription would reflect that in native *Salmonella*. Growth and infection curves provided evidence of phage defence conferred by the system in this format. This provided initial evidence showing (i) the system was being expressed from the plasmid construct; (ii) the system is active in *E. coli* and; (iii) that the system provided defence against coliphages. In order to display growth in the presence of TB34, a low MOI was required. This is partially reflective of relatively low levels of phage defence conferred by

BREX. Additionally, growth and infection assays were carried out in small volumes and in 96-well plate format, producing slow growth rates even from DH5α (figure 3.2, left panel) which were further inhibited by addition of chloramphenicol for plasmid maintenance (figure 3.2, middle and right panels). As a result, phage infection at higher MOIs potentially outpaced bacterial replication causing rapid culture collapse. Notably, DH5α infected with TB34 partially recovered after around 10 h while those infected with T7 did not. This suggests that host mutations conferring resistance to TB34 were less costly to produce and/or maintain compared to those required to escape T7 infection. Sequencing of escape phage genomes and comparison to wild type phages may provide further insight into how phages escape BREX activity and conversely, provide insight into BREX function [150]. It should also be considered that phage T7 naturally encodes the anti-restriction protein, Ocr, reported to trigger Abi by PARIS [243]. Despite this, no defence activity is evident against T7. This is unusual, as the type II PARIS system from *E. coli* B185 was shown to be active against phage T7. The type I PARIS system from *E. coli* O42, a homologue of which is encoded elsewhere in the *Salmonella* genome, was shown to be inactive against phage T7, despite reducing plaque size. It is possible that the *Salmonella* BREX-associated PARIS system here has a different molecular trigger to that in *E. coli*, or perhaps the BREX-associated type II PARIS system is inactive and its function has been replaced by the type I system. In chapter five, PARIS phage defence activity is probed with Ocr and its *Salmonella* homologue, Gp5, and discussed further.

Following confirmation of *Salmonella* BREX activity in this format, the analysis was extended to EOP assays and comparison of resistance profiles to both the pEFER and *E. coli* type I BREX systems. The differing resistance profiles produced against our phage library does not seem to be attributable to presence of or number of recognition motif in the genome. Instead, phages seem to possess mechanisms of escaping species specific BREX systems. For example, TB34 is resistant to pEFER BREX but susceptible to both *E. coli* and *Salmonella* BREX systems. Interestingly, previous work has shown that TB34 is susceptible to an as-yet uncharacterised defence system on the *E. fergusonii* pEFER plasmid [183], demonstrating again the importance of considering interaction and complementation provided by phage defence systems in each organism, or contained on each mobile genetic element. The lack of activity against phages Titus, Mak and Bam, from pBrxXL$_{Sty}$, each with a single recognition motif, is intriguing and it would be interesting to introduce motifs into these genomes to determine whether they then become susceptible to BREX. The species specificity observed could also be attributed to defence systems targeting phages they are exposed to in nature, as both mutation rates and evolutionary turnover are higher in bacterial reproduction than phage replication [354,362]. In-line with this, the *E. coli* BREX system conferred resistance against more of the coliphages (eight) than either the *Salmonella* (four) or pEFER (five) BREX systems. It may then be beneficial to look for phages infecting

a target strain in collections of phages of loosely related strains as the clinical strain would not have optimised its defence systems to target such phages. Overall, EOP assays paint a complex picture and suggest that predicting phage infectivity will require far greater detail on phage bacteria interactions than investigating whether phage genome has appropriate recognition motifs. Towards this end, genomic comparison of phages here alongside aforementioned genetic profiling of spontaneous resistance mutants will be valuable.

3.7.4 Conclusions

In summary, this chapter has provided detailed bioinformatic analysis of the *Salmonella* D23580 BREX operon in its local and complete genomic context. The BREX operon forms a small part of the total phage defence arsenal available to *Salmonella* D23580, amounting to at least 13 putative phage defence systems. The BREX operon sits on a larger defence island in the hypervariable region of the *Salmonella* D23580 genome, which was possibly introduced by a now deteriorated integrase as part of a MGE [283,354]. Recent structural and biochemical studies [318] combined with analysis of putative functions of BREX components (Table 3.2) paint a complex picture of BREX defence, involving at least two BREX complexes translocating along host and/or target DNA. Towards further study of components and complexes, each gene in the BREX operon was individually subcloned into overexpression vectors and protein expression was demonstrated from each, apart from pSAT-*ariB*. These constructs will be invaluable in the further structural and biochemical investigation of BREX components, as well as composition and function of BREX complexes and interactions. Optimisation of expression and purification of PglZ and PglX towards this purpose will be detailed further in chapter five. The entire BREX operon was subcloned onto a single plasmid and shown to be active in *E. coli* against four members of our coliphage library. No activity was evident against phage T7, despite this phage naturally encoding the Ocr protein reported to trigger PARIS phage defence. Comparison of phage defence profiles of three type I BREX systems (*Salmonella* D23580, pEFER and *E. coli*) demonstrated a complex response to phages, with phages seemingly able to produce species-specific mechanisms of inhibiting BREX function. Genomic analysis of phages in this library, together with genomic analysis and comparisons to spontaneously arising phage escape mutants, may shed further light on the mechanisms by which phages inhibit or escape BREX systems and potentially improve our understanding of the mechanism of BREX defence. Together, these data will contribute to furthering our understanding of the complex and intricate phage-bacteria interactions towards both their use as therapeutics and as potential sources of biotechnological tools.

Chapter 4 – Function and essentiality of individual BREX genes

4.1 Introduction

In Chapter 3, the *Salmonella* BREX system was subcloned into the pBrxXL$_{Sty}$ plasmid and was shown to provide phage defence in DH5α against coliphages from the Durham Phage Collection [62]. The defence profiles produced by the *Salmonella*, pEFER and *E. coli* BREX systems differed, with defence conferred not correlative to the number of respective recognition motifs in the phage genome, making it difficult to predict defence response to new phages. The components of type I BREX systems essential for phage restriction and host methylation from different host species have also previously been reported to differ [281,314], adding further complexity to BREX function. In Chapter 4, the essentiality of individual components of the *Salmonella* BREX system are investigated for comparison by producing individual gene knockouts of each gene in the BREX operon. Function is analysed by growth and infection curves and EOP assays. The utility of nanopore sequencing [363,364] is tested for the analysis of host genome methylation by BREX systems and benchmarked against PacBio sequencing. This also involved establishing an effective bioinformatic pipeline for the identification of N6mA modifications within BREX motifs from nanopore sequencing datasets.

Comparison of phage defence responses from the *Salmonella*, pEFER and *E. coli* BREX systems against the Durham Phage Collection also suggested that phages encode mechanisms for evading phage defence from species specific BREX systems. Phage escape mechanisms are varied and can be highly specific to defence systems [150] or more general, such as the anti-restriction protein Ocr, encoded by phage T7 [285,287]. Ocr mimics the shape and charge of around 20 bp of DNA and competitively inhibits RM systems [291,325]. More recently, Ocr was shown to inhibit the *E. coli* BREX system through direct interaction with PglX [294]. In this chapter, inhibition of the *Salmonella* BREX system by Ocr and by the *Salmonella* phage homologue, Gp5, is investigated. Understanding and predicting phage response to defence systems and the wider intricacies of phage-bacteria interactions will be essential in designing effective phage therapeutics.

4.2 Generation of BREX operon gene knockouts on pBrxXLSty.

Having established that the *Salmonella* D23580 BREX system provides defence against coliphages in an *E. coli* background, this model was used to further probe the roles of individual proteins within the *Salmonella* BREX system. Studies in *Acinetobacter* and *E. coli* have shown some discrepancy in which BREX components are essential for host methylation and phage restriction [281,314]. Investigation of the essentiality of BREX components in the *Salmonella* system would provide a valuable comparison. This required the production of individual gene knockouts of each of the six BREX genes in the operon. In order to test whether any of the phage defence activity we observed in Chapter 3 was the result of activity from the inserted PARIS phage defence system, individual gene knockouts of each of the PARIS genes as well as a double gene knockout of both *ariA* and *ariB* genes were also produced. Production of these mutants was partially simplified by the previous transfer of the BREX operon to a plasmid backbone as this allowed the use of standard *E. coli*-based cloning methods. A modified version of Gibson assembly (GA) was used allowing the creation of sequence independent, scarless gene deletions [323]. To prevent any remnant activity from truncated protein forms, clean deletions were produced in all cases where this would not interfere with the ribosome binding site of subsequent genes. Areas encoding putative promoter sequences (3′ region of *pglX*, 5′ region of *ariA*, 3′ region of *ariB*), as identified in Section 3.1, were retained in relevant knockouts and additional stop codons were introduced to prevent any errant translation. Design of the GA knockouts is shown in Figure 4.1 A and explained in detail in Chapter 2. In brief, the 17.9 kb pBrxXL$_{Sty}$ plasmid was split into two or three fragments to allow PCR amplification. Primers were designed to contain homologous regions matching either side of the gene to be removed. Fragments were then amplified and gel purified to remove incorrect amplicons and free NTPs. Amplicons were then combined and treated with a standard GA protocol; T4 exonuclease chewed back bases at 3′ ends allowing annealing of homologous sequences and ligation. Colonies produced after transformation and selection on chloramphenicol plates were then confirmed by sequencing. Using this method, four out of nine gene knockouts were successfully produced (pBrxXL$_{Sty}$-Δ*brxC*, pBrxXL$_{Sty}$-Δ*pglX*, pBrxXL$_{Sty}$-Δ*brxL* and pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB*). Despite several rounds of redesigning primers and fragment regions, the remaining knockouts remained elusive, often producing incorrect and incomplete assemblies. These knockouts were instead produced by Genscript. In all, nine new constructs were available for restriction and methylation studies (Figure 4.1 B–J).

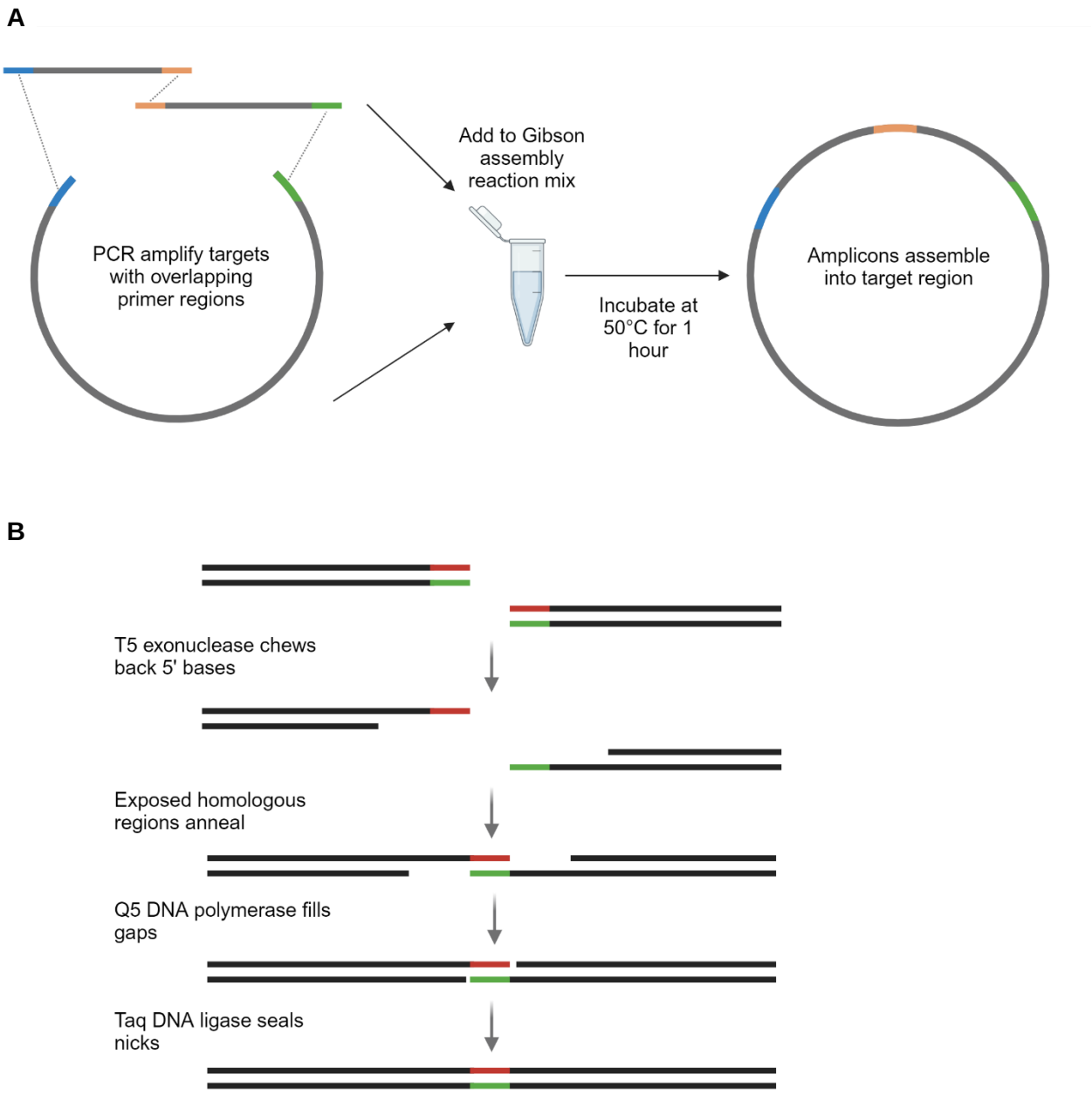

Figure 4.1 – Gibson assembly produced individual gene knockouts of BREX genes from pBrxXL$_{Sty}$. A; Gibson assembly is a homology-based cloning method for the scarless assembly of multiple DNA sequences in a single reaction. Target sequences are amplified with primers which introduce homologous regions, amplicons are combined and assembly proceeds at 50°C. B; T5 exonuclease chews back bases at exposed 5’ ends allowing annealing of homologous regions . A DNA polymerase then fills in gaps and the heat resistant Taq ligase seals DNA nicks.

4.3 Liquid culture infection and EOP assays show BREX genes essential for restriction.

The investigation of the new knockout constructs began by producing growth and infection curves against phages TB34 and T7, as carried out in Chapter 3. As previously demonstrated in Chapter 3, both phages are able to infect and kill *E. coli* DH5α but the presence of the pBrxXL$_{Sty}$ plasmid provides protection against TB34, but not phage T7 (Figure 3.6). Growth and infection curves for knockout strains were carried out using the same method, with the same low MOI (10^{-6}). Results demonstrated that *brxA*, *brxB*, *brxC*, *pglX* and *pglZ* are essential for phage restriction by the *Salmonella* D23580 BREX system (Figure 4.2 A–E). Conversely, *brxL* is not required for restriction (Figure 4.2 F). Knockouts of either *ariA* and *ariB*, and the double knockout of both genes demonstrated that the PARIS system is not providing active defence against TB34 (Figure 4.2 G–H), and that defence conferred by this operon is attributable to BREX.

As with the initial testing of pBrxXL$_{Sty}$ in Chapter 3, further resolution was added to this data through EOP assays (Table 4.1). EOP values were again calculated against the DH5α p507 empty vector strain. The results confirmed that *brxA*, *brxB*, *brxC*, *pglX* and *pglZ* are essential for restriction (Table 4.1). In liquid culture assays deletion of *brxL* produced similar growth in uninfected and TB34 infected cultures (Figure 4.2 F), suggesting the BREX system was still active. In EOP assays, deletion of *brxL* appeared to enhance the BREX phenotype, causing an ~10,000-fold reduction in EOP against TB34 (Table 4.1). Consistent with growth curve data and previous EOP data (Chapter 3), no activity was observed by the *pglX* knockout against T7, nor indeed against TB34, again indicating no apparent phage defence activity by the PARIS system. Strains with either or both of the PARIS system genes removed showed slight increases in phage defence activity against TB34. Again, no phage defence activity was seen against phage T7 by any of the deletion strains.

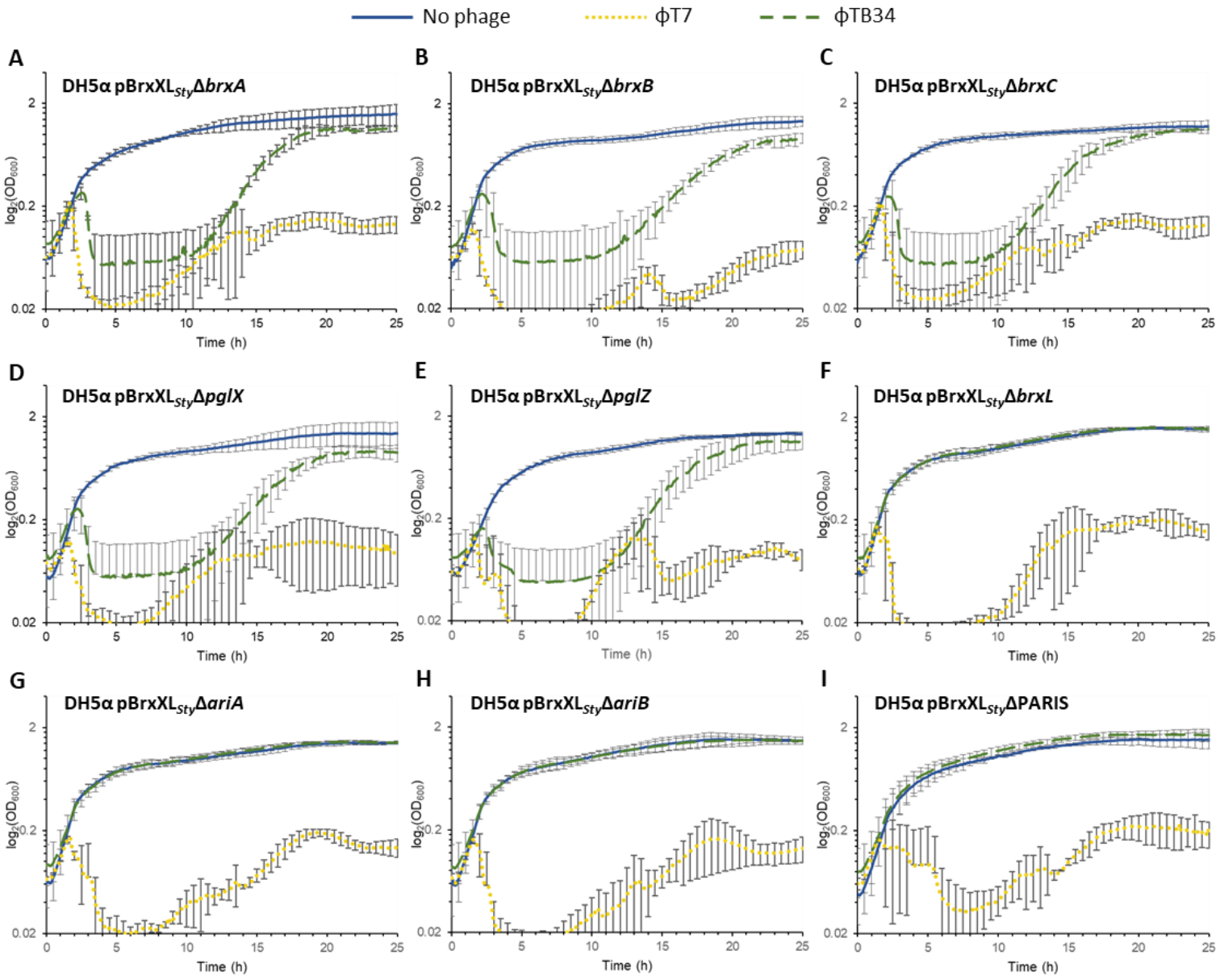

Figure 4.2 – Knock-out analysis of the S. Typhimurium D23580 BREX phage defence island. *(A-I) Growth of E. coli DH5α strains harbouring sub-cloned wild type and mutant BREX locus plasmids, in the absence or presence of phages TB34 and T7 at phage multiplicity of infection of 10^{-6}. Data are shown in triplicate and error bars represent the standard deviation. For comparison, DH5α, DH5α pTRB507 and DH5α pBrxXL$_{Sty}$ are available in Figure 3.6.*

Table 4.1 – EOPs against pBrxXL$_{Sty}$ and mutant derivatives.[1]

Construct	TB34	T7
DH5α pBrxXL$_{Sty}$ WT	$2.86 \times 10^{-2} \pm 1.54 \times 10^{-2}$	1.36 ± 0.30
DH5α pBrxXL$_{Sty}$-Δ*brxA*	0.47 ± 0.30	0.73 ± 0.08
DH5α pBrxXL$_{Sty}$-Δ*brxB*	0.44 ± 0.39	0.83 ± 0.42
DH5α pBrxXL$_{Sty}$-Δ*brxC*	0.49 ± 0.37	1.06 ± 0.20
DH5α pBrxXL$_{Sty}$-Δ*pglX*	0.57 ± 0.39	0.70 ± 0.23
DH5α pBrxXL$_{Sty}$-Δ*pglZ*	0.28 ± 0.13	0.56 ± 0.18
DH5α pBrxXL$_{Sty}$-Δ*brxL*	$7.50 \times 10^{-6} \pm 3.46 \times 10^{-6}$	0.94 ± 0.45
DH5α pBrxXL$_{Sty}$-Δ*ariA*	$3.65 \times 10^{-3} \pm 2.59 \times 10^{-3}$	0.91 ± 0.25
DH5α pBrxXL$_{Sty}$-Δ*ariB*	$8.15 \times 10^{-3} \pm 3.01 \times 10^{-3}$	0.87 ± 0.29
DH5α pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB*	$1.06 \times 10^{-2} \pm 2.96 \times 10^{-3}$	0.76 ± 0.50

[1]Green indicates EOPs ~1, yellow indicates EOPs ~0.01-0.1, orange indicates EOPs ~0.001-0.01, red indicates EOPs <1 x 10^{-4}.

No previous study has reported BREX restriction in the absence of *brxL* [281,314]. To determine whether this phenotype is specific to defence against TB34, the EOP assay analysis was extended to the entire Durham Phage Collection (Table 4.2). To determine whether the increased resistance phenotype from PARIS knockouts was also consistent across different phage species, the pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* construct was included in this analysis. The EOP data for the pBrxXL$_{Sty}$ generated in Chapter 3 is also included for comparison.

The reduction in EOP seen for pBrxXL$_{Sty}$-Δ*brxL* against TB34 was not exceeded (Table 4.2). For the phages susceptible to *Salmonella* BREX defence, most show further reductions in EOP when assayed against pBrxXL$_{Sty}$-Δ*brxL*, ranging from around a 2-fold reduction from CS16 to a reduction of around three orders of magnitude for Sipho (Table 4.2). Conversely, phage BB1 demonstrated an increase in plaquing on this strain, though plaque assays for this phage were highly variable and the pinprick plaques were otherwise difficult to quantify. Phages Paula and PATM had previously demonstrated marginal reductions in EOP when infecting the pBrxXL$_{Sty}$ strain but produced substantial reductions in EOP when infecting pBrxXL$_{Sty}$-Δ*brxL*.

***Table 4.2 – EOP of the Durham Phage Collection against* Salmonella *BREX and selected knockouts.*[1]**

Phage	*Salmonella* BREX sites	pBrxXL$_{Sty}$	pBrxXL$_{Sty}$-*ΔbrxL*	pBrxXL$_{Sty}$-*ΔariAΔariB*
TB34	120	2.86 x10^{-2} ± 1.54 x10^{-2}	7.50 x10^{-6} ± 3.46 x10^{-6}	1.06 x10^{-2} ± 2.96 x10^{-3}
Trib	0	1.19 ± 0.11	0.70 ± 0.27	0.94 ± 0.23
BHP	0	1.02 ± 5.89 x10^{-2}	3.16 ± 3.61	0.17 ± 0.20
Baz	0	1.35 ± 0.46	1.34 ± 0.81	2.08 ± 1.67
Alma	110	2.64 x10^{-2} ± 3.21 x10^{-3}	2.7 x10^{-4} ± 1.61 x10^{-4}	6.35 x 10^{-3} ± 4.39 x10^{-3}
Pau	84	0.19 ± 0.13	3.73 x10^{-4} ± 2.65 x 10^{-4}	0.14 ± 0.27
PATM	83	0.23 ± 5.07 x10^{-2}	3.27 x10^{-3} ± 6.45 x10^{-4}	9.64 x10^{-2} ± 2.57 x10^{-3}
BB1	51	0.11 ± 5.25 x10^{-2}	0.27 ± 0.91	4.29 x10^{-8} ± 1.61 x10^{-8}
Jura	0	2.58 ± 0.11	1.95 ± 1.37	1.82 ± 0.53
SPSP	0	2.01 ± 5.12 x10^{-1}	0.99 ± 0.13	1.03 ± 0.24
AL25	0	1.90 ± 0.51	0.79 ± 0.14	1.26 ± 0.22
Titus	1	1.30 ± 0.26	1.69 ± 0.78	0.78 ± 0.19
Mak	1	0.78 ± 0.10	3.77 ± 2.90	0.67 ± 0.12
Bam	1	1.06 ± 0.60	1.71 ± 0.69	0.59 ± 0.16
CS16	13	3.25 x10^{-2} ± 2.30 x10^{-2}	1.45 x10^{-2} ± 2.39 x10^{-3}	6.17 x10^{-2} ± 1.86 x 10^{-2}
Mav	13	0.15 ± 2.16 x10^{-2}	5.82 x10^{-2} ± 4.08 x10^{-2}	9.4 x10^{-2} ± 9.89 x10^{-2}
Sipho	84	3.74 x10^{-2} ± 1.19 x10^{-2}	2.19 x10^{-5} ± 2.57 x10^{-4}	5.81 x10^{-3} ± 2.06 x10^{-2}

[1]Green indicates EOPs >0.1, yellow indicates EOPs 10^{-2}-10^{-1}, orange indicates EOPs 10^{-3}-10^{-2}, pink indicates EOPs 10^{-3}-10^{-4} red indicates EOPs <1 x 10^{-4}.

For EOP assays against the PARIS knockout strain, pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB*, slight reductions in EOP are observed for phages Alma, CS16 and Mav, in line with those seen for TB34. Phage BB1 had a reduction in EOP of around seven orders of magnitude compared to pBrxXL$_{Sty}$ strains (Table 4.2). The discrepancy between the results for Pau and PATM is interesting: the apparent amplification of phage defence provided by the pBrxXL$_{Sty}$-Δ*brxL* strain suggested that both phages were susceptible to *Salmonella* BREX, however, while the EOP on pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* by PATM is reduced, the EOP for Paula is not. At no point was there a significant increase in EOP from pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* compared to pBrxXL$_{Sty}$, suggesting that the PARIS system is not active against any of our selected phages.

4.4 Nanopore sequencing determines the BREX genes required for host methylation.

Using the same constructs, BREX genes essential for host methylation were investigated. Initially, the methylation deficient *E. coli* 2796 was selected as a host strain as this would provide a clean, non-methylated background [365]. Previously, our lab had utilised PacBio sequencing for methylation profiling when studying the pEFER BREX system [183]. To investigate alternative avenues for sequencing, we utilised nanopore sequencing as a potential in-house sequencing method that would reduce lead times and costs [363,364]. As such, it was first necessary to develop an applicable library prep and bioinformatic pipeline for methylation analysis. Initial trials and optimisation of this process were carried out on pEFER BREX genomic DNA, as this was already confirmed as having N6mA methylation at the pEFER BREX motif [183]. Full details of the library prep and bioinformatic pipeline used for this study are detailed in Chapter 2. For all nanopore sequencing, extracted genomes were barcoded for parallel sequencing on a MinIon Mk1C sequencer (Oxford Nanopore).

For basecalling of modified DNA, several options are available [366]. In this study, a combination of three packages were utilised, allowing cross-referencing of results between tools and providing complementary data. Tombo allows *de novo* identification of modified motifs from sequencing data and was one of the first tools developed for basecalling of modified DNA [367]. nanodisco is a relatively newly available tool designed for basecalling of modified DNA in bacteria [367]. Further, nanodisco allows *de novo* detection of modified motifs and can specify which individual base within that motif is modified. Megalodon was

developed by Oxford Nanopore for modified basecalling and can provide per-base methylation data (https://github.com/nanoporetech/megalodon), and it performs well in benchmarking against other tools [366]. Thus, three complementary tools are utilised; Tombo and nanodisco can both identify the modified motif, nanodisco can specify which base within the motif is modified and megalodon can provide information of what percentage of sites within the genome are modified.

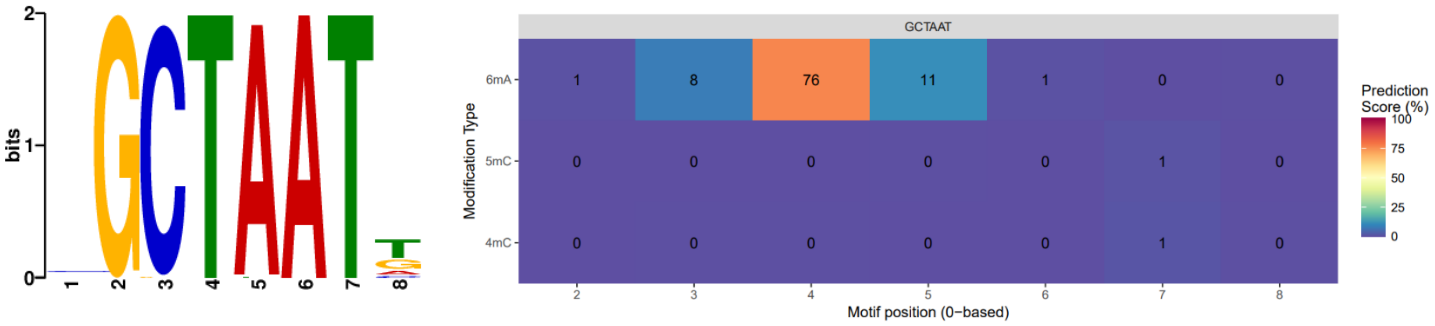

Figure 4.3 – Tombo (left) and nanodisco (right) show that pEFER BREX methylates the fifth adenine residue of the GCTAAT motif.

In order to establish and benchmark our bioinformatic pipeline, analysis began with processing of pEFER BREX data from sequencing of ER2796 pBrxXL$_{Efer}$ genomic DNA. Both Tombo and Nanodisco correctly identified the GCTAAT motif as being modified within the host genome, with Nanodisco further identifying the fifth adenine base as being modified and that the modification type was N6mA (Figure 4.3), as expected from a BREX system. Megalodon indicated that 92% of the 1699 pEFER BREX motif sites were methylated.

With the bioinformatic pipeline established for pEFER BREX, the ER2796 pBrxXL$_{Sty}$ data was processed in the same way. Throughout preparation of the deletion strains and the knockout derivatives using the ER2796 background, poor growth and transformation efficiencies were a common theme. Upon processing the ER2796 pBrxXL$_{Sty}$ dataset, no methylated motifs could be identified by any of the three detection packages. As all functional data had been carried out in DH5α and *Salmonella* BREX was thus known to be active in this context, and strains in this host were transformed and grown more efficiently, we attempted sequencing and analysis in DH5α instead. DH5α contains native Dam methylation at GATC motifs [368]. As there was a risk that Dam methylation would interfere with correct calling of modified *Salmonella* BREX motifs, predicted to be GATCAG [315], an additional whole genome amplification (WGA) sample was

included as a negative control as such a sample should be inherently unmodified. Sequencing of the DH5α pBrxXL$_{Sty}$ genome allowed identification of the GATCAG motif as modified by both Tombo and nanodisco, with the latter indicating that the 5th adenine residue had N6mA modification (Figure 4.4). Megalodon then confirmed that 78.8% of the genomic motif sites were methylated (Table 4.3). The methylation level of the WGA sample by comparison was 12.87% indicating a high level of noise (Table 4.3).

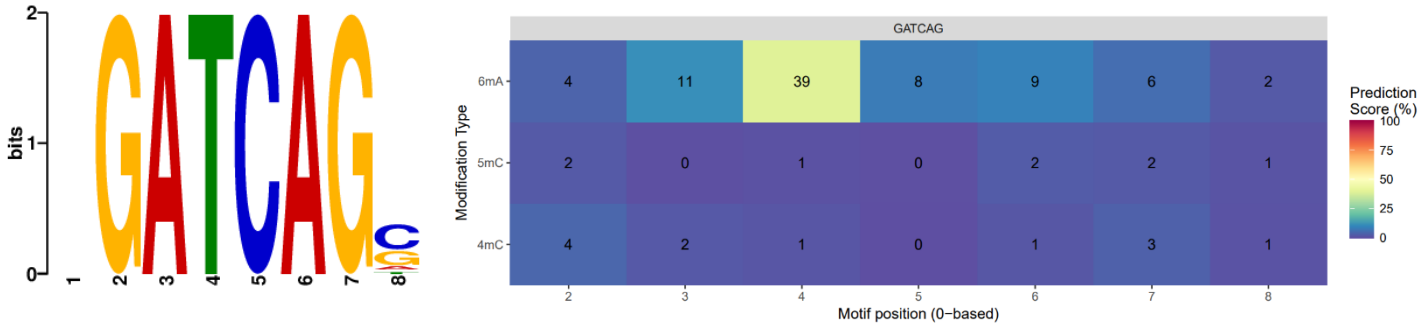

Figure 4.4 – Tombo (left) and nanodisco (right) show that* Salmonella *D23580 BREX methylates the fifth adenine residue of the GATCAG motif.

With the sequencing and bioinformatic pipeline established and DH5α pBrxXL$_{Sty}$ genome methylation confirmed at GATCAG motifs, the analysis was rolled out to the pBrxXL$_{Sty}$ knockout mutants (Table 4.3). The DH5α p507 empty vector strain included as a negative control displayed genome methylation levels of 41.83%, suggesting that the Dam methylation at GATC sites was indeed interfering with modified basecalling and validating the decision to include a WGA control sample. As with restriction, *brxA*, *brxB*, *brxC*, *pglZ* and *pglX* are required for host methylation (Table 4.3). *brxL* meanwhile is not required for methylation, as demonstrated previously in both the *Acinetobacter* and *E. coli* BREX systems [281,314]. As expected, removal of either or both of the PARIS genes did not interfere with host methylation. As this pipeline was developed as an in-house method for potential use in future studies, samples from the same strains used for nanopore sequencing were also sent for sequencing and analysis by PacBio to ensure accuracy. PacBio results were more definitive than nanopore sequencing, with both the WGA and DH5α p507 strains showing no sign of methylation at GATCAG sites, indicating a much lower level of background noise and higher specificity at differentiating between BREX and Dam methylation sites. Knockouts of *brxA*, *brxB*, *brxC*, *pglZ* and *pglX* also showed no methylation at GATCAG sites, while pBrxXL$_{Sty}$, pBrxXL$_{Sty}$-Δ*brxL*, pBrxXL$_{Sty}$-Δ*ariA*, pBrxXL$_{Sty}$-Δ*ariB* and pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* strains showed near 100% motif methylation.

Together, these results show that nanopore sequencing and analysis of pBrxXL$_{Sty}$ strains is sufficient for determining whether a strain is methylated or not, but that PacBio is more suitable for quantitative analysis of methylation levels.

Table 4.3 - Detection of GATCAG N6mA motifs made by pBrxXL$_{Sty}$ and knockout mutants.[1]

Genome Sample	Nanopore Average methylation (%)	PacBio Average methylation (%)
pBrxXL$_{Sty}$ WGA[2]	12.87	0
Empty vector p507	41.83	0
pBrxXL$_{Sty}$	78.78	97.8
pBrxXL$_{Sty}$-Δ*brxA*	39.86	0
pBrxXL$_{Sty}$-Δ*brxB*	44.39	0
pBrxXL$_{Sty}$-Δ*brxC*	41.01	0
pBrxXL$_{Sty}$-Δ*pglX*	47.15	0
pBrxXL$_{Sty}$-Δ*pglZ*	49.09	0
pBrxXL$_{Sty}$-Δ*brxL*	87.54	100
pBrxXL$_{Sty}$-Δ*ariA*	68.45	99.8
pBrxXL$_{Sty}$-Δ*ariB*	87.30	99.8
pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB*	86.87	97.8

[1]Green indicates >95% methylation, light green indicates 75-95% methylation, yellow indicates 50-75% methylation, orange indicates 25-50% methylation, pink indicates 1-25% methylation, and red indicates 0% methylation.

[2]WGA, whole genome amplification.

4.5 BREX activity is blocked by Ocr and Gp5, but neither activate PARIS phage defence.

Phages can encode specific mechanisms for inhibiting phage defence systems [150,334]. Based on EOP results, it is likely that several of the phages in the Durham Phage Collection encode mechanisms for either avoiding or inhibiting *Salmonella* D23580 BREX function (Figure 3.7). The T7 encoded restriction system inhibitor, Ocr, has been shown to inhibit phage defence activity of the *E. coli* BREX system [294]. Additionally, this molecule is reported to trigger Abi by the type II PARIS phage defence system [243], a homologue of which is present in the *Salmonella* BREX operon, yet no activity has been observed against phage T7, which naturally encodes Ocr. Following the production of individual gene knockouts, it was now possible to individually assay inhibition of BREX and activation of PARIS by Ocr. To determine whether Ocr inhibits BREX, vector pBAD30-*ocr* was generated that allows arabinose-inducible production of Ocr. Plaque assays were then carried out with DH5α pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* pBAD30-*ocr* (Table 4.4). EOP assays were carried out as previously, except *ocr* was induced with 0.2% arabinose 30 minutes prior to infection and 0.2% arabinose was included in both top and bottom agar layers. Expression of Ocr fully inhibited BREX defence (Table 4.4). As Ocr is a product of T7, a coliphage, this experiment was also repeated using an Ocr homologue, Gp5, encoded by *Salmonella* phage Sp6 [369]. Homology was inferred by protein sequence searches using BLAST (NP_853565.1: 78.6% sequence similarity, 88% coverage) followed by predictive modelling from protein sequence using AlphaFold. Structures of Ocr and Gp5 aligned with an root mean squared deviation (RMSD) of 0.91 Å (Figure 4.5). We again chose TB34 as a model phage. Results showed that Gp5 also fully inhibited phage defence from pBrxXL$_{Sty}$ (Table 4.4).

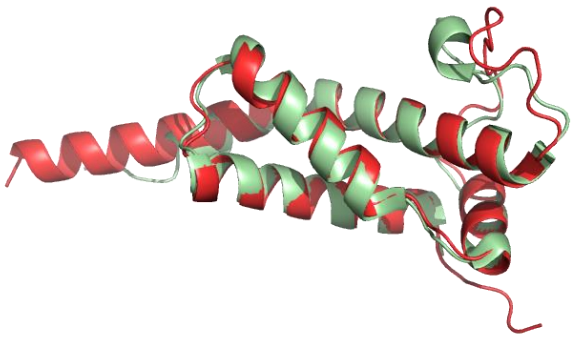

Figure 4.5 – The* Salmonella *phage, Sp6, encodes a structural homologue of Ocr. *Ocr (1S7Z) is shown in pale green, the Sp6 homologue, Gp5, is shown in red. Structures align with an RMSD of 0.91 Å. The Ocr structure was downloaded from PDB, the Gp5 structure was predicted using AlphaFold.*

As this method demonstrated inhibition of BREX by overexpression of the inhibitors Ocr and Gp5, it was postulated that the same induction system may elicit phage defence from the PARIS system. This time, the pBrxXL$_{Sty}$-Δ*pglX* strain was utilised for co-expression of Ocr as this strain is deficient for BREX phage defence. Since phage inhibition of defence systems can sometimes be specific to systems from a certain strain, this experiment was again repeated with the *Salmonella* phage Ocr homologue, Gp5 [369]. EOP assays were carried out in the same manner, but no defence activity was observed against TB34 (Table 4.4), even when altering incubation time post-induction between 0 and 60 minutes before infection.

