The Role of BrxR in Regulating Bacterial Phage-Defence Systems

DUFFNER, SAMUEL,JOHN

The Role of BrxR in Regulating Bacterial Phage-Defence Systems

Submitted for the degree of MSc by Research

Samuel John Duffner

College of St Hild and St Bede

Biological Sciences

Durham University

Table of contents **Page**

Abstract

Bacteria are under constant attack by bacteriophages, their natural predators that outnumber them 10-fold. The resulting selection pressure has given rise to a diverse range of phage-defence systems within bacteria. These systems are often clustered within ‘defence islands’ in the bacterial genome, facilitating coregulation and complementary action. In this thesis, regulation of an *Escherichia fergusonii* defence island containing both BREX and type IV restriction modification systems by BrxR protein is explored. Through LacZ assays, mutagenesis studies, and EOP assays, BrxR is functionally characterised as a ligand-binding transcriptional repressor of phage defence. The roles of the HTH and WYL domains found within BrxR are identified as likely DNA- and ligand-binding regions, respectively, and the groundwork is laid for future study of BrxR, including identification of its cognate phage-associated ligand.

Declaration

The work presented within this thesis was undertaken at Durham University between September and June 2023. All work presented is my own original research unless cited or otherwise indicated within the text. This work has not been submitted for any other qualification.

Statement of copyright

Publications linked to this work

The work presented in this thesis has also been presented in the following publications:

Picton, D. M. *et al.* A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. *Nucleic Acids Res* **50**, 5191–5207 (2022).

Went, S. C. *et al*. Comparative analysis of BREX phage defence islands using the Durham phage collection shows complex and broad diversity in phage response. (manuscript in review).

Acknowledgements

There are a number of people without whom the work presented in this thesis would not have been completed to nearly the same standard – or at all. For this I am sincerely grateful.

Foremost, I would like to thank Prof. Tim Blower. His inclusive and approachable attitude is reflected not only in his eagerness to discuss any topic, technical or otherwise, but also in the attitudes adopted by his colleagues. Within the Blower Lab there is a culture of open communication, camaraderie, and wholesome banter. These things make the Lab and office spaces a joy to work in and, in my opinion, are the signs of a lab run right.

To Dr David Picton, thank you for fostering such a fun environment to work in. Your ability to foster an environment that is educational and productive whilst also lively and entertaining is admirable and something that I hope to emulate in your example.

Thank you also to my co-supervisor Prof. Ehmke Pohl and progress committee meeting members Dr Gary Sharples and Prof. Adam Benham for their guidance and support throughout the length of my project.

Finally, thank you to all the members of Lab 234. I've already mentioned the cheerful and welcoming atmosphere present in the Lab, but it is worth emphasising that this was only possible with the active participation of its members. The daily 10:30 coffee-and-crossword ritual is just one example of the ever-present community spirit that pervaded all aspects of Lab life.

List of figures

List of tables

List of abbreviations

Abbreviation	Definition
5mC	5-methyl cytosine
5hmC	5-hydroxymethyl cytosine
Ap	Ampicillin
Cm	Chloramphenicol
D-glu	D-glucose
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
E. coli	*Escherichia coli*
EMSA	Electrophoretic Mobility Shift Assay
EOP	Efficiency Of Plating
g	Grams
glc-5hmC	Glucosyl-5-hydroxymethyl cytosine
HTH	Helix-Turn-Helix
IPTG	Isopropyl β-D-thiogalactopyranoside
Km	Kanamycin
L	Litre
L-ara	L-arabinose
mM	Millimolar
ONPG	O-nitrophenyl-β-D-galactopyranoside
pfu	Plaque forming units
TAE	Tris-acetate-EDTA
Tc	Tetracycline
WYL	Tryptophan-Tyrosine-Leucine
µL	Microlitre
µM	Micromolar

1. Introduction

Bacteria are under constant attack by bacteriophages — their natural predators that outnumber them 10-to-one[1]. The presence of $\geq 10^{30}$ bacteriophages in nature results in considerable selection pressure exerted upon their targets, giving rise to a wide array of bacterial defence systems used to protect against infection by phages. Study of these systems has led to characterisation of an ever-increasing multitude of such 'phage defence' systems, most notably restriction-modification (RM)[2] and CRISPR-Cas[3] systems, both of which have proven biotechnological applications of immense utility.

Such potential for biotechnological applications, as well as the attractiveness of phage therapy to combat increasingly prominent antibiotic resistance, has encouraged the identification of new phage defence systems, including (amongst many others) BstA[4], CBASS[5], and bacteriophage exclusion (BREX)[6], the latter of which will be the primary focus of this thesis.

The clustering of phage defence systems into 'defence islands' has been observed following comparative analysis of bacterial genomes[7], an observation which offers two-fold insight. Firstly, the grouping of known phage defence systems into islands facilitates the inference of function of genes of unknown function located nearby and has led to the discovery of novel defence systems[7–9]. Additionally, and of more interest within this thesis, the grouping of defence systems into islands raises questions regarding transcriptional regulation. The clustering of defence systems into defence loci implies a potential for coregulation, so as to activate these systems only when necessary as a way to preserve bacterial fitness.

1.1 Bacteriophage exclusion

The BREX system (fig. 1a) provides host bacteria with defence against phages and other mobile genetic elements on the basis of recognition of non-methylated DNA. BREX methyltransferases mark the host genome for immunity via hemi-methylation of specific non-palindromic 6 bp sequences on the N6 adenine nitrogen (N6mA)[6]. By this mechanism, the BREX system is able to discriminate between host and incoming phage DNA and, by an unknown mechanism, prevent its proliferation.

1.1.1 Discovery of BREX

The first observation of a BREX system was made in 2015 by Goldfarb *et al* on the basis of BREX component association with the *pglZ* gene[6]. It had been previously noted that *pglZ* was enriched within defence islands in addition to its essential role in conferring a unique phage-resistance phenotype in *Streptomyces coelicolor* as part of the phage growth limitation (Pgl) system[10]. Due to this enrichment of *pglZ*-domain genes in other bacterial species, they investigated the role of *pglZ*-domain genes in other phage defence systems, finding it to be present in ~10% of bacterial genomes sequenced. In over half of these cases, *pglZ*-domain genes were found to be associated with a six-gene cluster. Two genes of the cluster (*pglX* and *pglZ* itself) were found to have considerable homology to Pgl system components, whereas the other four genes comprised a gene of unknown function as well as genes encoding a putative protease, an ATPase-domain containing protein, and a putative RNA-binding protein (later recharacterised as the dsDNA-binding protein BrxA[11]). Hypothesising that this cluster constituted a novel phage defence system, they denoted it BREX (BacteRiophage EXclusion).

By transferring the BREX system from *Bacillus cereus* to *Bacillus subtilis*, Goldfarb *et al* demonstrated that BREX confers resistance to a wide range of *Bacillus* phages, confirming its role as a phage defence system. Given the difference between this phage resistance phenotype and that of the Pgl system, they inferred that the BREX system operated by a novel mechanism. Finally, Goldfarb *et al* observed that these *pglZ*-containing clusters could be divided into six subtypes on the basis of conserved gene order and composition.

1.1.2 BREX subtypes

For the major clades of the Pgl phylogenetic tree, Goldfarb *et al* examined the genes in the vicinity of *pglZ* to characterise each BREX subtype on the basis of gene order and composition (fig. 1a). They observed consistent sets of 4-8 genes within each clade and so hypothesised that each one represented a unique multi-gene system with conserved order and composition. Given the dominant prevalence of BREX systems over the subsequently identified variants (fig. 1b), they elected to name the phage

defence superfamily after it, with type 1 denoting the core six-gene cassette, type 2 denoting the Pgl system, and types 2-6 denoting the variants thereof (fig. 1a).

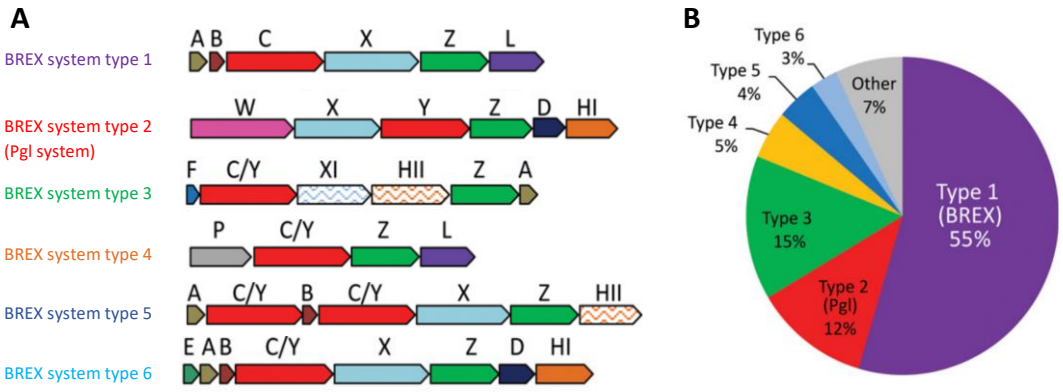

Figure 1- Visual representation of BREX subtypes. (**A**) Differing gene orders and compositions of BREX subtypes; (**B**) Prevalence of BREX subtypes within BREX superfamily. Panels taken/adapted from Goldfarb *et al*, 2015[6].

1.1.3 BREX mechanism of action

Although the exact mechanism by which the BREX system confers phage defence to its host is unknown, since its discovery in 2015 many individual BREX components have been characterised and some early ideas about their collective function have been clarified. Starting with the earliest characterisation of BREX, Goldfarb *et al* observed that the BREX system allows phage adsorption but blocks subsequent phage DNA replication. Furthermore, they showed that phage DNA is not cleaved, as in restriction-modification systems, or degraded in any way by BREX components[6]. These facts together give a credible basis for the hypothesis that the BREX system operates via a novel mechanism.

In 2019, Gordeeva *et al* showed for the first time that epigenetic modification was the basis for BREX-mediated phage defence. Specifically, they observed acquired BREX evasion in λ prophages induced from BREX-carrying cells and determined this to be the result of methylated adenine residues on motifs found within the phage genome which match those found within the host genome[12]. This conclusion is supported by research carried out the same year by Hui *et al*, who demonstrated the role of PglX within the BREX system via mutant studies. They found that m6A methylation of the 5'-ACRCAG-3' motif was eliminated in host cells carrying the Δ*pglX* mutation and further noted that these cells had a marked increase in plasmid acquisition capacity[13].

These items together depict epigenetic modification as the basis for the ability of BREX-carrying cells to differentiate and restrict entry to non-self-DNA and PglX as the BREX component responsible for enacting this modification.

Additionally, research carried out in 2022 by Zaworski *et al* characterised the roles of *brxC*, *pglZ*, and *brxL* within the StySA RM system, a BREX variant. They show that the BrxL protein is not required for epigenetic modification but is essential for restriction (in a manner that scales with the number of unmodified target sites). They also characterise the involvement of proteins PglX and BrxC in licensing BrxL action[14].

Furthermore, work carried out in 2022 by members of the Blower Lab characterised BrxA as a dsDNA binding protein somehow involved in BREX-mediate defence, albeit via an unknown mechanism[11]. The absence of *brxA* in BREX types 2 and 4 despite its conservation among BREX types 1, 3, 5, and 6 suggest that these subtypes may work via a different mechanism, although this has not yet been investigated.

1.2 The pEFER phage defence island and BREX

The Blower Lab recently characterised a novel phage defence island encoded on pEFER, a multidrug resistant plasmid found in *Escherichia fergusonii*[15]. The defence island encodes both a BREX system and the type IV restriction enzyme BrxU which, together, were shown to confer synergistic protection against a diverse array of coliphages.

The pEFER BREX system is a type 1 system comprising the canonical genes *brxA*, *brxB*, *brxC*, *pglX*, *pglZ*, and *brxL*, as well as three additional genes *brxR*, *brxS*, and *brxT* (fig. 2, Table 1). The complementary protection provided by this uncommonly large BREX operon was investigated via mutational analysis, comparing the response of modified version of the pEFER defence island to a suite of 30 coliphages isolated from environmental sources in Durham city centre. Efficiency of Plating (EOP) data (which quantifies the ability of a phage stock to form plaques on a bacterial lawn) showed differing patterns of phage defence among bacterial strains with the mutation *ΔpglX*, inactivating the BREX system, versus strains with the mutation *ΔbrxU*, inactivating the RM system. Notably, of the 30 tested, there were phages capable of evading either BREX-mediated defence or BrxU-mediated defence, whilst some

phages were resistant to both[15]. These data showed the protection conferred to bacterial hosts by the complementary defence island pEFER. PacBio sequencing of modified DNA isolated from bacterial hosts allowed further insight into the differential mechanisms by which BREX and BrxU mediate phage defence, revealing that PglX (of the BREX operon) was essential for N6mA modification of GCTAAT motifs, whereas the *ΔbrxU* mutation had minimal impact on the number of methylated motifs[15]. This work was followed by subsequent analysis of BrxU, revealing biochemical insight into the modification dependent DNA cleavage performed by GmrSD family RM systems.

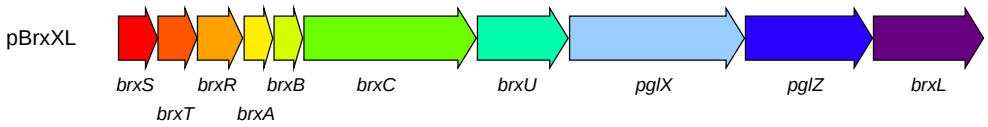

Figure 2- Linear representation of the phage defence island encoded on pEFER. Image taken from Picton et al, 2021[15]

Table 1-BREX operon components and their functions.

BREX Gene	Encodes	References
brxS	Putative IS3 family transposase	Goldfarb et al, 2015[6] Picton et al, 2021[15] Beck et al, 2022[11], Picton et al, 2022[16]
brxT	Hypothetical protein of unknown function	
brxR	WYL-domain containing transcriptional regulator	
brxA	dsDNA recognition protein	
brxB	Unknown	
brxC	ATP-binding motif containing protein	
pglX	Putative methylase	
pglZ	Predicted alkaline phosphatase	
brxL	Lon-like protease domain protein	

The *brxU* gene, present between *brxC* and *pglX* on the pEFER defence island, encodes BrxU, a modification-dependent restriction enzyme that recognises and cleaves DNA containing 5-methyl, 5-hydroxymethyl, or glucosyl-5-hydroxymethyl cytosine (5mC, 5hmC, or glc-5hmC, respectively)[15]. Thus, BREX and BrxU work in a synergistic manner to act against both phages with and without modified cytosines (fig. 3) whilst protecting the host genome from degradation via recognition of the N6mA modification. These phage defence systems were proposed to be coregulated together by BrxR, a protein containing WYL domains (so-named for three conserved amino acids).

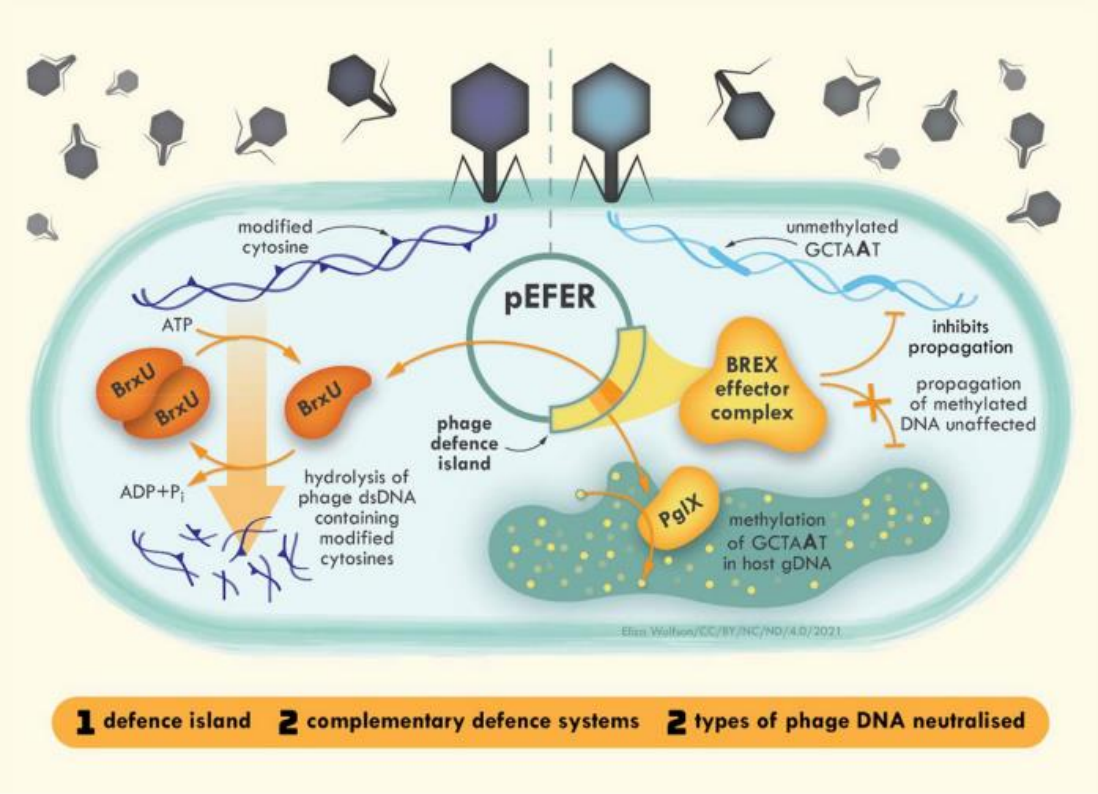

Figure 3- Complementary function of the BREX and BrxU systems found on the pEFER defence island, providing 'belt-and-braces' phage defence. Image taken from Picton et al, 2021[15] with license-holder's permission.

1.3 WYL-domain containing proteins

Of particular interest to this thesis is the BrxR protein encoded by *brxR*, found upstream of the canonical pEFER BREX genes (fig. 2). BrxR is predicted to contain a WYL domain as well as a helix-turn-helix (HTH) domain. HTH domains are structural motifs well known to be involved in DNA binding and transcriptional regulation[17]. This suggests that BrxR might be a regulator of the pEFER defence island. WYL domains (so named for three conserved amino acids within the domain) are found in prokaryotic

proteins and have been postulated to play a role as ligand-binding and sensor domains, with a potential role as part of transcriptional regulators of phage defence systems[18]. The ability of WYL domains to bind a wide range of biological molecules is well characterised[18] and plays a central role in their ability to differentiate conditions in which gene expression is advantageous/disadvantageous, i.e. these domains are the structural source of the 'switch' allowing regulatory proteins to activate/repress transcription.

Numerous recent studies have characterised the roles played by WYL-domain containing proteins in diverse prokaryotic processes, including DriD as a transcriptional up-regulator of DNA damage response mediators in *Caulobacter crescentus*[19], Sll7009 as a repressor of CRISPR phage defence in *Synechocystis*[20], and coupling of ATPase activity and DNA unwinding by PIF1 in *Thermotaga elfii*[21]. WYL-domain containing proteins were also recently observed to be involved in transcriptional regulation at a number of hotspots within a phage defence island also responsible for antibiotic resistance within *Vibrio cholerae*[22].

The presence of both WYL and Helix-Turn-Helix (HTH) domains together form the basis of the hypothesis that BrxR protein is a ligand-binding transcriptional regulator of the pEFER BREX operon. There is potential for a phage-derived signal to be recognised by BrxR, leading to activation of the defence island.

1.4 Aims

This thesis, together with its associated experimental work, aims to characterise the functional role of BrxR as a ligand-binding transcriptional regulator of the pEFER defence island. Techniques used in pursuit of this aim will include (**1**) variable mutagenesis of the *brxR* gene and functional investigation of mutant BrxR protein produced thereby via (**2**) its interaction with bacteriophages, explored by EOP assays, and (**3**) the BREX operon itself, explored by LacZ assays.

This thesis will also discuss the precedent set by BrxR as a model system for a larger family of ligand-binding transcriptional regulators of phage defence.