Table 4.4 – Ocr and Gp5 inhibit BREX but do not trigger phage defence from PARIS.

Strain	EOP
pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB*	$1.06 \times 10^{-2} \pm 2.96 \times 10^{-3}$
pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* pBAD30-*ocr*	1.281 ± 0.41
pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* pBAD30-*gp5*	1.146667 ± 0.33
pBrxXL$_{Sty}$-Δ*pglX* pBAD30-*ocr*	1.22 ± 0.8
pBrxXL$_{Sty}$-Δ*pglX* pBAD30-*gp5*	1.056667 ± 0.35

[1]Green indicates EOPs >0.1, yellow indicates EOPs 10^{-2}-10^{-1}

4.6 Discussion

4.6.1 Production of pBrxXL$_{Sty}$ knockout mutants

Genes essential for phage restriction and host methylation have shown some variation between type I BREX systems from different bacterial hosts [281,314]. The production of individual gene knockouts allowed direct comparison with previous studies [281,314]. Again, the size of the *Salmonella* BREX operon presented a considerable challenge, with the additional plasmid backbone sequence bringing the entire construct size to 17.9 kb. Use of traditional RM cut and paste cloning was not possible due to lack of appropriate cut sites and returning to modify the original GGA donor fragments would have been complex and in some cases left truncated BREX proteins. GA is more commonly used for inserting large regions of DNA rather than making deletions [323]. Nevertheless, utilisation of GA reaction mixes in combination with amplification of fragments of pBrxXL$_{Sty}$ allowed us to produce scarless, clean deletions. GA is often hailed as being sequence independent. In reality, GA relies heavily on the ability to design appropriate primers and is strongly inhibited by mispriming and secondary structure formation – likely the cause of difficulty in producing *brxA*, *brxB*, *pglZ*, *ariA* and *ariB* knockouts in this study. Thus, though a powerful molecular biology tool, alternative methods should be considered upon repeated cloning failure.

4.6.2 Functional analysis of pBrxXL$_{Sty}$ knockout mutants

Growth and infection curves for the deletion constructs presented an efficient method to determine which genes in the BREX operon are essential for phage restriction and host methylation. As with previous studies, *brxB*, *brxC*, *pglX* and *pglZ* are essential for both restriction and methylation [281,314]. *brxA* on the other hand, shown to be essential for restriction and methylation in *Salmonella* BREX (Figure 4.1 and Table 4.1, and Table 3, respectively), was shown to be dispensable for both in *E. coli* BREX [314] but essential for both in *Acinetobacter*[281]. It is interesting that *brxA* is predicted to relate to regulation of BREX, yet it is essential in *Acinetobacter* BREX, which encodes the BrxR regulatory protein, and dispensable in *E. coli* BREX, which contains no apparent transcriptional regulator proteins. Perhaps some trans-acting transcription factor is present in *E. coli* which substitutes for BrxA activity but is absent in *Salmonella* D23580. There

exists some conserved sequence within the region upstream of both the *Salmonella* and the *E. coli* BREX systems (Figure 4.6), suggesting that some form of promoter element might be encoded in this region. Analysis of transcription factor recognition sequences within this region was carried out using the online transcription factor search tool, SAPHIRE.CNN, which utilises convolutional neural networks to detect σ-factors using a model trained on *Salmonella enterica* transcriptomic datasets [370]. This implicated two regions in transcription factor recognition, both conserved between *Salmonella* D23580 and *E. coli*, one of which was within the -35 σ-70 transcription factor binding region (denoted by orange arrows) and one further upstream. Identification of BREX regulation in these two strains would provide further insight into new regulation mechanisms of BREX system and potentially allow guilt-by-association based discovery of new phage defence systems, as systems often share common regulators [279–281].

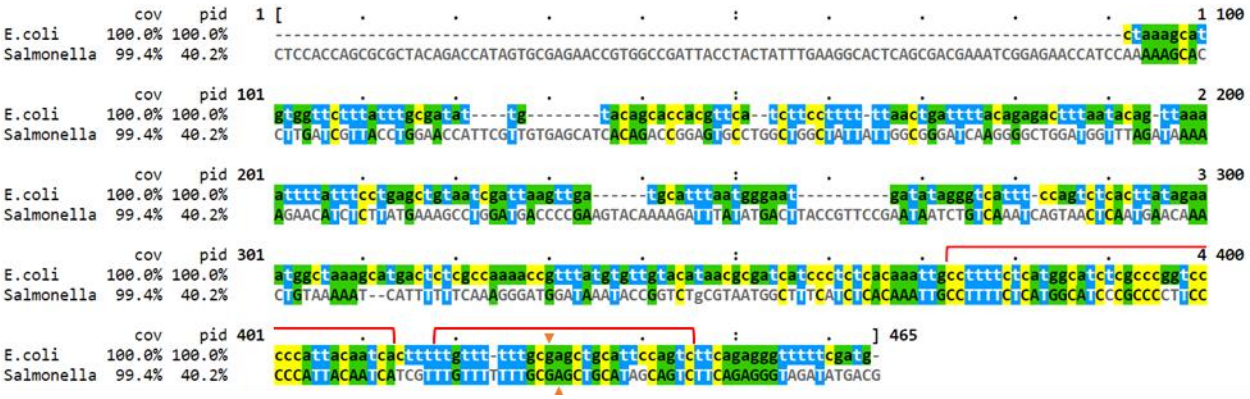

***Figure 4.6 – The 5' upstream region of BREX is conserved between* Salmonella *D23580 and* E. coli.** *Genetic sequence alignment of the regions upstream of the BREX systems from* E. coli *and* Salmonella *D23580. Alignments were generated using Clustal Omega. Regions pertaining to potential transcription factor recognition sequences are denotet by red lines, as predicted by the online σ-factor predictor tool, SAPHIRE.CNN. Orange triangles indicate -35 bases from the* brxA *transcription start site in respective sequences.*

BrxL was previously shown to be dispensable for host methylation [281,314], as was demonstrated in this study (Figure 4.1 and Table 4.1, and Table 3, respectively). On the other hand, BrxL was shown to be essential for phage defence in both *E. coli* and *Acinetobacter* BREX systems, in contrast to *Salmonella* BREX [281,314]. The persistence of phage defence in the absence of BrxL is puzzling. BrxL was recently shown to form a dimer of hexameric rings and forms a barrel-like structure which is able to bind to and translocate along DNA [318]. Further, others in our lab have demonstrated that it interacts directly with BrxB, BrxC, PglX and PglZ *in vitro* (unpublished data). Thus, BrxL appears to play a defined role in phage defence. The

EOP assay results for DH5α pBrxXL$_{Sty}$-Δ*brxL* strains only extenuate the apparent disparity, as EOP of phage TB34 increases by several orders of magnitude. This phenotype is consistent across our phage library for susceptible phages, though the scale of the reduction in EOP appears to be somewhat phage dependant. It is possible that BrxL in *Salmonella* somehow modulates or regulates activity of BREX defence. RM systems have been reported to be accompanied by restriction alleviation proteins which activate in times of stress, reducing restriction activity and increasing methylation activity – a phenotype characteristic of type I RM systems [371–373]. It is possible that BrxL plays some analogous role in BREX restriction and that the rate of restriction increases in its absence – but if that is the case, why is this phenotype not observed for *brxL* knockouts in *E. coli* or *Acinetobacter* BREX systems? Elsewhere, overexpression of a C-terminal fragment of BrxL has been shown to upregulate several genes elsewhere in the genome, including prophage genes [315]. Zaworski *et. al.* further postulate that the corresponding Lon-like domain in this BrxL fragment presents similarity to the Lon-related C-terminal domain of RadA, required for DNA branch migration in homologous recombination [374], and that BrxL may be implicated in inhibition of phage DNA replication at DNA forks, rather than restriction. This would be somewhat in keeping with the model of BrxL complexes translocating along DNA [318]. Deeper understanding of the key functions of BREX defence are required to understand this relationship and greater understanding of the role of BrxL may provide insight into the mechanism of phage restriction.

Phage defence activity from the PARIS knockouts, alongside absence of defence from the DH5α pBrxXL$_{Sty}$-Δ*pglX* construct, confirms that phage defence observed so far from this operon is from the BREX system. The slight increase in phage defence produced by the DH5α pBrxXL$_{Sty}$-Δ*ariA*, DH5α pBrxXL$_{Sty}$-Δ*ariB* and DH5α pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* strains was consistent in EOP assays against other phages in the Durham Phage Collection. It is possible that this increase is attributable to expression of the BREX genes. As discussed in Chapter 3, there are several promoter regions for the *Salmonella* BREX operon, including promoter regions within both *ariA* and *ariB*. Removal of additional promoters in this region may allow more efficient transcription and expression of the BREX system. BREX systems from both *E. coli* and pEFER encode two discreetly transcribed regions – *brxA* to *pglX* and *pglZ* to *brxL*, with only a single promoter for the second transcribed unit (chapter 3). It is possible that insertion of the PARIS system into the BREX operon partially disrupted expression of the system and that removal of the PARIS system and the additional promoters restored previous activity.

4.6.3 Establishing a methylation profiling pipeline

Establishing a methylation profiling pipeline for analysis of host genome methylation levels in conjunction with nanopore sequencing provided an in-house alternative to PacBio, potentially reducing lead times and providing greater flexibility in experimental design and analysis for future studies. The experiences of this study, however, show that this approach is not straightforward. Initial outlay on a sequencing device and the computational equipment for data processing, combined with costs of reagents and consumables required for process optimisation, and knowledge of using community created command line tools, presented a significant barrier to entry. Recent development of new flow cell chemistry and library preparation kits, however, are reported to improve efficiencies and quality of reads [375,376]. Further, ONT have recently discontinued support for the Tombo and Megalodon in favour of the new all-in-one Dorado basecalling package (https://github.com/nanoporetech/dorado). Though this also requires access to high power computing resources, it may contribute to increased modified basecalling accuracy.

Initial trials and optimisation on the pEFER BREX system provided an important positive control for establishing our workflow and analysis pipeline. Though performed using PacBio rather than nanopore sequencing, it has previously been shown that the host genome of ER2796 pBrxXL$_{Efer}$ strains carried N6mA modification at the fifth base of almost 100% of GCTAAT. Using the same strain, Tombo and nanodisco were able to replicate this result, though Megalodon suggested that fewer motif sites within the genome were methylated than PacBio (92% vs. 99%, respectively). This apparent underestimation of motif site modification continued throughout methylation profiling. Application of this same pipeline to ER2796 pBrxXL$_{Sty}$ strains was less successful. None of the three tools used were able to identify methylation and poor transformation efficiencies and growth suggested toxicity and made initial sample preparation problematic. It would seem that the lack of methylation in *E. coli* ER2796 incurs some sort of cost when expressing *Salmonella* BREX but not pEFER BREX, with *Salmonella* BREX perhaps targeting the host chromosome before full methylation, or some other off-target activation at other sites which would otherwise be protected. The fact that toxicity is not observed in DH5α is likely attributable to the similarity between the *Salmonella* BREX motif sites (GATCAG) and native Dam methylation motifs (GATC). It would be interesting to see whether pBrxXL$_{Sty}$ is similarly toxic in a DH5α *dam*$^{-}$ strain, or whether the *E. coli* BREX system is similarly toxic in ER2796.

Overall, the sequencing and bioinformatic pipeline established here proved sufficient for identification of host methylation from both *Salmonella* and pEFER BREX systems in DH5α, and identification of methylation by the pEFER BREX system in ER2796. The choices of Tombo, nanodisco and Megalodon were designed to both provide redundancy and also complementary analysis – i.e., recognition motif, position of methylation and total genomic methylation level. The high background signal in the WGA sample and the underestimation of methylation by Megalodon compared to PacBio would however indicate that this method is ill-suited for quantitative comparisons. Further, the high background methylation present in the p507 plasmid control suggests that these tools have difficulty distinguishing between modifications in close proximity. It is important to consider however that both Tombo and nanodisco identify modified bases by comparison of signal to signal from a negative control sample. In this case, the WGA sample allowed accurate identification of the correct modified motif, while results using the DH5α p507 data as a control group was more ambiguous. Megalodon, however, requires no control sample for comparison, instead, providing modified base statistics at a per-base level. As such, while the Tombo and nanodisco results were less affected by the native Dam methylation due to the WGA, presence of GATC site methylation may still have inhibited modified basecalling by Megalodon. Together, these data demonstrate the importance of redundancy in DNA modification analysis, with separate sequencing and analysis methods providing confidence in accuracy of conclusions. This approach will be further utilised in Chapter 6 when screening for methylation by PglX mutants.

4.6.4 Using Ocr and Gp5 to inhibit BREX and probe for phage defence by PARIS

Ocr is a DNA mimic that has been reported to inhibit a range of phage defence systems [286,334], including *E. coli* BREX [294]. Ocr forms a dimer that mimics 20 base pairs of DNA and binds non-specifically to DNA binding proteins [286,291,325,377]. Inhibition of *Salmonella* BREX is unsurprising as Ocr functions through competitive inhibition and is an example of a non-specific defence system inhibitor utilised by phages. The close *Salmonella* phage homologue, Gp5, is predicted to dimerise and function in the same manner, and similarly inhibits phage defence from DH5α pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* (Table 4.4). As the assay for this inhibition is based on overexpression of Ocr or Gp5 rather than native expression, this inhibition is likely not representative of the dynamics of BREX inhibition in phage infection, which would require subcloning either of these genes into a susceptible phage. Use of phage T7, which natively encodes Ocr, is not possible for DH5α pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* as T7 lacks *Salmonella* BREX recognition motifs.

Though no apparent activity has been demonstrated by the type II PARIS system within the BREX operon so far, it was postulated that PARIS defence may be provoked by exposure to overexpressed Ocr, which was previously shown to activate type II PARIS phage defence in *E. coli* B185. As phage inhibition of defence systems can be as specific as defence system activity, the Gp5 protein was also assayed [150,334]. Neither overexpressed inhibitor produced defence against TB34, however. The PARIS system within the BREX operon is likely activated by some other phage-borne molecular trigger, or the Abi mechanism is not active in *E. coli*. The latter hypothesis can be tested by performing further EOP assays against phage SP6 in *Salmonella* D23580, or other environmental phages resistant to *Salmonella* BREX activity and encoding a homologue of Ocr. Beyond this, EOP assays against a wider range of coliphages is necessary to identify a trigger. It seems unlikely that this PARIS system is inactive as it is somewhat conserved in *Salmonella* genomes [315].

4.6.5 Conclusion

The creation of individual gene knockouts in pBrxXL$_{Sty}$ has allowed identification of the genes essential for both phage restriction and host methylation in the *Salmonella* BREX system. Essential genes differ between the *Salmonella*, *E. coli* and pEFER BREX systems [281,314]: BrxA was essential for methylation and restriction in both *Acinetobacter* and *Salmonella* systems but not in *E. coli*; BrxL was essential for both host modification and phage restriction in *Acinetobacter* BREX, essential for only restriction in *E. coli* BREX but dispensable for both restriction and host modification in *Salmonella* BREX. BrxA is predicted to function in some regulatory capacity and the variation in essentiality likely relates to differing regimes of regulation between BREX systems. BrxL was recently shown to form large complexes which are able to bind to and translocate along DNA [318] and has long been predicted to be an effector protein in BREX activity [308,314]. EOP analysis demonstrated that deleting *brxL* further reduced the EOP of BREX susceptible phages in the Durham Phage Collection by varying degrees, often by multiple orders of magnitude. The cause of this phenotype is puzzling, and two hypotheses are suggested. First, BrxL may play a role analogous to restriction alleviation proteins [372,373] in which BrxL functions to mitigate the toxic effects and growth costs of BREX and that deleting *brxL* effectively removes this limiter. Second, deletion of the C-terminal of *brxL* in *Salmonella* Typhimurium LT2 was shown to upregulate gene expression across the host genome, particularly in prophage regions [315]. Thus, deletion of *brxL* may be inadvertently upregulating factors which influence phage infection, and potentially some other phage defence elements within the *Salmonella*

D23580 genome (Table 3.1) as phage defence elements are commonly encoded on prophage regions [243]. Understanding this phenotype can only be attained with a greater knowledge of the mechanisms of BREX phage defence and detailed investigation of DH5α pBrxXL$_{Sty}$-Δ*brxL* strains.

Deletion of the PARIS genes genes from pBrxXL$_{Sty}$ did not result in loss of phage defence activity against any of the phages tested in from the Durham Phage Collection and deletion of *pglX* abolished all phage defence activity, together demonstrating that no phage defence is being conferred by PARIS against these phages. Interestingly, no phage defence activity was demonstrated by the PARIS system against phage T7, which was reported to activate PARIS-mediated Abi in *E. coli* B185 [243]. It is possible that the BREX associated system in *Salmonella* D23580 is activated by some factor not encoded by these phages, is not active in *E. coli*, or is redundant and its activity has been replaced by the other PARIS system in the *Salmonella* D23580 genome (Table 3.1). Attempts to elicit activity by overexpression of the phage encoded RM inhibitor, Ocr, reported to trigger PARIS activity in *E. coli* B185, were similarly unsuccessful. Further attempts at triggering activity from the PARIS system will need to be extended to modified phages in the Durham Phage Collection and to testing for phage defence against *Salmonella* phages in a native *Salmonella* host strain. Ocr had previously been shown to inhibit activity of *E. coli* BREX [294] and was similarly effective at inhibiting *Salmonella* BREX function here, as was the *Salmonella* phage encoded homologue, Gp5. These data demonstrate that phages encode conserved mechanisms for broad spectrum phage defence inhibition.

Chapter 5 – Structural investigation of core BREX proteins

5.1 Introduction

So far, work in this study has provided bioinformatic analysis of the BREX system from *Salmonella* D23580, characterised phage defence against the Durham Phage Collection for direct comparison to phage defence conferred by the *E. coli* and pEFER BREX systems and explored which individual components of the *Salmonella* BREX system are essential for host methylation and phage restriction. In order to fully understand the functional mechanisms of BREX, greater molecular detail is required for both individual components and complexes. Towards this, Chapter 5 will outline attempts to produce structural models of PglZ and PglX using x-ray crystallography. In addition, initial *in vitro* biochemical and biophysical characterisation of PglX was performed.

PglZ, as part of the Pgl system, was initially used to discover the wider BREX phage defence family [308,313] and is the only component common to all BREX subtypes [308]. Recently, the structure of the PglZ homologue, PorX, was described and phosphodiesterase activity was observed [317]. Subsequently, phosphodiesterase activity was demonstrated from both *Salmonella* and pEFER PglX *in vitro* by other members of our lab (Unpublished data). Obviously then, the phosphodiesterase activity of PglZ is essential to BREX function as PglZ has been demonstrated to be essential to both host methylation and host restriction here and in the *E. coli* and *Acinetobacter* BREX systems [281,314]. This marks PglZ as a potentially valuable target for structural investigations. Another hallmark of BREX systems is host DNA methylation at the fifth adenine residue of non-palindromic six base pair motif [314]. This is carried out by PglX and is common to all BREX system apart from the type IV subtype [308]. PglX has also been shown to be essential to phage restriction in addition to DNA modification and is thus also intrinsic to BREX function [308,314,378]. A structural model of PglX in complex with DNA would provide valuable insight into both motif recognition and methylation. This would be of particular use in this study as one of the aims was to rationally mutate residues involved in motif recognition. Conversely, a structural model of PglX in complex with Ocr would provide insight into the mechanisms by which Ocr allows phages to escape BREX defence, as demonstrated in Chapter 4.

5.2 Large scale protein expression and purification

Cloning work carried out in Chapter 3 produced constructs for the T7-driven overexpression of tagged BREX proteins in the pSAT1 plasmid backbone. These constructs will allow biochemical and structural investigation into BREX system components. Our initial small scale (25 ml) trial expression showed expression from all six BREX genes. This study aimed to further characterise the methyltransferase, PglX, biochemically, biophysically and structurally. In addition, work towards producing a structural model of the phosphodiesterase, PglZ, was also carried out. In order to perform these analyses, sufficient recombinant protein yields are required at purities appropriate for crystallographic screening.

5.2.1 Optimisation of PglX protein expression and purification

Initial trial expression of PglX is shown in Chapter 3 and details of trial expression protocols are available in Chapter 2. Scale up in the pSAT1 expression system (12 L) demonstrated relatively linear scale up in product yields, though expression was sporadic. A detailed large scale expression protocol is available in Chapter 2. The pSAT1 overexpression plasmid produces cleavable His-Sumo tagged fusion protein, allowing for multiple stages of purification. Optimisation of each of these stages was carried out in an attempt to produce sufficient quantity of protein at sufficient purity for crystallography. PglX was bound to Ni-NTA resin and elution was carried out first using a full gradient of imidazole (10 mM to 250 mM), then, once binding affinity was established, using isocratic elution steps (50 mM, 90 mM, 150 mM and 250 mM), allowing removal of many nickel-binding contaminants. Next, 90 mM and 150 mM imidazole elutions were subjected to anion exchange chromatography, allowing separation by protein charge, and this step was optimised through NaCl gradient elution (100 mM to 1000 mM). The tagged protein eluted from anion exchange was then cleaved with Senp protease, removing the 6x Histidine-SUMO tag, leaving untagged PglX. This allowed for a second purification step with a nickel column; this time removing remaining nickel binding contaminants and the tagged protease. Finally, PglX was purified using (SEC). This purification protocol provided high yields of PglX but not at sufficient purity for crystallisation trials (Figure 5.1A). As PglX is predicted to bind DNA, a heparin column was introduced into the workflow in place of anion exchange. This improved purity but again, did not produce PglX samples clean enough for crystallography (Figure 5.1B).

In an attempt to improve purity further, a streptavidin tag was added to the N-terminal of the tagged PglX construct by GA – using the same methods used in Chapter 4 and described in greater detail in Chapter 2 – resulting in a pSAT1-His-Strep-Sumo-*pglx* construct and allowing an additional affinity purification step. Again, PglX purity was markedly improved after just two affinity steps, but a distinct contaminant band remained at around 100 kDa (Figure 5.1C). Returning to previous samples, this band also seemed to co-purify with PglX on nickel, heparin, anion exchange and SEC purification steps, suggesting that the contaminant was an N-terminal fragment of PglX, containing both affinity tags, retaining DNA binding ability and presenting similar physio biochemical properties to the full-length protein. Inclusion of proteases in purification buffers did nothing to alleviate the co-purification of the N-terminal PglX fragment. Thus, PglX appeared to be either cleaved *in vivo*, or was being partially transcribed or translated. To determine whether the latter was causing truncation, the His-Strep-Sumo-*pglx* construct was subcloned into the pBAD30 overexpression plasmid by LIC as previously described for pSAT-LIC – resulting in a pBAD30-His-Strep-Sumo-*pglx* construct. Protein expression from pBAD30 is controlled by the araBAD promoter, allowing induction with L-arabinose. Expression from araBAD promoters utilise native transcription machinery as opposed to T7 promoters in which transcription is performed by the powerful T7 polymerase. As such, transcription occurs at a slower rate, allowing more efficient coupling with protein translation [379]. Indeed, expression and purification from pBAD30-His-Strep-Sumo-*pglx* displayed less contamination and also prevented the sporadic expression observed from pSAT1 constructs. This allowed the omission of the streptavidin column, which provided weak binding to tagged PglX and resulted in a substantial loss of product. The final purification workflow devised for the purification of PglX is shown in Figure 5.1D. Protein was either used immediately for screening of crystallisation conditions or dialysed into storage buffer (20 mM Tris pH 7.9, 150 mM NaCl, 70% glycerol), snap frozen in liquid nitrogen and stored at -80 °C for future use. Yields from a standard 12 L PglX expression typically ranged between 2 – 10 mg.

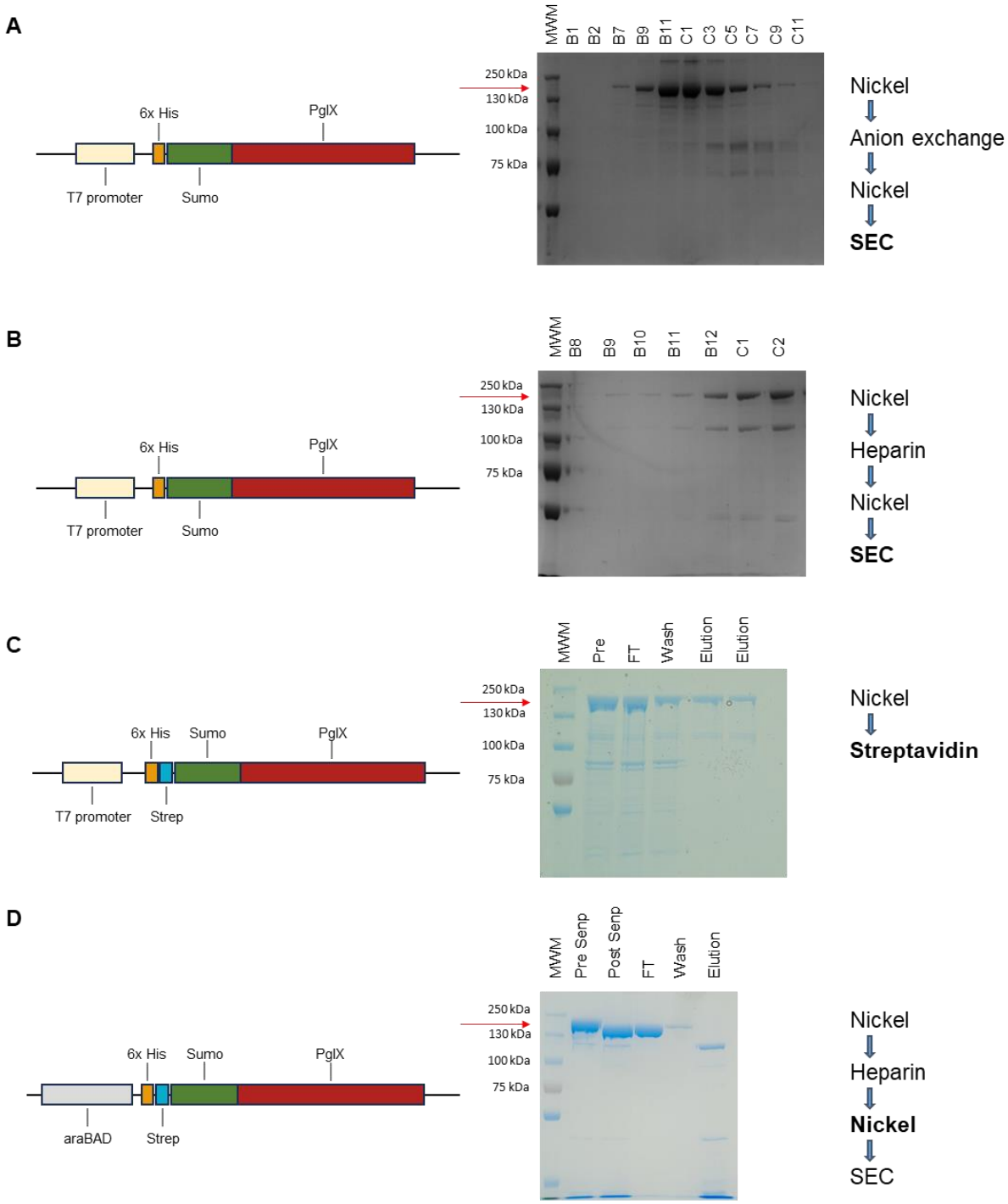

Figure 5.1 – Protein purification optimisation for PglX. *Different protein expression and tag systems are indicated on the left of each panel. Representative final purified product from each system is indicated by SDS-PAGE gels in the centre and the purification workflow for each system is shown on the right of each panel. Bold text in the purification workflow indicates the purification stage shown by the gel. The protein bands corresponding to PglX are denoted by a red arrow. For panels A and B, gel labels indicate selected fractions from size exclusion chromatography. For panel C, Pre indicates sample before application to column. For panels C and D, FT indicates column flow through. For panel C, elution from streptavadin resin was using 2.5 mM D-desthiobiotin. For panel D, elution from nickel resin was using 250 mM imidazole.*

5.2.2 Optimisation of PglZ protein expression and purification

Trial expressions had demonstrated expression of PglZ from the pSAT1-*pglZ* overexpression vector (Chapter 3). Scale up of PglZ expression was carried out in the same manner as for PglX. Purification utilised the same tag system and purification workflow as described above for initial protein expression from the pSAT1-*pglX* plasmid and is described in detail in Chapter 2. Tagged PglZ displayed a weaker affinity for the nickel purification resin relative to that seen for PglX. Following subsequent anion exchange, tag cleavage and second nickel column, sample displayed a high purity sufficient for crystallography (Figure 5.2). Final application to SEC however, resulted in fragmentation of PglZ. As such, the protocol was repeated for a fresh protein expression and the SEC step was omitted. Protein was either used immediately for screening of crystallisation conditions or dialysed into storage buffer, snap frozen in liquid nitrogen and stored at -80 °C for future use.

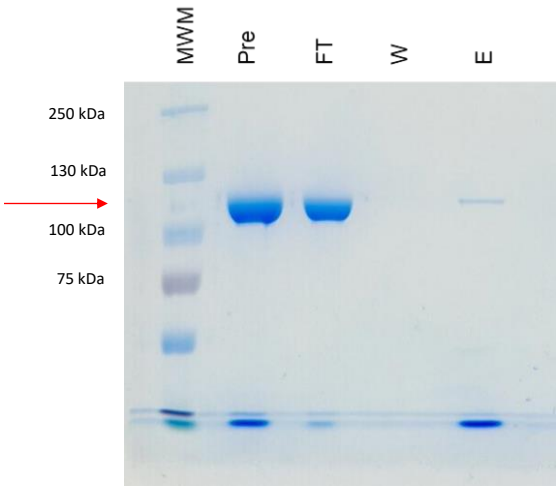

Figure 5.2 – Purified recombinantly expressed PglZ. *PglZ was recombinantly expressed from pSAT1-*pglZ*. Gel displays fractions from a second nickel affinity purification after tag cleavage. Pre, before application to column; FT, flow through; W, wash; E, 250 mM imidazole elution. The protein bands corresponding to PglZ is denoted by a red arrow.*

5.3 Biochemical and biophysical characterisation of PglX

5.3.1 PglX shows no methyltransferase activity *in vitro*.

With expression and purification methods established, biophysical and biochemical characterisation of PglX could be pursued. PglX is predicted to interact with DNA and produce SAM-dependant N6mA methylation at non-palindromic GATCAG motifs [314]. First, biochemical function of PglX was probed. A commercial SAM-dependant methyltransferase assay (Promega) was used to assess the ability of purified PglX to methylate DNA *in vitro* in a 96-well plate format. Full details of the methyltransferase assay kit and protocol can be found in Chapter 2. Using DH5α genomic DNA, known to contain the target BREX motif, as a substrate, PglX was incubated for 30 minutes at room temperature in a buffer containing the S-Adenosyl-L-methionine (SAM) methyl group donor, with methyltransferase activity measured indirectly via the reaction product, S-Adenosyl-L-homocysteine (SAH). Following incubation and addition of SAH detection agents, Luminescence was measured on a Biotek Synergy 2 plate reader. No methylation was apparent from PglX in this format (Figure 5.3). As BREX activity likely requires the formation of protein complexes, it was hypothesised that PglX may require the presence of other BREX components to produce methyltransferase function. As shown by other members of our lab (Unpublished data), PglX interacts with both PglZ and BrxB in pulldown assays. As such, the methyltransferase assay was repeated, this time with the inclusion of equimolar amounts of purified BrxB and/or PglZ. Using the same protocol, no methylation was seen by any combination of BREX proteins (Figure 5.3). It is important to state however that these data are results from duplicate experiments and further replicates are required for confirmation.

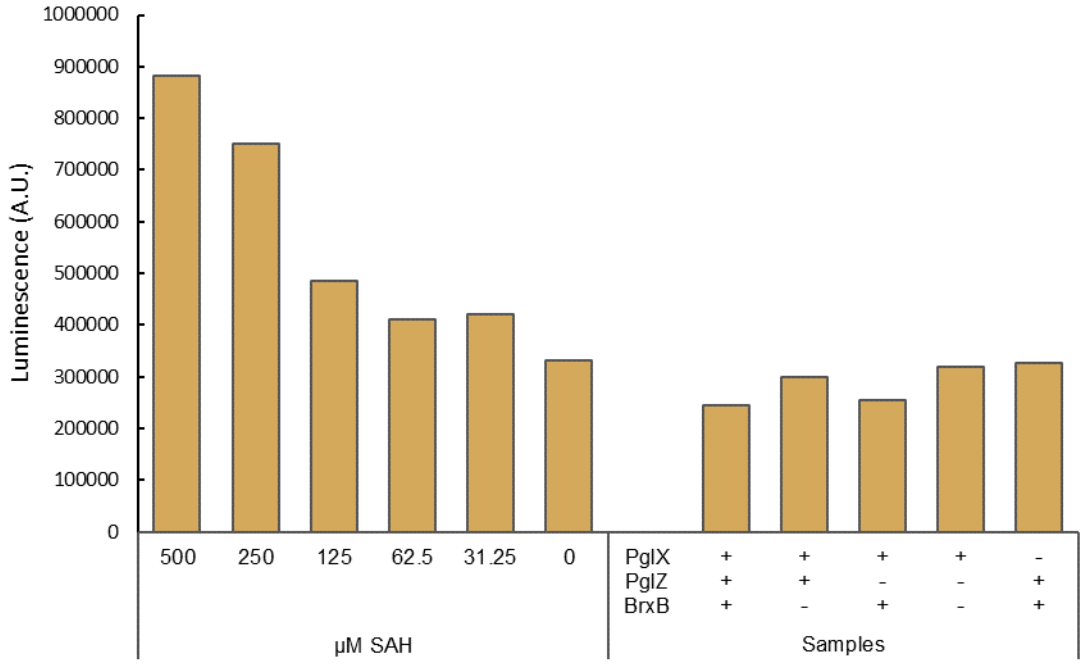

***Figure 5.3 – PglX shows no methyltransferase activity* in vitro.** *Results for SAM dependant methlytransferase assay (Promega) detecting the SAH reaction product of methylation. Standards of SAH are shown on the left and PglX with different combinations of PglZ and/or BrxB are shown on the right.Data are representative of two biological replicates.*

5.3.2 PglX directly interacts with Ocr

The solution state of native PglX was determined using analytical SEC (A-SEC). Size calibration of the Superose 6 Increase 10/300 column (GE) was performed as per manufacturers instruction using a High Molecular Weight calibration kit (Figure 5.4A), allowing estimation of sample size from the relative partition coefficient (K_{av}), calculated from the elution volume [380]. Full protocol details and buffer conditions are available in Chapter 2. PglX eluted from the column after 15.55 ml, indicating a suggested size of around 150 kDa (Figure 5.4B). The calculated weight of PglX is 143 kDa. SEC supports PglX existing as a monomer in solution.

The phage defence inhibitor protein, Ocr, has been shown to inhibit *E. coli* BREX through direct interaction with PglX [294]. In Chapter 4, it was demonstrated that Ocr also inhibits phage defence from *Salmonella* BREX (Figure Table 4.4). To determine whether Ocr directly interacts with the *Salmonella* PglX in the same

way as *E. coli*, A-SEC was performed for PglX combined with Ocr. Purified recombinant Ocr was kindly provided by Dr David Dryden [325]. The Ocr sample was first examined by A-SEC in isolation (Figure 5.4C). Whilst the Ocr SEC profile appeared to have multiple species, there was a dominant peak at 15.9 ml and a shoulder at 18 ml. Ocr is known to be a dimer in solution [291,325], which would be 27.6 kDa and correspond to the 18 ml peak, leaving the 15.9 ml peak unidentified. PglX and Ocr were combined at a 1 : 2 ratio, mixed thoroughly and incubated for 30 minutes on ice before loading onto the column (Figure 5.4D). The combined sample produced additional peaks beyond those from the PglX sample (Figure 5.4B) and the Ocr sample (Figure 5.4C). Of particular interest was the peak at an elution volume of 14.2 ml. According to the calibration curve, this would indicate a complex of approximately 379 kDa and would implicate at least two copies of PglX and potentially several Ocr dimers.

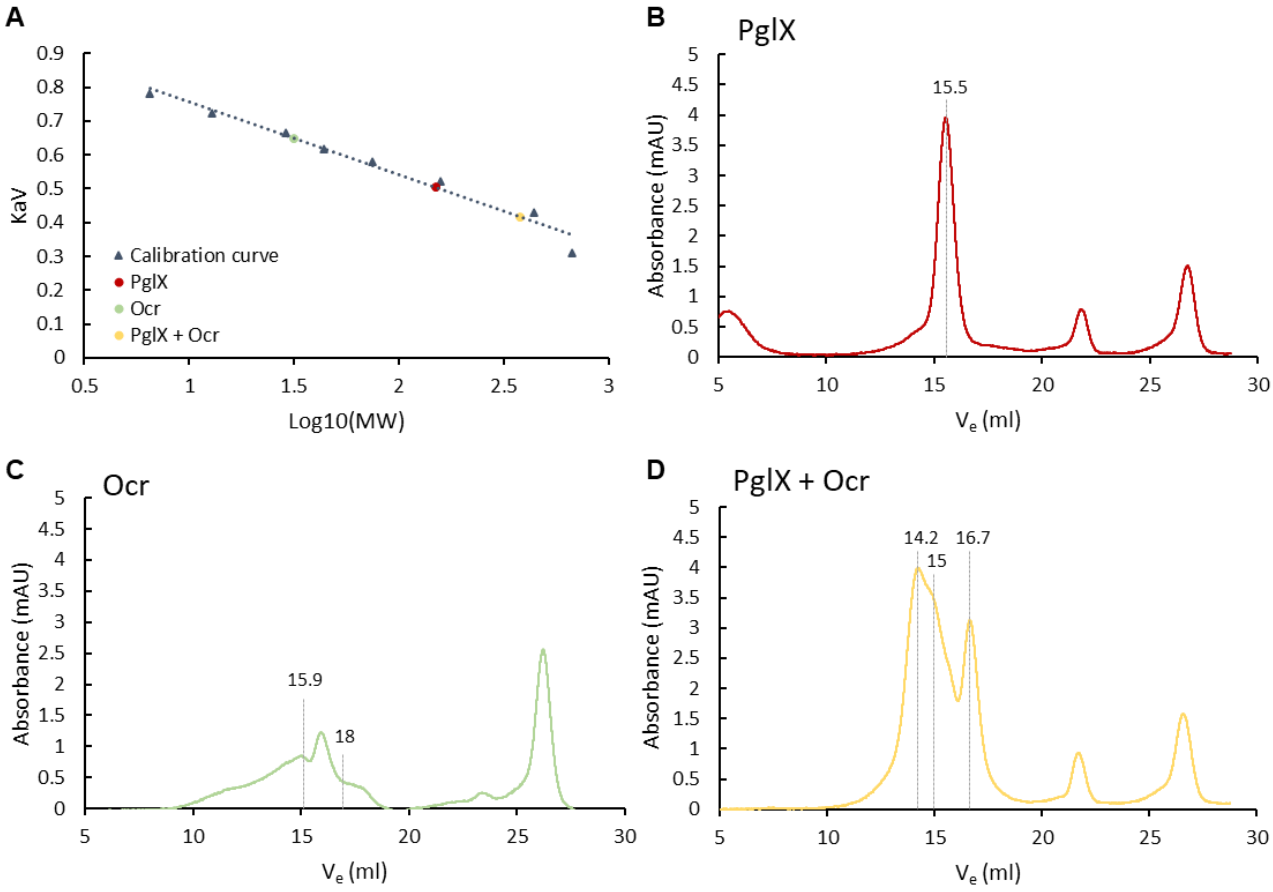

***Figure 5.4 – PglX is bound by Ocr* in vitro.** *Analytical size exclusion chromatography for: A, PglX; B, Ocr; C, PglX and Ocr combined. Grey text indicates elution volumes of selected absorbance peaks. D; Positions of the peak from A (PglX), the 18 ml peak from B (Ocr dimer) and the14.2 ml peak from C (PglX-Ocr heterodimer), relative to the positions of commercially available calibration proteins.*

Elution volume is dependent on protein molecular weight, but also on the shape and size of the protein molecule itself. Therefore, elucidation of protein sizes using this method assumes that the protein of interest presents the same size and shape as that of the protein standards used for calibration. Two proteins with similar molecular weights however could display drastically different surface areas able to interact with the column bed. Instead, protein elution volumes in A-SEC chromatography are better attributed to the hydrodynamic radius of the protein. The hydrodynamic radius of the PglX-Ocr complex seen on A-SEC can be calculated from the observed K_{av} value [380], which can then be compared to the calculated hydrodynamic radius of predicted PglX-Ocr complex models produced by AlphaFold [331]. A model of two copies of PglX and two copies of Ocr produced by AlphaFold produced a predicted hydrodynamic radius of 58.3 Å compared to a calculated hydrodynamic radius of 63.9 Å for the observed A-SEC peak, suggesting this additional peak at 14.2 ml represents a PglX-Ocr heterotetramer in solution.