2. Materials and Methods

2.1 Bacterial strains and culture conditions

Total genomic DNA (gDNA) was obtained for *E. fergusonii* ATCC 35469 from ATCC. *Escherichia coli* strain DH5α (ThermoFisher Scientific) was grown at 37 °C, either on agar plates or shaking at 220 rpm for liquid cultures. Luria broth (LB) was used as the standard growth media for liquid cultures and was supplemented with 0.35% (w/v) or 1.5% (w/v) agar for semi-solid and solid agar plates, respectively. Growth was monitored using a spectrophotometer (WPA Biowave C08000) measuring optical density at 600 nm (OD_{600}). When necessary, growth media was supplemented with ampicillin (Ap, 50 µg/mL), tetracycline (Tc, 10 µg/mL), isopropyl-D-thiogalactopyranoside (IPTG, 1 mM), L-arabinose (L-ara, 0.1% or 0.01% w/v), or D-glucose (D-glu, 0.2% w/v).

2.2 Use of environmental coliphages

Escherichia coli phages Pau, Trib and Baz were isolated previously from freshwater sources in Durham, UK[15]. To make lysates, 10 µL of phage dilution was mixed with 200 µL of *E. coli* DH5α overnight culture and mixed with 4 mL of sterile semi-solid 'top' LB agar (0.35% agar) in a sterile plastic bijou. Samples were poured onto solid LB agar plates (1.5% agar) and incubated overnight at 37 °C. Plates showing a confluent lawn of plaques were chosen for lysate preparations and the semi-solid agar layer was scraped off into 3 mL of phage buffer. 500 µL of chloroform was added and samples were vigorously vortexed and incubated for 30 min at 4 °C. Samples were centrifuged at 4000 × g for 20 min at 4 °C and the supernatant was carefully transferred to a sterile glass bijou. 500 µL of chloroform was added and lysates were kept at 4 °C for long term storage.

2.3 Recombinant DNA techniques

DNA-handling techniques (e.g. plasmid miniprep, transformation) were performed as per standard methodology[23]. Oligonucleotide primers were sourced from Integrated DNA Technologies.

2.4 DNA extraction and visualisation

2.4.1 Plasmid extraction

Plasmid DNA was extracted with use of the NEB Monarch® Plasmid MiniPrep kit as per manufacturer's instructions, eluted in 30 µL dH_2O and stored at -20 °C.

2.4.2 Agarose gel electrophoresis

Agarose gels were prepared as described in Table 4. 6x loading dye was added to samples to a final 1x concentration and mixed prior to gel loading. Samples were loaded alongside the NEB Quick-Load Purple 1 kb DNA ladder (NEB #N0552) and run at 120 V until bands were sufficiently separated. Bands containing desired DNA were excised and DNA was extracted with use of the NEB Monarch® Gel Extraction kit as per manufacturer's instructions, eluted in 30 µL dH_2O and stored at -20 °C.

Table 2- Media used in this study

Medium	Components (/L)
Luria-broth	10 g Casein digest peptone 5 g Yeast extract 5 g NaCl
Luria-broth agar	10 g Casein digest peptone 5 g Yeast extract 5 g NaCl 15 g Agar
2 x YT broth	16 g Casein digest peptone 10 g Yeast extract 5 g NaCl

Table 3- Antibiotics and supplements used in this study

Chemical	Stock solution	Working concentration
Antibiotic		
Ampicillin (Ap)	1000 x stock, 50 mg/mL in dH_2O, stored at -20 °C	50 µg/mL
Tetracycline (Tc)	1000 x stock, 10 mg/mL in dH_2O, stored at -20 °C	10 µg/mL
Kanamycin (Km)	1000 x stock, 50 mg/mL in dH_2O, stored at -20 °C	50 µg/mL
Chloramphenicol (Cm)	1000 x stock, 34 mg/mL in dH_2O, stored at -20 °C	34 µg/mL
Supplement		
Isopropyl β-D-thiogalactopyranoside (IPTG)	1000 x stock, 1 M in dH_2O, stored at -20 °C	1 mM
D-glucose	100 x stock, 20% in dH_2O	0.2% w/v
L-arabinose	100 x stock, 10% in dH_2O	0.1% w/v

Table 4- Solutions used in this study

Solution	Components
50x TAE Buffer (per L)	242 g Tris base [tris(hydroxymethyl)aminomethane] 57.1 mL Glacial acetic acid (17.4 M) 18.61 g EDTA, disodium salt pH 8.0
6x DNA loading dye	NEB #N0552
Agarose gel mix (per L)	10 g agarose in 1 L TAE 50 µL Ethidium bromide
β-galactosidase assay master mix	60 mM Na_2HPO_4 40 mM NaH_2PO_4 10 mM KCl 1 mM $MgSO_4$ 36 mM β-mercaptoethanol 0.1 mg/mL T7 lysozyme 1.1 mg/mL Ortho-nitrophenyl-β-galactoside (ONPG) 6.7% PopCulture Reagent (Merck Millipore)

Table 5- Bacterial strains used in this study

Strain	Genotype	Source
Escherichia coli DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1	Invitrogen

2.5 Preparation and use of heat shock competent cells

Heat-shock competent cells were prepared as per standard methodology or using an adaptation of the Inoue method for cases in which especially high competency was desired[24]. Appropriate quantities of DNA were added to 50 µL aliquots of competent cells, mixed and incubated on ice for 1 hour. Mixtures were heat shocked at 42 °C for 2 minutes and recovered in 1 mL Luria broth at 37 °C for 1 hour before plating.

2.6 Polymerase chain reaction (PCR)

Materials and methodology for polymerase chain reaction outlined below. Individual primer melting temperature (Tm) was estimated using the online NEB Tm Calculator (https://tmcalculator.neb.com/#!/main).

Table 6- Polymerase chain reaction materials and methods

Component	Volume (µL)
dH_2O	37.5
10 x Q5 buffer	5
10 µM Forward primer	2.5
10 µM Reverse primer	2.5
10 mM dNTP mixture	1
DNA template	1
Q5 polymerase	0.5

Stage (2-4 repeated)	Temperature (°C)	Duration (s)
1 – First denaturation	95	30
2 – Denaturation	95	10
3 – Primer annealing	Specific to primers	30
4 - Extension	72	30
5 – Final extension	72	120
6 - Holding	10	N/A

2.7 Mutagenesis and cloning

brxR mutant R17A was generated by QuikChange™ methodology described below using primers TRB1987 and TRB1988 listed in Table 7. These primers result in substitution of the arginine at position 17 within BrxR protein with alanine. The remaining mutants R53A, R151A, and R178A were acquired by commercial commission from GenScript Biotech.

2.8 Sequencing and sequence analysis

Plasmids extracted as per the above-described method were sequenced in-house (DBS Genomics, Durham University Biosciences Department) with the ABI 3730 DNA sequencer using primers listed in Table 7. Sequence confirmation and analysis were carried out using SnapGene viewer (https://www.snapgene.com/snapgene-viewer) and BLASTN software

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch).

Table 7- Primers used in this study

Primer	Sequence	Notes
TRB1197	TTATTTGCACGGCGTCAC	FWD pPBAD30
TRB1198	TTTATCAGACCGCTTCTGC	RV pBAD30
TRB1987	AGCCAAGCACAGAGAGAAGCACTCGCTCATATTGATTTC	FWD QC BrxR R17A
TRB1988	GAAATCAATATGAGCGAGTGCTTCTCTCTGTGCTTGGCT	RV QC BrxR R17A
TRB1989	GCAACACAAGATTTTGCGGCGTACAAAGCGTTGGCTCCT	FWD QC BrxR R53A
TRB1990	AGGAGCCAACGCTTTGTACGCCGCAAAATCTTGTGTTGC	RV QC BrxR R53A
TRB1991	ATTAATATTGAGTACACCGCATTGTCGAGTGGTCATGGG	FWD QC BrxR S143A
TRB1992	CCCATGACCACTCGACAATGCGGTGTACTCAATATTAAT	RV QC BrxR S143A
TRB1993	TCGAGTGGTCATGGGAGCGCGCAAATAGTTCCTCATACC	FWD QC BrxR R151A
TRB1994	GGTATGAGGAACTATTTGCGCGCTCCCATGACCACTCGA	RV QC BrxR R151A
TRB1995	GTCCGAGCATTCGATAGAGCGCATAGAGAATTTAGGGAT	FWD QC BrxR K173A
TRB1996	ATCCCTAAATTCTCTATGCGCTCTATCGAATGCTCGGAC	RV QC BrxR K173A
TRB1997	CGAGCATTCGATAGAAAGGCTAGAGAATTTAGGGATTTT	FWD QC BrxR H174A
TRB1998	AAAATCCCTAAATTCTCTAGCCTTTCTATCGAATGCTCG	RV QC BrxR H174A
TRB1999	AGAAAGCATAGAGAATTTGCGGATTTTGTTTTAACCAGA	FWD QC BrxR R178A
TRB2000	TCTGGTTAAAACAAAATCCGCAAATTCTCTATGCTTTCT	RV QC BrxR R178A
TRB2001	AGGGATTTTGTTTTAACCGCAATAAGTGAAGTTGAGTTG	FWD QC BrxR R184A
TRB2002	CAACTCAACTTCACTTATTGCGGTTAAAACAAAATCCCT	RV QC BrxR R184A

Table 8- Plasmids used in this study

Plasmid	Backbone	Resistance	Inserted region/gene	Notes	Source
	pRSF	Km	None	Dual expression vector	Blower Lab
	pBAD30	Ap	None	Empty arabinose-inducible expression vector	[25]
	pRW50	Tc	None	Promoterless *lacZ* reporter vector	[26]
	pCOLA	Ap	None	Transformation efficiency marker	N/A (not used)
pTRB451	pBAD30	Ap	$BrxR_{Efer}$ N-term 6His	Arabinose-inducible *brxR* expression vector	Blower Lab
pTRB452	pRW50	Tc	R4	Putative promoter P_{BrxR} fragment	This study
pTRB454	pRW50	Tc	R1	Putative promoter P_{BrxR} fragment	This study
pTRB455	pRW50	Tc	R3	Putative promoter P_{BrxR} fragment	This study
pTRB464	pRW50	Tc	R2	Putative promoter P_{BrxR} fragment	This study
pTRB465	pRW50	Tc	R5	Putative promoter P_{BrxR} fragment	This study
pTRB466	pRW50	Tc	R6	Putative promoter P_{BrxR} fragment	This study
pTRB658	pRW50	Tc	R7	Putative promoter P_{BrxR} fragment	This study
pTRB659	pRW50	Tc	R8	Putative promoter P_{BrxR} fragment	This study
pTRB660	pRW50	Tc	R9	Putative promoter P_{BrxR} fragment	This study
pTRB661	pRW50	Tc	R10	Putative promoter P_{BrxR} fragment	This study
pTRB662	pRW50	Tc	R11	Putative promoter P_{BrxR} fragment	This study
pTRB663	pRW50	Tc	R12	Putative promoter P_{BrxR} fragment	This study
pTRB668	pRW50	Tc	R7 IR1-IR1	Putative promoter P_{BrxR} fragment with duplicated inverted repeat 1	This study