5.4 Crystallisation trials of PglX

5.4.1 Crystallisation of PglX with SAM

A greater understanding of the function of PglX, both as a methyltransferase and as part of the wider BREX system, could be achieved through elucidation of the structure of the protein. To date, no structure of PglX from any BREX system has been published. The closest structural homologue to the Alphafold predicted structure of PglX in the protein databank (PDB) database is the type IIL RM, MmeI, which aligns poorly with low sequence coverage (RMSD = 7.13 Å, 60% sequence coverage). Thus, the structure of *Salmonella* PglX was sought through X-ray crystallography. Purified PglX was dialysed to crystal buffer (50 mM Tris pH 7.9, 150 mM NaCl, 2.5 mM DTT), concentrated and used to screen for crystallisation – at 1 : 1 and 2 : 1 ratios of PglX to screen conditions, respectively – across twelve commercially available screens (Molecular dimensions; Table 1). For all following crystallisation screens, PglX was added from a stock concentration of 12 mg/ml. Despite multiple attempts using different batches of PglX, no crystals were produced (Table 5.1).

Table 5.1 – Crystallisation screens used in this study. *Commercially available crystallisation screens (Molecular dimensions) used in this study are shown on the left. Screens which produced crystallisation for a given protein are marked with a '+'. Screens which were not set for a given protein are marked as 'N/A'.*

Screen	PglX	PglX + SAM	PglX + SAM + DNA	PglX + SAM + Ocr	PglZ
Pact premier	-	-	-	-	+
JCSG+	-	-	-	+	-
Clear strategy I	-	+	-	+	+
Clear strategy II	-	-	-	-	+
Morpheus I	-	+	+	+	-
Morpheus II	-	+	-	-	-
Morpheus III	N/A	N/A	N/A	-	-
Morpheus Fusion	N/A	+	-	-	-
LMB	-	-	+	+	-
Midas	-	-	-	+	-
Structure 1 + 2	-	-	-	-	-
PGA screen	-	-	N/A	-	-

It was postulated that native PglX may be inherently flexible; Alphafold modelling had predicted two distinct domains joined by a central hinge region (Figure 3.2). It was hypothesised that PglX bound to SAM may bring about a more rigid structure and that this may be more amenable to crystallisation. The panel of commercial crystallisation screens were reset but this time PglX was first incubated in crystal buffer with 1 mM SAM for 30 minutes on ice. Using this method crystallisation was apparent, first in condition H10 of the clear strategy I screen (0.2 M KBr, 0.1 M Tris pH 8.5, 8% w/v PEG 20000, 8% w/v PEG 500 MME) where very small needle-like crystals had formed. These crystals were deemed of insufficient size for X-ray diffraction analysis and further optimisation was pursued around this condition (Table 5.2).

Table 5.2 – PglX crystallisation optimisations. *Two rounds of optimisation conditions for crystallisation of PglX in well H10 of the commercially available crystallisation screen, Clear strategy I (Molecular Dimensions).*

	Optimisation conditions						
Component	*Round 1*				*Round 2*		
KBr (mM)	0.1	**0.2**	0.3			**0.2**	
Tris pH	8	**8.5**	9		7.5	**8**	8.5
PEG 20000 (% w/v)	6	7	**8**	9		**8**	
PEG 500 MME (% w/v)	6	7	**8**	9	5	**6**	7

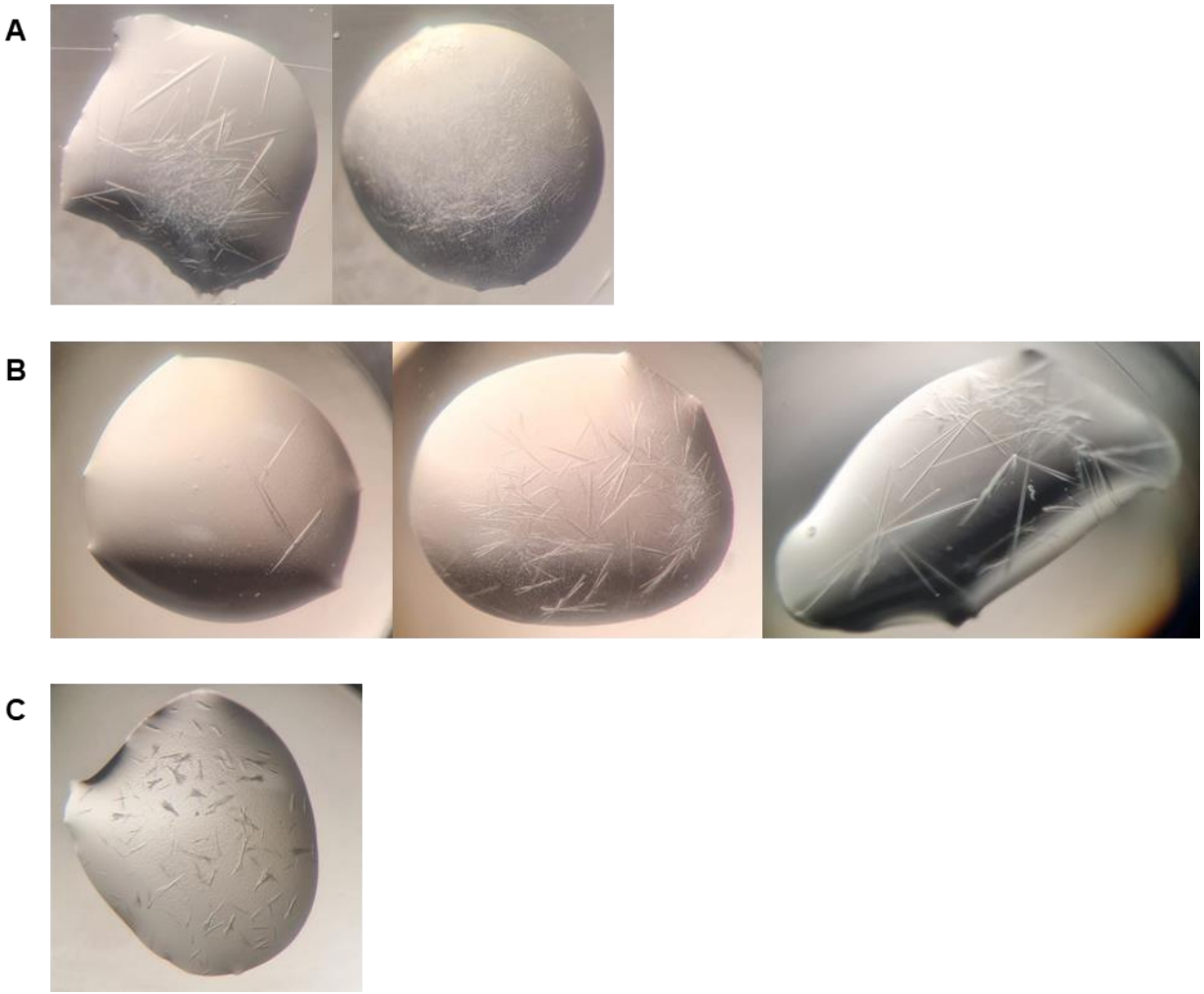

Figure 5.5 – PglX + SAM protein crystals. *A selection of protein crystals from crystallisation screens using PglX co-incubated with Ocr. A; crystals from CSI optimisation round 1. B; crystals from CSI optimisation round 2. C; crystals representative of those produced by the commercially available screens, Morpheus I, Morpheus II and Morpheus Fusion (Molecular Dimensions).*

The first round of optimisation produced crystals across around two thirds of the plate. Though most of these crystals were of the similar small needle-like morphology as those in the initial screen hit, one condition (0.2 M KBr, 0.1 M Tris pH 8, 6% PEG 500 MME, 8% PEG 20000) produced larger crystals potentially big enough for extraction and analysis (Figure 5.5A). A second round of optimisation, using larger drop sizes, was devised around these conditions (Table 5.2). This produced larger needle crystals again (Figure 5.5B), though in some conditions that within the previous round of optimisation has produced no crystals. Further probing of commercially available screens saw additional crystallisation in Morpheus I, Morpheus II and Morpheus Fusion, predominantly in the form bundles of small needles (Figure 5.5C).

5.4.2 Crystallisation of PglX with SAM and DNA oligos

As a DNA methyltransferase, PglX is expected to bind DNA. A structural model of PglX bound to DNA would provide great insight into the mechanism of methylation and, importantly for this study, allow identification of residues and interactions involved in BREX motif recognition. Towards this end, two DNA oligos were designed and synthesised (Figure 5.6). Oligo 1 was based on the DNA oligos crystallised bound to the PglX structural homologue, MmeI, but with the *Salmonella* D23580 BREX recognition motif substituted in at position 3 and the thymine base at position 10 substituted for a cytosine. Oligo 2 is based on the DNA oligos crystallised bound to the *Clostridium difficile* orphan methylase, CamA – also shown to be a structural homologue of PglX – which contained T/A overhangs which stabilised crystal structure formation, again with the *Salmonella* D23580 motif substituted in. It was not known whether binding of PglX to DNA required prior SAM binding to PglX, therefore PglX was first mixed and incubated with SAM for 30 minutes on ice. As PglX methylates at non-palindromic sites and is a monomer in solution, it is expected that a single PglX molecule will bind to a single recognition motif. Thus, annealed DNA oligos were then mixed with PglX at a 1.2 : 1 ratio, allowing a slight excess of DNA to encourage homogenous binding, and incubated on ice for a further 30 minutes. Screening against commercial crystal screens produced protein crystals in several conditions for PglX bound to Oligo 1 across Morpheus I and LMB screens (Table 5.1), including different crystal morphologies to those for PglX + SAM (Figure 5.7). Very small needles were produced for several conditions containing PglX bound to Oligo 2. Unfortunately, due to time constraints full optimisation of these conditions was not possible and conditions producing crystals with Oligo 1 were prioritised.

DNA Oligo 1:

5' -TAGATC**A**GACAAC- 3'

3' -ATCTAGTCTGTTG- 5'

DNA Oligo 2:

5' -TTCGATC**A**GTCCCA -3'

3' - AGCTAGTCAGGGTA-5'

Figure 5.6 – DNA oligonucleotides for PglX crystallisation. Oligonucleotides were designed based on the DNA molecule bound to MmeI (5HR4) and CamA (7LT5) for oligo 1 and oligo 2, respectively. ###state why the A is bold highlighted

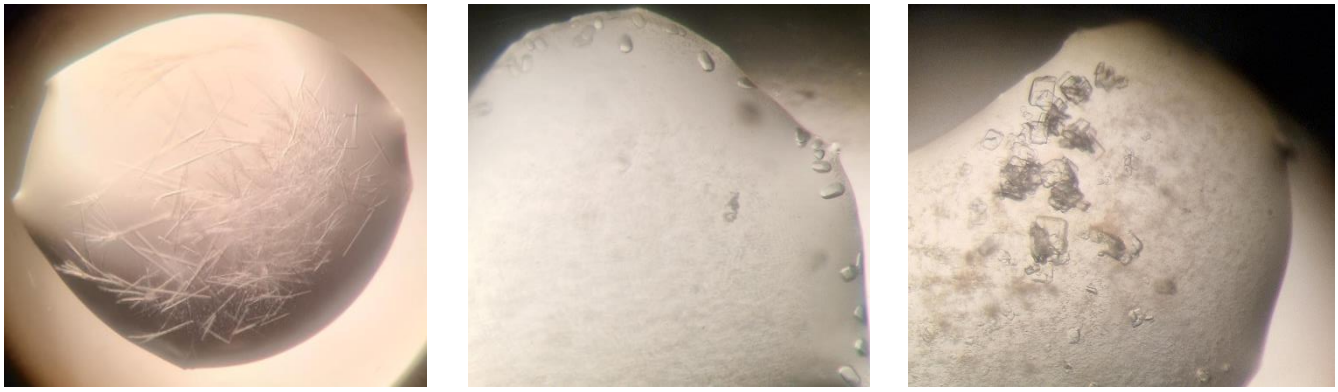

Figure 5.7 – PglX + SAM + DNA crystals. *Selection of protein crystals produced by commercially available crystallisation screens (Molecular Dimensions).*

5.4.3 Crystallisation of PglX with SAM and Ocr

It was previously shown that PglX directly interacts with the phage encoded inhibitor, Ocr, and produced a multimeric complex likely at a 2 : 2 stoichiometry (Figure 5.4). To provide further insight into the mechanism of BREX inhibition by Ocr, efforts were made to produce a structural model of this complex through X-ray crystallography. It was not known whether binding of PglX to Ocr required prior SAM binding to PglX, therefore PglX was first mixed and incubated with SAM for 30 minutes on ice, as previously. PglX and Ocr were then mixed at a 1 : 2 molar ratio (final concentrations: PglX, 12 mg/ml; Ocr, 2.74 mg/ml) and incubated on ice for a further 30 minutes, and screens were set as previously. These conditions produced crystals in a different range of screens to the PglX + SAM and PglX + SAM + DNA screens (Table 5.1) with markedly different morphologies. CSI-F11 (Figure 5.8E) produced two different crystal morphologies within the same drop. LMB-C1 produced small ovaloid crystals (not shown), while LMB-F4 produced one very large crystal along one edge of the drop (Figure 5.8D). Midas-F11 and Midas-G2 produced small, square plate-like crystals protruding from a central point (Figure 5.8C). Crystals seen in JCSG and Morpheus I conditions appeared similar to PglX + SAM crystal produced in these screens, albeit in different chemical conditions (Figure 5.8A and B). Crystallisation was also apparent in conditions used in the CSI optimisation screens used for PglX + SAM with a similar crystal morphology (Figure 5.8 F).

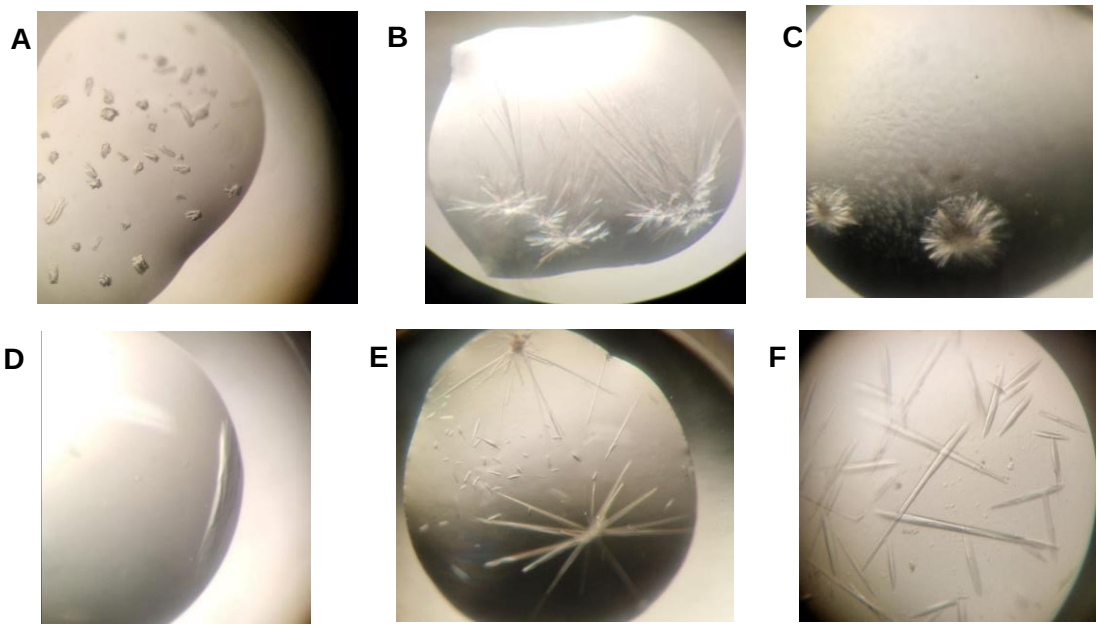

Figure 5.8 – PglX + SAM + Ocr crystals. *Selection of protein crystals produced by commercially available crystallisation screens (Molecular dimensions.)*

5.5 Crystallisation trials of PglZ

PglZ is the only BREX component common to all BREX subtypes and was initially used to discover BREX phage defence systems [308]. Elucidating the structure of PglZ would provide great value in understanding the seemingly vital function of PglZ in BREX defence and would allow comparison to the recently described phosphodiesterase homologue, PorX [317]. Initial crystallisation screening of PglZ was performed similarly to screening with PglX. PglZ was dialysed into crystal buffer and concentrated to around 12 mg/ml and drops were set at 1 : 1 and 2 : 1 ratios of PglZ to screen condition. Crystallisation was observed in six conditions in CSI (E8, F2, F11, G8, G10, H10), two conditions in Pact Premier (E8, H4) and one condition in CSII (F12) (Table 1). Though the pH, precipitant and salt conditions in each were varied, a common theme was the presence of a sulphur moiety in the salt additive, either in the form of lithium sulphate (Li_2SO_4), potassium thiocyanate (KSCN) or sodium sulphate. In an attempt to increase the size and quality of the protein crystals produced, pH and precipitant optimisation screens were devised using either Li_2SO_4 or KSCN salts (Table 5.3). Larger crystals seemed to form around the central conditions of the Li_2SO_4 screens, so a second optimisation screen was devised using a narrower range of conditions and larger drop sizes were utilised to promote large crystal growth. Indeed, large crystals were seen in several of the Li_2SO_4 conditions, presenting as long, rectangular prisms (Figure 5.9D). The small hexagonal crystals observed in some KSCN-based conditions (Figure 5.9B and C) did not increase in size and only appeared sporadically. Others in our lab had shown that the phosphodiesterase activity displayed by PglZ *in vitro* was dependant on binding to Zn^{2+}. Purified PglZ was incubated with 0.5 mM $ZnCl_2$ for 1 hour on ice and screening was repeated. PglZ incubated with Zn^{2+} formed crystals in the same range of crystal conditions and no new crystal morphologies were produced.

Table 5.3 – PglZ crystal optimisation screens. *Conditions used for two rounds of crystallisation optimisation conditions used for PglZ.*

	Optimisation conditions					
Component	*Round 1*			*Round 2*		
Li_2SO_4/KSCN (mM)	0.1	**0.2**	0.3	0.15	0.2	0.25
pH	7	**8**	9	7	8	9
PEG 4000 (w/v)	10	**15**	20	13	15	17

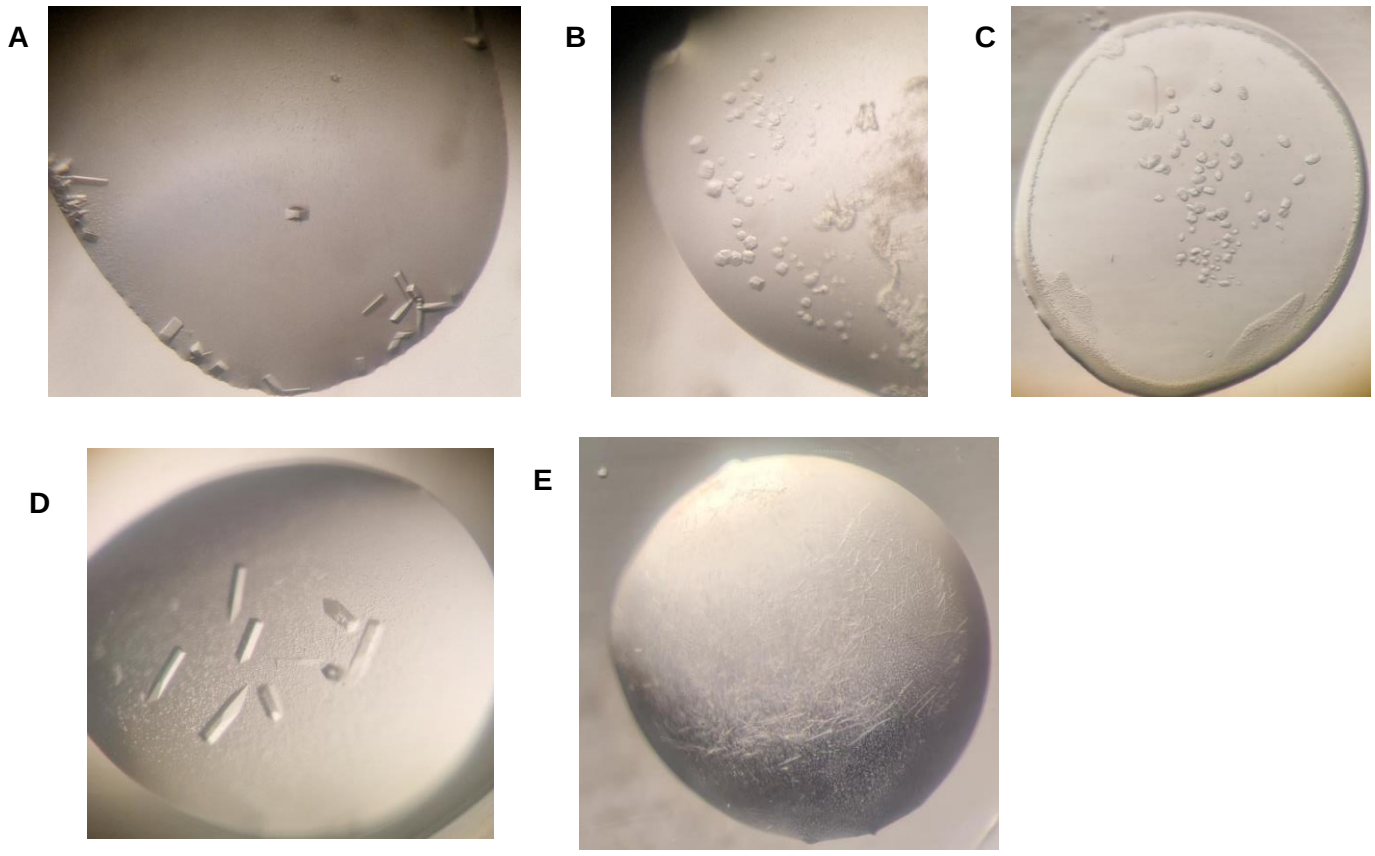

Figure 5.9 – PglZ crystals. *Selection of protein crystals produced by commercially available crystallisation screens (Molecular dimensions.)*

5.6 Structure of PglX

5.6.1 Data collection and structure elucidation

A selection of PglX + SAM crystals were mixed with a cryoprotectant buffer, looped and flash frozen in liquid nitrogen from both the second round CSI optimisation screens and Morpheus screens and sent to the Diamond synchrotron, Oxford, for X-ray diffraction. Crystals were shot on beamlines I04 or I23 and data was processed using xia2 Dials [326]. A majority of crystals from either CSI optimisation or Morpheus screens produced poor diffraction (5 – 8 Å), or no diffraction at all. A single crystal produced in the second optimisation round of CSI conditions displayed the best diffraction (~3.5 Å) and multiple datasets were collected. A combined dataset of the highest resolution collections was combined with xia2 Dials to produce a single dataset with resolution to 3.4 Å (Table 5.4). As the resolution of the dataset was too low to produce an *ab initio* solution, molecular replacement was performed with the AlphaFold generated PglX model using Phaser in the CCP4i2 [327] software package. The resulting structure underwent further cycles of building and refinement using Coot [328] and Phenix [324], respectively (Table 5.4).

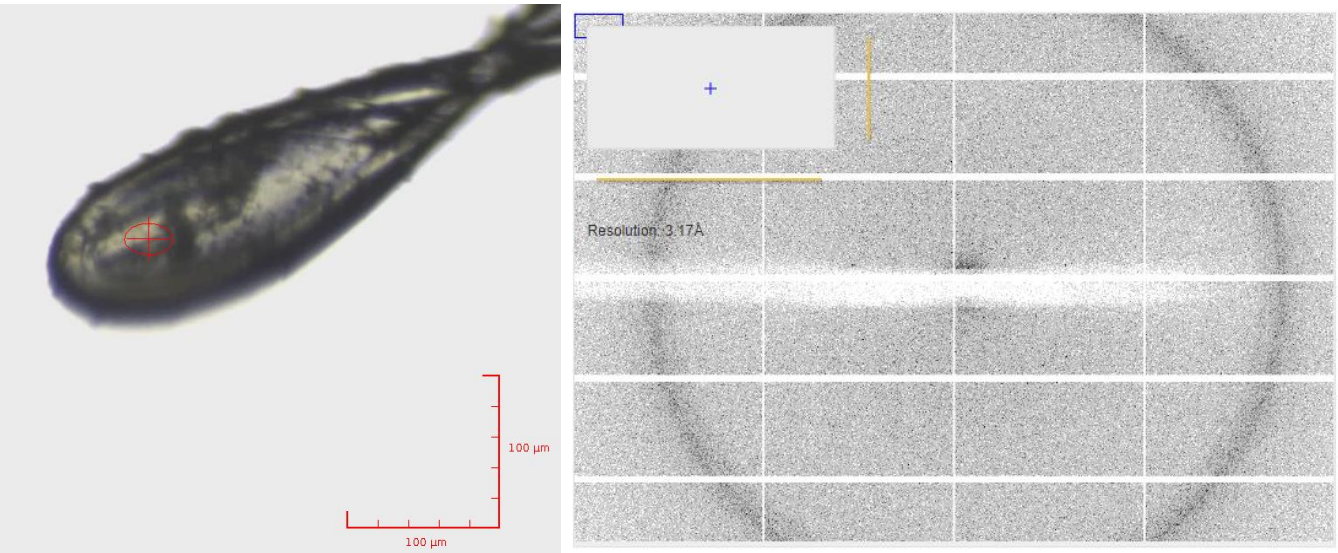

Figure 5.10 – PglX crystal and diffraction. *Left; looped PglX crystal, Right; diffraction pattern used for structural solution.*

Table 5.4 – Data Collection and Refinement of PglX+SAM*. Numbers in parentheses represent highest resoution shell.*

	PglX+SAM
Wavelength	0.9795
Resolution range	48.98 - 3.402 (3.523 - 3.402)
Space group	P 41 21 2
Unit cell	138.539 138.539 407.956 90 90 90
Total reflections	104405
Unique reflections	55611 (5460)
Multiplicity	1.9
Completeness (%)	87.15 (15.55)
Mean I/sigma(I)	8 (0.1)
Wilson B-factor	159.6
R-merge	0.047
R-meas	0.067 (2.142)
R-pim	0.047 (1.515)
CC1/2	0.999 (0.214)
Reflections used in refinement	48492 (849)
Reflections used for R-free	2444 (43)
R-work	0.2745 (0.4253)
R-free	0.2992 (0.4026)
Number of non-hydrogen atoms	19848
macromolecules	19848
Protein residues	2432
RMS(bonds)	0.005
RMS(angles)	0.91
Ramachandran favored (%)	90.36
Ramachandran allowed (%)	9.64
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	12.25
Average B-factor	169.33
macromolecules	169.33

5.6.2 PglX presents two distinct domains with a bound SAM co-factor

The crystal structure contains two copies of PglX in the asymmetric unit. However, the arrangement of the two copies allows only weak interactions that likely support formation of the crystal, rather than being biologically significant. The A-SEC data also strongly supports PglX being monomeric (Figure 5.4B). The architecture of PglX is highly similar to the model produced by Alphafold (Figure 3.2), presenting two distinct domains, N-terminal and C-terminal, joined by a central hinge region (Figure 5.11A). Due to absence of available density, two short loop regions were unable to be modelled (G53 to D56; D418 to F420). Also visible in the structure is the SAM co-factor bound to a pocket in the N-terminal domain near the hinge region (Figure 5.11A, inset). PoseView [381] allows detection of important interactions between protein residues and bound ligands and was used to infer interactions between PglX residues and SAM (Figure 5.11B). SAM binding is attributed to several hydrogen bonds, shown by dashed lines, and three hydrophobic interactions, shown by green lines (Figure 5.11B). A close-up view of SAM bound to PglX is shown in Figure 5.11A inset, with residues identified as forming hydrogen bonds shown in pale orange and residues forming hydrophobic interactions shown in pale green. S316 forms hydrogen bonds with C5 and C6 within the methionine moiety at one end, N354 forms a hydrogen bond with one of the hydroxyl groups of the ribose component and S452 forms two hydrogen bonds with N11 within the adenine. Meanwhile, the interaction is further stabilised by hydrophobic interactions from F541 and I355 with the adenine moiety and between P511 and the methionine moiety.

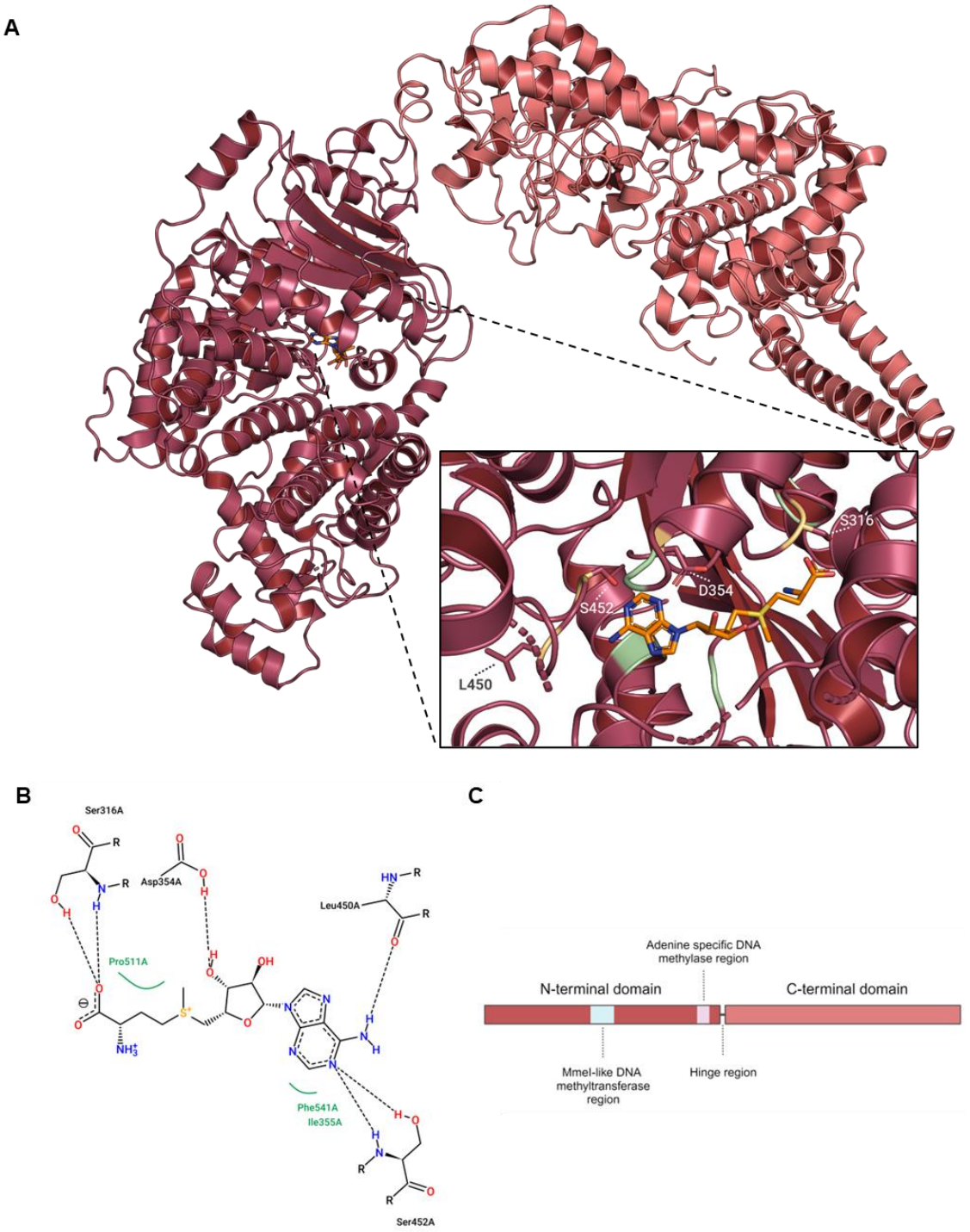

Figure 5.11 – Structure of PglX*. A; the crystal structure of PglX, with the position of the bound SAM molecule shown inset. PglX displays distinct N-terminal and C-terminal domains, shown in raspberry and pink, respectively. Residues coloured pale orange form hydrogen bonds with SAM. Residues coloured green form hydrophobic interactions with SAM. B; Interactions between PglX residues and SAM, as indicated by PoseView* [381]*. Dashed lines indicate hydrogen bonds. Green lines indicate hydrophobic interactions. C; Linnear view of the N-terminal and C-terminal domains of PglX with the positions of predicted methyltransferase regions*.

5.6.3 The core methyltransferase domain of PglX is conserved

The closest structural homologue for the solved PglX structure as designated by the DALI server remains the type IIL RM system, MmeI, thought the similarity is lower (Z-score 23.8 for predicted structure, Z-score 20.3 for solved structure). This is perhaps unsurprising as AlphaFold would have used MmeI in structure prediction. MmeI demonstrates both N6mA DNA methyltransferase and DNA restriction modalities (see Chapter 3) but the MmeI structure provides only 60.8% sequence coverage (1225 residues and 745 residues for PglX and MmeI, respectively) and aligns with an RMSD of 7.13 Å (Figure 5.12A). A majority of this alignment seems to fall within the N-terminal domain of PglX and bridges the hinge region, extending slightly into the C-terminal domain. PglX contains a SAM-dependent methyltransferase superfamily domain (residues 80 – 718), predominantly within the N-terminal region (Figure 5.12B). Within the N-terminal region of PglX (Figure 5.11C) sit the predicted MmeI-like DNA methyltransferase region (residues 306 – 381, shown in pale blue) containing the S316, N354, N354 and I355 residues implicated in SAM binding, and an adenine specific DNA methylase region (residues 500 – 641, shown in pale pink) containing F541, which together broadly encircle the site of SAM binding (Figure 5.12B). S452, implicated in hydrogen bond formation with N11 within the adenine moiety of SAM sits between these two domains. Looking at conserved residues using ConSurf [382], the MmeI-like DNA methyltransferase region appears highly conserved (Figure 5.12C). The adenine specific DNA methylase region on the other hand shows far less conservation. Outside of identified functional domains, the core of the N-terminal domain of PglX appears fairly conserved with the exception of some variable loop and helix regions at the beginning of the protein and between the two identified methyltransferase related functional domains. Conservation within the N-terminal domain other than the functional methyltransferase domains hints at some significant role in BREX function, possibly through interactions with other BREX components. Similarly, the C-terminal domain extends 745 residues beyond the end of the alignment with MmeI and shows a high degree of conservation, suggesting an important role in BREX function. Together these data show that the N-terminal domain and hinge region of PglX has conserved methyltransferase modality while the C-terminal domain potentially presents some unknown but similarly important function.

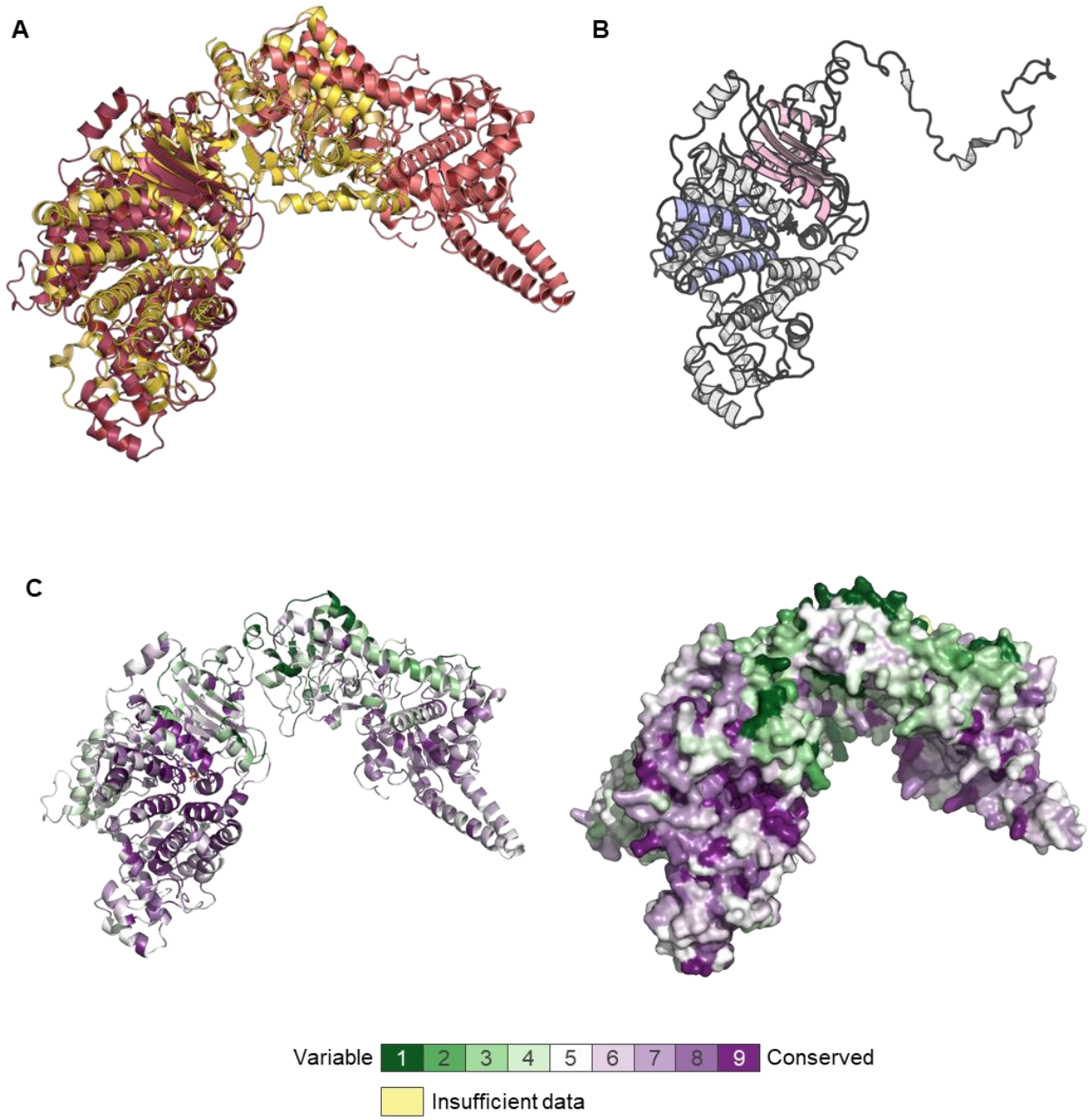

Figure 5.12 – The PglX methyltransferase domain is conserved. *A; alignment of PglX with the closest structural homologue, MmeI. N-terminal and C-terminal domains of PglX are, shown in raspberry and pink, respectively, MmeI is shown in yellow. The DNA molecule bound to MmeI is shown in orange. B; Identified SAM dependant methyltransferase superfamily domain isolated from PglX. The functional MmeI-like methyltransferase domain and domains adenine specific methylase domain was inferred by InterPro and are shown in pale blue and pale pink, respectively. C and D; Ribbon and surface views, respectively, of conserved regions in PglX, as calculated by ConSurf* [382].

5.6.4 PglX displays a negatively charged central groove for potential DNA binding

PglX has been shown to bind DNA *in vitro* by our collaborators in the Severinov lab, who used electrophoretic mobility shift (EMSA) assays (Dr Artem Isaev, personal communication) DNA binding proteins often bind DNA through charge-based interactions and demonstrate highly positive surfaces around their DNA binding sites promoting binding to negatively charged DNA backbones [383]. The surface charge of PglX was calculated using APBS software plugin [384] and modelled in PyMOL to attempt to predict a DNA binding position (Figure 5.13A). Notably, PglX displayed a large positively charged surface area in the hinge region between the N-and C-terminals, extending further along the inside of the C-terminal. The structural homologue, MmeI, was solved in a DNA bound state and aligns over the hinge region of PglX. Superimposing these two structures and removing the MmeI molecule, the DNA molecule sits within this positively charged hinge region (Figure 5.13B). Further to this, the DNA molecule from the MmeI structure contained an adenine base which had been flipped out of the DNA molecule for methylation. Looking at the position of the superimposed MmeI DNA molecule, this adenine base is positioned close to the SAM molecule in PglX (Figure 5.13B, inset). Together, these data suggest that PglX binds DNA within this hinge region in a very similar conformation to that seen in MmeI, though the exact orientation of the DNA molecule may shift around the position of the adenine base. This is likely, as the donated methyl group of the SAM is not quite positioned correctly for transfer to the flipped adenine. The C-terminal domain of PglX remains removed from the DNA molecule, however. Binding of DNA may require, or produce, a conformational change in PglX, bringing this domain closer to the DNA molecule.

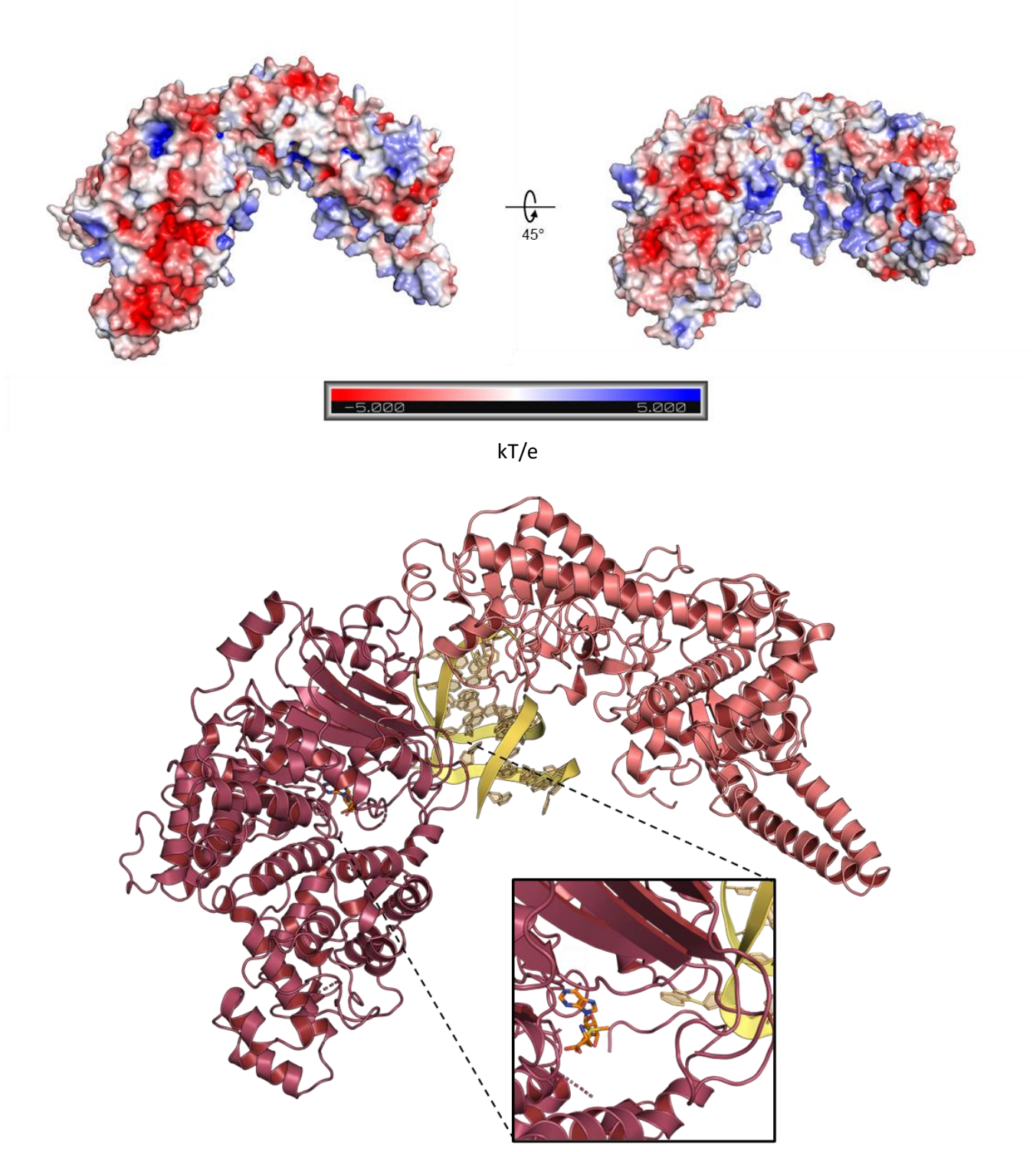

Figure 5.13 – PglX possesses a potential DNA binding groove. *A; Surface views of PglX with surface charge calculated by APBS* [384]. *Red indicates negative charge. Blue indicates positive charge. B; Alignment of the DNA molecule from MmeI (yellow) aligned with PglX with the position of the flipped out adenine base from this structure relative to the SAM molecule in PglX shown inset. N-terminal and C-terminal domains of PglX are, shown in raspberry and pink, respectively.*

5.7 Structure of PglX bound to OCR

5.7.1 Data collection and structure elucidation

A selection of PglX + SAM + Ocr crystals were mixed with a cryoprotectant buffer, looped and flash frozen in liquid nitrogen and sent to the Diamond synchrotron, Oxford, for X-ray diffraction. Crystals were chosen both from conditions distinct from PglX + SAM and PglX + SAM + DNA condition, and from crystals formed in CSI optimisation conditions used for PglX + SAM. Almost all crystals analysed produced no diffraction, regardless of morphology. One exception was the large crystal produced by LMB-F4 (Figure 5.8; Figure 5.14), which produced diffraction datasets to around 4 – 6 Å (Figure 5.14). Overall, forty-seven datasets were collected from this crystal, and combinations of the best datasets were merged using xia2 Dials [326] to attempt to find one which provided sufficient resolution and quality. Finally, a 3.5 Å dataset was produced and the structure was solved by molecular replacement with the previously generated PglX + SAM solution (Section 5.4) and the PDB structure of Ocr (1S7Z). The solution was further built and refined using Coot [328] and Phenix [324], respectively (Table 5.5).

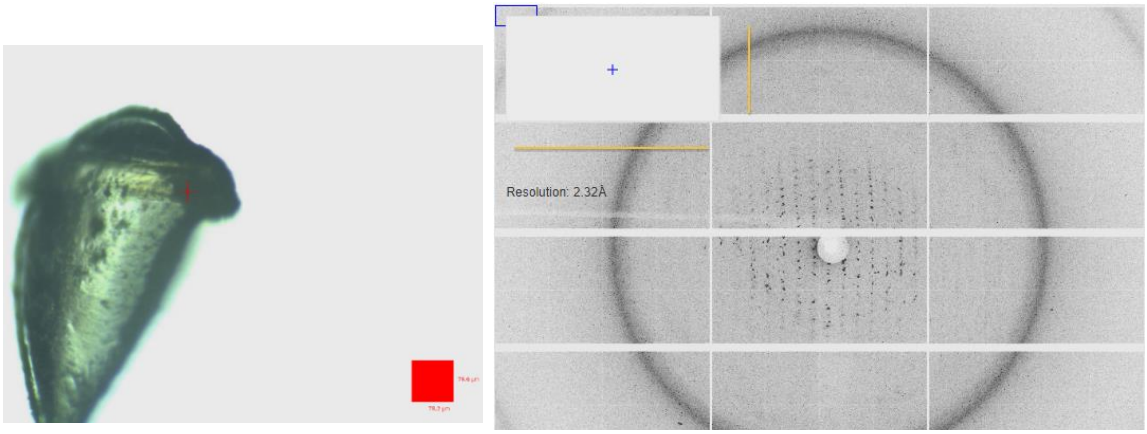

Figure 5.14 – PglX + Ocr crystal and diffraction. *Left; looped PglX + Ocr crystal, Right; diffraction pattern used for structural solution.*

Table 5.5 – Data Collection and Refinement of PglX+SAM+Ocr. *Numbers in parentheses represent highest resoution shell.*

	PglX-Ocr
Wavelength	0.9795
Resolution range	59.61 - 3.5 (3.625 - 3.5)
Space group	C 1 2 1
Unit cell	238.458 60.7861 146.637 90 114.889 90
Total reflections	47094 (8532)
Unique reflections	24556 (2426)
Multiplicity	1.9
Completeness (%)	97.84 (80.53)
Mean I/sigma(I)	3.8 (0.3)
Wilson B-factor	127.38
R-merge	0.028
R-meas	0.092 (0.756)
R-pim	0.092 (0.756)
CC1/2	0.995 (0.378)
Reflections used in refinement	24038 (1957)
Reflections used for R-free	1922 (144)
R-work	0.2462 (0.4074)
R-free	0.2917 (0.4202)
Number of non-hydrogen atoms	10776
macromolecules	10747
ligands	49
solvent	2
Protein residues	1318
Nucleic acid bases	0
RMS(bonds)	0.004
RMS(angles)	0.78
Ramachandran favored (%)	91.6
Ramachandran allowed (%)	8.4
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.26
Clashscore	15.24
Average B-factor	138.5
macromolecules	138.54
ligands	126.14
solvent	113.43
Number of TLS groups	0

5.7.2 PglX and Ocr form a heterodimer in the asymmetric unit

Within the asymmetric unit, Ocr binds to PglX as a 1 : 1 ratio, with a single protomer of Ocr binding along the negatively charged C-terminal region of a single PglX protomer (Figure 5.15A). The structure has similar resolution and R-values to the PglX structure (Table 5.5 and Table 5.4, respectively). Within PglX, there are again two regions of the sequence which could not be modelled due to insufficient density (Q54 to F55; T413 to 420). The latter is an extended gap in the same region as a smaller gap in the PglX + SAM structure (D418 to F420), suggesting flexibility in this region. The gap is located in a loop within a non-conserved helical bundle between the two methyltransferase functional regions (Figure 5.13). Again, visible in the structure is a bound SAM molecule, in the same ligand binding position as seen in the apo PglX structure (Figure 5.15A). The exact orientation of ribose and methionine components of the molecule does vary slightly, though this is likely due to variation in manual positioning of the molecule during refinement, as well as the resolution. The PglX molecules from the apo PglX and Ocr-bound PglX structures align closely with an RMSD of 1.34 Å, and the resulting alignment suggests that binding of Ocr does not elicit any domain movement (Figure 5.16). Ocr binding to type I RM complexes has previously been shown to elicit domain movement similar to DNA binding, suggesting that PglX domain movement is reliant on interactions with other BREX components. This is consistent with the lack of methyltransferase activity *in vitro* in the absence of other BREX components (Figure 5.3) or from PglX alone *in vivo* [314].

Published SEC data suggested that Ocr exists as a dimer in solution, not a monomer, and it has previously been shown to bind in this conformation, with a dimer of Ocr mimicking bases of DNA [325], often at a 2 : 1 ratio of Ocr to target molecule [287,291]. A-SEC experiments had suggested that PglX and Ocr form a heterotetramer based on elution volume and hydrodynamic radius calculations (Figure 5.4). When crystallographic symmetry was applied to the observed model, two apparent heterodimers came together to form a heterotetrameric complex, with PglX protomers independently bound to either protomer of a bridging Ocr dimer (Figure 5.15B). Important residue interactions for Ocr binding were inferred using EMBL PISA [385]. The complex is stabilised by a number of hydrogen bonds between Ocr and the C-terminal domain of PglX (Table 5.6). Additionally, six salt bridges are produced between arginine (R79), asparagine (N35, N42, N62, N76) and glutamine (Q109) groups on Ocr and asparagine (N1213) and lysines (K1201, K1097, K1070, K1110, K516) on PglX (Table 5.6).

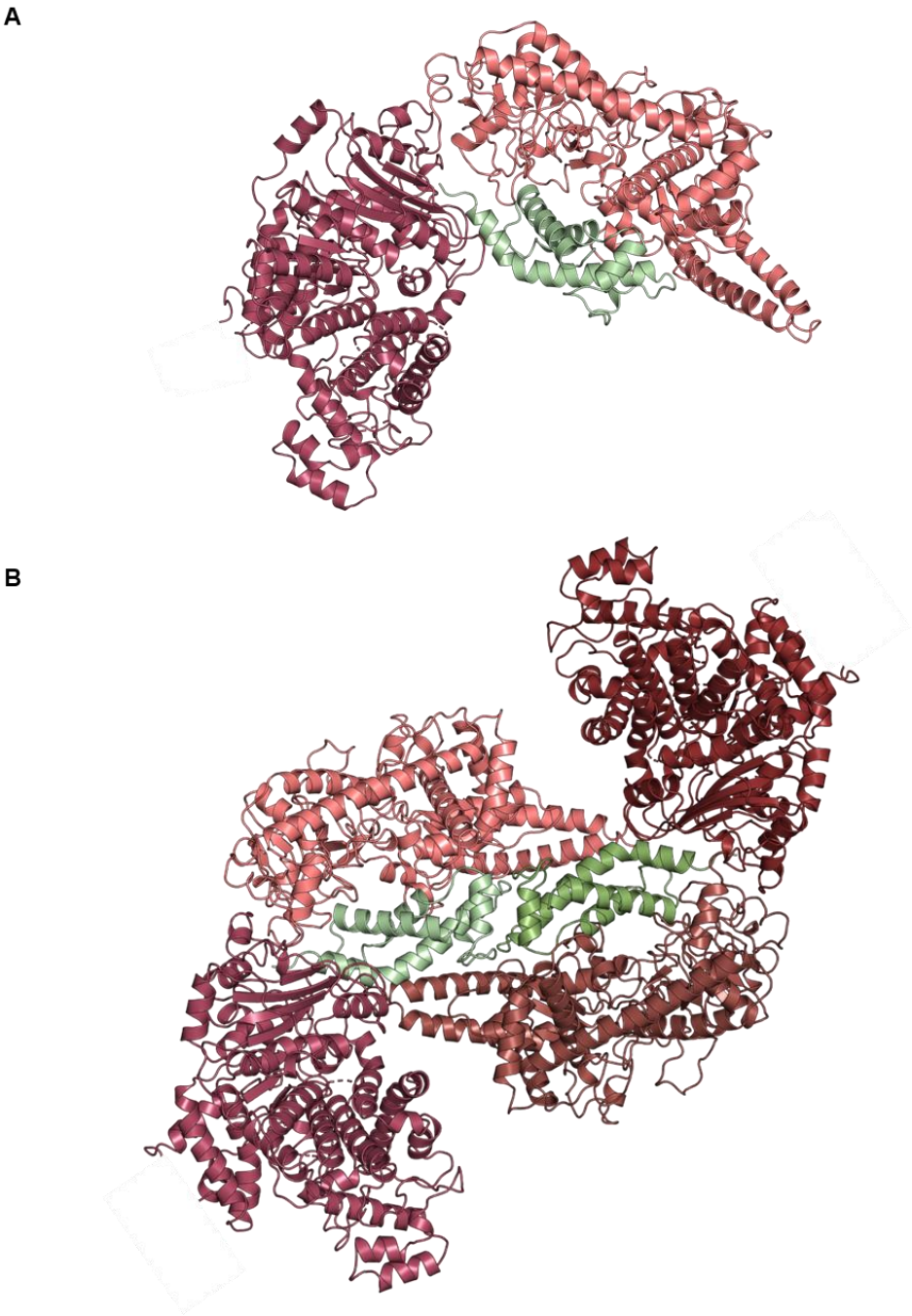

Figure 5.15 – Crystal structure of PglX bound to Ocr. *A; the unit cell view of on molecule of PglX bound to one molecule of Ocr. B; Heterotetramer complex of PglX and Ocr formed across the asymmetric unit. N-terminal and C-terminal domains of PglX are, shown in raspberry and pink, respectively. Ocr is shown in pale green and smudge green for for respective molecules in the Ocr dimer.*

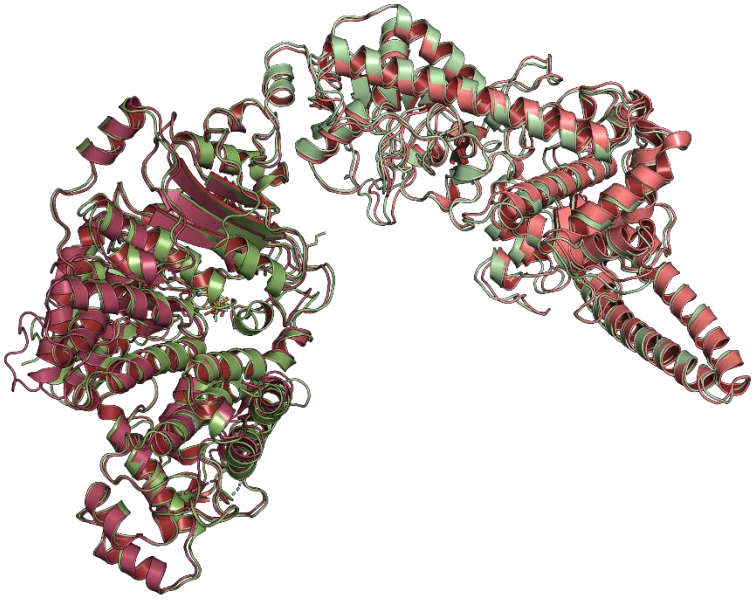

Figure 5.16 – Alignment of the solved structure of PglX bound to SAM and of PglX from the structure bound to SAM and Ocr. *N-terminal and C-terminal domains of PglX are, shown in raspberry and pink, and Smudge and pale green respectively.*

Ocr has been shown to mimic the structure of 20 – 24 bp of B-form DNA [325], as shown by the binding of both molecules to the EcoKI methyltransferase complex [282,381]. Using the DNA bound (2Y7H) and Ocr bound (2Y7C) complexes of EcoKI, the Ocr and DNA molecules were superimposed onto each other (Figure 5.16A) and then the Ocr molecule in the PglX-Ocr structure was aligned with the Ocr molecule in 2Y7C, effectively aligning the B-form DNA from 2Y7H to the Ocr molecule in PglX-Ocr structure (Figure 5.16B). The angle of the 2Y7C DNA molecule differs from the angle of the superimposed DNA molecule from MmeI (Figure 5.13C). There does however appear to be enough space for an extended DNA molecule to pass through the groove in the hinge region in this orientation (Figure 5.13B). This raises the possibility of an alternative DNA binding orientation and implicates the C-terminal domain specifically in motif recognition. In addition, the 2Y7C DNA molecule is 20 bp long compared to the 13 bp MmeI DNA molecule. If the position of Ocr binding along the C-terminal arm of PglX is representative of DNA binding, DNA binding may in turn require a longer DNA molecule than those utilised in section 5.3.2.

Table 5.6 - Residues implicated in PglX:Ocr binding interface, as calculated by EMBL-PISA.

Hydrogen bonds		
Ocr	**Dist. (Å)**	**PglX**
B:ARG 79[NH1]	3.80	A:ASP1213[OD2]
B:ARG 79[NH2]	3.04	A:THR1085[O]
B:ASP 35[OD2]	3.79	A:LYS1201[NZ]
B:GLU 66[O]	3.51	A:SER1169[OG]
B:GLU 66[OE2]	3.19	A:SER1169[OG]
B:HIS 47[O]	3.61	A:GLN1086[NE2]
B:LEU 105[O]	2.36	A:LYS 516[NZ]
B:LYS 75[N]	3.49	A:LYS1108[O]
B:MET 39[O]	3.80	A:LYS1097[NZ]
B:MET 56[O]	2.57	A:LYS1068[NZ]
B:SER 58[O]	3.89	A:LYS1068[NZ]
B:SER 68[OG]	2.35	A:GLU1109[OE2]
B:THR 90[O]	3.38	A:ARG1165[NH1]
B:TYR 24[O]	2.96	A:ASN 691[ND2]
B:TYR 48[OH]	2.39	A:LYS 964[NZ]
B:TYR 49[OH]	3.23	A:LYS1089[NZ]
Salt bridges		
B:ARG 79[NH1]	3.80	A:ASP1213[OD2]
B:ASP 35[OD2]	3.79	A:LYS1201[NZ]
B:ASP 42[OD2]	3.30	A:LYS1097[NZ]
B:ASP 62[OD2]	3.79	A:LYS1070[NZ]
B:ASP 76[OD2]	3.05	A:LYS1110[NZ]
B:GLU 109[OE2]	3.83	A:LYS 516[NZ]

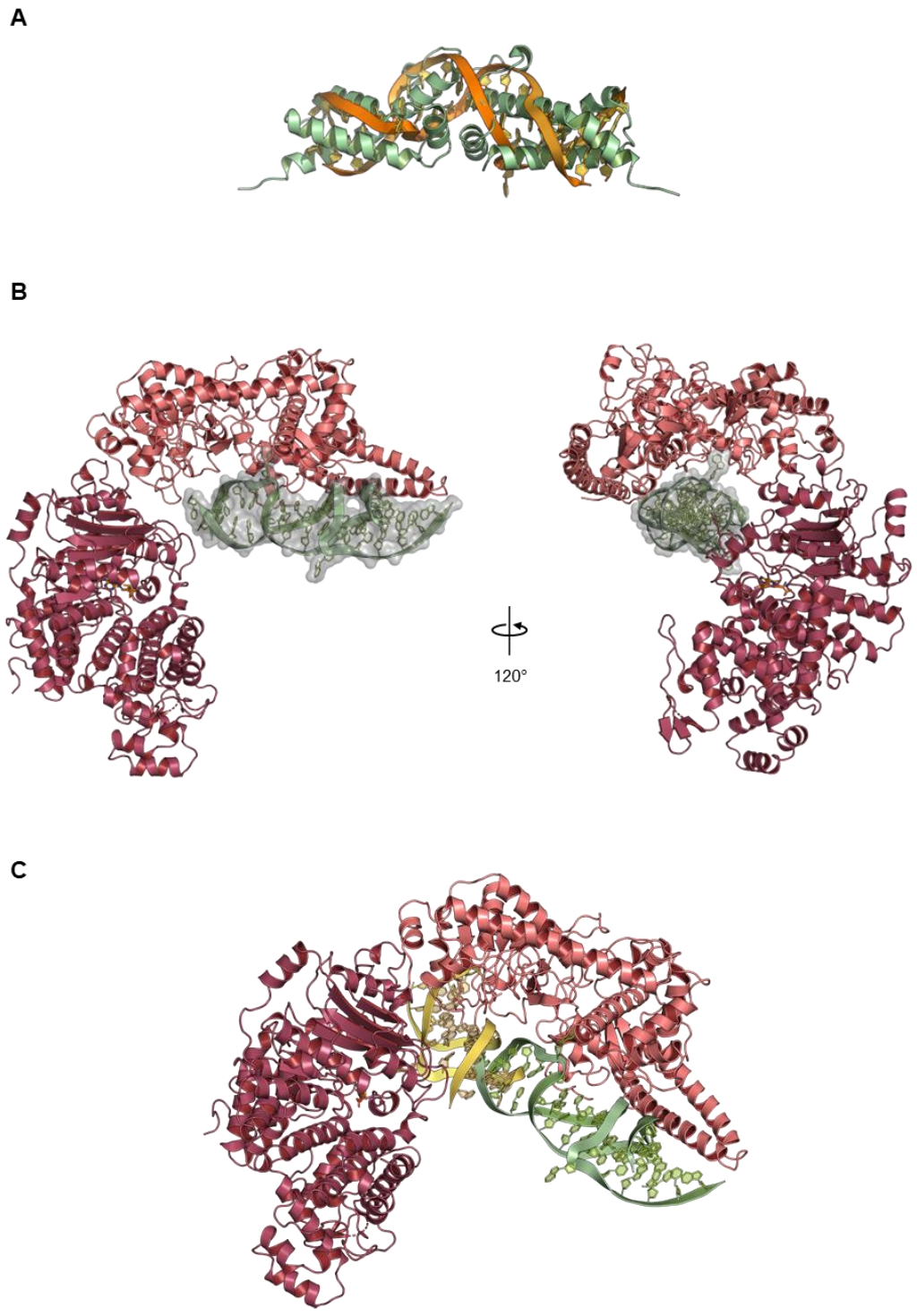

Figure 5.17 – Ocr binding implies an alternative DNA binding orientation. *A; Alignment of the Ocr dimer with a slightly bent B-form DNA molecule from 2Y7C. B; Superimposition of this DNA molecule (green) into the binding site of Ocr. C; Comparison of the positions of the superimposed MmeI (yellow) and Ocr DNA (green).*

5.8 PglX + SAM + DNA and PglZ crystals did not produce sufficient diffraction.

A selection of PglX + SAM + DNA crystals were mixed with a cryoprotectant buffer, looped, flash frozen in liquid nitrogen and sent to the Diamond synchrotron, Oxford, for X-ray diffraction. PglX + SAM + DNA screens had produced a variation of crystal morphologies (Figure 5.7) and a representative selection of each were screened for diffraction. No diffraction was observed from any crystals.

By the same process, a selection of PglZ crystals were sent to the Diamond synchrotron for x-ray diffraction. PglZ crystals produced weak diffraction to around 4 – 10 Å. As PglZ phosphodiesterase activity had been shown to be dependent on Zn^{2+} binding, crystal screens had been repeated with $ZnCl_2$ to allow potential Zn^{2+} binding and crystals produced were harvested using the same method as previously. These crystals behaved similarly to previous PglZ crystal, demonstrating weak diffraction. In addition, no X-ray absorbance peak was detected at the k-edge of Zinc (9660.5 – 9666 eV), suggesting that Zn^{2+} was not present in the crystals. In all, over 100 diffraction datasets were collected on PglZ crystals formed with and without a $ZnCl_2$ pre-incubation step. Datasets were filtered for quality and combinations were combined using xia2 Dials. The highest resolution achieved from combined datasets was 3.4 Å. Attempts were made to produce a solution by molecular replacement using Phaser with an Alphafold generated model of PglZ. Despite using different combinations of full length PglZ, dimers of PglZ and individual domains, unfortunately no solution could be found.

5.9 Discussion

5.9.1 Optimisation of protein expression and purification

Producing high quality structures of BREX components and BREX complexes will be of great value towards gaining a broader understanding BREX phage defence functionality. The design of efficient overexpression and purification protocols is an essential step towards both structural and biochemical investigations of BREX proteins. The initial difficulty encountered when attempting expression and purification of PglX is demonstrative of protein expression varying with expression system. Expression of PglX from an arabinose vector was significantly more stable than expression in the T7-driven pSAT1 plasmid backbone. Conversely, expression of PglZ from pSAT1-*pglZ* provided high yields of high purity protein. Screening expression of a protein of interest in various expression systems and plasmid backbones may require initial time and expense but identifying optimum expression conditions early can avoid greater loss of time later. As such, it is recommended that optimisation of overexpression protocols for a new protein of interest include plasmid backbones and expression systems as factors, in addition to typical expression condition parameters.

The poor expression in T7-based systems compared to arabinose systems may be attributed to several reasons. Perhaps the most obvious is that T7 expression systems can occasionally be “leaky”, whereby a single T7 RNA polymerase molecule erroneously expressed before induction can be sufficient to express significant quantities of protein. If the target protein is toxic, as is the case with PglX, this can be sufficient to cause cell damage/death sufficient to allow any cells which block expression of PglX a growth advantage. An arabinose-based system is less susceptible to such occurrences as the promoter is more tightly controlled, minimising pre-induction expression. Another potential explanation my lie in the close coupling of transcription and translation in prokaryotes [379]. In Coupled Transcription-Translation (CTT), host ribosomes will begin translating the messenger RNA (mRNA) transcript before full transcription of the CDS is complete [379]. For some genes, this coupling can be essential in maintaining mRNA stability, preventing targeting by ribonucleolytic enzymes and preventing R-loop formation and Rho-mediated translation termination of untranslated transcripts. T7 polymerase transcribes rapidly compared to most host polymerases (230 nt/s) and often host ribosomes cannot keep pace with their translation (14 – 17 AA/s).

For genes for which coupling is essential, this can be problematic. As transcription with host polymerases would be slower and more controlled (42 – 49 nt/s), this may be allowing CTT necessary for correct translation of PglX. The lack of putative fragmented forms of PglX may be indicative of this as transcripts are only partly translated before failing or degrading. It was interesting that PglZ did not display any similar issues with fragmentation or truncation. This is suggestive of PglX transcripts either containing some important secondary structure elements which aid in CTT or that transcripts contain some signalling factor which targets them for degradation. Overexpression of PglX has been shown to inhibit growth, both in this study and in the study of PglX from the pEFER BREX system by other members of our lab (data not shown). It would make sense then that excess PglX transcripts would be targeted internally for degradation to minimise the growth costs of unnecessary PglX translation and activity. PglZ on the other hand did not appear to inhibit growth rate.

5.9.2 Biophysical and biochemical characterisation

Purified PglX was observed to be a monomer in solution, as demonstrated by A-SEC experiments. The phage encoded restriction inhibition protein, Ocr, has been shown to inhibit BREX function through direct interaction with PglX [294], and exists in solution as a dimer that mimics 20 – 24 bp of B-form DNA [325]. It was predicted then that one dimer of Ocr would bind to a single DNA binding site and thus bind PglX at a 2 : 1 ratio. A-SEC experiments however suggested formation of a larger complex, accounting for two molecules of PglX and two molecules of Ocr (Figure 5.4).

PglX has previously been reported to be the methylating agent of BREX systems [281,308,314]. This has been further supported for the *Salmonella* BREX system through bioinformatic and knockout analyses in Chapters 3 and 4, respectively. The lack of methyltransferase activity demonstrated by recombinant PglX *in vitro* (Figure 5.3) is perhaps unsurprising due to the evidence of BREX systems forming multicomponent complexes [318]. It was hoped however that inclusion of additional recombinant BREX proteins would elicit some activity from PglX. While SAM is provided in the reaction buffer, it could be the case that some other co-factor is required by one of the other BREX components to elicit activity and that complex formation alone is insufficient for methylation. PglZ, for example, requires Zn^{2+} for function and may only be able to complex with PglX in a Zn^{2+} bound state. Alternatively, methylation by PglX may rely on the biochemical

function of PglZ and thus a certain PglZ substrate would need to be included in the reaction buffer. It is also possible that BREX function may require the presence of another BREX component in addition to BrxB, PglZ and PglX. This is unlikely to be BrxL as this was demonstrated to be dispensable for host methylation in Chapter 4. BrxA is known to bind DNA and was predicted to be involved in some form of regulation of BREX. It could be possible that BrxA is involved in correct assembly or loading of PglX and/or the methylation complex. Further probing of PglX for methylation activity will include BrxA in case of this scenario. The methylation kit used here provides an indirect measurement of DNA methylation through the SAH reaction product. SAH directly binds to SAM-dependant methyltransferases, providing negative feedback through competitive inhibition of SAM binding [386,387]. Finally, it is possible that SAH produced by PglX activity is then bound by PglX and prevented from producing a signal and PglX is in turn prevented from producing further DNA methylation. This is unlikely as the methylation detection kit is designed for SAM-dependent methyltransferases, through as the detection reaction is proprietary it is difficult to know for certain. As an alternative method to the indirect measurements often utilised by commercially available methylation detection kits, methylation could be measured directly by re-purifying the DNA substrate and utilising the nanopore sequencing pipeline developed in Chapter 4. This however would be a far less efficient use of money and resources.

5.9.3 Crystallisation trials and X-ray diffraction

Extensive screening of crystallisation conditions for PglX yielded no results and only on addition of SAM were crystals produced. The absence of crystallisation from PglX-only samples in the same conditions which reproducibly produced crystallisation with PglX + SAM samples suggests that SAM binding stabilises PglX structure, reducing flexibility. The poor diffraction produced by PglX crystals in SAM, SAM + DNA and SAM + Ocr bound states suggests that PglX might be inherently flexible, as reflected by the low signal to noise ratios produced for both PglX + SAM and PglX + SAM + Ocr datasets (Tables 5.4 and 5.5, respectively). This effect was consistent across crystallisation conditions and crystal morphologies and is reflected in the fact that final datasets used for structural solutions of both PglX + SAM and PglX + SAM + Ocr came from the largest crystal morphologies. The lack of diffracting crystals in PglX + SAM + DNA crystallisation conditions may be related to the design of the DNA oligos. It was assumed that DNA would bind within the groove, similar to that seen in the MmeI structure, and short DNA oligos were designed accordingly. The binding position of Ocr along the C-terminal arm of PglX however suggests a different DNA binding modality (Figure

5.16B). Perhaps then design of longer nucleotides of around 20 bp which mimic Ocr would allow oligo binding along the C-terminal in a similar manner to that seen by Ocr. Indeed, though the MmeI structure from PDB contains a 13-mer of DNA, a 29-mer of DNA was used for crystallisation and diffraction [192]. Further, elongating the oligo and mirroring it to include two BREX motifs may allow a single DNA oligo to bind two PglX molecules in a similar fashion to the Ocr dimer binding two PglX molecules across symmetric units. Though no methylation has been detected by PglX *in vitro*, it may also be beneficial to crystallise PglX bound to DNA with a non-hydrolysable analogue of SAM, such as sinefungin [192]. The weak diffraction shown by PglX will likely continue however and identification of new crystallisation conditions, crystal morphologies along with optimisation for larger protein crystals should all be prioritised where possible.

Crystallisation trials of PglZ produced three distinct crystal morphologies (Figure 5.9). Optimisation of the small hexagonal crystals did not succeed in increasing crystal size, and these were not pursued further. The rectangular prism-shaped crystals (Figure 5.9D) on the other hand, were optimised to a far larger size and presented prime candidates for X-ray diffraction. Unfortunately, though these crystals consistently diffracted (unlike PglX crystals), resolution was poor. The combined datasets produced resolutions of between 3.2 – 3.8 Å of varying quality and no molecular replacement solution was found using AlphaFold generated models. The lack of solution at this resolution may indicate that the AlphaFold prediction is not representative of the true PglZ structure, or simply that a higher resolution is necessary. Further crystallisation conditions will be explored, and crystallisation of metal bound PglZ will be revisited. In parallel to this work, pursuit of the crystal structure will include selenomethionine labelled PglZ to potentially allow phasing of a labelled diffraction dataset with the native dataset. Alternatively, crystallisation of PglZ in complex with other BREX components will be sought. This approach may also be applied for cryo-EM structure solutions. Indeed, cryo-EM may also be applicable to producing structural models of larger BREX complexes as variations in complex compositions will likely preclude crystallisation.

5.9.4 PglX structures

Molecular replacement using the AlphaFold (Figure 5.11A) predicted structure of PglX was successful in providing a structural solution for PglX. This demonstrates the utility of AlphaFold in allowing molecular replacement-based solutions to lower resolution datasets – in some cases negating the need for alternative

methods such as selenomethionine labelling. In contrast, the failure of the same method for the solution of PglZ demonstrates that this method is not infallible and that, for now, traditional approaches to producing structural solutions for lower resolution datasets will continue to be necessary.

PglX displays two obvious domains joined by a central hinge region. The N-terminal region contains the core methyltransferase regions, binds SAM and appears to be highly conserved. The SAM molecule is sandwiched between the two identified methyltransferase regions of PglX in a binding pocket close to the central hinge region and is stabilised by hydrogen bonds and hydrophobic interactions with adjacent residues (Figure 5.11). The C-terminal region shows little sequence overlap with MmeI but also demonstrates conserved regions, particularly along the inside edge of the two long C-terminal alpha helices. This would imply some other vital component of PglX function is housed in the C-terminal domain. Searching for structural homologues of the C-terminal domain (residues 672 – 1221) alone using DALI predominantly returns type II RMs and type I RM specificity subunits (Table 5.7). The type II RM homologues correspond to a region of the C-terminal domain closest to the central hinge. Further trimming to the C-terminal long alpha helices and adjacent helical bundle (residues 1049 – 1241) returns only very low scoring structural homologues (Z-score ≤6) with diverse functions likely unrelated to BREX (e.g. pyocin, tRNA synthetase, SMC proteins). The binding of Ocr along the inside face of this region is also suggestive of the C-terminal domain being required for DNA binding and possibly motif recognition. It is however as yet unclear whether DNA binds in this orientation or whether it binds in a similar orientation to that suggested by overlaying DNA molecule from the aligned MmeI structure. This is certainly plausible as the flipped-out adenine base within the MmeI DNA molecule lies in close proximity to the SAM molecule in PglX. It is also possible that conserved regions within the C-terminal domain are related to interactions with other BREX proteins and complex formation. Elucidation of a crystal structure of PglX bound to DNA would be of great value in deciphering the purpose of these conserved regions in BREX function.

Table 5.7 – DALI structural homologues of the PglX C-terminal domain.