Table 8 continued- Plasmids used in this study

pTRB682	pRW50	Tc	R7 IR1-IR2c	Putative promoter P_{BrxR} fragment with polyC tract-replaced inverted repeat 2	This study
pTRB683	pRW50	Tc	R7 IR1c-IR2c	Putative promoter P_{BrxR} fragment with both polyC tract-replaced inverted repeats	This study
pTRB684	pRW50	Tc	R7 IR2-IR2	Putative promoter P_{BrxR} fragment with duplicated inverted repeat 2	This study
pTRB685	pRW50	Tc	R7 IR1c-IR2	Putative promoter P_{BrxR} fragment with both polyC tract-replaced inverted repeat 1	This study

2.9 Efficiency of Plating (EOP) assays

Escherichia coli DH5α were transformed with pBAD30- *his6-brxR* and transformants were used to inoculate overnight cultures. Serial dilutions of phages Pau, Trib and Baz[15] were produced ranging from 10^{-3} to 10^{-10}. 200 µL of overnight culture and 10 µL of phage dilution were added to 3 mL top LB agar and plated on solid LB agar supplemented with 0.2% D-glu or 0.1% L-ara, to repress or induce *brxR* expression from pBAD30 constructs, respectively. Plates were incubated overnight before plaque forming units (pfu) were counted on each plate. EOP values were calculated by dividing the pfu of the L-ara-containing plates by the pfu of the D-glu-containing plates. Data are the mean and standard deviation of three independent replicates.

2.10 β-galactosidase (LacZ) assays

Putative promoter regions (R1-12, or mutants thereof) were ligated into the promoterless *lacZ* fusion plasmid, pRW50[26]. *Escherichia coli* DH5α was then co-transformed with one of the *lacZ* reporter constructs (or pRW50 as a vector control) and either pBAD30, pBAD30-*his6-brxR* or pBAD30-*his6-brxR*-R17A. Transformants were used to inoculate overnight cultures, supplemented with 0.2% D-glu or 0.01% L-ara, to repress or induce *brxR* expression from pBAD30 constructs, respectively. These were then used to seed 80 µL microplate cultures at an OD_{600} of either 0.05 (for cultures containing D-glu) or 0.1 (for cultures containing L-ara). These cultures were then

grown to mid-log phase in a SPECTROstar Nano (BMG Labtech) plate reader at 37 °C with shaking at 500 rpm. Cultures were then supplemented with 120 µL master mix (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 36 mM β-mercaptoethanol, 0.1 mg/ml T7 lysozyme, 1.1 mg/mL ONPG, and 6.7% PopCulture Reagent (Merck Millipore)). Initial OD_{600} readings were taken, and OD_{420} and OD_{550} readings were then taken every minute for 30 min, at 37 °C with shaking at 500 rpm. Miller Units (mU) were generated as described[27]. The plotted data are the normalised mean and standard deviation of three independent replicates.

2.11 Transformation efficiency assays

Three batches of competent cells (empty pBAD30 expression vector, pBAD30-*brxU*, and pBAD30-*brxU*-R102A – *brxU* loss-of-function mutant) were produced as previously described. Growth media supplemented with 0.2% w/v D-glu during this process to prevent unintended expression from the pBAD30 vector. 50 µL aliquots of each competent cell type were mixed with 60 ng DNA (PCR-prepped pRSF, PCR-prepped 5hmC-pRSF, and supercoiled pRSF individually) and plated on LB agar plates supplemented with kanamycin and ampicillin (Ap, 50 µg/mL) and kanamycin (Km, 50 µg/mL). Plates were incubated at 37 °C overnight and colonies counted the following morning. Data are the mean and standard deviation of three independent replicates.

3. Results

3.1 Regulation of the pEFER phage defence island by BrxR

The presence of a helix-turn-helix domain within the protein encoded by the third open reading frame of the phage defence island found on pEFER was predicted computationally using PHYRE 2.0[28]. It was hypothesised that the protein expressed from this open reading frame, subsequently named BrxR, bound DNA to act as a transcriptional regulator. Further analysis of the defence island suggested a promoter upstream of *brxR* (P_{brxR}), presenting a potential binding site for the BrxR protein (fig. 4a). Transcriptional control of other BREX systems is mediated by promoters upstream of *brxA* and *pglZ*[6] here denoted P_{brxA} and P_{pglZ}. Another promoter, P_{brxS}, was hypothesised to exist upstream of *brxS* and *brxT*, so as to facilitate independent expression of these other BREX system components (fig. 4a).

As a preliminary investigation into the role of BrxR in phage defence, gene *brxR* was cloned into pBAD30[25] in order to facilitate L-arabinose-inducible expression of His_6-BrxR, yielding pBAD30-*his*$_6$-*brxR*. This would allow testing against selected phages, to check whether BrxR alone had any impact on phage replication. *E. coli* DH5α was transformed with pBAD30-*his*$_6$-*brxR* and used to inoculate overnight cultures. Serial dilutions of phages Pau, Trib, and Baz[15] ranging from 10^{-3} to 10^{-10} were produced, from each of which a 10 µL aliquot was taken along with 200 µL of overnight culture and added to 3 mL of top LB agar. This mixture was plated immediately on solid LB agar containing 0.2% D-glu or 0.1% L-ara to repress or induce *brxR* expression, respectively. After incubation overnight, plaque-forming units (pfu) were counted and used to calculate EOP values. BrxR alone had no significant impact on the ability of phages to form plaques, as evidenced by the EOP values being roughly equal to one in each case (Table 9). This is the expected result, if indeed BrxR is a transcriptional regulator rather than an effector of phage defence.

Table 9- EOP values for two BREX-sensitive phages (Pau, Trib) and one BREX-resistant phage (Baz) against DH5α pBAD30-*his$_6$-brxR* with induced plasmid as test strain and uninduced as control. Values shown are biological triplicate means and standard deviations.

Φ	BrxR EOP ± SD
Pau	0.99 ± 0.05
Trib	1.12 ± 0.51
Baz	0.80 ± 0.31

To investigate the ability of BrxR to interact with the putative promoters in the capacity of a transcriptional regulator, 12 regions of the defence island, designed to variably overlap the putative promoters and denoted 'regions 1-12' (R1-12) (fig. 4b), were chosen and cloned into pRW50[26], a promoter-less *lacZ*-encoding construct. *E. coli* transformed with pBAD30-*his$_6$-brxR*, as previously described, were co-transformed with either pRW50 vector-only control, or pRW50-R1-12 reporter plasmids. The resulting dual-plasmid strains were cultured in liquid media containing either D-glu or L-ara and both promoter activity and BrxR interactivity of regions R1-12 were determined by measuring β-galactosidase activity (fig. 4c).

No expression was observed from the putative P$_{brxS}$ (R1-5) or P$_{pglZ}$ (R11-12) promoters. Strong expression was observed from P$_{brxR}$ (R6-8) and weaker expression was observed from P$_{brxA}$ (R6, R9-10). In all cases of measurable expression from a putative promoter, the L-ara-induced expression of *his$_6$-brxR* reduced β-galactosidase activity (fig. 4c). The ability of empty pBAD30 vector to repress expression from R7 was compared to that of pBAD30-*his$_6$-brxR*, to confirm that expression of His$_6$-BrxR was the source of transcriptional repression (fig. 4d). Finally, plasmids pRW50-P$_{abiEi}$ and pRW50-P$_{rv2827c}$ that were generated in a previous Blower Lab study on transcriptional regulators unrelated to BrxR[27], were selected as likely negative controls for BrxR control. They were used in this project to test the ability of His$_6$-BrxR to repress expression in a non-sequence-specific manner (fig. 4e). His$_6$-BrxR was only able to repress expression from P$_{brxR}$, suggesting that BrxR activity is sequence dependent. These data collectively indicate that transcriptional expression of the pEFER phage defence island is negatively regulated by His$_6$-BrxR in a sequence-specific manner.