PDB	Name	Function	Z-score
5HR4	MmeI	Type IIL RM	12.6
3S1S	BpuSI	Type IIG RM	11.1
1YDX	MG438	Type I RM specificity subunit	9.6
2IH4	M.TaqI	Adenine specific DNA methyltransferase	9.6
2Y7C	EcoKI	Type I RM specificity protein	9.3
5FFJ	LlaGI	Type ISP RM	8.8
1YF2	*gi 15669898*	Type I RM specificity protein	8.8

As discussed, Ocr mimics B-form DNA and competitively inhibits DNA binding [325]. Much of this interaction is explained by the highly negatively charged surface of Ocr (Figure 5.17). DNA binding proteins often have both non-specific and specific mechanisms of interactions, with the former produced by electrostatic interactions allowing scanning for a specific motif which is then bound specifically through hydrogen bonds formed between DNA bases and amino acids protruding into the major and minor grooves of the DNA molecule [388,389]. Ocr provides an initial negatively charged surface for interaction and then binds PglX through a series of hydrogen bonds and salt bridges, preventing DNA binding. The ability of Ocr to bind two molecules of PglX provides a significant advantage, allowing sequestering of more of an essential BREX component and inhibiting phage defence. Interestingly, calculating the hydrodynamic radius of the solved PglX-Ocr complex produces a near identical radius to the A-SEC measured figure (63.8 Å and 63.9 Å, respectively), further supporting that this complex is freely occurring in solution and not a crystal artefact. The Ocr dimer is representative of around 20 – 24 bp of a slightly bent B-form DNA molecule. As the PglX recognition motif is non-palindromic and methylation occurs on only one strand of DNA, it is not expected that two molecules of PglX would be able to bind to a single recognition motif. It is unclear at this point whether PglX binding either requires or causes DNA bending represented by Ocr. DNA binding by RMs can require energetically unfavourable bending of the DNA substrate to allow access to binding pockets [186]. The inherently bent conformation of the Ocr dimer allows more energetically favourable binding relative to DNA. It can therefore be hypothesised that in the same way PglX binding with Ocr would be more energetically favourable and would inhibit displacement by DNA, ensuring that phage defence is tightly repressed.

5.9.5 Conclusion

In this chapter biochemical and biophysical investigations were carried out into the core BREX system components, PglX and PglZ. Overexpression and purification optimisation of PglX ultimately provided quantities and purities sufficient for crystallography, though this required the tighter regulation provided by an arabinose-based expression system. PglX is a monomer in solution and shows no *in vitro* methyltransferase activity, even when combined with BrxB and PglZ. Extensive crystallographic screening and optimisations provided several conditions producing several crystal morphologies, though all diffracted weakly and only the largest crystals provided sufficient diffraction for higher resolutions. The first structure of PglX from a BREX phage defence system has been produced to a resolution of 3.4 Å. In addition, the structure of PglX bound to the RM inhibitor Ocr – previously shown to inhibit BREX defence [294] – was solved to a resolution of 3.5 Å. As suggested by A-SEC, Ocr binds PglX as a heterotetramer at a 2 : 2 ratio.

The core methyltransferase domain of PglX is conserved and flanks the SAM binding site. The C-terminal domain also showed conserved regions along an electropositive surface but does not align with the structural homologue, MmeI. Ocr interacts with this electronegative region through electrostatic interactions, allowing the formation hydrogen bonds and salt bridges. The binding position of Ocr and the position of the superimposed DNA molecule from aligned MmeI present two potential orientations of DNA binding and potentially implicate the C-terminal in DNA binding and motif recognition. Crystallisation trials of PglX in complex with DNA did not provide diffracting protein crystals. As the structure of PglX bound to DNA would be of great value in further deciphering PglX activity and BREX function, further trials towards diffracting protein crystals will be carried out, this time taking into account the binding position of Ocr.

PglZ was previously shown to be a metal dependant phosphodiesterase by others in our lab. Crystallisation trials and optimisations of PglZ produced large protein crystals which only weakly diffracted. This resolution was insufficient to provide a solution using molecular replacement with an AlphaFold generated model. As PglZ is the only BREX component common to all BREX subtypes [308], elucidation of a crystal structure may provide great insight into the function of diverse BREX modalities and allow further identification of common molecular themes across these subtypes. As such, attempts towards this goal will continue.

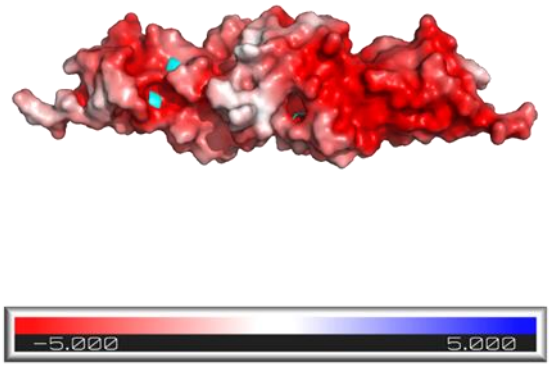

Figure 5.18 – Surface charge of Ocr. *As calculated by APBS.*

Chapter 6 – Rational modification of BREX target recognition

6.1 Introduction

Preceding work in this study has provided bioinformatic analysis of the *Salmonella* D23580 BREX system, characterised defence activity against the Durham Phage Collection, demonstrated essential genes for both methylation and restriction, and provided structures of PglX in complex with SAM, and PglX in complex with SAM and the BREX inhibitor Ocr. This chapter will document attempts to rationally modify the recognition motif of PglX, altering the specificity of the *Salmonella* BREX system.

PglX is the methyltransferase component of BREX systems that methylates the host genome and is a feature of all BREX subtypes except type IV systems. PglX recognises non-palindromic 6 bp DNA motifs and generates N6mA modifications at the fifth adenine residue [308,314]. While it has been shown that BrxA [316], BrxC (see bioinformatic analysis in Chapter 3) and BrxL [318] possess DNA binding modalities, only PglX has demonstrated specificity for DNA motifs to date [314]. The structure of PglX was described for the first time in Chapter 5, providing new insight into this essential BREX component. Work on producing the structure of the closest structural homologue to PglX currently available in the PDB, the type IIL RM MmeI, included rationally modifying the recognition motif of this enzyme, allowing methylation and cleavage of different DNA targets [192,194]. This was achieved first through sequence alignments using homologues with known or predicted recognition motifs [193] and modelling conserved residues onto the molecular structure of MmeI bound to DNA [192]. Though no structural model of PglX bound to DNA is available, it was postulated that using sequence alignments the recognition motif of PglX could be similarly altered. Rational design of RMs with novel target motifs has long been pursued by the biotechnology industry. Success in altering the recognition motif of PglX would potentially provide valuable insight. For instance, this would demonstrate whether PglX is indeed the sole determinant of BREX system specificity or if some other factor can also recognise BREX motifs. Rational retargeting of PglX could also change both BREX methylation activity, and also BREX defence, so that it provides defence against a new subset of phages.

6.2 Design of PglX mutants towards altered DNA motif specificity

6.2.1 Identification of target residues and mutations

Rational engineering of RM recognition motifs would be highly beneficial to molecular biology and biotechnology industries and several studies have been conducted towards this goal with some success [192–194]. Modification of motif recognition is simpler when a crystal structure of the enzyme bound to DNA is available as the residues interacting with DNA bases can be readily identified [192]. Rational modification of motif recognition however remains possible in the absence of a crystal structure through sequence alignments of enzymes with known recognition motifs. This method was applied for the type IIL RM, MmeI, a structural homologue of PglX [194], allowing the production of REs with new specificity. As such, the same approach was applied to attempt to modify the DNA recognition sequence of PglX, with the help of Dr Rick Morgan. Protein sequences from BREX related methyltransferases with assigned DNA recognition motifs were collected and added to the sequences of BREX methyltransferases identified in the REBASE RM database [160], using those that displayed high sequence similarity scores to PglX in BlastP (<E100), resulting in 32 distinct sequences (Figure 6.1). Most of the predicted motifs from REBASE are inferred by matching the BREX methyltransferase to an observed N6mA modification in genomic sequencing data. MmeI is the closest structural homologue of PglX and the residues essential for motif recognition have been identified from structural data [192]. As with PglX, MmeI recognises a 6 bp motif (TCCR**A**C) and produces N6mA modifications at the 5th adenine base. Structural alignments of MmeI and PglX allowed identification of the residues of PglX that aligned with the residues involved in MmeI motif recognition and provided regions in which to focus the search for covariation in BREX methyltransferase sequence alignments. It is important to note that the design of PglX mutants was carried out prior to elucidation of the crystal structure of PglX and thus used a predicted PglX structural model produced by Phyre2 [390]. Candidate residues and alterations were then chosen based on these alignments. For example, for motif position -1 (relative to the modified adenine base); lysine is conserved at residue 802 for enzymes recognising cytosine at this position, or histidine is conserved at residue 838 for enzymes recognising guanosine at this position, or asparagine is conserved at residue 838 for enzymes recognising adenine at this position (Figure 6.1). This resulted in the design of 23 mutants targeted at altering all five of the non-adenine bases in the PglX recognition motif (Table 6.1).

```
Sal_PglX____GATCAG   752  E-IRNFKFENGKTRS------AVRNDEYYFREGITWSKISQGN-FCVRYRPKGFVFDD-TGRC-GFSNNKNELLY--AAGLMCTPVVNHYLSILAPTLSFTSGELASVPYPE----IEDEIIE---LV  860
pEFER_PglX___GCTAAT   755  E-LRQYQGS------------FLRGEKFYFKAGLTWSKVSSGI-LSFRLFDEGFLFDT-GGLC-AFSDNIEY-----IAAILNSKVSLDIMSILAPTLNFTVGTVSSLPIVE-------GIPRLTENA  854

Sen6480IV____GTTCAT   754  RICQHSGA------------YPRNKDKYFTKGLAYTNISSAK-FAARYTDTGFIFDQ-KGSM-FFSEKQNGIQI--AASLLHSSYGTSLLEAICPTLDFNPGSLGNLPILD------VDIDKISNIF  857
SenPU131IV__GAANCAG   748  E-LKDFVEELNKLRPG----GRLKNQEYYFLECINYSSLSSGF-FSARYTNTGFLFDT-KGSG-IFPKDGNVKA---FLSLLNSNVIQDFLDILCPTLDYSSIGINSLPVKL----------LSIPQI  854
Sen373III____CANCATC 748  K-IRNFYNDKGKLRS------RPQNIQFYCKEGLTWTSLTISS-LSMRYVPNGYIFDA-KGPM-CFPINAADIWD--ILGYTNSKIINIFLKHLAPTMDYSQGPVGNVPFKS-------PTRNITNII  856
Cko11077IV___TGACAG   749  Q-IKAIPHS------------VVANESNYFKPGITFSTVSSSG-YGFRLFDNGFIFDN-KGAS-IFIDGEKRLY---LLAILNSKIFELTVHSITPTISLQPGDVAKLPVIE---GDSTQMEKVISLS  854
Eco4174I_____GCACAG  751  AIIANGKA------------YPRSKDYYFKESLTYSATSSSY-FGIRYSNPGFIFDA-KGSS-CFSDSTTLKL---GLGFLSSQLASFLLKAINPTIEFQTGDINILPFKK--------VCEIEHFV  851
Eco8620I____CRARCAG  751  E-LLDFAASLYGSPTR-----TIKNIPFYFREGATWSTISSSD-FSIRYSPTGFISET-KGAV-CFADDKDILLS--ILGFGNSKLVNYFLKSLSPTLDYHEGPIGKLPFKG------KLKAQVLENV  861
SenSARA26III_ACRCAG  751  E-LKSFADETTGRIRS-----HNYNGNYAFREGFSWSGISSGS-FAVRHVSPGFMFDA-KGPM-GYVNNKNDLYP--IEAFLNSTVANHLIKMLAPTLDFKLGHILNLPFIE----AKEDIITLEKLT  863
Sen5722III__GNGGCAG  746  A-IKNQEAKTSGDRGW-----RATSEEFYGKDGITWGGLTNAF-LTFRWSDYGALFDTNKGPM-MFPQENAYY----LLGFLNSPLAIEYCKLLNPTISFQNADINRIPFFLPSHDTNNKIQE---KV  858

Kor51II_____RTCGAG   749  RYSKPSNG------------SLRNKEYYFKECIAWADVSSEN-YSFSKFSPGCIANS-TSHF-MILKDNKYANL--VLSFLNSNYAKYVFNIINPSIHLNPGEVASLPIPK------NASEHDWSHT  852
Eco9699II_____TAGARC 740  A-IRHEQNS------------GVGPEELLFKECITWTDVTWR--VSARFLPEGHLSDH-AGPC-AYFDEKDKLYT--ALAVFNTPLGENWSNLLNPTLHFQAGDFKKLPYPN-----NISHSDLLPLV  843
Kpn156V_____CRTGATT  750  A-IKNFKDENGKLRS------RPQGLDLKFKEMLSWTSLSSTY-LGVRYYPNGFISDQ-NGNF-LVPNNRFENDIYFILGWLCSSTANRIVKMLNPTMHILVSNISNLPIPM----DEVTIKEVCDIT  863
Kpn9178I____GNGCGAG  739  NHYKNNHSS------------RIIDEKFWYLPGITWTDITSSG-TGFRYLPVNTTYDT-TGIS-FFLANNENIPK--FLGALNSKPATHILSTINPTLHANLVDVKSLPIPSLTDYDSN-------IA  842
Ror431III____CTRGAG  751  E-IKNCVDENGKLRS------RQQNSTFYFRESVSWSDVTSGD-NAFRYIPKGFIFDA-TGHS-SFYNNDSDLHL--CLAFLNNIYCSYITKILNPTIHFHVGYFNRLPFPE---DIKKHSDEIISLA  863

Ecl388I_____CGGNAAG  751  E-MKDAVIKRYNGGSYT---KEIRSEDRYFKDSITWSALTAGT-PSFRLSTYGAIFDS-AGSS-MFPIENTYE----ILGLLNSNVSAYILKLLNPTLNYGAGTVANIPVAL------PRSNALYTMI  861
CfrMH16VI____CTAAAG  749  E-IRNYSDENGKQRS------RPQNIAYYRKKSITWSFISSSY-FGARYSDTHAIFDV-AGSS-AFPRKDCIFN---YTAYMCSSVAHYFMKLMNPTLNFQVGNVSNLPVPS-----QMNESLINGIA  858
SenWT8IV_____CCAAAT  754  D-IKCETIEKYPQLSWDNLGWKITNEPDFFKKSITWSFISSSN-FGVRCSLGGAIFDV-SGSS-AFPKDSDYFT---TAGFLCSRVAYEFLKVLNPTLNFQVINLNSLPWIS----PKEKSSSIEQIV  870
Sen5800I_____CCAACAC  749  D-IKCETLEKYPQLSWDNLGWKITNEPDFFKKSITWSFISSSN-FGVRCSLGGAIFDV-GGSS-AFPKDSVYFT---TAGFLCSRVAYEFLKVLNPTLNFQVINLNSLPWIS----PKEKSSSIEKIV  865
Eco11117Put__CGAAAC  749  DLMQTFSGH------------RHDGKSHYFKEGVTWTFISSSN-FAARYSSPGFVFDV-SGST-FFVNSPRA-----FTAFLCSKISETVLKMLNPTLNFQVGNIKSLPILD---GNNIFSDSDYLLA  853
SmaUMH8I_____GCGAACB 736  SHYERMG-------------GIYPSKFRDKVGICWSKITSGT-VSFRLKQSDFEYDS-ASPV-IFKKDFSFDPY--ILSLLNSKPYIFILNALNPSMNTQVADVLSLPVIN---LNADNKKTLNEIE  841
Sen5794III___ACGAACB 755  ELQNTLHPSGNRIWA------HNFVLDSIFKESIVWSKITSGK-PCFRYSPKGFLFDDASGVC-TFNEGTKEF----LIGLLCSNINTSLQEIINPTLNIQPANIRDIPIPK------DISSFPKNTV  864
Yru10476I____AGGAAG  753  E-LFNMQGNGYKVGS------TNHNLEYIFKPAIVFSKITSST-PHFRYAPQGFLFDDASGLC-AIKDEKQTFK---LLAFLCSGVCSVFNNIINPTLNLQPGNVANLPAPL--------INESEYIA  860
Cdu23823II___GTGAAG  744  D-IKNHSKGA-----------AIRNEKYYFKSCITWSRISSGS-ISFRYLDKGFVHND-ASCFISLPSEESKYT---ILGYLNSPLKKPLLEDLNPTINLLPGMLNETPYID----------NYCPEV  844
EcoNIH6II____ATGAAG  737  E-IKDYVVDRYPYLNGNY-ALVVKNEATYFQNGILTSRITSGG-LGFRIKDACELYSD-ACTA-AFPKDSNV-----ILGILCSKVS-YLMQQLNPTINFQASDLEKIPVIY-----PENSQCFSELV  848
Eco9010II____GGTAAG  742  ELRKNKA-------------NLRNKDMYFQEGGTWTVVSTTG-FSMRYMPKGFLFDQ-GGSA-VFCENNDELSIYNILACMNSKYINYSASLICPTLNFTTGDVRKFPVIK-----NNHLED---LA  844
KpnNIH50I____GCYAAG  733  LYYDSHG-------------GLSNSKFWNKLGITWSLIGTKS-VSFRIKPKHLQYSS-GSPV-IFCEDLNQTYV--TLAFLNSTVGQYYLSAISPTINTTVNDVLSLPVPELDHMHSDKIIS---LT  838

Vdi96II_____GNCYTAG  737  LHYRKDKIS------------RITSTKSQDLPGITWSIISSSSGRAFRHLDETSLFNS-VSPS-LFAKDKTDVND--WLTYLNSNIASYLLELINPTLATNVGNVLALPVPH----------RIPEA  837

Kcr10483I____GGCYAC  743  E-LKKFSGF------------ELRNEAYFFQKGLTWGDVSSGD-FSMRYHDEKGLFEG-RGPM-AFSNDFEY-----LLGFMNSNIAKAFLDLFCPTLTYNVGDVVKVPLSFTPQKMKNEIIN---NT  846
EcoC9964I____ACCYAC  749  K-IRNFGIENGKVRS------HNYNLDFIFRKGITWSDVTSGS-NAFRVLPQGFLFDG-RGSS-GFCEDNLRPT---ILCLLNSKLSYNIINIINPTIAINVGEIAKIPLHN---NLMNNKNNIENLT  860
EcoA23Put____AGCYAG  746  AITTKKHPTENRIWA------TNFNLDYIFKPNVNWNSIASSK-LSFRFSPGGELFAS-TGLA-CFPTNNFEF----IISYLNSSIAEYFLSIFAPTLSANPGDIARLPIYD------INKFKHQQQV  854

Kae10004II___GCCKAG  733  AFYASNG-------------GMINSKFLNRTGVTWQTVTSKK-NAFRLKQKSSIYSS-VSPC-LFFKNDNENDILNTLASMNSEVIAYVMKAINPTLQANPGDVLKLPVCN---------ITNKKNI  834
```

Figure 6.1 – Example of sequence alignments of PglX homologues. *Sequences are sorted by the DNA base at position -1 of the recognition motif (relative to the modified adenine). Residues aligning with the residues required for DNA binding of MmeI at this position are highlighted in green. The same method was performed for each of the non-adenine bases in the* Salmonella *BREX recognition motif. The potential significance of K782 and I783 in PglX is discussed in 6.5.2 and these residues are highlighted in yellow.*

Table 6.1 – PglX mutations for altering PglX recognition specificity.

Mutant number	Motif position[1]	Current motif base	Mutations	Predicted change	Resulting motif
1	-1	C	T802T; S838H	G	GATGAG
2			T802N; S838H	G	GATGAG
3			T802A; S838N	A	GATAAG
4			T802G; S838N	A	GATAAG
5			T802V; S838A	T	GATTAG
6	-2	T	K782F; D801V	A	GAACAG
7			K782R; D801D	G	GAGCAG
8			K782R; D801S	G	GAGCAG
9			K782D; D801S	C	GACCAG
10			K782A; D801S	N	GANCAG
11	-3	A	A684L; S687R	C	GCTCAG
12			A684G; S687K	C	GCTCAG
13			A684K; S687D	G	GGTCAG
14			A684R; S687A	G	GGTCAG
15			A684H; S687S	G	GGTCAG
16			A684V; S687Q	T	GTTCAG
17			A684T; S687Q	T	GTTCAG
18	-4	G	A766R; R768Q	C	CATCAG
19			A766R; R768D	C	CATCAG
20			A766H; R768F	A	AATCAG
21			A766T; R768H	A	AATCAG
22			A766V; R768A	T/N	T(N)ATCAG
23	+1	G	Swap entire loop with pEFER PglX loop (AAs 591 - 600)	T	GATCAT

[1] Relative to modified adenine base.

6.2.2 Developing a complementation system to assay phage defence activity from PglX mutants

Following the design of the PglX mutants, an assay system was required to test their function. Generating each of the mutants individually in the 17.9 Kbp pBrxXL$_{Sty}$ plasmid would be costly and time consuming. Instead, a complementation system was designed, utilising the pBrxXL$_{Sty}$-Δ*pglX* construct produced in Chapter 3. The *Salmonella* BREX *pglX* gene was subcloned into a pBAD30 plasmid backbone by LIC, this time without the additional fusion tags and with the incorporation of a ribosome binding site 7 bp upstream of the *pglX* start codon. pBAD30 encodes complementary origin of replication and resistance markers, allows induction of *pglX*, and has been shown to function in co-expression with pBrxXL$_{Sty}$ in previous assays in Chapter 4 (Table 4.4). Complementing the pBrxXL$_{Sty}$-*ΔpglX* construct with the pBAD30-*pglX* plasmid in EOP assays provided phage defence against TB34, albeit slightly lower than that seen from the DH5α pBrxXL$_{Sty}$ construct (Table 6.2). Full details of the methodology used for complementation EOP assays can be found in Chapter 2. Next, a marker was required to indicate whether the recognition motif had been modified. Again, it was preferable to initially test this through functional EOP assays as sequencing for methylation changes caused by all 23 mutants would be time consuming and expensive. Fortunately, the activity of pBrxXL$_{Sty}$ had already been characterised against the Durham Phage Collection and phages in this collection had been sequenced to allow enumeration of BREX recognition motifs. This allowed the identification of one phage, Trib, which was susceptible to both *E. coli* and pEFER BREX systems but contained no native *Salmonella* D23580 BREX recognition motifs(Figure 3.6). Trib did however encode all of the predicted modified motifs designed in 6.2.1 (Table 6.1). In this way, all mutants could first be screened for phage defence activity against phage Trib and the recognition motif of active mutants could be determined by sequencing.

6.3 EOP assays for phage Trib against PglX mutant strains

EOP assays were carried out in triplicate for all 23 pBAD30-*pglX* mutants co-expressed with the pBrxXL$_{Sty}$-Δ*pglX* construct in DH5α. Mutant 4 consistently produced poor overnight growth and failed to provide sufficient bacterial lawns for plaque enumeration, even after increasing the inoculum volume. Mutant 3

appeared to provide around 10-fold protection against Trib, similar to phage defence levels provided by the *E. coli* and pEFER BREX systems against this phage (Figure 3.6). Mutants 8, 10, 15 and 22 showed sporadic reductions in EOP, usually around two-fold. Remaining mutants demonstrated no noticeable reduction in plaquing efficiency. To confirm whether the BREX system was still functional, mutants 3, 8, 10, 15 and 22 were also assayed against phage TB34. Mutant 3 showed a reduction in EOP similar to that shown against Trib, though around two-fold higher than produced by the DH5α pBrxXL$_{Sty}$ strain (Table 6.2). The remaining mutants did not show any reduction in EOP against TB34 and were deemed to be inactive. The discrepancy between activity against TB34 from the DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3) strain compared to that seen in the DH5α pBrxXL$_{Sty}$ strain, despite a high number of each motif in the TB34 genome, alongside the difference in EOP values between the DH5α pBrxXL$_{Sty}$ and DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX* complementation system, suggested that the BREX system may not be as effective in this complementation format (Table 6.2). As such, the T802A and S838N mutations in mutant 3 were also generated directly in the *pglX* gene in pBrxXL$_{Sty}$, resulting in pBrxXL$_{Sty}$(*pglX* mut.3) and the new construct was similarly assayed against both TB34 and Trib. This format reduced the EOP values further for both TB34 and Trib against DH5α pBrxXL$_{Sty}$(*pglX* mut.3), though still not quite as low as the wild type system against TB34 (Table 6.2).

Table 6.2 – EOP results for wild type* pglX *and* pglX *mutant 3 against phages TB34 and Trib*[1]*. *The number of recognition motifs in each phage genome is shown for the wild type* Salmonella *BREX system and for the predicted recognition motif of* pglX *mutant 3.*

		TB34	*Trib*
# Genomic Motifs	Wild type (GATCAG)	120	0
	pglX mut.3 (GATAAG)	93	83
EOPs	DH5α pBrxXL$_{Sty}$	$2.86 \times 10^{-2} \pm 1.54 \times 10^{-2}$	1.19 ± 0.11
	DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*	$9.63 \times 10^{-2} \pm 6.14 \times 10^{-2}$	$7.52 \times 10^{-1} \pm 0.035$
	DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3)	$1.45 \times 10^{-1} \pm 1.33 \times 10^{-1}$	$9.98 \times 10^{-2} \pm 1.02 \times 10^{-1}$
	DH5α pBrxXL$_{Sty}$(*pglX* mut.3)	$1.03 \times 10^{-1} \pm 9.13 \times 10^{-2}$	$4.28 \times 10^{-2} \pm 4.33 \times 10^{-2}$

[1] Green indicates EOPs >0.5, yellow indicates EOPs ~0.1 – 0.5, orange indicates EOPs ~0.05 – 0.1, red indicates EOPs < 0.05.

6.4 Methylation profiling of PglX mutant 3

Next, the host genomes of DH5α pBrxXL$_{Sty}$(p*glX* mut.3) and DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3) strains were sequenced and genomic methylation levels were assessed using the previously designed nanopore sequencing and bioinformatic pipeline from Chapter 4. Unfortunately, nanopore sequencing did not generate sufficient data for required genomic coverage and results were therefore not of sufficient quality to be conclusive. Neither Tombo nor nanodisco could identify any recognition motifs with this low coverage dataset, even from positive controls. Analysis with Megalodon produces per site methylation data, however, and individual sites with sufficient coverage could be assessed. Megalodon suggested that the genomes of DH5α BrxXL$_{Sty}$(p*glX* mut.3) and DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3) strains were methylated at the predicted GATAAG motif however, it also suggested that these genomes were still methylated at the wild type GATCAG *Salmonella* BREX motif site. GATTAG sites demonstrated no methylation while GATGAG sites had demonstrated high methylation signal from all high coverage datasets tested over the course of this study, even WGA samples. As such, it appeared that *pglX* mutant 3 was producing modification at GAT**M**AG motifs. Unfortunately, time and funding would not allow further repetition of this analysis by nanopore sequencing. To provide a more conclusive examination of genomic methylation in DH5α pBrxXL$_{Sty}$(p*glX* mut.3) and DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3) strains, these samples were also sent for PacBio sequencing (Dr Andrew Nelson, Prof Darren Smith, Northumbria University) as this had demonstrated more sensitive and accurate analysis than nanopore sequencing in Chapter 4. PacBio sequencing confirmed the preliminary results provided by Megalodon and showed that both the DH5α pBrxXL$_{Sty}$(p*glX* mut.3) and DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3) genomes were methylated at almost 100% of GATMAG motifs (Table 6.3). Furthermore, the DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX* included as a control also demonstrated almost 100% methylation at GATCAG sites, demonstrating efficient methylation from the complementation system.

Table 6.3 – Genomic methylation by* pglX *mutant 3. *Sequencing was performed on genomic DNA by PacBio.*

Construct	Motif	Motif sites in *E. coli* DH5α genome	Methylation percentage (%)
DH5α pBrxXL$_{Sty}$	GATCAG	2947	97.8
DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*			99.4
DH5α pBrxXL$_{Sty}$-Δ*pglX* + pBAD30-*pglX*(mut.3)	GATMAG	5293	99.6
DH5α pBrxXL$_{Sty}$(p*glX* mut.3)			99.8

6.5 Discussion

6.5.1 Design of PglX mutants

Rational modification of RM system specificity is a desirable capability for molecular biology and biotechnology and would greatly expand the utility of RMs. Here, the rational modification of the recognition motif of the BREX methyltransferase, PglX, has been demonstrated and was shown to redirect BREX phage defence specificity and methylation targeting. That the specificity of PglX has been broadened rather than discreetly altered shows the complexity of the interactions involved in DNA motif recognition and the importance of having either closely related homologues for the application of sequence alignment-based methods or structural models of the protein in complex with DNA. Greater success and understanding were possible when altering the recognition motif of MmeI using structural data rather than sequence alignments [192,194]. For the initial alteration of MmeI specificity, sequence alignments were performed using other RMs from the MmeI-like family, all of which display high similarity and have documented recognition motifs [174]. BREX methyltransferases and related methyltransferases used for this study however are comparatively poorly characterised and, in many cases, the recognition motifs of these enzymes were inferred in the absence of empirical data. This is perhaps mirrored in the lack of success in switching recognition motifs for the other 22 mutant PglX variations. Further success using this method would likely first require proper characterisation of PglX homologues. For RM enzymes, this would simply require restriction digests of target sequences and analysis on an agarose gel [174]. BREX systems do not cleave DNA however [308,314] and PglX has not demonstrated methyltransferase activity *in vitro*

(Figure 5.3). As such, identifying recognition will require analysis of methylation motifs through sequencing methods, similar to those described in this study. Perhaps an easier approach would be to produce a structure of PglX bound to its target DNA motif, though this has so far proven elusive, as discussed in Chapter 5. As an alternative approach to rational design, directed evolution-based approaches could be applied, allowing the production of PglX variants with new recognition motifs independent of structural and sequence data input [391]. Libraries of PglX mutants could be screened for activity by assaying against phage Trib to ensure BREX phage restriction has not been impaired and then active mutants could be screened for host methylation by sequencing, as has been described here.

6.5.2 PglX mutant 3

Of all the mutants designed, only mutant 3 demonstrated active phage defence against phage Trib. EOP was reduced by between 10 and 20-fold, roughly in line with phage defence provided by the *E. coli* and pEFER BREX systems against the same phage (Figure 3.6). The EOP against DH5α pBrxXL$_{Sty}$(pglX mut.3) was reduced by more than half compared to DH5α pBrxXL$_{Sty}$-ΔpglX + pBAD30-pglX(mut.3), suggesting that phage restriction does not function as efficiently in the complementation system, in which PglX is overexpressed. This is further supported by the reduced defence activity conferred by the DH5α pBrxXL$_{Sty}$-ΔpglX + pBAD30-pglX strain in this system compared to the DH5α pBrxXL$_{Sty}$ strain. Phage defence then may be disrupted slightly by the over availability of PglX molecules, implying that the balance of BREX complex formation and composition is important for effective defence. Host genome modification on the other hand proceeds as normal and the overabundance of PglX molecules does not appear to inhibit methylation activity (Table 6.3).

While a structure of PglX bound to DNA is not currently available, the unbound structure of PglX produced in Chapter 5 does provide the location of the mutated residues in mutant 3 (Figure 6.2). This allows comparison of these positions relative to the two predicted DNA orientations provided by the aligned MmeI DNA molecule and the 20 bp DNA molecule represented by Ocr produced in Chapter 5 (Figure 6.2A). Both the T802A and S838N mutations are located in the C-terminal region close to the hinge region. This precludes identification of potential interactions from the DNA represented by the Ocr molecule as this DNA molecule does not extend to this region. It remains possible however that this molecule is

representative of DNA binding orientation and that the C-terminal arm is required for initial DNA binding, but motif recognition is produced by residues closer to the hinge region. Together these data may help design oligos used for further structural studies of PglX bound to DNA. Binding of PglX to the Ocr-aligned DNA molecule may be representative of the DNA binding orientation but this sequence would need to be extended further into the hinge region to allow motif binding. This approach may allow stabilisation of the PglX-DNA structure in a way the shorter 12-13 bp oligos utilised in Chapter 5 did not and, as a result, be more amenable to crystallisation and X-ray diffraction.

Further insights can be drawn from comparison of the residues mutated when modifying the motif recognition of MmeI at position -1 (Figure 6.2C). MmeI recognises guanine at this position through a combination of two residues: R810 forms a hydrogen bond with guanine in the major groove, and K489 protrudes into the minor groove and forms nonspecific hydrogen bonds with either guanine or adenine bases. An A774L mutant was shown to prevent binding of an A-T base pairing at position -1 through steric interference, switching specificity from R:Y to G:C [192,194]. The T802A and S838N mutations produced in PglX mutant 3 correspond to the positions of the A774 and R810 residues in MmeI, respectively (Figure 6.2C, highlighted in cyan and green, respectively). It is interesting that MmeI motif binding at this position is stabilised by interactions from opposite sides of the DNA molecule in both the major and minor grooves. It stands to reason then that another residue in PglX on the N-terminal side of the hinge region is required for motif recognition. The K489 position of MmeI provides this function but does not have a corresponding PglX mutation. The K516 residue of PglX aligns closely (Figure 6.2C, shown in orange) but no obvious covariation is evident in sequence alignments (Figure 6.3). Mutations at the T801 residue of PglX corresponded to the A774 residue in MmeI, but T801 is far removed from the MmeI DNA molecule in the aligned structures. K782 and I783 residues in PglX are located closer to the A774 residue in MmeI and protrude towards the DNA molecule in a similar fashion. It is possible that these residues play a similar role to A774 in MmeI and that residues with bulkier side chains maintain binding specificity, though no obvious co-variation is evident here either (Figure 6.1). It would be interesting to produce T802A and S838N mutations separately to determine whether both are responsible for differential motif recognition or if one mutation is redundant. Once again, this analysis is accompanied by the caveat that some of the recognition motifs of the methyltransferases used for sequence alignments were inferred. While direct investigation of recognition motifs by these proteins and expansion of the collection of methyltransferases utilised may

improve the effectiveness of this methodology, acquiring a structure of PglX bound to DNA would allow identification of interactions and would likely provide greater success [192,194].

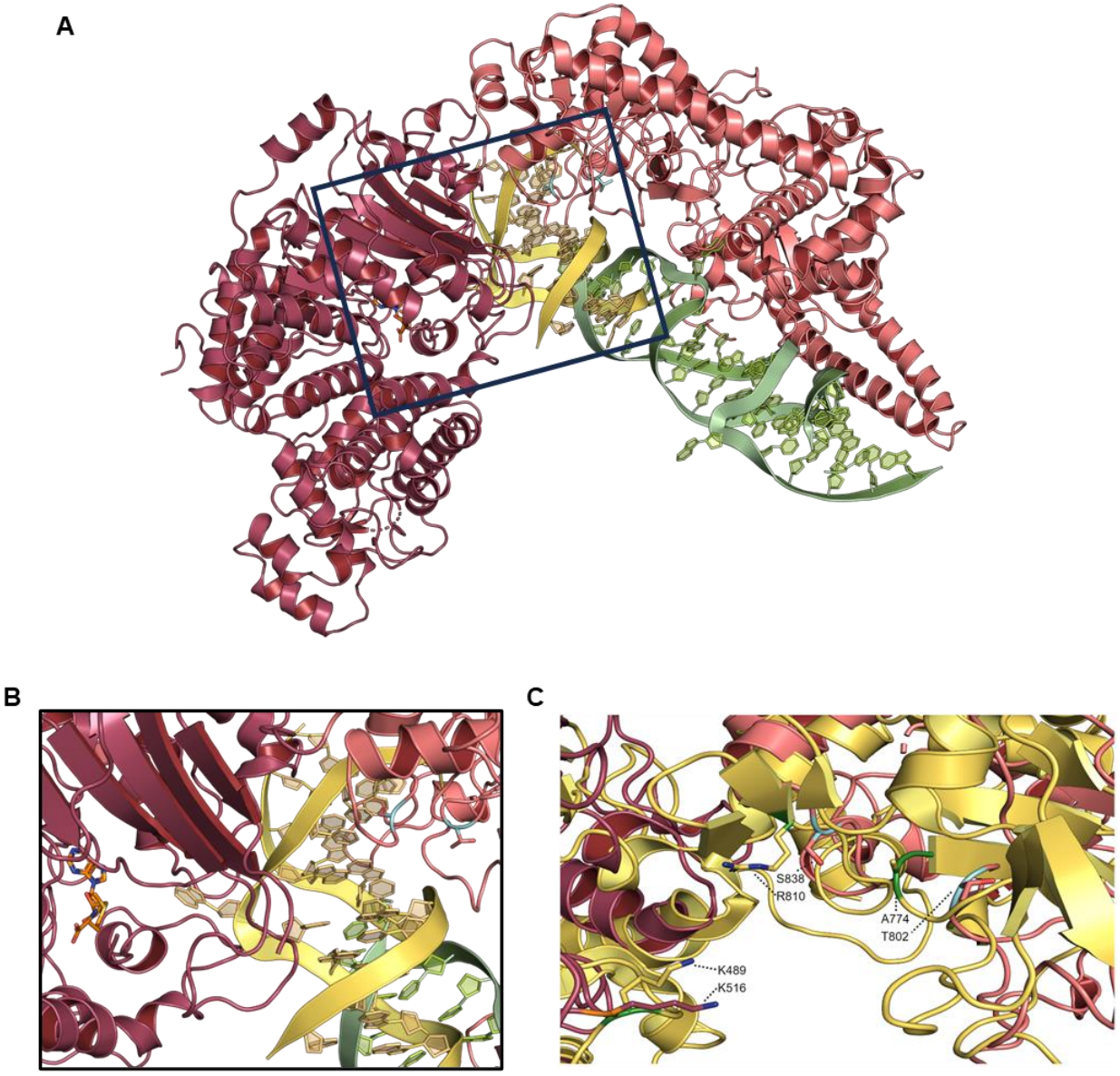

Figure 6.2 – Position of the mutated residues in* pglX *mutant 3. *A; Structure of PglX produced in Chapter 5 aligned with the DNA molecule from MmeI (shown in yellow) and the aligned Ocr DNA molecule from Figure 5.16 (shown in green). Mutated residues are shown in cyan. Black outline indicates the region magnified in panel B. B; Magnified view of A showing the positions of mutated residues relative to the aligned MmeI DNA molecule. C; Structure of PglX aligned with MmeI showing the relative positions of bases required for motif recognition (shown in cyan and green for PglX and MmeI, respectively). K489 of MmeI does not have a corresponding mutation in PglX but aligns closely with K516 in PglX, shown in orange.*

Mutation of PglX and subsequent alteration of the recognition motif successfully elicited phage defence activity against phage Trib, broadening the target range of BREX by allowing recognition of GATAAG motifs within the Trib genome, while producing additional host methylation at GATAAG motifs. Together these data demonstrate that PglX is the sole specificity factor of BREX phage defence and is responsible for discriminating between self and non-self in both phage restriction and methylation complexes. Change in specificity of BREX mediated phage defence through just three point mutations in a single component would allow quick adaptation to new phage threats. That PglX recognises non-palindromic DNA motifs on a single strand further facilitates alterations in BREX system specificity as fewer residues are implicated in motif binding. Rapid adaptability and evolution are vital factors in the phage-bacteria arms race and increase survivability of the local population [392]. Indeed, phase variation is common in *pglX* genes, but not other BREX components [308,393], implying that phase variation in PglX is a means of altering BREX defence specificity.

6.5.3 Conclusion

In this chapter, rational mutation of PglX allowed broadening of BREX phage defence to target a previously resistant phage. In absence of structural detail from a model of PglX bound to DNA, target mutations were designed by aligning the protein sequences of PglX homologues and looking for covariation at residues aligning with those involved in motif recognition by the structural homologue, MmeI. Twenty-three PglX mutants were designed aimed at altering DNA base recognition at a different position in the 6 bp BREX recognition motif. Mutants were then assayed for phage defence activity against phage Trib, which does not encode any of the wild type *Salmonella* BREX motifs (GATCAG) in its genome but does contain each of the predicted new recognition motifs produced by the PglX mutants. Recognition was then confirmed by sequencing of the host genome and identification of methylated motifs. Only one mutant provided defence against phage Trib and demonstrated methylation at GATMAG motifs, effectively broadening target specificity rather than discreetly altering it to a new motif. Thus, PglX is the sole specificity factor in BREX phage defence and provides motif recognition in both phage restriction and host methylation. This explains why phase variation is observed in PglX but not in other components, allowing adaptation and evolution for targeting of new phage threats. Further attempts in the retargeting of BREX defence through modification of the PglX recognition motif would benefit greatly from a DNA-bound structure of PglX as this would confirm which specific bases are involved in binding to the wild type recognition sequence.

Sequence alignment-based methods for altering DNA motif specificity require a library of highly similar homologues with confirmed motif recognition. The ability to produce restriction enzymes with desired specificities would be highly beneficial to molecular biology studies.

```
Sal PglX      GATCAG   507  VANPPYMGGKGMNSELKEFAKNNFPDSKADLFAMFMQNAFSLLKENGFNAQVNMQSWMFLSSYEALRNWLLDNKTFITMAHLGARAFGQIS-GEVVQTTAWVIKNQHSERYQPVFFRLID--GREEVK  631
pEFER PglX    GCTAAT   508  IANPPYMGSKYQTAEVKMFLKDNFKGYEKDLFSAYIVRNLQLAKEHGQLGFMTPFVWMFISSYEQLRSTLIDDEIISTLIQLEYSGFD----GATVPICTFTLTKGHIPEYIGSYIRLSDFKGAANQA  631

SenPU131IV__GAANCAG   505  VANPPYMGGKGMNGELKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRGWLLANKTLITMAHLGARAFGQIS-GEVVQTTAWVISNNHTERYQPVFFRLIE--GNEEQK  629
Cko11077IV___TGACAG   505  VANPPYMGGKGMNGELKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRGWLLDNKTFITMAHLGARAFGQIS-GEVVQTTAWVMSNNHFERYQPVFFRLIE--GNEELK  629
Eco8620I____CRARCAG   507  VANPPYMGGKGMNGELKEFAKNNFPDSKADLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRGWLLNNKTFITMAHLGARAFGQIS-GEVVQTTVWVVNNNHTEFYRPVFFRLIN--GSEEEK  631
Eco4174I_____GCACAG   507  VANPPYMGGKGMNGELKEFAKNHFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSFELFRKWLIDSKTLITMAHLGPKAFSQIS-GEIVQTTAWVIHNIHFNYYKPVFFRLID--GNEEAK  631
Sen6480IV____GTTCAT   507  VANPPYMGGKFQSPDVKRHLKDNFNGYEKDLFSAFVIRNLKFSKENGQLGFMTPFVWMFISSYEELRKRLIEKELVTSLIQLEYSGFDGAT----VPICTFTLQKTHIDNFTGSYIRLSDFRGSVNQG  630
Sen373III____CANCATC  507  VANPPYMGGNYMEAELKNYVSTKYPQGKADLYSSFMIRLIHQTKKNGALSLMTPFTWMNLSSFEEARKEILTTCCLHSLVQPEYHSFFESA---YVPICAFTIFKFQ-SSWNATFFDLSEFYGERNQA  630
SenSARA26III_ACRCAG   507  VANPPYMGNGGMNSDLKIFAKTYYPDSKSDLFAIFMQHAFSLLKKHGFNAQINMQSWMFLSSYEALRIWLLDKKTIISMAHLGARAFSQIN-GEIVQTSAWIIQSIHIINHMPTFFRLVE--GNEEEK  631
Sen5722III__GNGGCAG   507  VANPPYMGNGGMNNELKEFAKNNFPDSKTDLFAMFMQHAFSLLKENGFNTQINMQSWMFLSSYEPLRDWLLDNKTFITMAHLGARAFGQIS-GEIVQTTAWVICNRHEKNYRPDFFRLTD--GNEKQK  631

Ecl388I_____CGGNAAG  507  VANPPYMGGKGMNGELKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRNWLLDNKTFITMAHLGARAFGQIS-GEVVQTTTWVINNSRTDSYQPVFFRLIE--GNEENK  631
Yru10476I____AGGAAG  509  VANPPYMGGKGMNGELKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRSWLLDNKTFVTMAHLGPRAFGQIS-GEVVQTTAWVINNNHVAYYQPVFFRLID--GNEENK  633
CfrMH16VI____CTAAAG  509  VANPPYMGGKGMNGELKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRSWLLDNKTFVTMAHLGARAFGQIS-GEVVQTTAWVINNSHAEFYQPVFFRLID--GSEQQK  633
Eco9010II____GGTAAG  506  VANPPYMGGKGMNGDLKEFAKKQFPDSKSDLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRGWLLDNKTFITMAHLGARAFGQIS-GEVVQTTAWVIKNNHSGFYKPVFFRLVD--DNEEHK  630
SenWT8IV_____CCAAAT  507  VANPPYMGSKYQSLSVKKHLKDNFTGYDKDLFSAFIVRNLELSKQGAQLGFMTPFVWMFITTYEELRKKLIEQEIITSLVQLEYSGFDGAT----VPICTFTLQKGHINHFTGSYIRLSDFRGASNQG  630
Sen5800I_____CCAACAC  507  VANPPYMGGKGMNGELKEFAKNNFPDSKSDLFAMFIERGFCWLKDAGFNSMVTMQSWMFLSSFENMRENILNNYTIETMVHMG-NGVMKIA----FGTNATIFRNSHISPYKGSFSYVENNDINEDGY  629
Eco11117Put__CGAACAC  507  VANPPYMGSKGMNGELKEFAKNNFPDSKSDLFAMFIERGFLWLKNAGFNSMVTMQSWMFLSSFENMRKNILNNYTIETMVHMG-NGVMKIA----FGTNATVFRNKHAPVFQGSYSFTDNDDINNVGY  629
Sen5794III___ACGAACB  507  VANPPYMGSNWMNEMLKVYAKRYFPNSKSDLFSMFIEQSFFMVKKGGNAGLIVMEAWMFLASYEKFRIDIINNKNIHDLIHMPYEGKGKTPLGINFGTSAVIFENSYIKHKKTHFSCIRYYEIDESGI  634
SmaUMH8I_____GCGAACB  507  VANPPYMGNGGMNNELKEFAKAYFPESKNDLFSMFIQHGFSLLKVNSLISMIVMESWMFLASFERMRSRIINTKTIFDLIHMPYEGKGRTPLGINFGTSAFIISNNSLPLHKSHFSFIRHYEINENGT  634
KpnNIH50I____GCYAAG  507  VANPPYMGGKGMNSELKEFAKNNFPDSKADLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSFEELREWLIENKTFVNMIHLGSRAFAEIS-GEIVQTTAWVMNNYEINKYQPIFFRLIE--GNEFQK  631
Cdu23823II___GTGAAG  507  VANPPYMGNGGMNNELKEFAKNNFPDSKSDLFSMFMQHAFFLLKENGFNAQINMQSWMFLTSFESLRRWLLENVTLITMAHLGARAFGQIS-GEVVKTCTYVFLKKEHLSFNASYQDLKD--GNESLK  631
EcoNIH6II____ATGAAG  508  IANPPYMGSKGMNAGLKDFAKKNYPNSKSDLFAIFMERAFKLLSQYGFNAQINMQSWMFLSSYEQLRSNLLEDHTFITMAHLGARAFSQIS-GEVVQTTAWIIRNQNINKYQPTFYRLID--GNEEEK  632

Kor51II_____RTCGAG   506  LANPPYMNSKGMNDDLKDFAKNLFPASKSDLFAMFMQHAFYLLKDYGFNAQINMQVWMFLSTFKSLREDILKNKNIISMLHLGARAFEQIT-GEVVQTTAFIITKEKITDYLPTFIKLTS--MNTKEK  630
Eco9699II_____TAGARC 740  A-IRHEQNS-----------GVGPEELLFKECITWTDVTWR--VSARFLPEGHLSDH-AGPC-AYFDEKDKLYT--ALAVFNTPLGENWSNLLNPTLHFQAGDFKKLPYPN-----NISHSDLLPLV  843
Kpn156V_____CRTGATT  507  VANPPYMGGKGMNGELKEFAEINFSKAKSDLFSLFIERSISFCHQNSILSMITPYVWMFIGSFEDLRKEILRRHSISTLIQLEYNAFAPAC----IPVAAFTLTQDSLREYKGTFIKLSDFKGVDSQS  630
Kpn9178I____GNGCGAG  507  VANPPYMGNGGMNNELKEFAKNHSPDSKSDLFAIFMQHAFSLLKENGYNAQVNMQSWMFLSSFESLRKWLIDNKTLITMAHLGPKAFSQIS-GEVVQTTAWVINNAHFNYLKPVFFRLIE--GNEEDK  631
Ror431III____CTRGAG  507  VANPPYMGTGGMNSDLKIFAKTYYPDSKSDLFAMFMQRAFSLLKENGFNAQINMQSWMFLSSYETLRYWLLNNKTFTTMAHLGARAFGQIS-GEVVQTTAWVIKNSYIVNNQPVFFRLIE--GSEEQK  631

Vdi96II_____GNCYTAG  504  IANPPYMGGKGMNADLKDFAKENYPNSKSDLFAIFMERAFKLLSQYGFNAQINMQSWMFLSSYEQLRSNLLEDHTFITMAHLGARAFGQIS-GEVVQTTAWVIRNNYINRYQPIFFRITD--GSENEK  628

EcoC9964I____ACCYAC  507  VANPPYMGSKGMNGELKEFAKNNFPDSKSDLFAMFIERGFLWLKNAGFNSMVTMQSWMFLSSFENMRENILTNYTIETMVHMG-NGVMKIA----FGTNATVFRNNHISTYQGSFSYVENGNINQEGN  629
EcoA23Put____AGCYAG  507  VANPPYMGGKGMNGELKEFAKNNFPDSKADLFAMFMQHAFSLLKENGFNAQVNMQSWMFLSSYEALRGWLLNNKTFITMAHLGARAFGQIS-GEVVQTTVWVVNNNHTEFYRPVFFRLIN--GSEEEK  631
Kcr10483I____GGCYAC  507  VANPPYMGSKGMNGELKEFAKNNFPDSKSDLFAMFIERGFCWLKDVGFNSMVTMQSWMFLSSFENMRENILSNYTIETMVHMG-NGVMKIA----FGTNATIYRKISLQNFRGYFNYISQEDINKHGE  629

Kae10004II___GCCKAG  507  VANPPYMGNGGMNSELKDFAKNYFPDSKSDLFAMFMQRAFSLLKENGFNSQINMQAWMFLSSYEQLRKWIITQKDIITMAHLGARAFTQIS-GEVVQTTAFICRNNKTKEYIPTFLRLVK--GNEEHK  631
.
```

Figure 6.3 – Alignment of protein sequences of PglX homologues around K516. *Sequences are sorted by the DNA base at position -1 of the recognition motif (relative to the modified adenine). Residues aligning with the residues required for DNA binding of MmeI at this position are highlighted in green.*

Chapter 7 – Final discussion

7.1 Project overview

This project began with the goal of providing new insight into the function of BREX bacteriophage resistance systems by applying structural and functional biology techniques to the BREX system from *Salmonella* D23580. Subcloning of the *Salmonella* BREX system onto a transferrable plasmid backbone allowed functional characterisation and direct comparison of defence activity against the Durham Phage collection between the *Salmonella*, *E. coli* and pEFER BREX systems [183,314]. The creation of knockout mutants for each of the genes in the *Salmonella* BREX operon then allowed comparison of genes essential for methylation in *Salmonella* BREX compared to those in *E. coli* and *Acinetobacter* BREX systems [281,314]. Structural studies focussed on the core methyltransferase component of BREX defence systems, PglX, and the phosphodiesterase, PglZ, which is the only component common to all BREX subtypes [308,317]. Following the subcloning of individual genes into protein overexpression plasmids, optimisation of expression and purification and subsequent crystallisation trials, X-ray crystallography was applied to attempt to decipher the molecular structure of these BREX defence components. Ultimately, this led to the elucidation of the first structure of PglX bound to its SAM cofactor and of PglX bound to the phage encoded RM inhibitor, Ocr.

Further, this study aimed to rationally mutate the methyltransferase component of the *Salmonella* BREX system to attempt to retarget phage defence against a new subset of phages. Rational design of RM enzymes with novel recognition motifs is highly desirable in biotechnology and has been demonstrated with limited success, including for the PglX structural homologue, MmeI [192]. In collaboration with Dr Rick Morgan (NEB), residues essential for DNA motif recognition by PglX were predicted through sequence alignments to PglX homologues. PglX mutants were then designed aimed at altering essential residues to effectively allow recognition of alternative DNA bases.

7.2 Summary of results

7.2.1 Bioinformatic characterisation

Initial bioinformatic analysis provided predicted functions of *Salmonella* BREX components via sequence and structural homology searches and allowed understanding of the context of the BREX system within the wider phage defence landscape of the *Salmonella* D23580 genome. BREX components are predicted to encode diverse biochemical functions presenting a complex mechanism of both host methylation and phage restriction. In type I BREX systems, three components have either been demonstrated to (BrxL and PglX) or predicted to (BrxC) bind to DNA and have in turn shown interaction with multiple other BREX components [318]. This paints a picture of multiple heteromeric BREX complexes translocating along DNA strands, though the exact composition and mechanistic function of these complexes remains elusive.

The increasingly apparent cooccurrence of phage defence systems within an organism's genome [134,136,243,335,353] has raised the possibility of potential interaction and synergistic effects [334]. Analysis of the *Salmonella* D23580 genome revealed the presence of twelve additional phage defence systems (Table 3.1), including two PARIS systems, a Lamassu-Fam system, a BstA system, a Mokosh system, a retron system, a Cas system a PD-T4-1 system and four RM systems, either found in isolation or clustered together on discreet defence islands. The *Salmonella* BREX system is encoded within a hypervariable region of the *Salmonella* D23580 genome on a defence island containing a Mokosh system, a type IV RM system, and a type II PARIS system. The synergistic function of BREX and type IV RM systems has previously been described [183]. The PARIS system has at some point been inserted into the BREX system between the *pglX* and *pglZ* genes. PARIS systems have previously been shown to provide Abi-based phage defence triggered by exposure to the T7 encoded RM inhibitor, Ocr [243]. Ocr has in turn previously been demonstrated to inhibit *E. coli* BREX function by direct interaction with PglX [294]. This presents the intriguing possibility that the BREX and PARIS systems in *Salmonella* D23580 confer a synergistic defence, in which the BREX system restricts unmodified phage DNA, and the PARIS system triggers Abi if BREX is blocked by an Orc-like restriction inhibitor. Though no phage defence activity could be elicited from the PARIS system by exposure to phage T7, or recombinantly expressed Ocr or its *Salmonella* phage specific homologue Gp5 (Table 4.4),

further probing for phage defence from the PARIS system in its native *Salmonella* host against *Salmonella* phages may demonstrate activity.

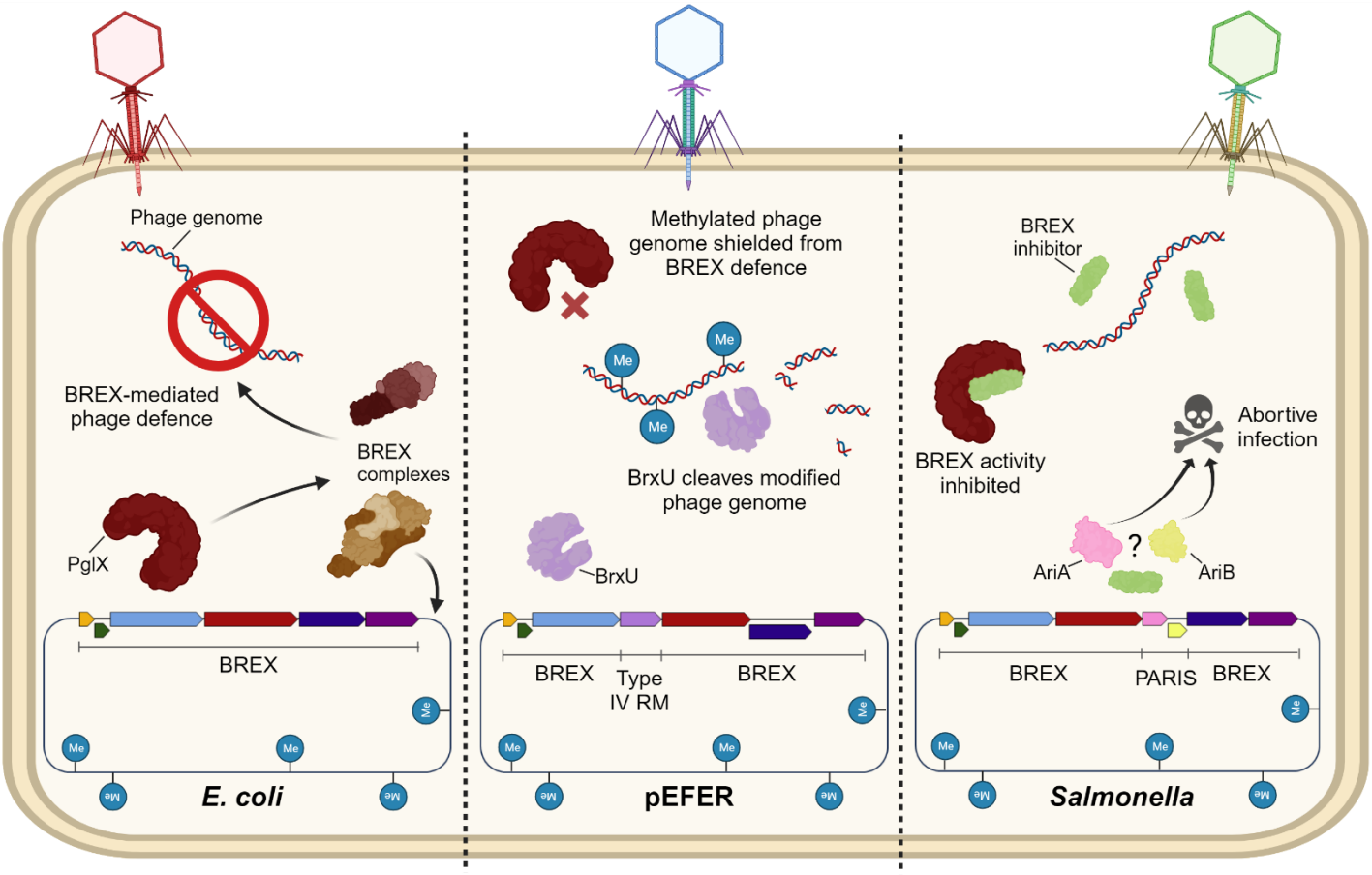

Figure 7.1 – Cartoon schematic of the layers of phage defence provided by BREX operons. Left; in the *E. coli* BREX operon, invading non-methylated phage DNA is targeted by BREX, directed via BREX motif recognition by PglX. PglX also directs methylation of the host genome at BREX motif sites (shown by blue circles), allowing discrimination between self and non-self. Middle; in the pEFER BREX operon, methylated phage genomes are shielded from BREX phage defence but are targeted and degraded by the type IV RM, BrxU. Right; in the *Salmonella* D23580 BREX operon, BREX activity is inhibited by phage encoded factors, possibly by direct inhibition of PglX function, activating the PARIS system and resulting in abortive infection.

7.2.2 Functional characterisation

Functional comparisons between BREX phage defence conferred by the *Salmonella*, *E. coli* and pEFER systems against the Durham Phage Collection showed varying levels of activity which did not correlate with the number of respective genomic motifs within each phage genome (Figure 3.6). Together, these data demonstrated that predicting the defence conferred by a particular BREX system is more complex and that phages encode methods of inhibiting or evading species-specific BREX systems.

Phage defence conferred by individual knockouts has been reported to vary between *E. coli* and *Acinetobacter* systems [281,314] and growth curves and EOP assays of gene knockouts from the *Salmonella* BREX system allowed additional comparison (Table 7.1). As with the *Acinetobacter* system, *brxA*, *brxB*, *brxC*, *pglX*, and *pglZ* were essential for both host methylation (Table 4.3) and phage restriction (Figure 4.2, Table 4.1 and Table 4.2), in contrast to the *E. coli* system in which *brxA* is dispensable for both. In all three systems, *brxL* was dispensable for methylation, but only in the *Salmonella* BREX system was it also dispensable for phage restriction. BrxA binds DNA and is predicted to provide some form of regulatory function, suggesting that differences in *brxA* essentiality is possibly due to key differences in regulation between the three BREX systems. Phage restriction in the absence of *brxL* is harder to explain and was further complicated by the dramatic decrease observed in EOPs across the phage library. The potential role of BrxL in modulating BREX activity points at a key functionality provided by the assembly and translocation of BrxL complexes along DNA strands [318]. The Lon-like domain of BrxL has shown similarity to the Lon-like C-terminal of RadA, involved in branch migration in homologous recombination, implicating BREX activity in targeting phage DNA replication forks [315,374]. Understanding the discrepancy in *brxL* essentiality between these three systems would likely contribute significantly to deciphering the mechanism of BREX phage restriction and is an important target for future study.

Table 7.1 – Essential components for phage restriction and host genome methylation by BREX systems from three different host strains.

Gene	*Salmonella* D23580		*E. coli*[1]		*Acinetobacter*[2]	
	Host methylation	**Phage restriction**	**Host methylation**	**Phage restriction**	**Host methylation**	**Phage restriction**
brxA	-	-	+	+	-	-
brxB	-	-	-	-	-	-
brxC	-	-	-	-	-	-
pglX	-	-	-	-	-	-
pglZ	-	-	-	-	-	-
brxL	+	+	+	-	+	-

[1] [314]

[2] [281]

7.2.3 Structural characterisation

Structural characterisation of PglX utilised X-ray crystallography and produced the first structure of the core BREX methyltransferase, PglX, to a resolution of 3.4 Å. PglX presents as two prominent domains connected by a central hinge region (Figure 5.11) and shows structural similarity to the methyltransferase and DNA binding domains of the type IIL RM, MmeI (Figure 5.12A). Indeed, the predicted methyltransferase domain of PglX is conserved and the SAM cofactor is visible within this region in the PglX structure (Figure 5.12B and C). The C-terminal region of PglX is also conserved but does not align with MmeI, though it does show structural similarity to several type II RMs and type I RM specificity subunits (Table 5.7). Alignment of the MmeI molecule to PglX shows the DNA molecule bound to MmeI sitting within the hinge region of PglX. This places the flipped out adenine base in proximity to the bound SAM molecule in the PglX structure suggests that DNA binds to PglX in this positively charged region.

X-ray crystallography was also successful in producing a structure of PglX bound to the RM inhibitor, Ocr, shown to inhibit the *Salmonella* BREX system (Table 4.4) through direct interaction with PglX [294]. Ocr mimics around 20 bp of slightly bent B-form DNA as a homodimer [287,291,325] and is able to bind PglX as

a heterotetramer (Figure 5.15). Ocr binds along the conserved and positively charged C-terminal region of PglX and suggests a slightly different orientation of DNA binding compared to that suggested by the DNA molecule from aligned MmeI (Figure 5.16). Though crystallisation trials of PglX bound to DNA produced a variety of crystal morphologies, diffraction was not observed from any crystals analysed. It is possible that this is due to the inherent lack of diffraction produced by all PglX crystals screened throughout this study. It could also be due to the design of the oligos used for PglX binding. The structures of PglX and PglX bound to Ocr may then allow better informed design of oligos for future crystallisation screening. It may also better inform the design of PglX mutants towards altering the PglX recognition motif.

7.2.4 Mutation of PglX recognition sequence

In conjunction with Dr Rick Morgan, the DNA recognition motif od PglX was rationally mutated from GAT**C**AG to GAT**M**AG. This allowed BREX phage defence against a previously resistant phage and produced additional host methylation at GAT**A**AG motifs. In the absence of structural data for DNA binding by PglX, sequence alignments of PglX homologues with known or inferred recognition motifs were used, as was described previously for the rational modification of the recognition motif of the PglX structural homologue, MmeI [194]. The residues required for motif binding in MmeI have been identified by structural analysis [192] and the residues in PglX which aligned with these residues in MmeI were used as focus points for identifying covariation with base recognition at each of the non-adenine bases in the BREX recognition motif. Out of the 23 PglX mutants designed, only one produced altered motif recognition: PglX mutant 3, which encoded T802A and S838N. Though these mutations allowed recognition of adenine at position -1 (relative to the modified adenine), they did not negate cytosine binding, resulting in broadening of specificity rather than discreet retargeting. Future attempts at altering PglX recognition motif would likely benefit from availability of a DNA bound structure [192]. Nevertheless, retargeting of BREX phage defence through reprogramming of PglX demonstrates that PglX is the sole specificity factor in determining host-phage recognition by BREX systems in both methylation and restriction. PglX displays a high rate of phase variation and mutation, which is likely the mechanism by which bacteria adapt to new phage threats.

7.3 Future work

Work carried out in this study has advanced our understanding of type I BREX systems, but further study is required to decipher the molecular mechanisms and functions involved in both host methylation and phage restriction. This study has also highlighted additional avenues of investigation through bioinformatic and mutational analysis. Recommended foci of future study are discussed below, and suggested methodologies and techniques are described.

7.3.1 Functional characterisation

Work in Chapter 3 described the functional characterisation of the *Salmonella* D23580, *E. coli* and pEFER BREX systems through EOP assays against the Durham Phage Collection. This work showed that the level of phage defence did not correlate directly with the number of BREX motifs in the phage genome. Furthermore, work here suggested that a single motif was insufficient to provide phage defence. This raises the question of how many motifs are required in a phage genome to elicit BREX phage defence? Further, does the positioning and orientation of these motifs relative to each other produce variation in phage defence? Understanding these factors will allow greater insight into the function of BREX in phage restriction, providing insight into how different BREX complexes assemble on phage DNA, how many complexes assemble or are necessary and whether BREX motif orientation of DNA strand location are significant in restriction. A phage containing a single *Salmonella* BREX motif has already been identified in this study and introduction of BREX motifs at different locations and orientations could be performed. Several techniques for phage genome mutagenesis are available [394]. The most common methods are based on homologous recombination and involve transformation of host bacteria with plasmids [395,396] or retrons [397] containing the desired mutation flanked by regions homologous to the phage genome. Selection of correctly mutated phages can then be carried out either by inclusion of a marker gene [398] or through elimination of incorrect phages by propagation on host strains containing CRISPR-Cas targeting incorrect sequences [399,400].

The dramatic increase in phage defence conferred by DH5α pBrxXL$_{Sty}$-Δ*brxL* strains is puzzling and was not observed in previous studies on the *E. coli* and *Acinetobacter* BREX systems [281,314]. The function and mechanisms of BrxL in phage defence are not currently understood. BrxL has been shown to bind to DNA as a dimer of hexamers [318] and interact with several other BREX components in A-SEC analysis (unpublished data). Based on previous data in *E. coli* and *Acinetobacter*, it would stand to reason then that these complexes are required for phage restriction, though not methylation [281,314], and thus, any phage defence activity observed in its absence is unexpected. Confirmation of this phenotype should be pursued by both EOP assays in different *E. coli* host strains and by producing a genomic *brxL* knockout in *Salmonella* D23580 for EOP assays against *Salmonella* phages. No functional assays have been carried out for *brxL* knockouts in the pEFER BREX system either and this experiment would be simple to produce. Together, these experiments would determine whether this phenotype was host strain or system specific.

A previous study on the *Salmonella* BREX system demonstrated that C-terminal deletions of *brxL* resulted in large variations in transcription levels elsewhere in the genome [315]. It would be interesting to determine whether this holds true within an *E. coli* DH5α host and, if so, for which genes is transcription being altered. No protein annotated DH5α genome is currently available but the NEB5α strain is a close derivative, with a single *fhuA2* gene deletion rendering it resistant to phages which require the FhuA outer membrane protein for adsorption and injection [401,402]. NEB5α contains several predicted phage defence systems, including Cas, Druantia, RnlAB, Hachiman, Mokosh and several RM systems (Table 7.2). It is possible that one or more of these systems are being activated by deletion of *brxL* and resulting variations. It is also possible that the BREX system interacts with some other system or systems in DH5α, either directly or indirectly, and that removing *brxL* restores activity of some other system or combination of systems. Producing a DH5α strain with each of these systems deleted would then negate this phenotype.

Table 7.2 – Phage defence systems in* E. coli *NEB5α, as determined by the DefenseFinder Python package.

Type	Subtype	System start	System end	No. of genes
Cas	Cas Class 1 Subtype I-E	AOO72682.1_2746	AOO71003.1_2753	8
Druantia	Druantia III	AOO72512.1_4356	AOO72513.1_4357	2
Lit	Lit	AOO69417.1_1051	AOO69417.1_1051	1
RnlAB	RnlAB	AOO70893.1_2634	AOO70894.1_2635	2
Hachiman	Hachiman	AOO70890.1_2631	AOO70891.1_2632	2
Mokosh	Mokosh Type II	AOO72474.1_4314	AOO72474.1_4314	1
RM	RM Type IV	AOO69432.1_1071	AOO69432.1_1071	1
RM	RM Type IV	AOO72514.1_4358	AOO72519.1_4364	3
RM	RM Type I	AOO72517.1_4361	NEB5A_22430_4363	3

7.3.2 Biochemical and biophysical characterisation

This study has provided preliminary investigations into the function of PglX. PglX presents as a monomer in solution and shows no apparent methyltransferase activity *in vitro*, either alone or in conjunction with BrxB and/or PglZ. Ample evidence of PglX being the effector of host methylation has been provided in this study through a combination of the absence of methylation it the absence of PglX, the predicted functional methyltransferase domains, the close structural alignment with the methyltransferase domain of MmeI and the presence the methyl group donor SAM bound to PglX in the crystal structure. Methyltransferase functionality then likely requires some combination of BREX components and/or cofactors. PglX activity may in turn be dependent on the activity of (rather than simply the presence of) PglZ, which is in turn dependent on the presence of Zn^{2+} or Mg^{2+} (unpublished data); these could easily be provided in future methyltransferase assays. Additionally, BrxA was found to be required for methylation in the *Salmonella* BREX system and could also be included. It is also possible that the methyltransferase activity assay is not sensitive enough to detect PglX activity and methylation could instead be assayed directly through

sequencing of DNA from methyltransferase reactions, either through nanopore sequencing or PacBio. Regardless, probing of PglX for methyltransferase activity should continue and will provide vital information on the composition of the methyltransferase complex and the process of BREX host methylation.

PglZ has recently demonstrated phosphodiesterase activity *in vitro* and presents the only permanent factor of all BREX subtypes. The significance of phosphodiesterase activity in BREX function however remains unknown. The PglZ phosphodiesterase domain is commonly found in response regulator genes of two component signalling systems which are involved in signalling and response to environmental factors [403]. PglZ can cleave cyclic and linear oligos (unpublished data), commonly used as signalling molecules and are produced by both CRISPR-Cas type III [228] and cBASS phage defence systems [229,255]. In cBASS systems, cyclic oligoadenylates activate factors which induce Abi [229,255] while in CRISPR-Cas type III systems these molecules activate ancillary enzymes, predominantly nucleases [228]. Neither Abi nor nuclease activity are implicated in BREX phage defence. It stands to reason that cyclic and oligos may be acting on or activating some other BREX factor and that PglZ is removing these molecules and preventing overactivation. It would be interesting to determine whether any other BREX components interact with these signalling molecules and if so, how that interaction modulates activity or complex formation.

The formation of methylation and/or restriction complexes appears to be essential to BREX function. Detail on the composition of these complexes will allow further understanding on how these complexes facilitate BREX function. Initial work on identifying BREX complexes and complex composition is currently being undertaken by other members of the lab using A-SEC techniques discussed in section 5.3.2. This will be supported by Mass photometry analysis using a Refeyn system. Mass photometry allows accurate measurement of single molecules in solution and will provide detail on BREX complex stoichiometries in a native state [404]. Molecular interactions in complex formation will also be studied using structural methodologies, as described in section 7.3.3.

7.3.3 Structural studies

X-ray crystallography allowed elucidation of the molecular structure of PglX and demonstrated how Ocr binds to and inhibits PglX function. This study was unsuccessful in the production of structures of PglX bound to DNA and of PglZ, however. These structures remain obvious and vital targets for determining the molecular mechanisms of BREX phage defence. PglZ crystal optimisation should continue, aiming to identify new crystal morphologies and crystallisation conditions, potentially utilising seeding techniques [405]. X-ray crystallography will also be explored for PglZ bound to BrxB, potentially providing greater stability and different crystallisation conditions which may in turn produce higher resolution diffraction. Work on producing crystal structures of PglX bound to DNA will continue, using redesigned, longer oligos, as discussed in Chapter 5. Provision of a structure of PglX bound to DNA will also allow better informed design of PglX mutants towards retargeting of the *Salmonella* BREX system. Structural work on PglZ and BrxB, PglX and DNA, and larger BREX complexes can also be pursued by cryo-EM, which is more suitable for larger proteins and protein complexes. Structural work on larger BREX complexes will complement the A-SEC and Refeyn methodologies described in 7.3.2.

7.3.4 Altering PglX recognition motifs

Initial attempt to alter PglX specificity found limited success. There are three potential approaches to the continuation of this work. Perhaps the most straightforward would be to produce a DNA bound PglX structure as this would allow conclusive identification of the PglX residues involved in DNA base recognition at each position in the BREX recognition motif [192]. From here, mutations of these residues could be designed aimed at forming the necessary contacts and bonds with desired DNA bases. A DNA bound structure of PglX has so far proven elusive however, as described in Chapter 5 and outlined in 7.2.3 and 7.3.3. Should PglX-DNA complexes continue to prove recalcitrant to structural biology techniques, additional attempts at modifying the recognition motif of PglX could be pursued using the sequence alignment techniques performed here. The rational engineering of MmeI recognition motifs using this method utilised a well characterised family of highly similar RMs [174]. In contrast, the PglX homologues used for sequence alignments here have not been functionally characterised and in many cases the recognition motif was inferred by matching appropriate N6mA genomic modifications to a homologous

type II RM or *pglX* gene. Subcloning these genes within their corresponding BREX systems into DH5α and sequencing the host genomes would allow confirmation of the recognition motif of each of the homologues and EOP assays would allow confirmation of phage defence activity. Better characterisation would then potentially translate to more effective identification of appropriate mutations for altering motif specificity.

Directed evolution-based methods present an alternative approach to rational design of PglX specificity [391]. Mutations produced using this method which led to altered motif recognition could then be identified and thus, residue involved in DNA binding could be inferred. Directed evolution can be performed independently of sequence or structural data, negating the need for further structural and characterisation work required for rational design. Mutation *in vitro* can easily be performed by error prone PCR amplifications [406] or DNA shuffling techniques [407]. Mutated genes could then be subcloned into pBAD30 and assayed as previously described in Chapter 6. Far greater diversity can be achieved by utilising *in vivo* continuous evolution methods, though this requires some method of continuous selection linked to protein activity [391]. This would be difficult in the case of PglX as many techniques either require isolation of genes of interest or would not be able to be targeted at a specific gene in an operon, and PglX activity can only be detected in the presence of an active BREX system. T7-targeted dCas9-limited in vivo mutagenesis (T7-DIVA) allows targeted mutagenesis, with target regions delimited by the insertion of a T7 promoter sequence on the antisense strand and recognition site of a catalytically inactive Cas9 [408]. T7 polymerase fused to a deaminase then catalyses the mutation of bases from C to T and A to G. A T7 promoter site could theoretically be inserted following the *pglX* gene as *pglZ* and *brxL* transcription would be maintained by promoters in the PARIS genes. Activity could then be screened by infection of mutated cultures with phage Trib, as cultures containing mutations in *pglX* which did not alter specificity would be infected and killed, then positive results could be assayed for motif recognition as described in Chapter 6. Directed evolution techniques introduce a far greater diversification and allows for alterations to protein secondary and tertiary structure in a way that rational design of individual mutations does not.

7.3.5 Interaction and cooccurrence with other defence systems

It is becoming increasingly apparent that phage defence systems should be considered within the context of the wider phage defence landscape, both within the host genome and within the local microbial

community [136,274,332,334,353]. Identification of defence systems within the *Salmonella* genome revealed that the BREX defence operon sits within one of several defence islands. Within the BREX defence island are a type IV RM system, a Mokosh system and the previously discussed PARIS system within the BREX operon itself. The potential for synergistic interaction between BREX and PARIS will first require confirmation of active phage defence from the PARIS system. The *ariA* and *ariB* genes have already been cloned into separate inducible plasmids and also both together onto a single plasmid backbone. This will allow more effective probing of the Durham Phage Collection for PARIS activity and can also be extended to phages with modified genomes which were not included in this study. In case the molecular trigger for the PARIS system is specific to *Salmonella* phages, defence activity can be assessed against *Salmonella* phages using a *Salmonella* D23580 Δ*pglX* strain, both of which were isolated and produced, respectively, by Dr David Picton in a previous study in our lab [321]. Complementary phage defence from BREX and type IV RM enzymes has previously been shown [183] and it would be interesting to see whether this effect is mirrored in *Salmonella* D23580. Subcloning of the type IV RM from the *Salmonella* genome onto a plasmid backbone and assaying for activity against phages with modified genomes in the Durham Phage Collection could be complemented by *in vitro* restriction assays using purified genomic DNA from modified phages. Interaction with the Mokosh system is more difficult to predict as the molecular mechanism of phage defence is not known. Mokosh systems encode RNA helicase and nuclease-like domains and provide protection against T-even phages [134], potentially through recognising phage RNA or RNA/DNA intermediates. A full study targeting this system will be necessary to establish functional mechanisms, using many of the functional, biochemical and structural techniques described here.

Phage defence systems often have common features, such as universal regulatory proteins [279–281], which may aid horizontal gene transfer and integration of phage defence systems in phage defence islands. This possibility is intriguing as it implies a degree of modularity akin to that displayed in phage genomes. Both the *Salmonella* and *E. coli* BREX operons lack an obvious universal regulatory protein, but the upstream region shows sequence conservation (Figure 4.6). Identification of common regulatory features would provide new insight into universal regulation mechanisms of phage defence systems, but also potentially allow identification of novel defence systems through guilt-by-association methodologies. Analysis of transcription levels using SalComMac [337] allowed identification of promoter regions within the 3’ region of *pglX*, the 5’ region of *ariA* and the 3’ region of *ariB*. The positioning of the promoters in *pglX* and *ariB* are interesting as they have facilitated the integration of the PARIS system. The promoter within

the 3’ region of *pglX* enabled the transcription of *ariA* while the promoter within the 3 ’ region of *ariB* ensured that integration did not disrupt the function of local genes – in this case, *pglZ* and *brxL*. Again, phage defence systems appear to encode features which facilitate horizontal gene transfer within bacterial defence islands, suggesting a modularity which mirrors that of their phage counterparts. Transcriptional and bioinformatic analyses within the defence islands of other species and strains would help to determine whether this is a common feature of phage defence systems or whether this phenomenon is limited to the *Salmonella* BREX operon. Defence islands encoding BREX and/or PARIS systems present a good starting point for this analysis.

7.4 Final conclusions

This study aimed to functionally and structurally characterise the BREX phage defence system from the clinically relevant *Salmonella* D23580, and to attempt to manipulate the recognition motif of the BREX system by mutation of the methyltransferase component, PglX. Functional EOP assays against coliphages in the Durham Phage Collection (Figure 3.5, Table 3.3, Figure 3.6), and comparison against EOP assays against *E. coli* and pEFER BREX system (Figure 3.6), demonstrated that phage defence conferred by BREX is difficult to predict, with no correlation between the number of recognition motifs within the phage genome and phages encoding methods for inhibiting species specific BREX systems. Production of individual gene knockouts and development of a nanopore sequencing and bioinformatics pipeline allowed determination of genes essential for both phage restriction (Figure 4.2, Table 4.2) and host methylation (Table 4.1). Again, differences were observed between essential genes in *Salmonella* BREX and those in previous studies on *E. coli* and *Acinetobacter* BREX systems. Together, these data add further complexity to a system whose mechanisms remain elusive.