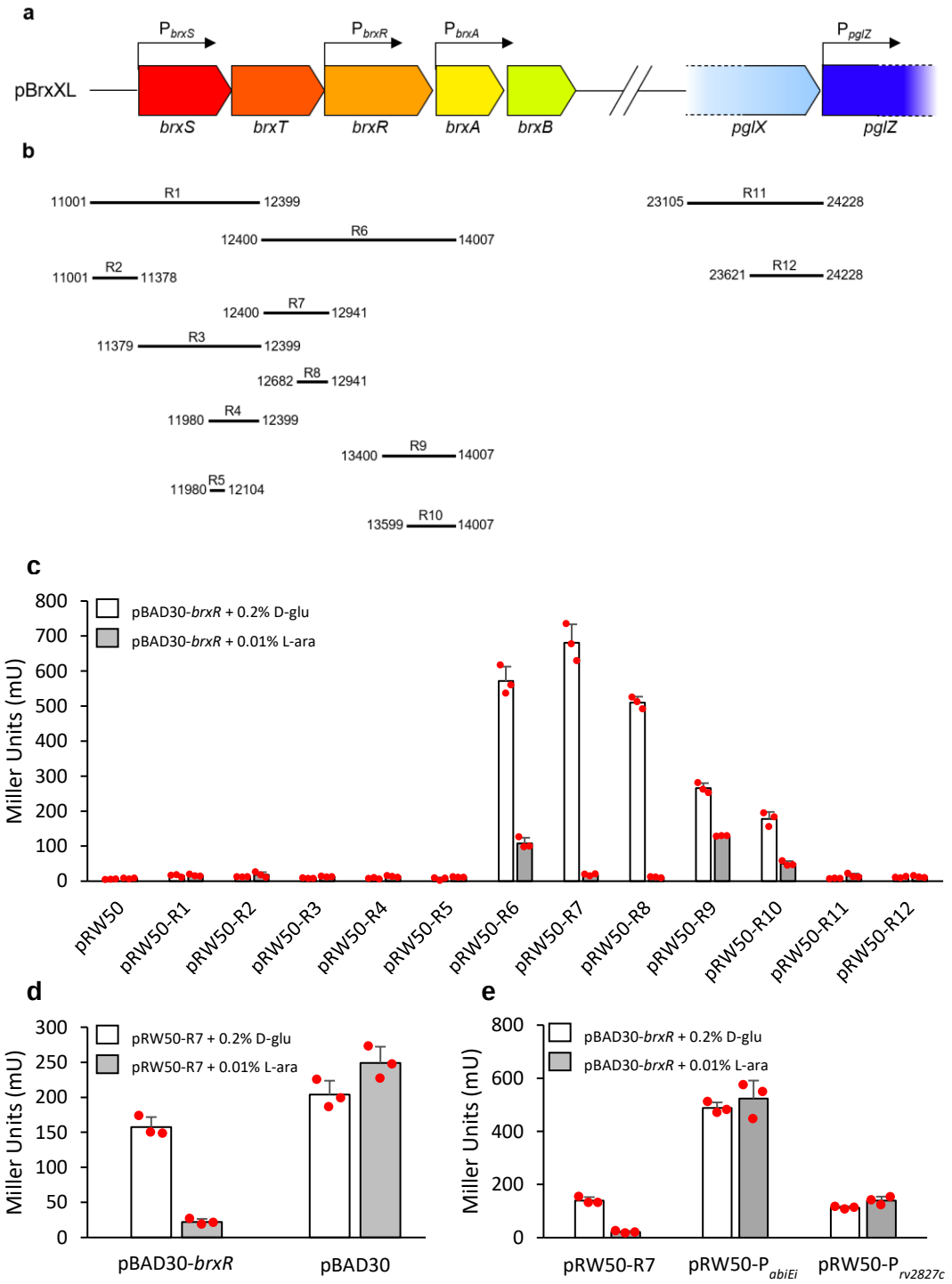

Figure 4- The pEFER phage defence island is regulated by BrxR at the transcriptional level. (**A**) Linear representation of the phage defence island of pEFER. (**B**) Transcriptional organization of the pEFER phage defence island, showing putative promoters P_{brxS}, P_{brxR}, P_{brxA}, and P_{pglZ}, with an accurate alignment of experimental test regions R1-12 that were cloned into the promoterless *lacZ*-reporter plasmid, pRW50. (**C**) LacZ-reporter assays using constructs pRW50-R1-12 with and without the induction of His_6-BrxR from pBAD30-*his*$_6$-*brxR*, showing activity from P_{brxR} and P_{brxA}, and repression by His_6-BrxR. (**D**) LacZ-reporter assays using pRW50-R7 with and without induction of pBAD30-*his*$_6$-*brxR* or a pBAD30 vector control. (**E**) LacZ-reporter assays using active pRW50 promoter constructs with and without induction of His_6-BrxR from pBAD30-*his*$_6$-*brxR*. Data are shown in biological triplicate, and error bars represent standard deviation of the mean. Images generated in this study, also presented in Picton et al, 2022[16].

Having confirmed BrxR activity in a model system, members of the Blower Lab then attempted to make a deletion of *brxR* in the native defence island locus. Unfortunately, attempts to generate a *brxR* knockout mutants both in-house and by commission (i.e. from GenScript Biotech) were unsuccessful due to toxicity from overexpression of the BREX locus, limiting our ability to further test the role of BrxR in regulating phage defence by the pEFER defence island.

3.2 BrxR binds two inverted repeat sites within the pEFER phage defence island

As further *in vivo* work on the involvement of BrxR in phage defence was not immediately accessible, subsequent work focussed on characterisation of the BrxR-DNA interaction. Previously characterised HTH transcriptional regulators have been shown to interact with inverted DNA repeats at their binding sites[27,29]. Closer examination of regions R7 and R9 led to the discovery of an imperfect 11 bp inverted repeat between P_{brxR} and *brxR* within R7 (positioned 12,820-12,846 bp) (fig. 4b). Electrophoretic Mobility Shift Assay (EMSA) data generated by other members of the Blower Lab (fig. 5) indicated that BrxR likely forms a stable dimer in solution and binds the inverted repeat site in this state[16].

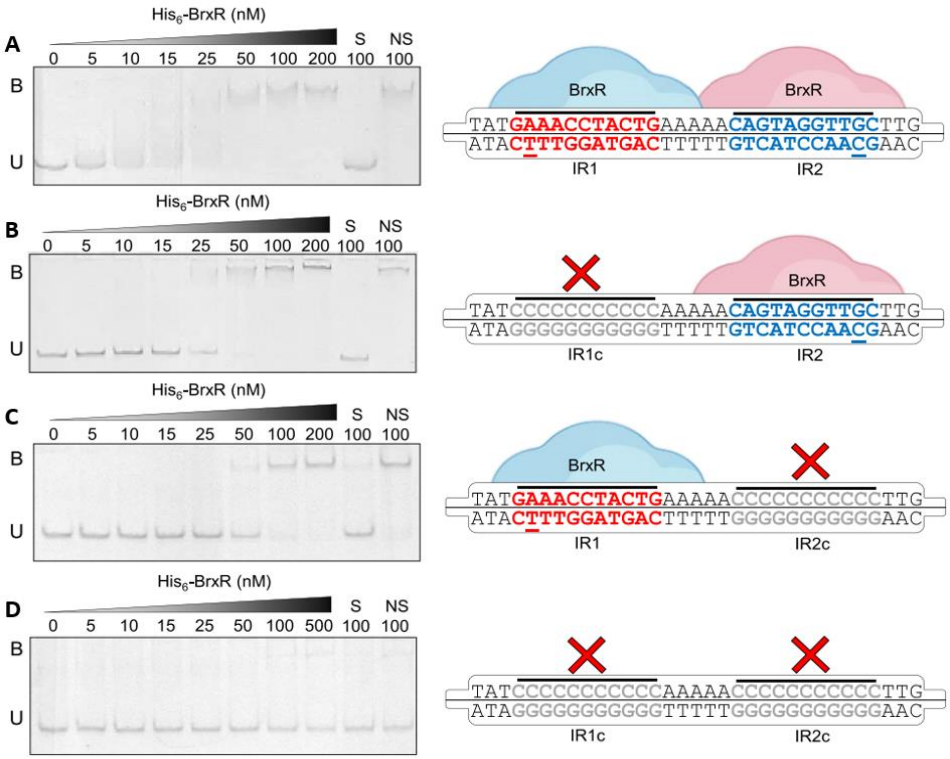

Figure 5- His_6-BrxR binds DNA inverted repeats *in vitro*. EMSAs of titrated His_6-BrxR protein with dsDNA probes spanning the R-BOX1 region of pEFER (nucleotides 12,801-12,870). Target probes were produced containing fluorescein and either the WT promoter inverted repeat region IR1-IR2 (**A**) or modified regions wherein IR1 (**B**), IR2 (**C**), or both IR1 and IR2 (**D**) were substituted for polycytosine residues. Binding events (B – bound, U – unbound) are shown with increasing protein concentration and compared to control lanes containing unlabelled specific (S) and non-specific (NS) DNA. Data and figures taken from Picton et al, 2022[16].

To investigate the roles of the IR1 and IR2 regions in BrxR binding, the following suite of constructs was created by variably replacing IR1 and IR2 with a polyC tract and cloning the result into the pRW50 vector: pRW50-R7-IR1c-IR2, pRW50-R7-IR1-IR2c, pRW50-R7-IR1c-IR2c, pRW50-R7-IR1-IR1, and pRW50-R7-IR2-IR2. This construct suite, in combination with pBAD30-*his*$_6$-*brxR*, allowed us to compare the affinity of BrxR for IR1 and IR2 individually to that for pRW50-R7, which contains the wild-type IR1-IR2 sequence. IR1-IR1 and IR2-IR2 double constructs were included to investigate the relative importance of the IR1 and IR2 repeat regions for BrxR binding. All changes made within each member of the construct suite, other than the IR1-IR1 duplication, resulted in measurably decreased baseline promoter activity, whilst none of the changes notably diminished the ability of BrxR protein to repress promoter activity upon binding (fig. 6). This demonstrated that BrxR repression was possible with only one of the inverted repeats present, suggesting that BrxR might bind as a stable dimer. This was corroborated through size exclusion analysis of BrxR in solution[16], and by a single binding event in EMSAs (fig. 5). Owing to the EMSA data it was expected that an IR1c-IR2c mutant could not be repressed by BrxR due to a lack of binding. As a result of the mutations to both the IR1 and IR2 regions, however, levels of transcription from the promoter region were sufficiently reduced that it was not easily possible to detect repression by BrxR. In other words, the IR1c-IR2c may have prevented BrxR binding, but this was unobservable due to a coincidental decrease in baseline transcription levels.

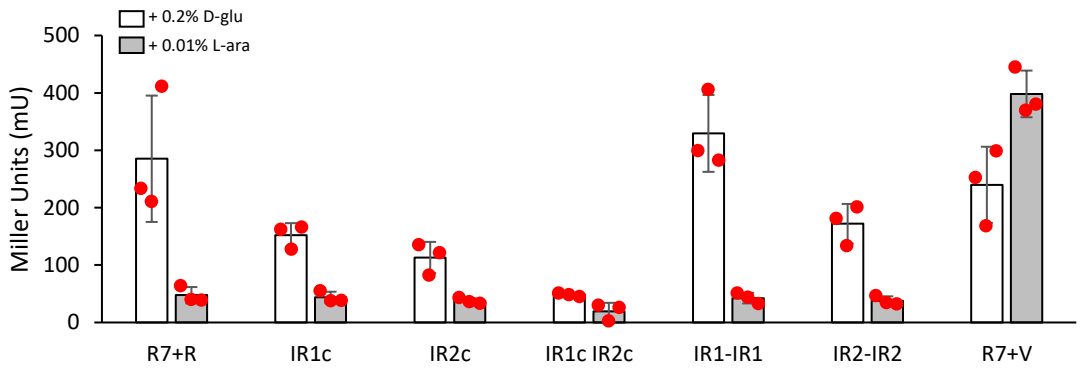

Figure 6- LacZ-reporter assays using constructs pRW50-R7 and variants of the R-BOX inverted repeats ('c' denoting replacement with a poly-C tract). pRW50-R7 plasmid constructs (denoted in shorthand by x-axis labels referring to the IR configurations described previously) were tested with and without induction of pBAD30-*his*$_6$-*brxR* or a pBAD30 vector control (R7 + V).