The application of X-ray crystallography allowed the elucidation of the first crystal structure of PglX, responsible for methylation of the host chromosome and ultimately, self from non-self recognition. PglX displays distinct N-terminal and C-terminal domains joined by a central hinge region, predicted to be the binding site of DNA (Figure 5.13). In the crystal structure, PglX is bound to its SAM co-factor, the methyl

group donor of methyl transfer (Figure 5.11) and the methylation domains surrounding SAM are highly conserved (Figure 5.12). Though no methyltransferase activity was observed from PglX (Figure 5.3), it was shown to bind directly to the RM inhibitor, Ocr (Figure 5.4), which had previously been shown to inhibit *Salmonella* BREX phage defence (Table 4.4). A crystal structure of PglX bound to Ocr was produced, showing that an Ocr dimer was able to bind two molecules of PglX forming a heterotetramer (Figure 5.15). Ocr binds along a conserved C-terminal arm of PglX (Figure 5.15), presenting an alternative DNA binding orientation and implicating this region in DNA binding and motif recognition (Figure 5.16). Attempts to modify the recognition motif were partially successful, with specificity broadened to GATMAG motifs rather than switched completely from GATCAG motifs. This resulted in a broader phage target range and demonstrated how small mutations in a gene with reportedly high phase variation can quickly alter phage susceptibility. The structural data produced in this study together provides valuable insight into the functionality and organisation of PglX in DNA binding and methylation as well as demonstrating how a common phage escape strategy inhibits a complex phage defence system.

Overall, this study provides valuable new data to add to the growing literature on BREX phage resistance systems. Much work remains to be done before the complex and multifaceted molecular mechanisms of this system are understood. Indeed, as more and more phage defence systems and modalities are discovered and mechanisms of phage escape are documented, the complexity of these phage bacteria interactions and their extensive impact as drivers of evolution in microbial communities becomes more apparent. Mapping and characterising these interactions will require concerted and multidisciplinary collaboration and will be essential in the effective design and application of phages as therapeutic agents.

References

1. Clokie MR, Millard AD, Letarov A V, Heaphy S: **Phages in nature**. *Bacteriophage* 2011, **1**:31–45.

2. Weinbauer MG, Rassoulzadegan F: **Are viruses driving microbial diversification and diversity?** *Environ Microbiol* 2004, **6**:1–11.

3. Hurwitz BL, Hallam SJ, Sullivan MB: **Metabolic reprogramming by viruses in the sunlit and dark ocean**. *Genome Biol* 2013, **14**:R123.

4. Stern A, Sorek R: **The phage-host arms-race: Shaping the evolution of microbes**. *Bioessays* 2011, **33**:43.

5. Wilhelm SW, Suttle CA: **Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs**. *Bioscience* 1999, **49**:781–788.

6. Salmond GPC, Fineran PC: **A century of the phage: past, present and future**. *Nat Rev Microbiol* 2015, **13**:777–786.

7. Zrelovs N, Dislers A, Kazaks A: **Motley Crew: Overview of the Currently Available Phage Diversity**. *Front Microbiol* 2020, **11**:579452.

8. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC: **Uncovering Earth's virome**. *Nature 2016 536:7617* 2016, **536**:425–430.

9. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, et al.: **Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses**. *Nature 2016 537:7622* 2016, **537**:689–693.

10. Hampton HG, Watson BNJ, Fineran PC: **The arms race between bacteria and their phage foes**. *Nature* 2020, **577**:327–336.

11. Ofir G, Sorek R: **Contemporary Phage Biology: From Classic Models to New Insights**. *Cell* 2018, **172**:1260–1270.

12. Lin DM, Koskella B, Lin HC: **Phage therapy: An alternative to antibiotics in the age of multi-drug resistance**. *World J Gastrointest Pharmacol Ther* 2017, **8**:162.

13. Twort FW: **An investigation on the nature of ultra-microscopic viruses by Twort FW, L.R.C.P. Lond., M.R.C.S. (From the Laboratories of the Brown Institution, London)**. *Bacteriophage* 1915, **1**:127–129.

14. D'Herelle F: **On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917**. *Res Microbiol* 1917, **158**:553–554.

15. Chanishvili N: **Phage Therapy—History from Twort and d'Herelle Through Soviet Experience to Current Approaches**. *Adv Virus Res* 2012, **83**:3–40.

16. Abedon ST, Kuhl SJ, Blasdel BG, Kutter EM: **Phage treatment of human infections**. *Bacteriophage* 2011, **1**:66–85.

17. Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H, et al.: **Antibiotic resistance-the need for global solutions**. *Lancet Infect Dis* 2013, **13**:1057–1098.

18. O'Neil J: **Tackling drug-resistant infections globally: Final report and recommendations**. 2016, **2020**.

19. Pendleton JN, Gorman SP, Gilmore BF: **Clinical relevance of the ESKAPE pathogens**. *Expert Rev Anti Infect Ther* 2013, **11**:297–308.

20. Bours PH, Polak R, Hoepelman AI, Delgado E, Jarquin A, Matute AJ: **Increasing resistance in community-acquired urinary tract infections in Latin America, five years after the implementation of national therapeutic guidelines**. *Int J Infect Dis* 2010, **14**:e770-4.

21. Gordillo Altamirano FL, Barr JJ: **Phage Therapy in the Postantibiotic Era**. *Clin Microbiol Rev* 2019, **32**:e00066-18.

22. Nobrega FL, Costa AR, Kluskens LD, Azeredo J: **Revisiting phage therapy: new applications for old resources**. *Trends Microbiol* 2015, **23**:185–191.

23. Hyman P, Abedon ST: **Chapter 7 - Bacteriophage Host Range and Bacterial Resistance**. In *Advances in Applied Microbiology*. . Academic Press; 2010:217–248.

24. Chan BK, Abedon ST, Loc-Carrillo C: **Phage cocktails and the future of phage therapy**. *Future Microbiol* 2013, **8**:769–783.

25. Amorena B, Gracia E, Monzón M, Leiva J, Oteiza C, Pérez M, Alabart JL, Hernández-Yago J: **Antibiotic susceptibility assay for Staphylococcus aureus in biofilms developed in vitro**. *J Antimicrob Chemother* 1999, **44**:43–55.

26. Curtin JJ, Donlan RM: **Using Bacteriophages To Reduce Formation of Catheter-Associated Biofilms by Staphylococcus epidermidis**. *Antimicrob Agents Chemother* 2006, **50**:1268–1275.

27. Flemming H-C, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S: **Biofilms: an emergent form of bacterial life**. *Nat Rev Microbiol* 2016, **14**:563–575.

28. Bedi MS, Verma V, Chhibber S: **Amoxicillin and specific bacteriophage can be used together for eradication of biofilm of Klebsiella pneumoniae B5055**. *World J Microbiol Biotechnol* 2009, **25**:1145.

29. Mushtaq N, Redpath MB, Luzio JP, Taylor PW: **Treatment of experimental Escherichia coli infection with recombinant bacteriophage-derived capsule depolymerase**. *J Antimicrob Chemother* 2005, **56**:160–165.

30. Luepke KH, Suda KJ, Boucher H, Russo RL, Bonney MW, Hunt TD, Mohr III JF: **Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications**. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 2017, **37**:71–84.

31. Le P, Kunold E, Macsics R, Rox K, Jennings MC, Ugur I, Reinecke M, Chaves-Moreno D, Hackl MW, Fetzer C, et al.: **Repurposing human kinase inhibitors to create an antibiotic active against drug-resistant Staphylococcus aureus, persisters and biofilms**. *Nat Chem* 2020, **12**:145.

32. Pahil KS, Gilman MSA, Baidin V, Clairfeuille T, Mattei P, Bieniossek C, Dey F, Muri D, Baettig R, Lobritz M, et al.: **A new antibiotic traps lipopolysaccharide in its intermembrane transporter**. *Nature 2024* 2024, doi:10.1038/s41586-023-06799-7.

33. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, et al.: **Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus**. *Nat Med* 2019, **25**:730–733.

34. Jault P, Leclerc T, Jennes S, Pirnay JP, Que Y-A, Resch G, Rousseau AF, Ravat F, Carsin H, Le Floch R, et al.: **Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by**

Pseudomonas aeruginosa (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. *Lancet Infect Dis* 2019, **19**:35–45.

35. Rahman M, Nabi A, Asadulghani M, Faruque SM, Islam MA: **Toxigenic properties and stx phage characterization of Escherichia coli O157 isolated from animal sources in a developing country setting**. *BMC Microbiol* 2018, **18**:98.

36. Allison HE: **Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens**. *Future Microbiol* 2007, **2**:165–174.

37. Yang Q, Le S, Zhu T, Wu N: **Regulations of phage therapy across the world**. *Front Microbiol* 2023, **14**.

38. Jones JD, Trippett C, Suleman M, Clokie MRJ, Clark JR: **The Future of Clinical Phage Therapy in the United Kingdom**. *Viruses* 2023, **15**.

39. Lurias SE, Delbrock M: **MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE**. *Genetics* 1943, **28**:491–511.

40. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, et al.: **Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene**. *Nature* 1976, **260**:500–507.

41. Hershey AD, Chase M: **Independent functions of viral protein and nucleic acid in growth of bacteriophage**. *J Gen Physiol* 1952, **36**:39–56.

42. Benzer S: **FINE STRUCTURE OF A GENETIC REGION IN BACTERIOPHAGE**. *Proc Natl Acad Sci U S A* 1955, **41**:344–354.

43. Crick FHC, Barnett L, Brenner S, Watts-Tobin RJ: **General Nature of the Genetic Code for Proteins**. *Nature* 1961, **192**:1227–1232.

44. Studier FW, Moffatt BA: **Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes**. *J Mol Biol* 1986, **189**:113–130.

45. Roberts RJ: **How restriction enzymes became the workhorses of molecular biology**. *Proceedings of the National Academy of Sciences* 2005, **102**:5905–5908.

46. Sauer B: **Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae**. *Mol Cell Biol* 1987, **7**:2087–2096.

47. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al.: **Real-Time DNA Sequencing from Single Polymerase Molecules**. *Science (1979)* 2009, **323**:133–138.

48. Sorek R, Lawrence CM, Wiedenheft B: **CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea**. *Annu Rev Biochem* 2013, **82**:237–266.

49. Mimmi S, Maisano D, Quinto I, Iaccino E: **Phage Display: An Overview in Context to Drug Discovery**. *Trends Pharmacol Sci* 2019, **40**:87–91.

50. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface**. *Science (1979)* 1985, **228**:1315–1317.

51. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M: **Filamentous bacteriophage: biology, phage display and nanotechnology applications**. *Curr Issues Mol Biol* 2011, **13**:51–76.

52. Watters KE, Kirkpatrick J, Palmer MJ, Koblentz GD: **The CRISPR revolution and its potential impact on global health security**. *Pathog Glob Health* 2021, **115**:80.

53. Knott GJ, Doudna JA: **CRISPR-Cas guides the future of genetic engineering**. *Science* 2018, **361**:866.

54. Carroll D: **Focus: Genome Editing: Genome Editing: Past, Present, and Future**. *Yale J Biol Med* 2017, **90**:653.

55. Ackermann HW: **5500 Phages examined in the electron microscope**. *Arch Virol* 2007, **152**:227–243.

56. Turner D, Kropinski AM, Adriaenssens EM: **A Roadmap for Genome-Based Phage Taxonomy**. *Viruses* 2021, **13**.

57. Dion MB, Oechslin F, Moineau S: **Phage diversity, genomics and phylogeny**. *Nat Rev Microbiol* 2020, **18**:125–138.

58. Marvin DA: **Filamentous phage structure, infection and assembly**. *Curr Opin Struct Biol* 1998, **8**:150–158.

59. Loh B, Kuhn A, Leptihn S: **The fascinating biology behind phage display: filamentous phage assembly**. *Mol Microbiol* 2019, **111**:1132–1138.

60. Webb JS, Thompson LS, James S, Charlton T, Tolker-Nielsen T, Koch B, Givskov M, Kjelleberg S: **Cell Death in Pseudomonas aeruginosa Biofilm Development**. *J Bacteriol* 2003, **185**:4585–4592.

61. Rice SA, Tan CH, Mikkelsen PJ, Kung V, Woo J, Tay M, Hauser A, McDougald D, Webb JS, Kjelleberg S: **The biofilm life cycle and virulence of Pseudomonas aeruginosa are dependent on a filamentous prophage**. *ISME J* 2009, **3**:271–282.

62. Kelly A, Went SC, Mariano G, Shaw LP, Picton DM, Duffner SJ, Coates I, Herdman-Grant R, Gordeeva J, Drobiazko A, et al.: **Diverse Durham collection phages demonstrate complex BREX defense responses**. *Appl Environ Microbiol* 2023, doi:10.1128/AEM.00623-23.

63. Menichi B, Buu A: **Integration of the overproduced bacteriophage T5 receptor protein in the outer membrane of Escherichia coli**. *J Bacteriol* 1983, **154**:130–138.

64. Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D: **Differences in Lactococcal Cell Wall Polysaccharide Structure Are Major Determining Factors in Bacteriophage Sensitivity**. *mBio* 2014, **5**:e00880-14.

65. Rakhuba D V, Kolomiets EI, Dey ES, Novik GI: **Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell**. *Pol J Microbiol* 2010, **59**:145–155.

66. Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov V V., Rossmann MG: **The tail structure of bacteriophage T4 and its mechanism of contraction**. *Nat Struct Mol Biol* 2005, **12**:810–813.

67. Maghsoodi A, Chatterjee A, Andricioaei I, Perkins NC: **How the phage T4 injection machinery works including energetics, forces, and dynamic pathway**. *Proc Natl Acad Sci U S A* 2019, **116**:25097–25105.

68. Plisson C, White HE, Auzat I, Zafarani A, São-José C, Lhuillier S, Tavares P, Orlova E V: **Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection**. *EMBO J* 2007, **26**:3720–3728.

69. Hu B, Margolin W, Molineux IJ, Liu J: **Structural remodeling of bacteriophage T4 and host membranes during infection initiation**. *Proceedings of the National Academy of Sciences* 2015, **112**:E4919–E4928.

70. Davison J: **Pre-early functions of bacteriophage T5 and its relatives**. *Bacteriophage* 2015, **5**:e1086500.

71. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S: **Switches in Bacteriophage Lambda Development**. *Annu Rev Genet* 2005, **39**:409–429.

72. Wadhwa D: **Design Principles of Lambda's Lysis/Lysogeny Decision vis-a-vis Multiplicity of Infection**. *bioRxiv* 2020, doi:10.1101/146308.

73. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S, et al.: **Communication between viruses guides lysis–lysogeny decisions**. *Nature* 2017, **541**:488–493.

74. Greenblatt J, Mah T-F, Legault P, Mogridge J, Li J, Kay LE: **Structure and Mechanism in Transcriptional Antitermination by the Bacteriophage λ N Protein**. *Cold Spring Harb Symp Quant Biol* 1998, **63**:327–336.

75. Ubeda C, Maiques E, Knecht E, Lasa I, Novick RP, Penadés JR: **Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci**. *Mol Microbiol* 2005, **56**:836–844.

76. Imamovic L, Ballesté E, Martínez-Castillo A, García-Aljaro C, Muniesa M: **Heterogeneity in phage induction enables the survival of the lysogenic population**. *Environ Microbiol* 2016, **18**:957–969.

77. Marr MT, Datwyler SA, Meares CF, Roberts JW: **Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins**. *Proceedings of the National Academy of Sciences* 2001, **98**:8972–8978.

78. Aksyuk AA, Rossmann MG: **Bacteriophage Assembly**. *Viruses 2011, Vol 3, Pages 172-203* 2011, **3**:172–203.

79. Dokland T: **Scaffolding proteins and their role in viral assembly**. *Cellular and Molecular Life Sciences* 1999, **56**:580–603.

80. Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C: **The bacteriophage ϕ29 portal motor can package DNA against a large internal force**. *Nature 2001 413:6857* 2001, **413**:748–752.

81. Riemer SC, Bloomfield VA: **Packaging of DNA in bacteriophage Heads: Some considerations on energetics**. *Biopolymers* 1978, **17**:785–794.

82. Camacho A, Jiménez F, Viñuela E, Salas M: **Order of assembly of the lower collar and the tail proteins of Bacillus subtilis bacteriophage phi 29.** *J Virol* 1979, **29**:540.

83. Young R: **Phage lysis: three steps, three choices, one outcome**. *J Microbiol* 2014, **52**:243.

84. Altman E, Altman RK, Garrett JM, Grimaila RJ, Young R: **S gene product: identification and membrane localization of a lysis control protein.** *J Bacteriol* 1983, **155**:1130.

85. Dewey JS, Savva CG, White RL, Vitha S, Holzenburg A, Young R: **Micron-scale holes terminate the phage infection cycle**. *Proc Natl Acad Sci U S A* 2010, **107**:2219.

86. White R, Chiba S, Pang T, Dewey JS, Savva CG, Holzenburg A, Pogliano K, Young R: **Holin triggering in real time**. *Proc Natl Acad Sci U S A* 2011, **108**:798–803.

87. Rajaure M, Berry J, Kongari R, Cahill J, Young R: **Membrane fusion during phage lysis**. *Proc Natl Acad Sci U S A* 2015, **112**:5497–5502.

88. Pang T, Fleming TC, Pogliano K, Young R: **Visualization of pinholin lesions in vivo**. *Proc Natl Acad Sci U S A* 2013, **110**:E2054.

89. Young R: **Phage lysis: Do we have the hole story yet?** *Curr Opin Microbiol* 2013, **16**:790–797.

90. Caspar DL, Klug A: **Physical principles in the construction of regular viruses.** *Cold Spring Harb Symp Quant Biol* 1962, **27**:1–24.

91. Brüssow H, Hendrix RW: **Phage Genomics: Small Is Beautiful**. *Cell* 2002, **108**:13–16.

92. Lu S, Le S, Tan Y, Zhu J, Li M, Rao X, Zou L, Li S, Wang J, Jin X, et al.: **Genomic and proteomic analyses of the terminally redundant genome of the Pseudomonas aeruginosa phage PaP1: establishment of genus PaP1-like phages**. *PLoS One* 2013, **8**:e62933–e62933.

93. van Zyl LJ, Nemavhulani S, Cass J, Cowan DA, Trindade M: **Three novel bacteriophages isolated from the East African Rift Valley soda lakes**. *Virol J* 2016, **13**:204.

94. Comeau AM, Bertrand C, Letarov A, Tétart F, Krisch HM: **Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery**. *Virology* 2007, **362**:384–396.

95. Botstein D: **A THEORY OF MODULAR EVOLUTION FOR BACTERIOPHAGES***. *Ann N Y Acad Sci* 1980, **354**:484–491.

96. Lima-Mendez G, Toussaint A, Leplae R: **A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules**. *Res Microbiol* 2011, **162**:737–746.

97. Fortier L-C, Sekulovic O: **Importance of prophages to evolution and virulence of bacterial pathogens**. *Virulence* 2013, **4**:354–365.

98. Du Toit A: **Phages ensure their survival**. *Nat Rev Microbiol* 2019, **17**:529.

99. Labrie SJ, Samson JE, Moineau S: **Bacteriophage resistance mechanisms**. *Nat Rev Microbiol* 2010, **8**:317–327.

100. Grant WD, Sutherland IW, Wilkinson JF: **Exopolysaccharide colanic acid and its occurrence in the Enterobacteriaceae**. *J Bacteriol* 1969, **100**:1187–1193.

101. Ryu J-H, Beuchat LR: **Biofilm formation by Escherichia coli O157:H7 on stainless steel: effect of exopolysaccharide and Curli production on its resistance to chlorine**. *Appl Environ Microbiol* 2005, **71**:247–254.

102. Vidakovic L, Singh PK, Hartmann R, Nadell CD, Drescher K: **Dynamic biofilm architecture confers individual and collective mechanisms of viral protection**. *Nat Microbiol* 2018, **3**:26–31.

103. Sutherland IW: **Polysaccharide lyases**. *FEMS Microbiol Rev* 1995, **16**:323–347.

104. Hughes KA, Sutherland IW, Jones M V: **Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase**. *Microbiology (N Y)* 1998, **144**:3039–3047.

105. Riede I, Eschbach ML: **Evidence that TraT interacts with OmpA of Escherichia coli**. *FEBS Lett* 1986, **205**:241–245.

106. Pedruzzi I, Rosenbusch JP, Locher KP: **Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein**. *FEMS Microbiol Lett* 1998, **168**:119–125.

107. Destoumieux-Garzón D, Duquesne S, Peduzzi J, Goulard C, Desmadril M, Letellier L, Rebuffat S, Boulanger P: **The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val11-Pro16 beta-hairpin region in the recognition mechanism**. *Biochem J* 2005, **389**:869–876.

108. Reyes-Robles T, Dillard RS, Cairns LS, Silva-Valenzuela CA, Housman M, Ali A, Wright ER, Camilli A: **Vibrio cholerae Outer Membrane Vesicles Inhibit Bacteriophage Infection**. *J Bacteriol* 2018, **200**.

109. Tzipilevich E, Habusha M, Ben-Yehuda S: **Acquisition of Phage Sensitivity by Bacteria through Exchange of Phage Receptors**. *Cell* 2017, **168**:186-199.e12.

110. Susskind MM, Wright A, Botstein D: **Superinfection exclusion by P22 prophage in lysogens of Salmonella typhimurium: II. Genetic evidence for two exclusion systems**. *Virology* 1971, **45**:638–652.

111. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL: **Prophages mediate defense against phage infection through diverse mechanisms**. *ISME J* 2016, **10**:2854–2866.

112. Lu M-J, Henning U: **Superinfection exclusion by T-even-type coliphages**. *Trends Microbiol* 1994, **2**:137–139.

113. Lu MJ, Stierhof YD, Henning U: **Location and unusual membrane topology of the immunity protein of the Escherichia coli phage T4**. *J Virol* 1993, **67**:4905–4913.

114. Sun X, Göhler A, Heller KJ, Neve H: **The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus thermophilus and Lactococcus lactis**. *Virology* 2006, **350**:146–157.

115. Bebeacua C, Lorenzo Fajardo JC, Blangy S, Spinelli S, Bollmann S, Neve H, Cambillau C, Heller KJ: **X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target**. *Mol Microbiol* 2013, **89**:152–165.

116. Harms A, Brodersen DE, Mitarai N, Gerdes K: **Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology**. *Mol Cell* 2018, **70**:768–784.

117. Durmaz E, Higgins DL, Klaenhammer TR: **Molecular characterization of a second abortive phage resistance gene present in Lactococcus lactis subsp. lactis ME2**. *J Bacteriol* 1992, **174**:7463–7469.

118. Garvey P, Fitzgerald GF, Hill C: **Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40**. *Appl Environ Microbiol* 1995, **61**:4321–4328.

119. Su P, Harvey M, Im HJ, Dunn NW: **Isolation, cloning and characterisation of the abiI gene from Lactococcus lactis subsp. lactis M138 encoding abortive phage infection**. *J Biotechnol* 1997, **54**:95–104.

120. Emond E, Dion E, Walker SA, Vedamuthu ER, Kondo JK, Moineau S: **AbiQ, an abortive infection mechanism from Lactococcus lactis**. *Appl Environ Microbiol* 1998, **64**:4748–4756.

121. Durmaz E, Klaenhammer TR: **Abortive Phage Resistance Mechanism AbiZ Speeds the Lysis Clock To Cause Premature Lysis of Phage-Infected Lactococcus lactis**. *J Bacteriol* 2007, **189**:1417–1425.

122. Parma DH, Snyder M, Sobolevski S, Nawroz M, Brody E, Gold L: **The Rex system of bacteriophage lambda: tolerance and altruistic cell death**. *Genes Dev* 1992, **6**:497–510.

123. Gentile GM, Wetzel KS, Dedrick RM, Montgomery MT, Garlena RA, Jacobs-Sera D, Hatfull GF: **More Evidence of Collusion: a New Prophage-Mediated Viral Defense System Encoded by Mycobacteriophage Sbash**. *mBio* 2019, **10**.

124. Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull GF: **Yet More Evidence of Collusion: a New Viral Defense System Encoded by Gordonia Phage CarolAnn**. *mBio* 2019, **10**.

125. Page R, Peti W: **Toxin-antitoxin systems in bacterial growth arrest and persistence**. *Nat Chem Biol* 2016, **12**:208–214.

126. Chopin M-C, Chopin A, Bidnenko E: **Phage abortive infection in lactococci: variations on a theme**. *Curr Opin Microbiol* 2005, **8**:473–479.

127. Samson JE, Spinelli S, Cambillau C, Moineau S: **Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin–antitoxin system**. *Mol Microbiol* 2013, **87**:756–768.

128. Samson JE, Bélanger M, Moineau S: **Effect of the abortive infection mechanism and type III toxin/antitoxin system AbiQ on the lytic cycle of Lactococcus lactis phages**. *J Bacteriol* 2013, **195**:3947–3956.

129. Bélanger M, Moineau S: **Mutational Analysis of the Antitoxin in the Lactococcal Type III Toxin-Antitoxin System AbiQ**. *Appl Environ Microbiol* 2015, **81**:3848–3855.

130. Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC: **A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism**. *Nucleic Acids Res* 2014, **42**:4590–4605.

131. Usher B: **Slowing down to get ahead: functional and structural characterisation of toxin-antitoxin systems from Mycobacterium tuberculosis**. 2021,

132. Hampton HG, Jackson SA, Fagerlund RD, Vogel AIM, Dy RL, Blower TR, Fineran PC: **AbiEi Binds Cooperatively to the Type IV abiE Toxin–Antitoxin Operator Via a Positively-Charged Surface and Causes DNA Bending and Negative Autoregulation**. *J Mol Biol* 2018, **430**:1141–1156.

133. Hayes F, Kędzierska B: **Regulating Toxin-Antitoxin Expression: Controlled Detonation of Intracellular Molecular Timebombs**. *Toxins (Basel)* 2014, **6**:337–358.

134. Millman A, Melamed S, Leavitt A, Kacen A, Amitai G, Correspondence RS, Doron S, Bernheim A, Hö J, Garb J, et al.: **An expanded arsenal of immune systems that protect bacteria from phages**. *Cell Host Microbe* 2022, **30**:1556-1569.e5.

135. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A: **Systematic and quantitative view of the antiviral arsenal of prokaryotes**. *Nature Communications 2022 13:1* 2022, **13**:1–10.

136. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R: **Systematic discovery of antiphage defense systems in the microbial pangenome**. *Science (1979)* 2018, **359**.

137. Kelly A, Arrowsmith TJ, Went SC, Blower TR: **Toxin–antitoxin systems as mediators of phage defence and the implications for abortive infection**. *Curr Opin Microbiol* 2023, **73**:102293.

138. Ramage HR, Connolly LE, Cox JS: **Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution**. *PLoS Genet* 2009, **5**:e1000767-.

139. LeRoux M, Srikant S, Teodoro GICC, Zhang T, Littlehale ML, Doron S, Badiee M, Leung AKLL, Sorek R, Laub MT: **The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA**. *Nat Microbiol* 2022, **7**:1028–1040.

140. Zhang T, Tamman H, Wallant KC, Kurata T, LeRoux M, Srikant S, Brodiazhenko T, Cepauskas A, Talavera A, Martens C, et al.: **Direct activation of an innate immune system in bacteria by a viral capsid protein**. *bioRxiv* 2022, doi:10.1101/2022.05.30.493996.

141. Jankevicius G, Ariza A, Ahel M, Ahel I: **The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA**. *Mol Cell* 2016, **64**:1109–1116.

142. Guegler CK, Laub MT: **Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection**. *Mol Cell* 2021, **81**:2361-2373.e9.

143. Takada K, Hama K, Sasaki T, Otsuka Y: **The hokW-sokW locus encodes a type I toxin–Antitoxin system that facilitates the release of lysogenic Sp5 phage in enterohemorrhagic escherichia coli O157**. *Toxins (Basel)* 2021, **13**.

144. Guo Y, Tang K, Sit B, Gu J, Chen R, Lin J, Lin S, Liu X, Wang W, Gao X, et al.: **Dual control of lysogeny and phage defense by a phosphorylation-based toxin/antitoxin system**. *bioRxiv* 2022, doi:10.1101/2022.09.05.506569.

145. Lampson BC, Inouye M, Inouye S: **Reverse transcriptase with concomitant ribonuclease H activity in the cell-free synthesis of branched RNA-linked msDNA of Myxococcus xanthus**. *Cell* 1989, **56**:701–707.

146. Rychlik I, Sebkova A, Gregorova D, Karpiskova R: **Low-molecular-weight plasmid of Salmonella enterica serovar Enteritidis codes for retron reverse transcriptase and influences phage resistance**. *J Bacteriol* 2001, **183**:2852–2858.

147. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R: **Bacterial Retrons Function In Anti-Phage Defense**. *Cell* 2020, **183**:1551-1561.e12.

148. Palka C, Fishman CB, Bhattarai-Kline S, Myers SA, Shipman SL: **Retron reverse transcriptase termination and phage defense are dependent on host RNase H1**. *Nucleic Acids Res* 2022, **50**:3490–3504.

149. Bobonis J, Mitosch K, Mateus A, Karcher N, Kritikos G, Selkrig J, Zietek M, Monzon V, Pfalz B, Garcia-Santamarina S, et al.: **Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems**. *Nature* 2022, **609**:144–150.

150. Stokar-Avihail A, Fedorenko T, Hör J, Garb J, Leavitt A, Millman A, Shulman G, Wojtania N, Melamed S, Amitai G, et al.: **Discovery of phage determinants that confer sensitivity to bacterial immune systems**. *Cell* 2023, **186**:1863-1876.e16.

151. Cai Y, Usher B, Gutierrez C, Tolcan A, Mansour M, Fineran PC, Ciarán Condon, Neyrolles O, Genevaux P, Blower TR: **A nucleotidyltransferase toxin inhibits growth of Mycobacterium tuberculosis through inactivation of tRNA acceptor stems**. *Sci Adv* 2020, **6**:6651–6680.

152. Boyer H: **GENETIC CONTROL OF RESTRICTION AND MODIFICATION IN ESCHERICHIA COLI**. *J Bacteriol* 1964, **88**:1652–1660.

153. Vasu K, Nagaraja V: **Diverse functions of restriction-modification systems in addition to cellular defense**. *Microbiol Mol Biol Rev* 2013, **77**:53–72.

154. Loenen WA, Dryden DT, Raleigh EA, Wilson GG: **Type I restriction enzymes and their relatives**. *Nucleic Acids Res* 2014, **42**:20–44.

155. Kennaway CK, Obarska-Kosinska A, White JH, Tuszynska I, Cooper LP, Bujnicki JM, Trinick J, Dryden DTF: **The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein**. *Nucleic Acids Res* 2009, **37**:762–770.

156. Liu YP, Tang Q, Zhang JZ, Tian LF, Gao P, Yan XX: **Structural basis underlying complex assembly and conformational transition of the type I R-M system**. *Proc Natl Acad Sci U S A* 2017, **114**:11151–11156.

157. Gao Y, Cao D, Zhu J, Feng H, Luo X, Liu S, Yan XX, Zhang X, Gao P: **Structural insights into assembly, operation and inhibition of a type I restriction–modification system**. *Nature Microbiology 2020 5:9* 2020, **5**:1107–1118.

158. Yuan R, Hamilton DL, Burckhardt J: **DNA translocation by the restriction enzyme from E. coli K**. *Cell* 1980, **20**:237–244.

159. Kennaway CK, Taylor JE, Song CF, Potrzebowski W, Nicholson W, White JH, Swiderska A, Obarska-Kosinska A, Callow P, Cooper LP, et al.: **Structure and operation of the DNA-translocating type I DNA restriction enzymes**. *Genes Dev* 2012, **26**:92–104.

160. Roberts RJ, Vincze T, Posfai J, Macelis D: **REBASE—a database for DNA restriction and modification: enzymes, genes and genomes**. *Nucleic Acids Res* 2015, **43**:D298–D299.

161. Morgan RD, Luyten YA, Johnson SA, Clough EM, Clark TA, Roberts RJ: **Novel m4C modification in type I restriction-modification systems**. *Nucleic Acids Res* 2016, **44**:9413–9425.

162. Zhu J, Gao Y, Wang Y, Zhan Q, Feng H, Luo X, Li P, Liu S, Hou H, Gao P: **Molecular insights into DNA recognition and methylation by non-canonical type I restriction-modification systems**. *Nature Communications 2022 13:1* 2022, **13**:1–12.

163. Meselson M, Yuan R: **DNA Restriction Enzyme from E. coli**. *Nature 1968 217:5134* 1968, **217**:1110–1114.

164. Pingoud A, Wilson GG, Wende W: **Type II restriction endonucleases--a historical perspective and more**. *Nucleic Acids Res* 2014, **42**:7489–7527.

165. Pingoud A, Fuxreiter M, Pingoud V, Wende W: **Type II restriction endonucleases: Structure and mechanism**. *Cellular and Molecular Life Sciences* 2005, **62**:685–707.

166. Berkhout B, Van Wamel J: **Accurate scanning of the BssHII endonuclease in search for its DNA cleavage site**. *Journal of Biological Chemistry* 1996, **271**:1837–1840.

167. Jeltsch A, Pingoud A: **Kinetic Characterization of Linear Diffusion of the Restriction Endonuclease EcoRV on DNA†**. *Biochemistry* 1998, **37**:2160–2169.

168. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DTF, Dybvig K, et al.: **A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes**. *Nucleic Acids Res* 2003, **31**:1805–1812.

169. Engler C, Kandzia R, Marillonnet S: **A One Pot, One Step, Precision Cloning Method with High Throughput Capability**. *PLoS One* 2008, **3**:e3647.

170. Lee JH, Skowron PM, Rutkowska SM, Hong SS, Kim SC: **Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes**. *Genet Anal* 1996, **13**:139–145.

171. Chan SH, Stoddard BL, Xu SY: **Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity**. *Nucleic Acids Res* 2011, **39**:1–18.

172. Heiter DF, Lunnen KD, Wilson GG: **Site-Specific DNA-nicking Mutants of the Heterodimeric Restriction Endonuclease R.BbvCI**. *J Mol Biol* 2005, **348**:631–640.

173. Morgan RD, Bhatia TK, Lovasco L, Davis TB: **MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection**. *Nucleic Acids Res* 2008, **36**:6558–6570.

174. Morgan RD, Dwinell EA, Bhatia TK, Lang EM, Luyten YA: **The MmeI family: type II restriction–modification enzymes that employ single-strand modification for host protection**. *Nucleic Acids Res* 2009, **37**:5208.

175. Rao DN, Dryden DT, Bheemanaik S: **Type III restriction-modification enzymes: a historical perspective**. *Nucleic Acids Res* 2014, **42**:45–55.

176. Gupta YK, Yang L, Chan SH, Samuelson JC, Xu SY, Aggarwal AK: **Structural Insights into the Assembly and Shape of Type III Restriction–Modification (R–M) EcoP15I Complex by Small-Angle X-ray Scattering**. *J Mol Biol* 2012, **420**:261–268.

177. Wyszomirski KH, Curth U, Alves J, MacKeldanz P, Möncke-Buchner E, Schutkowski M, Krüger DH, Reuter M: **Type III restriction endonuclease EcoP15I is a heterotrimeric complex containing one Res subunit with several DNA-binding regions and ATPase activity**. *Nucleic Acids Res* 2012, **40**:3610–3622.

178. Van Aelst K, Tóth J, Ramanathan SP, Schwarz FW, Seidel R, Szczelkun MD: **Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to-head and tail-to-tail inverted repeat**. *Proc Natl Acad Sci U S A* 2010, **107**:9123–9128.

179. Peakman LJ, Szczelkun MD: **DNA communications by Type III restriction endonucleases—confirmation of 1D translocation over 3D looping**. *Nucleic Acids Res* 2004, **32**:4166–4174.

180. Meisel A, Mackeldanz P, Bickle TA, Krüger DH, Schroeder C: **Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis.** *EMBO J* 1995, **14**:2958.

181. Loenen WA, Raleigh EA: **The other face of restriction: modification-dependent enzymes**. *Nucleic Acids Res* 2014, **42**:56–69.

182. Bair CL, Black LW: **A Type IV Modification Dependent Restriction Nuclease that Targets Glucosylated Hydroxymethyl Cytosine Modified DNAs**. *J Mol Biol* 2007, **366**:768–778.

183. Picton DM, Luyten YA, Morgan RD, Nelson A, Smith DL, Dryden DTF, Hinton JCD, Blower TR: **The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses**. *Nucleic Acids Res* 2021, **49**:11257.

184. Sood AJ, Viner C, Hoffman MM: **DNamod: The DNA modification database**. *J Cheminform* 2019, **11**:1–10.

185. Weigele P, Raleigh EA: **Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses**. *Chem Rev* 2016, **116**:12655–12687.

186. Su TJ, Tock MR, Egelhaaf SU, Poon WCK, Dryden DTF: **DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping**. *Nucleic Acids Res* 2005, **33**:3235.

187. Hutinet G, Kot W, Cui L, Hillebrand R, Balamkundu S, Gnanakalai S, Neelakandan R, Carstens AB, Fa Lui C, Tremblay D, et al.: **7-Deazaguanine modifications protect phage DNA from host restriction systems**. *Nature Communications 2019 10:1* 2019, **10**:1–12.

188. Xiong X, Wu G, Wei Y, Liu L, Zhang Y, Su R, Jiang X, Li M, Gao H, Tian X, et al.: **SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities**. *Nat Microbiol* 2020, doi:10.1038/s41564-020-0700-6.

189. Wang L, Chen S, Vergin KL, Giovannoni SJ, Chan SW, DeMott MS, Taghizadeh K, Cordero OX, Cutler M, Timberlake S, et al.: **DNA phosphorothioation is widespread and quantized in bacterial genomes**. *Proceedings of the National Academy of Sciences* 2011, **108**:2963–2968.

190. Wang L, Jiang S, Deng Z, Dedon PC, Chen S: **DNA phosphorothioate modification—a new multi-functional epigenetic system in bacteria**. *FEMS Microbiol Rev* 2018, **43**:109–122.

191. Rimseliene R, Maneliene Z, Lubys A, Janulaitis A: **Engineering of Restriction Endonucleases: Using Methylation Activity of the Bifunctional Endonuclease Eco57I to Select the Mutant with a Novel Sequence Specificity**. *J Mol Biol* 2003, **327**:383–391.

192. Callahan SJ, Luyten YA, Gupta YK, Wilson GG, Roberts RJ, Morgan RD, Aggarwal AK: **Structure of Type IIL Restriction-Modification Enzyme MmeI in Complex with DNA Has Implications for Engineering New Specificities**. *PLoS Biol* 2016, **14**:e1002442.

193. Skowronek K, Boniecki MJ, Kluge B, Bujnicki JM: **Rational engineering of sequence specificity in R.MwoI restriction endonuclease**. *Nucleic Acids Res* 2012, **40**:8579.

194. Morgan RD, Luyten YA: **Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity**. *Nucleic Acids Res* 2009, **37**:5222.

195. Smith JS, Nikonowicz EP: **Phosphorothioate Substitution Can Substantially Alter RNA Conformation**. *Biochemistry* 2000, **39**:5642–5652.

196. Eckstein F: **Phosphorothioate analogs of nucleotides**. *Acc Chem Res* 1979, **12**:204–210.

197. Eckstein F, Gish G: **Phosphorothioates in molecular biology**. *Trends Biochem Sci* 1989, **14**:97–100.

198. Zhou X, Deng Z, Firmin JL, Hopwood DA, Kieser T: **Site-specific degradation of Streptomyces lividans DNA during electrophoresis in buffers contaminated with ferrous iron**. *Nucleic Acids Res* 1988, **16**:4341–4352.