Following this demonstration of the role of IR1 and IR2 in BrxR-DNA binding, regions R9/10 corresponding to the P_{brxA} promoter (fig. 4b) were searched for potential BrxR binding sites on the basis of sequence similarity to IR1 and IR2. A near-match was found at the position 13,612-13,638 bp. This second set of IR regions was named R-BOX2, whereas the original set of IR regions was retroactively named R-BOX1. While alignment of the IR2 regions of R-BOX1 and R-BOX2 shows only a single base-pair difference, the IR1 regions align poorly and there is also a single base-pair difference in the 5 bp spacer region separating them (fig. 7). The conservation of IR2 between R-BOX1 and R-BOX2 supports the hypothesis that IR2 also facilitates BrxR-mediated repression of expression from the P_{brxA} promoter, though this could be investigated more thoroughly by a repeat of the above LacZ-reporter assay using the IR1 and IR2 regions from R-BOX2. This was not performed due to time constraints.

Figure 7- Alignment of R-BOX1 and R-BOX2. Image taken from Picton et al, 2022[16].

3.3 Investigation of BrxR WYL- and HTH-domains using site-directed mutants

Following the solution of the X-ray crystallographic structure of BrxR by members of the Blower Lab[16], a virtual model of the protein with a DNA structure superimposed (fig. 8a) allowed prediction of how it may interact via HTH domains. Pertinent residues making up these domains were selected for site-directed mutagenesis on the basis that mutation may incur a loss of the ability of BrxR to bind DNA and, thus, a loss of ability to repress transcription from the P_{BrxR} promoter.

HTH-domain mutant constructs pBAD30-*his*$_6$-*brxR*-R17A and pBAD30-*his*$_6$-*brxR*-R53A, in addition to WYL-domain mutant constructs pBAD30-*his*$_6$-*brxR*-R151A and pBAD30-*his*$_6$-*brxR*-R178A (fig. 8b) were generated to test their abilities to repress

transcription from the R7 promoter region using the same LacZ assay protocol as used previously. Whilst expression of the pBAD30-*his*$_6$-*brxR*-R17A construct did not result in transcriptional repression, instead generating a similar outcome to expression of the pBAD30 vector-only control, expression of the pBAD30-*his*$_6$-*brxR*-R53A mutant revealed persistent partial transcriptional repression (fig. 8c), with the R53A mutation apparently only partially affecting the ability of BrxR protein to bind its target DNA site. As expected, expression of both the pBAD30-*his*$_6$-*brxR*-R151A and the pBAD30-*his*$_6$-*brxR*-R178A constructs resulted in full transcriptional repression relative to the expression of wild-type BrxR protein from the pBAD30-*his*$_6$-*brxR* construct (fig. 8c).

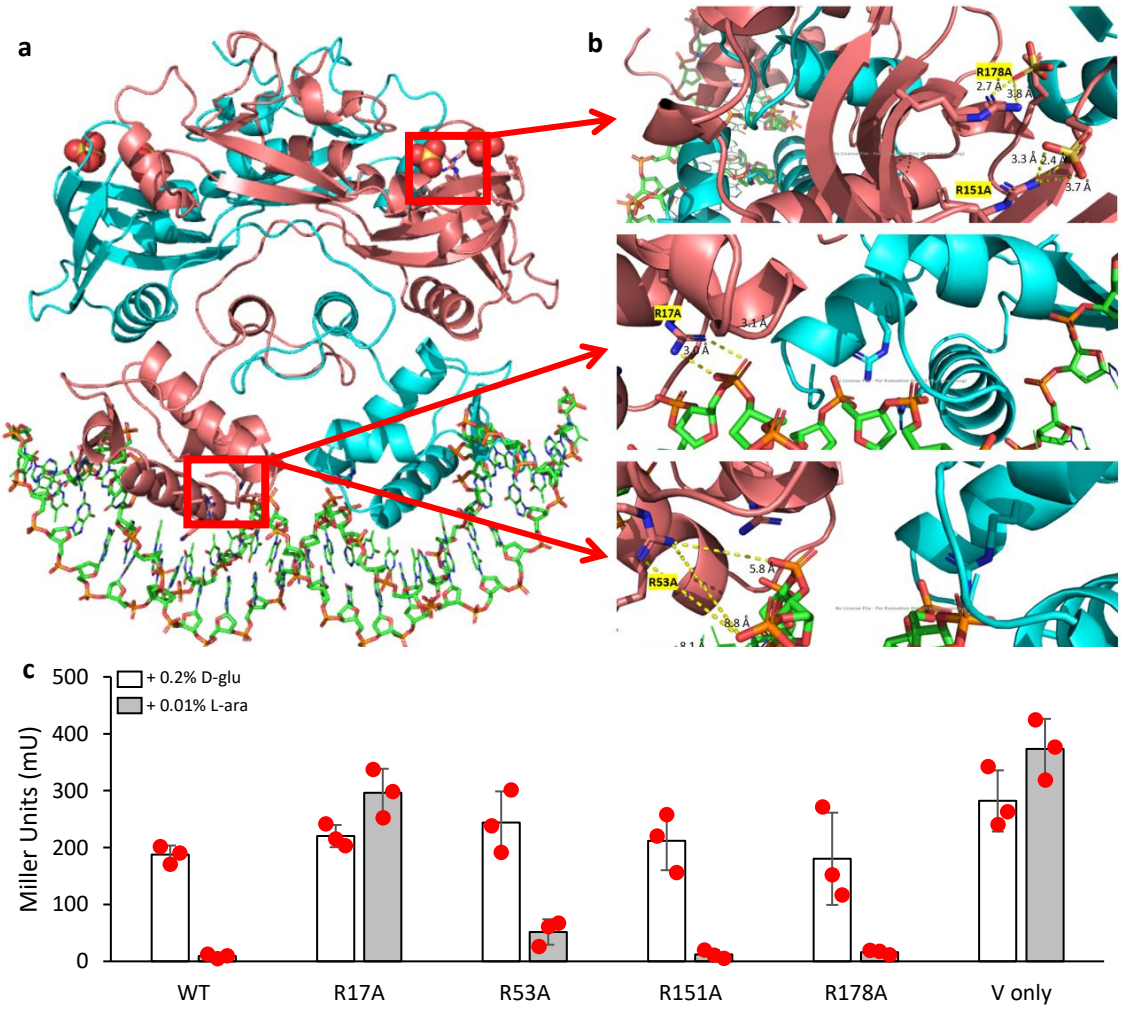

Figure 8- Roles of WYL and HTH domains within BrxR protein. (**A**) BrxR protein structure as derived by protein crystallography and superimposed on DNA double helix. (**B**) Magnified regions of BrxR protein structure illustrating the locations of WYL domain mutations R178A/R151A, and HTH-domain mutations R17A/R53A, respectively, and indicating their distance from the DNA double helix in angstroms. **(C)** LacZ-reporter assays using constructs PRW50-R7 and pBAD30-*his*$_6$-b*rxR* with varying WYL and HTH domain mutations. Structural data taken from Picton et al. 2022[16].

3.4 Investigation of BrxR interaction with a potential phage-associated ligand

We hypothesised that a phage-associated ligand might bind the BrxR WYL domain and de-repress the pEFER locus. This was initially planned to be investigated using the R7 promoter region, through use of a custom 'pRU' (fig. 9) plasmid containing the wild-type *brxU* gene downstream of the R7 promoter region. This combination was chosen due to the known strong promoter activity of the R7 region along with the restriction enzyme activity of BrxU protein, thus allowing for a strong measurable indicator of plasmid transcription (in the form of variable BrxU-sensitive DNA transformation efficiency) and subsequent repression by BrxR. Heat-competent *E. coli* DH5α cells containing both the pRU plasmid and pBAD30-*brxR* would either have been transformed with 5-hmC pCOLA or non-modified pCOLA (simply as a workable non-phage mobile genetic element with well characterised transformation efficiency). Alternatively, pRU pBAD30-*brxR* double plasmid strain would have been infected with known BrxU-sensitive phage Geo as a means of exploring the hypothesis that BrxR protein responds to a phage-associated ligand which would not be present in the cells transformed with pCOLA. The measurable output for these experiments would have been the variable transformation efficiency of pCOLA with or without 5hmC

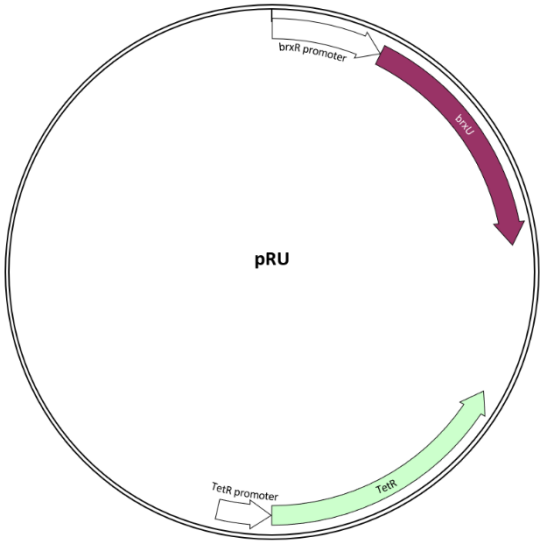

Figure 9- Mock-up plasmid map of pRU, showing TetR antibiotic resistance gene with promoter and *brxU* gene downstream of the P_{brxR} promoter, thereby allowing BrxR-mediated repression of BrxU production.

modifications targeted by BrxU, or changes in EOP for phage Geo as an indicator of *brxU* expression from the BrxR-repressed R7 promoter. Unfortunately, commercial delivery of the pRU plasmid was delayed and so an alternate experimental design using in-house materials was developed, as described below.

3.5 Investigation of BrxU interaction with mobile genetic elements

The alternate experimental design involved use of the in-house available pRSF (fig. 10) in place of pCOLA and pBAD30-*brxU* in place of pRU, and instead aims to characterise the interactions of BrxU protein with non-phage mobile genetic elements. Non-modified- and 5hmC-pRSF were transformed into three separate strains of competent cells, containing pBAD30 vector only, pBAD30-*brxU*, and pBAD30-*brxU*-R102A (a BrxU inactive mutant). Each of these three combinations was then split into two groups: one to be grown on solid LB agar containing 0.2% D-glu and another to be grown on solid LB agar containing 0.1% L-ara to repress or induce transcription from the pBAD30 constructs respectively. Given the conferred resistance to kanamycin from the pRSF plasmid to its host bacteria, the LB agar also contained 50 µg/ml kanamycin to select for pRSF. Resulting colonies were counted on each plate and compared to cultures grown from pBAD30 vector only, pBAD30-*brxR*, and pBAD30-*brxR*-R102A-containing strains that had been transformed with miniprepped pRSF as an indicator for baseline transformation efficiency. Theoretically, a decrease in the number of colonies would indicate the cleavage of incoming plasmid DNA by BrxU and should, thus, only be observed following conditions in which WT BrxU is expressed (i.e. in the strains that both carry pBAD30-brxU and are grown on an arabinose-supplemented medium) and should only be observed to occur to BrxU-sensitive genetic elements (i.e. those carrying the 5hmC modification).

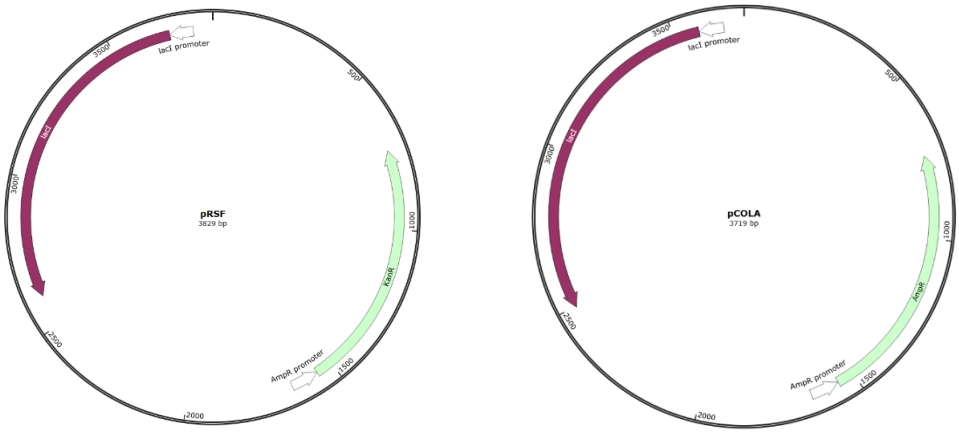

Figure 10- Plasmid maps of pRSF and pCOLA, showing antibiotic resistance genes and *lacI* downstream of its promoter.

Unfortunately, triplicate colony counts were highly inconsistent in most cases (Table 10). While this issue is likely due, in part, to inconsistency within and between the

three biological replicates used, it should also be noted that the transformation efficiency of miniprepped pRSF was multiple orders of magnitude greater than the transformation efficiency of the PCR-prepped non-modified- and 5hmC-pRSF despite the same mass of DNA being used for each transformation (33 ng) (Table 10). This is to be expected, due to the supercoiled structure of miniprepped plasmid DNA being better suited to passage through bacterial cell membranes than the circular non-supercoiled structure of PCR-prepped DNA, which was the final form of the 5hmC-pRSF. Should this experiment be repeated in the future, a greater amount of PCR-prepped 5hmC-pRSF plasmid should be used to improve transformation efficiency and thus improve the resolution of comparisons drawn between growth conditions. Despite these issues, there are features of the data that align with expectations, such as a lack of any colonies grown from the pBAD30-*brxU* strain when transformed with 5hmC-pRSF and grown on LB agar supplemented with L-ara (Table 10). No colonies would be expected in these conditions due to the ability of BrxU protein to cleave DNA possessing the 5hmC modification (following its expression from the PBAD30 vector, induced by the presence of L-ara in the growth medium). Despite this, no justifiable conclusion can be drawn about the action of BrxU protein in this instance due to a low number of colonies grown from the pBAD30-*brxU* strain when transformed with 5hmC-pRSF and grown on LB agar supplemented with D-glu (Table 10).

Table 10- EOP results for PCR-prepped modified (5hmC) and non-modified pRSF as well as miniprepped pRSF into competent bacterial strains containing pBAD30, pBAD30-*brxU*, and pBAD30-*brxU*-R102A *brxU* knockout variant with and without vector induction (+ A and + G, respectively)

	Average colony count ± SD		
Plasmid	pBAD30	pBAD30-*brxU*	pBAD30-*brxU*-R102A
5hmC + G	3.33 ± 0.94	1.00 ± 0.82	16.00 ± 6.16
5hmc + A	1.33 ± 0.47	0.00 ± 0.00	13.00 ± 2.83
Non-mod + G	48.67 ± 38.05	18.33 ± 19.29	232.00 ± 148.06
Non-mod + A	20.67 ± 13.30	12.00 ± 14.17	102.00 ± 83.83
pRSF miniprep	407.00 ± 448.63	632.00 ± 190.28	1633.00 ± 377.34

3.6 The effect of varying BREX systems on an array of phages

A separate final piece of work was also performed, as part of another study in the Blower Lab. The aim of this study was to investigate how varying BREX systems respond to a diverse suite of phages, and how the number of BREX recognition sites present in a given genome affects the extent of protection conferred by BREX systems against a given phage. While the full suite of phages comprises 17 BrxU-resistant strains, only six are presented in Table 11 below to reflect my personal contribution to the dataset. Specifically, my contribution is represented by the data reflecting the interactions between each of these six phages and a BREX system found in *E. coli* – pBREX-AL defence system (using pBTB-2 as a control), while data pertaining to phage interaction with the *Salmonella* and *E. fergusonii* BREX systems are included for the sake of completeness (the purpose of those experiments being to compare the effects of BREX systems found in different organisms). Each assay was performed as per the previously described efficiency-of-plating assay protocol. Whilst some correlation was predicted between the number of BREX recognition sites (GATCAG, GCTAAT, and GGTAAG motifs) within a given phage's genome and the scale of BREX-mediated defence against that phage, it is not observed in these results. This suggests the potential for anti-BREX genes within the genomes of the phages that escaped BREX susceptibility despite the presence of recognition sites in their genomes. As of the writing of this thesis, these data are within a study under review.

Table 11- EOP results of Durham collection phages against three BREX systems. Work done by members of the Blower Lab (unpublished data). Colour scale denotes preservation of EOP, from green (preserved) to red (diminished).

Source	***S. enterica* serovar Typhimurium ST313 strain D23580**				***E. fergusonii* ATCC 35469**		***E. coli* HS**	
BREX motif	**GATCAG**				**GCTAAT**		**GGTAAG**	
Phage	**BREX sites**	**pBrxXL$_{Sty}$**	**pBrxXL$_{Sty}$-*ΔbrxL***	**pBrxXL$_{Sty}$-*ΔariAΔariB***	**BREX sites**	**pBrxXL$_{Eferg}$**	**BREX sites**	**pBREX-AL**
Pau	84	0.19 ± 0.13	$3.73 \times 10^{-4} \pm 2.65 \times 10^{-4}$	0.14 ± 0.27	82	$3.00 \times 10^{-8} \pm 1.56 \times 10^{-8}$	90	0.11 ± 0.03
PATM	83	$0.23 \pm 5.07 \times 10^{-2}$	$3.27 \times 10^{-3} \pm 6.45 \times 10^{-4}$	$9.64 \times 10^{-2} \pm 2.57 \times 10^{-3}$	83	$2.00 \times 10^{-8} \pm 6.96 \times 10^{-9}$	91	0.81 ± 0.06
Jura	0	2.58 ± 0.11	1.95 ± 1.37	1.82 ± 0.53	74	1.00 ± 0.00	60	0.44 ± 0.30
Mak	1	0.78 ± 0.10	3.77 ± 2.90	0.67 ± 0.12	11	1.34 ± 0.73	41	$1.07 \times 10^{-3} \pm 3.10 \times 10^{-4}$
CS16	13	$3.25 \times 10^{-2} \pm 2.30 \times 10^{-2}$	$1.45 \times 10^{-2} \pm 2.39 \times 10^{-3}$	$6.17 \times 10^{-2} \pm 1.86 \times 10^{-2}$	23	$1.63 \times 10^{-3} \pm 1.06 \times 10^{-3}$	32	$8.16 \times 10^{-6} \pm 4.40 \times 10^{-6}$
Mav	13	$0.15 \pm 2.16 \times 10^{-2}$	$5.82 \times 10^{-2} \pm 4.08 \times 10^{-2}$	$9.4 \times 10^{-2} \pm 9.89 \times 10^{-2}$	23	0.15 ± 0.30	32	$1.05 \times 10^{-4} \pm 7.08 \times 10^{-5}$

4. Discussion

Knowledge of the presence of both a BREX and a type IV restriction-modification phage defence system on the *Escherichia fergusonii* plasmid pEFER, coupled with a lack of knowledge about the coregulation of these systems, warranted investigation into BrxR protein as a transcriptional regulator of the defence island.

This project has resulted in functional characterisation of BrxR as a transcriptional repressor acting at the P_{BrxR} promoter immediately upstream of gene *brxA*. EOP assay results have confirmed a lack of phage defence provided by BrxR in isolation, LacZ assay results have provide specific information on the interactions between BrxR and the pEFER phage defence island and the transcriptional effects thereof, and mutagenesis of BrxR has facilitated functional characterisation of the WYL and HTH domains of the protein, implicating them in its function as a ligand-binding transcriptional regulator.

4.1 BrxR solo impact on phage defence

Given the lack of change in EOP of phages Pau, Trib, and Baz (Table 9) brought about by solo expression of *brxR*, it can be concluded that BrxR protein is not a direct participant in phage defence. Subsequently, attempts to produce *brxR* knockout transformants in the context of the defence island following golden gate assembly, Gibson assembly, or by commercial means were all unsuccessful due to either substantial mutations elsewhere in the defence island or, in the case of commercial commission, an inability to generate the mutation outright. Whilst practically limiting, the lack of success in generating a *brxR* knockout mutant suggests that the role of BrxR within the BREX system may extend beyond regulation and may involve limiting a degree of toxicity implicit to constitutive expression of the defence system. This possibility represents an avenue for future study with a focus on the toxicity of other BREX components.