199. Zhou X, He X, Liang J, Li A, Xu T, Kieser T, Helmann JD, Deng Z: **A novel DNA modification by sulphur**. *Mol Microbiol* 2005, **57**:1428–1438.

200. Boybek A, Ray TD, Evans MC, Dyson PJ: **Novel site-specific DNA modification in Streptomyces: Analysis of preferred intragenic modification sites present in a 5.7 kb amplified DNA sequence**. *Nucleic Acids Res* 1998, **26**:3364–3371.

201. Xu T, Yao F, Zhou X, Deng Z, You D: **A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella**. *Nucleic Acids Res* 2010, **38**:7133–7141.

202. Hsu PD, Lander ES, Zhang F: **Development and Applications of CRISPR-Cas9 for Genome Engineering**. *Cell* 2014, **157**:1262–1278.

203. Adli M: **The CRISPR tool kit for genome editing and beyond**. *Nat Commun* 2018, **9**:1911.

204. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes**. *Science (1979)* 2007, **315**:1709–1712.

205. Koonin E V, Makarova KS, Zhang F: **Diversity, classification and evolution of CRISPR-Cas systems**. *Curr Opin Microbiol* 2017, **37**:67–78.

206. Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E: **The Biology of CRISPR-Cas: Backward and Forward**. *Cell* 2018, **172**:1239–1259.

207. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJ: **CRISPR-Cas: Adapting to change**. *Science (1979)* 2017, **356**.

208. McGinn J, Marraffini LA: **CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration**. *Mol Cell* 2016, **64**:616–623.

209. Wei Y, Terns RM, Terns MP: **Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation**. *Genes Dev* 2015, **29**:356–361.

210. Swarts DC, Mosterd C, van Passel MWJ, Brouns SJJ: **CRISPR Interference Directs Strand Specific Spacer Acquisition**. *PLoS One* 2012, **7**:e35888.

211. Díez-Villaseñor C, Guzmán NM, Almendros C, García-Martínez J, Mojica FJM: **CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli**. *RNA Biol* 2013, **10**:792–802.

212. Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R: **CRISPR adaptation biases explain preference for acquisition of foreign DNA**. *Nature* 2015, **520**:505–510.

213. Ivančić-Baće I, Cass SD, Wearne SJ, Bolt EL: **Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity**. *Nucleic Acids Res* 2015, **43**:10821–10830.

214. Wigley DB: **Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB**. *Nat Rev Microbiol* 2013, **11**:9–13.

215. Watson BNJ, Staals RHJ, Fineran PC: **CRISPR-Cas-Mediated Phage Resistance Enhances Horizontal Gene Transfer by Transduction**. *mBio* 2018, **9**:e02406-17.

216. Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF: **Phage mutations in response to CRISPR diversification in a bacterial population**. *Environ Microbiol* 2013, **15**:463–470.

217. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ, Severinov K: **Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence**. *Proceedings of the National Academy of Sciences* 2011, **108**:10098–10103.

218. Strotskaya A, Savitskaya E, Metlitskaya A, Morozova N, Datsenko KA, Semenova E, Severinov K: **The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies**. *Nucleic Acids Res* 2017, **45**:1946–1957.

219. Carte J, Wang R, Li H, Terns RM, Terns MP: **Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes**. *Genes Dev* 2008, **22**:3489–3496.

220. Gesner EM, Schellenberg MJ, Garside EL, George MM, MacMillan AM: **Recognition and maturation of effector RNAs in a CRISPR interference pathway**. *Nat Struct Mol Biol* 2011, **18**:688–692.

221. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul Ü, Wurm R, Wagner R, et al.: **Structural basis for CRISPR RNA-guided DNA recognition by Cascade**. *Nat Struct Mol Biol* 2011, **18**:529–536.

222. Xiao Y, Luo M, Hayes RP, Kim J, Ng S, Ding F, Liao M, Ke A: **Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System**. *Cell* 2017, **170**:48-60.e11.

223. Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, Farchione MD, Zhou S, Rajashankar K, Kurinov I, et al.: **Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation**. *Nat Struct Mol Biol* 2014, **21**:771–777.

224. Deng L, Garrett RA, Shah SA, Peng X, She Q: **A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus**. *Mol Microbiol* 2013, **87**:1088–1099.

225. Estrella MA, Kuo FT, Bailey S: **RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex**. *Genes Dev* 2016, **30**:460–470.

226. Niewoehner O, Garcia-Doval C, Rostøl JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M: **Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers**. *Nature* 2017, **548**:543–548.

227. Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V: **A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems**. *Science (1979)* 2017, **357**:605–609.

228. Athukoralage JS, White MF: **Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense**. *RNA* 2021, **27**:855.

229. Duncan-Lowey B, Kranzusch PJ: **CBASS phage defense and evolution of antiviral nucleotide signaling**. *Curr Opin Immunol* 2022, **74**:156–163.

230. Høyland-Kroghsbo NM, Maerkedahl RB, Svenningsen SL: **A quorum-sensing-induced bacteriophage defense mechanism**. *mBio* 2013, **4**:e00362-12.

231. Høyland-Kroghsbo NM, Paczkowski J, Mukherjee S, Broniewski J, Westra E, Bondy-Denomy J, Bassler BL: **Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system**. *Proc Natl Acad Sci U S A* 2017, **114**:131–135.

232. Patterson AG, Jackson SA, Taylor C, Evans GB, Salmond GPC, Przybilski R, Staals RHJ, Fineran PC: **Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems**. *Mol Cell* 2016, **64**:1102–1108.

233. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E: **CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III**. *Nature* 2011, **471**:602–607.

234. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, et al.: **Structures of Cas9 endonucleases reveal RNA-mediated conformational activation**. *Science (1979)* 2014, **343**:1247997.

235. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA: **Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage**. *Science (1979)* 2016, **351**:867–871.

236. Sternberg SH, LaFrance B, Kaplan M, Doudna JA: **Conformational control of DNA target cleavage by CRISPR–Cas9**. *Nature* 2015, **527**:110–113.

237. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E: **The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA**. *Nature* 2016, **532**:517–521.

238. Smargon AA, Cox DBT, Pyzocha NK, Zheng K, Slaymaker IM, Gootenberg JS, Abudayyeh OA, Essletzbichler P, Shmakov S, Makarova KS, et al.: **Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28**. *Mol Cell* 2017, **65**:618-630.e7.

239. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al.: **Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system**. *Cell* 2015, **163**:759–771.

240. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, et al.: **C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector**. *Science (1979)* 2016, **353**:aaf5573.

241. Liu L, Li X, Ma J, Li Z, You L, Wang J, Wang M, Zhang X, Wang Y: **The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a**. *Cell* 2017, **170**:714-726.e10.

242. Makarova KS, Wolf YI, Snir S, Koonin E V: **Defense islands in bacterial and archaeal genomes and prediction of novel defense systems**. *J Bacteriol* 2011, **193**:6039–6056.

243. Rousset F, Depardieu F, Miele S, Dowding J, Laval AL, Lieberman E, Garry D, Rocha EPC, Bernheim A, Bikard D: **Phages and their satellites encode hotspots of antiviral systems**. *Cell Host Microbe* 2022, **30**:740.

244. Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin E V., Zhang F: **Diverse enzymatic activities mediate antiviral immunity in prokaryotes**. *Science (1979)* 2020, **369**:1077–1084.

245. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT: **A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome**. *Nat Microbiol* 2022, **7**:1568–1579.

246. Johnson MC, Laderman E, Huiting E, Zhang C, Davidson A, Bondy-Denomy J: **Core defense hotspots within Pseudomonas aeruginosa are a consistent and rich source of anti-phage defense systems**. *Nucleic Acids Res* 2023, **51**:4995–5005.

247. Payne LJ, Meaden S, Mestre MR, Palmer C, Toro N, Fineran PC, Jackson SA: **PADLOC: a web server for the identification of antiviral defence systems in microbial genomes**. *Nucleic Acids Res* 2022, **50**:W541–W550.

248. Makarova KS, Wolf YI, van der Oost J, Koonin E V: **Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements**. *Biol Direct* 2009, **4**:29.

249. Joshua-Tor L, Hannon GJ: **Ancestral roles of small RNAs: an Ago-centric perspective**. *Cold Spring Harb Perspect Biol* 2011, **3**:a003772.

250. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJJ, et al.: **DNA-guided DNA interference by a prokaryotic Argonaute**. *Nature* 2014, **507**:258–261.

251. Willkomm S, Makarova KS, Grohmann D: **DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids**. *FEMS Microbiol Rev* 2018, **42**:376–387.

252. Kaya E, Doxzen KW, Knoll KR, Wilson RC, Strutt SC, Kranzusch PJ, Doudna JA: **A bacterial Argonaute with noncanonical guide RNA specificity**. *Proc Natl Acad Sci U S A* 2016, **113**:4057–4062.

253. Zeng Z, Chen Y, Pinilla-Redondo R, Shah SA, Zhao F, Wang C, Hu Z, Wu C, Zhang C, Whitaker RJ, et al.: **A short prokaryotic Argonaute activates membrane effector to confer antiviral defense**. *Cell Host Microbe* 2022, **30**:930-943.e6.

254. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R: **DISARM is a widespread bacterial defence system with broad anti-phage activities**. *Nat Microbiol* 2018, **3**:90–98.

255. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R: **Cyclic GMP–AMP signalling protects bacteria against viral infection**. *Nature 2019 574:7780* 2019, **574**:691–695.

256. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M, et al.: **CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection**. *Cell* 2020, **182**:38-49.e17.

257. Whiteley AT, Eaglesham JB, de Oliveira Mann CC, Morehouse BR, Lowey B, Nieminen EA, Danilchanka O, King DS, Lee ASY, Mekalanos JJ, et al.: **Bacterial cGAS-like enzymes synthesize diverse nucleotide signals**. *Nature 2019 567:7747* 2019, **567**:194–199.

258. Ye Q, Lau RK, Mathews IT, Birkholz EA, Watrous JD, Azimi CS, Pogliano J, Jain M, Corbett KD: **HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity**. *Mol Cell* 2020, **77**:709-722.e7.

259. Morehouse BR, Govande AA, Millman A, Keszei AFA, Lowey B, Ofir G, Shao S, Sorek R, Kranzusch PJ: **STING cyclic dinucleotide sensing originated in bacteria**. *Nature 2020 586:7829* 2020, **586**:429–433.

260. Duncan-Lowey B, McNamara-Bordewick NK, Tal N, Sorek R, Kranzusch PJ: **Effector-mediated membrane disruption controls cell death in CBASS antiphage defense**. *Mol Cell* 2021, **81**:5039-5051.e5.

261. Fatma S, Chakravarti A, Zeng X, Huang RH: **Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3',2'-cGAMP as the second messenger**. *Nature Communications 2021 12:1* 2021, **12**:1–9.

262. Tal N, Morehouse BR, Millman A, Stokar-Avihail A, Avraham C, Fedorenko T, Yirmiya E, Herbst E, Brandis A, Mehlman T, et al.: **Cyclic CMP and cyclic UMP mediate bacterial immunity against phages**. *Cell* 2021, **184**:5728-5739.e16.

263. Jenal U, Reinders A, Lori C: **Cyclic di-GMP: second messenger extraordinaire**. *Nature Reviews Microbiology 2017 15:5* 2017, **15**:271–284.

264. Latoscha A, Wörmann ME, Tschowri N: **Nucleotide second messengers in streptomyces**. *Microbiology (United Kingdom)* 2019, **165**:1153–1165.

265. Stülke J, Krüger L: **Cyclic di-AMP Signaling in Bacteria**. *https://doi.org/101146/annurev-micro-020518-115943* 2020, **74**:159–179.

266. Hengge R, Pruteanu M, Stülke J, Tschowri N, Kür¸sad Turgay K: **Recent advances and perspectives in nucleotide second messenger signaling in bacteria**. *microLife* 2023, **4**:1–18.

267. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L: **Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling**. *Nucleic Acids Res* 2015, **43**:10633–10654.

268. Makarova KS, Anantharaman V, Grishin N V., Koonin E V., Aravind L: **CARF and WYL domains: Ligand-binding regulators of prokaryotic defense systems**. *Front Genet* 2014, **5**:87188.

269. Lau RK, Ye Q, Birkholz EA, Berg KR, Patel L, Mathews IT, Watrous JD, Ego K, Whiteley AT, Lowey B, et al.: **Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity**. *Mol Cell* 2020, **77**:723-733.e6.

270. Schatz A, Jones D: **The Production of Antiphage Agents by Actinomycetes**. *Bulletin of the Torrey Botanical Club* 1947, **74**:9.

271. Asheshov IN, Strelitz F, Hall E, Flon Helen: **A Survey of Actinomycetes for Antiphage Activity.** *Antibiot Chemother (1971)* 1954, **4**:380–94.

272. Hardy A, Kever L, Frunzke J: **Antiphage small molecules produced by bacteria – beyond protein-mediated defenses**. *Trends Microbiol* 2023, **31**:92–106.

273. Kever L, Hardy A, Luthe T, Hünnefeld M, Gätgens C, Milke L, Wiechert J, Wittmann J, Moraru C, Marienhagen J, et al.: **Aminoglycoside Antibiotics Inhibit Phage Infection by Blocking an Early Step of the Infection Cycle**. *mBio* 2022, **13**.

274. Bernheim A, Sorek R: **The pan-immune system of bacteria: antiviral defence as a community resource**. *Nature Reviews Microbiology 2019 18:2* 2019, **18**:113–119.

275. Lucas X, Senger C, Erxleben A, Grüning BA, Döring K, Mosch J, Flemming S, Günther S: **StreptomeDB: a resource for natural compounds isolated from Streptomyces species**. *Nucleic Acids Res* 2013, **41**:D1130–D1136.

276. Kronheim S, Daniel-Ivad M, Duan Z, Hwang S, Wong AI, Mantel I, Nodwell JR, Maxwell KL: **A chemical defence against phage infection**. *Nature* 2018, **564**:283–286.

277. Wu Y, Hurk A van den, Aparicio-Maldonado C, Kushwaha SK, King CM, Ou Y, Todeschini TC, Clokie MRJ, Millard AD, Gençay YE, et al.: **Defence systems provide synergistic anti-phage activity in E. coli**. *bioRxiv* 2022, doi:10.1101/2022.08.21.504612.

278. Dupuis MÈ, Villion M, Magadán AH, Moineau S: **CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance**. *Nature Communications 2013 4:1* 2013, **4**:1–7.

279. Blankenchip CL, Nguyen J V., Lau RK, Ye Q, Gu Y, Corbett KD: **Control of bacterial immune signaling by a WYL domain transcription factor**. *Nucleic Acids Res* 2022, **50**:5239–5250.

280. Picton DM, Harling-Lee JD, Duffner SJ, Went SC, Morgan RD, Hinton JCD, Blower TR: **A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands**. *Nucleic Acids Res* 2022, **50**:5191–5207.

281. Luyten YA, Hausman DE, Young JC, Doyle LA, Higashi KM, Ubilla-Rodriguez NC, Lambert AR, Arroyo CS, Forsberg KJ, Morgan RD, et al.: **Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system**. *Nucleic Acids Res* 2022, **50**:5171–5190.

282. Botelho J: **Defense systems are pervasive across chromosomally integrated mobile genetic elements and are inversely correlated to virulence and antimicrobial resistance**. *Nucleic Acids Res* 2023, **51**:4385–4397.

283. Bishop AL, Baker S, Jenks S, Fookes M, Gaora PÓ, Pickard D, Anjum M, Farrar J, Hien TT, Ivens A, et al.: **Analysis of the hypervariable region of the Salmonella enterica genome associated with tRNAleuX**. *J Bacteriol* 2005, **187**:2469–2482.

284. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, Kauffman K, Rodino-Janeiro BK, Gavin H, Gomez A, et al.: **Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages**. *Science* 2021, **374**:488–492.

285. Studier FW: **Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host**. *J Mol Biol* 1975, **94**.

286. Bandyopadhyay PK, Studier FW, Hamilton DL, Yuan R: **Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7**. *J Mol Biol* 1985, **182**:567–578.

287. Roberts GA, Stephanou AS, Kanwar N, Dawson A, Cooper LP, Chen K, Nutley M, Cooper A, Blakely GW, Dryden DTF: **Exploring the DNA mimicry of the Ocr protein of phage T7**. *Nucleic Acids Res* 2012, **40**:8129.

288. Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM, Dryden DTF: **Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA**. *Mol Cell* 2002, **9**:187–194.

289. Powell LM, Dryden DTF, Willcock DF, Pain RH, Murray NE: **DNA Recognition by the EcoK Methyltransferase: The Influence of DNA Methylation and the Cofactor S-adenosyl-l-methionine**. *J Mol Biol* 1993, **234**:60–71.

290. Atanasiu C, Su TJ, Sturrock SS, Dryden DTF: **Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme**. *Nucleic Acids Res* 2002, **30**:3936–3944.

291. Ye F, Kotta-Loizou I, Jovanovic M, Liu X, Dryden DTF, Buck M, Zhang X: **Structural basis of transcription inhibition by the DNA mimic protein Ocr of bacteriophage T7**. *Elife* 2020, **9**.

292. Hausmann R, Messerschmid M: **Inhibition of gene expression of T7-related phages by prophage P1**. *MGG Molecular & General Genetics* 1988, **212**:543–547.

293. Kudryavtseva AA, Cséfalvay E, Gnuchikh EY, Yanovskaya DD, Skutel MA, Isaev AB, Bazhenov S V., Utkina AA, Manukhov I V.: **Broadness and specificity: ArdB, ArdA, and Ocr against various restriction-modification systems**. *Front Microbiol* 2023, **14**:1133144.

294. Isaev A, Drobiazko A, Sierro N, Gordeeva J, Yosef I, Qimron U, Ivanov N V., Severinov K: **Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence**. *Nucleic Acids Res* 2020, **48**:5397–5406.

295. Trasanidou D, Gerós AS, Mohanraju P, Nieuwenweg AC, Nobrega FL, Staals RHJ: **Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs**. *FEMS Microbiol Lett* 2019, **366**:fnz098.

296. Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, Wiedenheft B, Maxwell KL, Davidson AR: **Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins**. *Nature* 2015, **526**:136–139.

297. Knott GJ, Thornton BW, Lobba MJ, Liu JJ, Al-Shayeb B, Watters KE, Doudna JA: **Broad-spectrum enzymatic inhibition of CRISPR-Cas12a**. *Nat Struct Mol Biol* 2019, **26**:315–321.

298. Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y, Lee J, Edraki A, Shah M, Sontheimer EJ, Maxwell KL, et al.: **Naturally Occurring Off-Switches for CRISPR-Cas9**. *Cell* 2016, **167**:1829-1838.e9.

299. Nakamura M, Srinivasan P, Chavez M, Carter MA, Dominguez AA, La Russa M, Lau MB, Abbott TR, Xu X, Zhao D, et al.: **Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells**. *Nat Commun* 2019, **10**:194.

300. Bubeck F, Hoffmann MD, Harteveld Z, Aschenbrenner S, Bietz A, Waldhauer MC, Börner K, Fakhiri J, Schmelas C, Dietz L, et al.: **Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9**. *Nat Methods* 2018, **15**:924–927.

301. Alawneh AM, Qi D, Yonesaki T, Otsuka Y: **An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin–antitoxin module**. *Mol Microbiol* 2016, **99**:188–198.

302. Sberro H, Leavitt A, Kiro R, Koh E, Peleg Y, Qimron U, Sorek R: **Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning**. *Mol Cell* 2013, **50**:136–148.

303. Wen Z, Wang P, Sun C, Guo Y, Wang X: **Interaction of type IV toxin/antitoxin systems in cryptic prophages of Escherichia coli K-12**. *Toxins (Basel)* 2017, **9**.

304. Song S, Wood TK: **A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition**. *Front Microbiol* 2020, **11**:572034.

305. Song S, Wood TK: **Post-segregational killing and phage inhibition are not mediated by cell death through toxin/antitoxin systems**. *Front Microbiol* 2018, **9**:365078.

306. Blower TR, Chai R, Przybilski R, Chindhy S, Fang X, Kidman SE, Tan H, Luisi BF, Fineran PC, Salmond GPC: **Evolution of Pectobacterium Bacteriophage ΦM1 To Escape Two Bifunctional Type III Toxin-Antitoxin and Abortive Infection Systems through Mutations in a Single Viral Gene**. *Appl Environ Microbiol* 2017, **83**.

307. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC: **The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair**. *Proc Natl Acad Sci U S A* 2009, **106**:894.

308. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R: **BREX is a novel phage resistance system widespread in microbial genomes**. *EMBO J* 2015, **34**:169.

309. Chinenova TA, Mkrtumian NM, Lomovskaia ND: **[Genetic characteristics of a new phage resistance trait in Streptomyces coelicolor A3(2)]**. *Genetika* 1982, **18**:1945–1952.

310. Bedford DJ, Laity C, Buttner MJ: **Two genes involved in the phase-variable ϕC31 resistance mechanism of Streptomyces coelicolor A3(2)**. *J Bacteriol* 1995, **177**:4681–4689.

311. Laity C, Chater KF, Lewis CG, Buttner MJ: **Genetic analysis of the ϕC31 -specific phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2)**. *Mol Microbiol* 1993, **7**:329–336.

312. Sumby P, Smith MCM: **Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2)**. *Mol Microbiol* 2002, **44**:489–500.

313. Hoskisson PA, Sumby P, Smith MCM: **The phage growth limitation system in Streptomyces coelicolor A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity**. *Virology* 2015, **477**:100–109.

314. Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, Matlashov M, Truncaite L, Morgan RD, Ivanov N V., et al.: **BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site**. *Nucleic Acids Res* 2019, **47**:253.

315. Zaworski J, Dagva O, Brandt J, Baum C, Ettwiller L, Fomenkov A, Raleigh EA: **Reassembling a cannon in the DNA defense arsenal: Genetics of StySA, a BREX phage exclusion system in Salmonella lab strains**. *PLoS Genet* 2022, **18**:e1009943.

316. Beck IN, Picton DM, Blower TR: **Crystal structure of the BREX phage defence protein BrxA**. *Curr Res Struct Biol* 2022, **4**:211.

317. Schmitz C, Madej M, Nowakowska Z, Cuppari A, Jacula A, Ksiazek M, Mikruta K, Wisniewski J, Pudelko-Malik N, Saran A, et al.: **Response regulator PorX coordinates oligonucleotide signalling and gene expression to control the secretion of virulence factors**. *Nucleic Acids Res* 2022, **50**:12558–12577.

318. Shen BW, Doyle LA, Werther R, Westburg AA, Bies DP, Walter SI, Luyten YA, Morgan RD, Stoddard BL, Kaiser BK: **Structure, substrate binding and activity of a unique AAA+ protein: the BrxL phage restriction factor**. *Nucleic Acids Res* 2023, **51**:3513.

319. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, et al.: **Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype**. *Genome Res* 2009, **19**:2279–2287.

320. Picton DM: **Functional and Structural Insights into Novel Bacteriophage Defence Islands**. *Biosciences* 2021, **Doctor of Philosophy**.

321. Picton DM: **Functional and Structural Insights into Novel Bacteriophage Defence Islands**. 2021,

322. Cai Y, Usher B, Gutierrez C, Tolcan A, Mansour M, Fineran PC, Ciarán Condon, Neyrolles O, Genevaux P, Blower TR: **A nucleotidyltransferase toxin inhibits growth of Mycobacterium tuberculosis through inactivation of tRNA acceptor stems**. *Sci Adv* 2020, **6**.

323. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO: **Enzymatic assembly of DNA molecules up to several hundred kilobases**. *Nature Methods 2009 6:5* 2009, **6**:343–345.

324. Bailey TL: **Discovering Novel Sequence Motifs with MEME** . *Curr Protoc Bioinformatics* 2003, **00**.

325. Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM, Dryden DTF: **Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA**. *Mol Cell* 2002, **9**:187–194.

326. Winter G, Beilsten-Edmands J, Devenish N, Gerstel M, Gildea RJ, McDonagh D, Pascal E, Waterman DG, Williams BH, Evans G: **DIALS as a toolkit**. *Protein Science* 2022, **31**:232–250.

327. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al.: **Overview of the CCP4 suite and current developments**. *Acta Crystallogr D Biol Crystallogr* 2011, **67**:235.

328. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**:2126–2132.

329. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213.

330. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software**. *J Appl Crystallogr* 2007, **40**:658.

331. Fleming PJ, Fleming KG: **HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties**. *Biophys J* 2018, **114**:856–869.

332. Vassallo C, Doering C, Littlehale ML, Teodoro G, Laub MT: **Mapping the landscape of anti-phage defense mechanisms in the E. coli pangenome**. *bioRxiv* 2022, doi:10.1101/2022.05.12.491691.

333. Woudstra C, Granier SA: **A Glimpse at the Anti-Phage Defenses Landscape in the Foodborne Pathogen Salmonella enterica subsp. enterica serovar Typhimurium**. *Viruses* 2023, **15**.

334. Wu Y, Hurk A van den, Aparicio-Maldonado C, Kushwaha SK, King CM, Ou Y, Todeschini TC, Clokie MRJ, Millard AD, Gençay YE, et al.: **Defence systems provide synergistic anti-phage activity in E. coli**. *bioRxiv* 2022, doi:10.1101/2022.08.21.504612.

335. Makarova KS, Wolf YI, Snir S, Koonin E V.: **Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems**. *J Bacteriol* 2011, **193**:6039.

336. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC: **MacSyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems**. *PLoS One* 2014, **9**:e110726.

337. Srikumar S, Kröger C, Hébrard M, Colgan A, Owen S V., Sivasankaran SK, Cameron ADS, Hokamp K, Hinton JCD: **RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium**. *PLoS Pathog* 2015, **11**:e1005262.

338. Zhang Y, Lin J, Gao Y: **In silico identification of a multi-functional regulatory protein involved in Holliday junction resolution in bacteria**. *BMC Syst Biol* 2012, **6**:1–10.

339. Brown L, Villegas JM, Elean M, Fadda S, Mozzi F, Saavedra L, Hebert EM: **YebC, a putative transcriptional factor involved in the regulation of the proteolytic system of Lactobacillus**. *Scientific Reports 2017 7:1* 2017, **7**:1–11.

340. Wei L, Wu Y, Qiao H, Xu W, Zhang Y, Liu X, Wang Q: **YebC controls virulence by activating T3SS gene expression in the pathogen Edwardsiella piscicida**. *FEMS Microbiol Lett* 2018, **365**:137.

341. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al.: **InterPro in 2022**. *Nucleic Acids Res* 2023, **51**:D418.

342. Holm L, Laiho A, Törönen P, Salgado M: **DALI shines a light on remote homologs: One hundred discoveries**. *Protein Science* 2023, **32**:e4519.

343. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al.: **Highly accurate protein structure prediction with AlphaFold**. *Nature 2021 596:7873* 2021, **596**:583–589.

344. Xiong X, Wu G, Wei Y, Liu L, Zhang Y, Su R, Jiang X, Li M, Gao H, Tian X, et al.: **SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities**. *Nature Microbiology 2020 5:7* 2020, **5**:917–928.

345. Ozaki S, Kawakami H, Nakamura K, Fujikawa N, Kagawa W, Park SY, Yokoyama S, Kurumizaka H, Katayama T: **A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin**. *Journal of Biological Chemistry* 2008, **283**:8351–8362.

346. Iyer LM, Leipe DD, Koonin E V., Aravind L: **Evolutionary history and higher order classification of AAA+ ATPases**. *J Struct Biol* 2004, **146**:11–31.

347. Kulkarni M, Nirwan N, Van Aelst K, Szczelkun MD, Saikrishnan K: **Structural insights into DNA sequence recognition by Type ISP restriction-modification enzymes**. *Nucleic Acids Res* 2016, **44**:4396–4408.

348. Aslanidis C, de Jong PJ: **Ligation-independent cloning of PCR products (LIC-PCR).** *Nucleic Acids Res* 1990, **18**:6069.

349. Kaletta J, Pickl C, Griebler C, Klingl A, Kurmayer R, Deng L: **A rigorous assessment and comparison of enumeration methods for environmental viruses**. *Scientific Reports 2020 10:1* 2020, **10**:1–12.

350. Lynch MD, Gill RT: **Broad host range vectors for stable genomic library construction**. *Biotechnol Bioeng* 2006, **94**:151–158.

351. Aertsen A, Michiels CW: **Mrr instigates the SOS response after high pressure stress in Escherichia coli**. *Mol Microbiol* 2005, **58**:1381–1391.

352. Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, Zheng Y, Wilson GG, Cheng X: **Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix**. *Nucleic Acids Res* 2014, **42**:12092–12101.

353. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT: **A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome**. *Nature Microbiology 2022 7:10* 2022, **7**:1568–1579.

354. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, Kauffman K, Rodino-Janeiro BK, Gavin H, Gomez A, et al.: **Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages**. *Science (1979)* 2021, **374**:488–492.

355. Rocha EPC, Bikard D: **Microbial defenses against mobile genetic elements and viruses: Who defends whom from what?** *PLoS Biol* 2022, **20**:e3001514.

356. Ogura T, Wilkinson AJ: **AAA+ superfamily ATPases: Common structure-diverse function**. *Genes to Cells* 2001, **6**:575–597.

357. Wang L, Jiang S, Deng Z, Dedon PC, Chen S: **DNA phosphorothioate modification—a new multi-functional epigenetic system in bacteria**. *FEMS Microbiol Rev* 2019, **43**:109.

358. Duncan-Lowey B, Tal N, Johnson AG, Rawson S, Mayer ML, Doron S, Millman A, Melamed S, Fedorenko T, Kacen A, et al.: **Cryo-EM structure of the RADAR supramolecular anti-phage defense complex**. *Cell* 2023, **186**:987.

359. Whiteley AT, Eaglesham JB, de Oliveira Mann CC, Morehouse BR, Lowey B, Nieminen EA, Danilchanka O, King DS, Lee ASY, Mekalanos JJ, et al.: **Bacterial cGAS-like enzymes synthesize diverse nucleotide signals**. *Nature* 2019, **567**:194.

360. Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V: **A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems**. *Science* 2017, **357**:605–609.

361. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R: **Cyclic GMP-AMP signalling protects bacteria against viral infection**. *Nature* 2019, **574**:691–695.

362. Kupczok A, Neve H, Huang KD, Hoeppner MP, Heller KJ, Franz CMAP, Dagan T: **Rates of Mutation and Recombination in Siphoviridae Phage Genome Evolution over Three Decades**. *Mol Biol Evol* 2018, **35**:1147–1159.

363. Jain M, Olsen HE, Paten B, Akeson M: **The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community**. *Genome Biology 2016 17:1* 2016, **17**:1–11.

364. Deamer D, Akeson M, Branton D: **Three decades of nanopore sequencing**. *Nature Biotechnology 2016 34:5* 2016, **34**:518–524.

365. Anton BP, Mongodin EF, Agrawal S, Fomenkov A, Byrd DR, Roberts RJ, Raleigh EA: **Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of Escherichia coli K-12**. *PLoS One* 2015, **10**.

366. Liu Y, Rosikiewicz W, Pan Z, Jillette N, Wang P, Taghbalout A, Foox J, Mason C, Carroll M, Cheng A, et al.: **DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation**. *Genome Biol* 2021, **22**:1–33.

367. Stoiber M, Quick J, Egan R, Lee JE, Celniker S, Neely RK, Loman N, Pennacchio LA, Brown J: **De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing**. *bioRxiv* 2017, doi:10.1101/094672.

368. Palmer BR, Marinus MG: **The dam and dcm strains of Escherichia coli--a review**. *Gene* 1994, **143**:1–12.

369. Dobbins AT, George M, Basham DA, Ford ME, Houtz JM, Pedulla ML, Lawrence JG, Hatfull GF, Hendrix RW: **Complete Genomic Sequence of the Virulent Salmonella Bacteriophage SP6**. *J Bacteriol* 2004, **186**:1933.

370. Coppens L, Wicke L, Lavigne R: **SAPPHIRE.CNN: Implementation of dRNA-seq-driven, species-specific promoter prediction using convolutional neural networks**. *Comput Struct Biotechnol J* 2022, **20**:4969–4974.

371. Makovets S, Powell LM, Titheradge AJB, Blakely GW, Murray NE: **Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease?** *Mol Microbiol* 2004, **51**:135–147.

372. Thoms B, Wackernagel W: **Expression of ultraviolet-induced restriction alleviation in Escherichia coli K-12. Detection of a lambda phage fraction with a retarded mode of DNA injection**. *Biochim Biophys Acta* 1983, **739**:42–47.

373. Thoms B, Wackernagel W: **Genetic control of damage-inducible restriction alleviation in Escherichia coli K12: an SOS function not repressed by lexA**. *MGG Molecular & General Genetics* 1984, **197**:297–303.

374. Cooper DL, Lovett ST: **Recombinational branch migration by the RadA/Sms paralog of RecA in Escherichia coli**. *Elife* 2016, **5**.

375. Ni Y, Liu X, Simeneh ZM, Yang M, Li R: **Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing**. *Comput Struct Biotechnol J* 2023, **21**:2352.

376. Luo J, Meng Z, Xu X, Wang L, Zhao K, Zhu X, Qiao Q, Ge Y, Mao L, Cui L: **Systematic benchmarking of nanopore Q20+ kit in SARS-CoV-2 whole genome sequencing**. *Front Microbiol* 2022, **13**:973367.

377. McMahon SA, Roberts GA, Johnson KA, Cooper LP, Liu H, White JH, Carter LG, Sanghvi B, Oke M, Walkinshaw MD, et al.: **Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance**. *Nucleic Acids Res* 2009, **37**:4887.

378. Luyten YA, Hausman DE, Young JC, Doyle LA, Higashi KM, Ubilla-Rodriguez NC, Lambert AR, Arroyo CS, Forsberg KJ, Morgan RD, et al.: **Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system**. *Nucleic Acids Res* 2022, **50**:5171–5190.

379. Irastortza-Olaziregi M, Amster-Choder O: **Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises**. *Front Microbiol* 2020, **11**.

380. Verde V La, Dominici P, Astegno A: **Determination of Hydrodynamic Radius of Proteins by Size Exclusion Chromatography**. *Bio Protoc* 2017, **7**.

381. Stierand K, Rarey M: **Drawing the PDB: Protein-ligand complexes in two dimensions**. *ACS Med Chem Lett* 2010, **1**:540–545.

382. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N: **ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules**. *Nucleic Acids Res* 2016, **44**:W344.

383. Cherstvy AG: **Positively Charged Residues in DNA-Binding Domains of Structural Proteins Follow Sequence-specific Positions of DNA Phosphate Groups**. *Journal of Physical Chemistry B* 2009, **113**:4242–4247.

384. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, et al.: **Improvements to the APBS biomolecular solvation software suite**. *Protein Science* 2018, **27**:112–128.

385. Krissinel E, Henrick K: **Inference of Macromolecular Assemblies from Crystalline State**. *J Mol Biol* 2007, **372**:774–797.

386. Mato JM, Alvarez L, Ortiz P, Pajares MA: **S-adenosylmethionine synthesis: Molecular mechanisms and clinical implications**. *Pharmacol Ther* 1997, **73**:265–280.

387. Finkelstein JD: **Methionine metabolism in mammals**. *J Nutr Biochem* 1990, **1**:228–237.

388. Wang YM, Austin RH, Cox EC: **Single molecule measurements of repressor protein 1D diffusion on DNA**. *Phys Rev Lett* 2006, **97**:048302.

389. Kabata H, Kurosawa O, Arai I, Washizu M, Margarson SA, Glass RE, Shimamoto N: **Visualization of Single Molecules of RNA Polymerase Sliding Along DNA**. *Science (1979)* 1993, **262**:1561–1563.

390. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE: **The Phyre2 web portal for protein modeling, prediction and analysis**. *Nature Protocols 2015 10:6* 2015, **10**:845–858.

391. McLure RJ, Radford SE, Brockwell DJ: **High-throughput directed evolution: a golden era for protein science**. *Trends Chem* 2022, **4**:378–391.

392. Bikard D, Marraffini LA: **Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages**. *Curr Opin Immunol* 2012, **24**:15–20.

393. Sumby P, Smith MCM: **Phase Variation in the Phage Growth Limitation System of Streptomyces coelicolor A3(2)**. *J Bacteriol* 2003, **185**:4558.

394. Mahler M, Costa AR, van Beljouw SPB, Fineran PC, Brouns SJJ: **Approaches for bacteriophage genome engineering**. *Trends Biotechnol* 2023, **41**:669–685.

395. Shitrit D, Hackl T, Laurenceau R, Raho N, Carlson MCG, Sabehi G, Schwartz DA, Chisholm SW, Lindell D: **Genetic engineering of marine cyanophages reveals integration but not lysogeny in T7-like cyanophages**. *The ISME Journal 2021 16:2* 2021, **16**:488–499.

396. Namura M, Hijikata T, Miyanaga K, Tanji Y: **Detection of Escherichia coli with Fluorescent Labeled Phages That Have a Broad Host Range to E. coli in Sewage Water**. *Biotechnol Prog* 2008, **24**:481–486.

397. Ramirez-Chamorro L, Boulanger P, Rossier O: **Strategies for Bacteriophage T5 Mutagenesis: Expanding the Toolbox for Phage Genome Engineering**. *Front Microbiol* 2021, **12**:667332.

398. Sarkis GJ, Jacobs WR, Hatfulll GF: **L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria**. *Mol Microbiol* 1995, **15**:1055–1067.

399. Møller-Olsen C, Ho SFS, Shukla RD, Feher T, Sagona AP: **Engineered K1F bacteriophages kill intracellular Escherichia coli K1 in human epithelial cells**. *Scientific Reports 2018 8:1* 2018, **8**:1–18.

400. Hupfeld M, Trasanidou D, Ramazzini L, Klumpp J, Loessner MJ, Kilcher S: **A functional type II-A CRISPR–Cas system from Listeria enables efficient genome editing of large non-integrating bacteriophage**. *Nucleic Acids Res* 2018, **46**:6920–6933.

401. Anton BP, Raleigh EA: **Complete Genome Sequence of NEB 5-alpha, a Derivative of Escherichia coli K-12 DH5α**. *Genome Announc* 2016, **4**:1245–1261.

402. Langenscheid J, Killmann H, Braun V: **A FhuA mutant of Escherichia coli is infected by phage T1-independent of TonB**. *FEMS Microbiol Lett* 2004, **234**:133–137.

403. Galperin MY: **Structural Classification of Bacterial Response Regulators: Diversity of Output Domains and Domain Combinations**. *J Bacteriol* 2006, **188**:4169.

404. Young G, Hundt N, Cole D, Fineberg A, Andrecka J, Tyler A, Olerinyova A, Ansari A, Marklund EG, Collier MP, et al.: **Quantitative mass imaging of single biological macromolecules**. *Science (1979)* 2018, **360**:423–427.

405. D'Arcy A, Villard F, Marsh M: **An automated microseed matrix-screening method for protein crystallization**. *Acta Crystallogr D Biol Crystallogr* 2007, **63**:550–554.

406. Cadwell RC, Joyce GF: **Randomization of genes by PCR mutagenesis.** *Genome Res* 1992, **2**:28–33.

407. Stemmer WPC: **Rapid evolution of a protein in vitro by DNA shuffling**. *Nature 1994 370:6488* 1994, **370**:389–391.

408. Álvarez B, Mencía M, de Lorenzo V, Fernández LÁ: **In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9**. *Nature Communications 2020 11:1* 2020, **11**:1–14.