4.2 Interactions of BrxR protein with the BREX locus

The pEFER-encoded BREX and type IV restriction-modification systems were previously shown to complement their respective activities. The data presented here explore the role of *brxR* and its gene product within these systems, characterising $BrxR_{Efer}$ as a negative regulator of phage defence acting upstream of *brxR* and *brxA* (fig. 4c).

4.2.1 Expression from putative promoter sites and subsequent repression by BrxR

The initial LacZ assay data showing expression from the promoters P_{brxR} and P_{brxA} and subsequent repression by $BrxR_{Efer}$ (fig. 4c) concur with similar co-published findings exploring the action of homologues $BrxR_{Acin}$, found in *Acinetobacter*, and CapW, found in *Stenotrophomonas maltophilia*[30,31]. $BrxR_{Efer}$ represses expression from the P_{BrxR} promoter upstream of the first canonical BREX locus gene, *brxA*. While the primary point of interest is the observed expression from the P_{BrxR} promoter, there was also notable expression from the putative P_{brxA} promoter. This baseline expression was of a lesser degree than that from P_{BrxR} and was more weakly repressed by $BrxR_{Efer}$. This may be a secondary function of BrxR, involving regulation of transcription of the BREX locus from *brxA* onwards, but may also be explained by a degree of promiscuity between $BrxR_{Efer}$ and its binding sites with primary regulation of the P_{BrxA} promoter in control of an as yet unidentified factor.

4.2.2 brxR expression, not pBAD30 activation, is the source of transcriptional repression

The data supplementary to the primary LacZ assay (fig. 4d) show that activation of the empty pBAD30 vector had no repressive effect on expression of β-Galactosidase from the pRW50 vector containing the P_{BrxR} promoter within fragment R7. This permits the confident conclusion that BrxR expression, not pBAD30 activation, is the source of transcriptional repression observed within the primary assay data.

4.2.3 Repression by BrxR is sequence-specific

Assays performed using alternative inserts P_{abiEi} and $P_{rv2827c}$ within the pRW50 vector (fig. 4e), which are known to exhibit promoter activity within LacZ assays, showed no repression of β-Galactosidase transcription upon $BrxR_{Efer}$ expression. This permits the confident conclusion that repression by $BrxR_{Efer}$ is sequence specific.

4.2.4 Inverted repeats IR1 and IR2 are not individually essential for expression from P_{brxR} or subsequent repression by BrxR

Assays performed with adjusted R7 inserts for the PRW50 vector (fig. 4), including variable replacements of the IR1 and IR2 repeats with polyC tracts, show that neither IR1 nor IR2 are individually essential for expression from the P_{BrxR} promoter, and instead each provide a lesser degree of transcription on their own whilst still being

repressed to the same extent by BrxR$_{Efer}$. These data corroborate EMSA data produced by another member of the Blower Lab which indicate that His$_6$-BrxR can bind either IR1 or IR2 individually when the other is replaced with a polyC tract[16]. Unlike the EMSA data, however, the IR1c-IR2c mutant did not produce conclusive data. This is because LacZ assays rely on change from a baseline level of transcription and, as the mutations eliminated the baseline level of transcription from the P$_{BrxR}$ promoter, repression as a result of BrxR binding could not be observed.

4.2.5 Mechanism by which BrxR represses DNA transcription

Electrophoretic mobility shift assay (EMSA) studies carried out in parallel with the LacZ assays discussed showed that BrxR$_{Efer}$ exists as a stable dimer and binds inverted DNA repeats in this state[16]. Given that the inverted repeats to which BrxR$_{Efer}$ binds exist adjacent to the putative promoter P$_{brxR}$, the resulting repression is likely due to steric blocking of RNA polymerase. Following these studies, a second set of inverted repeats, named R-BOX2, was discovered upstream of P$_{brxA}$, offering a likely explanation for the observed ability of BrxR$_{Efer}$ to repress expression from regions R9 and R10 (fig. 4c). No further inverted repeat sequences were discovered, however the data show that either inverted repeat alone is sufficient to facilitate binding (fig. 5) and BrxR-mediated repression (fig. 6). Given this, it is possible that other BrxR binding sites exist within the defence island, in forms other than inverted repeats, awaiting discovery.

It is worth considering why the activity of BrxR is necessary at all, that is, why regulation of the defence island is necessary in the first place. As mentioned previously, all of our attempts to generate a *brxR* knockout were unsuccessful. Furthermore, we have observed toxicity of PglX$_{Efer}$ when overexpressed (unpublished data). These details together support the hypothesis that repression of the defence island is necessary to maintain host fitness outside of the context of phage infection, i.e. that unregulated expression of the defence island would incur some fitness penalty to the host. It is also worth exploring the possibility of non-BrxR transcriptional regulators, especially within BREX loci which do not contain BrxR-family homologues. Such studies could involve methods similar to those used to identify and characterise BrxR, i.e. examining genes of unknown function within BREX loci for the presence of ligand- or DNA-binding domains (e.g. WYL and HTH domains, in the case of BrxR).

4.3 Significance of the WYL and HTH domains within BrxR protein

Further to the LacZ assay data already discussed, the data detailing the function of BrxR DNA-loss binding HTH domain mutants R17A and R53A, as well as WYL-domain mutants R151A and R178A, illustrate the importance of both the HTH and WYL domains to the negative regulatory function of BrxR. Compared with wild-type His_6-BrxR, mutants His_6-BrxR-R151A and His_6-BrxR-R178A exhibited no loss in ability to repress expression from the R7 promoter region (fig. 8c). This is to be expected, as these strains contained versions of His_6-BrxR with mutations in the WYL domain, which is not involved in DNA-binding. Instead, the measurable effect produced by mutation within the WYL domain should be non-reversible repression of transcription from the P_{brxR} promoter resulting from loss of ligand-binding capability, although this was not attempted, as it would require identification of the ligand. However, having now observed DNA-binding *in vitro* through EMSAs, it may be possible to screen potential ligands for negative impacts on DNA binding.

4.3.1 R17A is a BrxR DNA loss binding mutant without the ability to repress transcription

The His_6-BrxR-R17A mutant exhibited a total loss of ability to repress expression from the R7 promoter region, whereas the His_6-BrxR-R53A mutant exhibited only a partial loss in ability to repress expression from the R7 promoter when compared to the vector-only control (fig. 8c). The notable difference between the data gathered from the His_6-BrxR-R17A and His_6-BrxR-R53A mutants, despite their both containing mutations within the DNA-binding-related HTH domain, may be attributed to the relative distances from the positively charged R17 and R53 amino acid residues to the negatively charged phosphate groups of the DNA backbone (fig. 8b). When DNA is bound to the protein, the R17 residue is roughly half the distance from the DNA backbone when compared to the R53 residue. This may account for the apparently greater impact of the R17A mutation on DNA binding.

The LacZ assay data relating to the effect of the R17A mutation within BrxR corroborate EMSA data generated by another member of the Blower Lab, in which His6-BrxR-R17A was shown to be incapable of binding an IR1-IR2 probe despite

correct folding[16]. These data together show that the R17 residue within BrxR$_{Efer}$ is essential for both DNA binding and transcriptional repression.

4.4 BrxR is a model system for a widespread family of WYL-domain transcriptional regulators

The aforementioned data characterising BrxR as a ligand-binding transcriptional repressor of phage-defence corroborate findings that describe similar activity of BrxR homologues found in other Gram-negative bacterial strains. Specifically, BrxR$_{Acin}$, found upstream of an *Acinetobacter* BREX system[30], and CapW, found in a CBASS system of *Stenotrophomonas maltophilia*[31] demonstrate other instances of BrxR-mediated transcriptional regulation of phage defence systems.

Additionally, comparative genomic analyses conducted by other members of the Blower Lab indicated that BrxR-family homologues are widespread within Proteobacteria. Nearly half of these homologues were identified in close proximity to known phage defence systems. These findings implicate BrxR as a model system for a widespread family of transcriptional regulators of phage defence. Furthermore, association of BrxR with conserved genes of unknown function affords the potential for characterisation of novel phage defence systems in the future.

4.5 BrxU interactivity with mobile genetic elements

While the experimental design used to generate the data presented in Table 10 had the potential to characterise BrxU interactivity with mobile genetic elements, a number of technical hurdles impeded its ability to do so. Whilst these data could have been used to preface a subsequent experiment comparing BrxU interactivity with mobile genetic elements of phage and non-phage origin, the high variance within biological triplicate sets as well as the discrepancy in transformation efficiency between mini-prepped non-modified and PCR-prepped modified pRSF render them largely unusable. While these issues could have been fully resolved given more time within the project research period, the data gathered are not entirely unavailing. The loss, and subsequent recovery, of transformation efficiency of 5hmC-pRSF into bacteria expressing pBAD30-brxU and pBAD30-brxU-R102A, respectively, validate the experimental design and indicate that a complete dataset could be gathered from future attempts.

Such a full dataset would ideally show the following (relative to mini-prepped pRSF transformation efficiency): total non-effect of pBAD30 vector-only expression on

plasmid transformation efficiency; selective reduction of 5hmC plasmid transformation efficiency upon expression from pBAD30-*brxU*; and total non-effect of pBAD30-*brxU*-R102A expression on plasmid transformation efficiency. This set of results would facilitate comparison of the effect of BrxU on non-phage mobile genetics elements to that on those of phage origin.

4.6 The effect of varying BREX systems on an array of phages

Investigation of the response of 17 phages from the Durham collection to the BREX systems found in *S. enterica*, E. *fergusonii*, and *E. coli* (Table 11) demonstrated a different phage profile of BREX resistance between different systems tested. These profiles often did not align with the number of BREX target sites present in each phage genome, potentially indicating alternative modes of BREX resistance within these phages that are yet unknown. Genomic analyses of the phages tested has been undertaken by Blower lab members and associates in a study under review as of the writing of this thesis, but did not identify any obvious anti-BREX components (e.g. Ocr or DarB-like proteins).

4.7 Implications and future research

Collectively, these data outline a model of BrxR action that involves phage-mediated de-repression of BREX transcription from the P_{BrxR} promoter. Binding upstream and, thus, repression of the P_{BrxR} promoter is mediated by the BrxR HTH domains, whereas putative ligand binding and, thus, dissociation and de-repression, are likely mediated by the BrxR WYL domains, a function that has been predicted previously[18]. The ligand bound by BrxR may be ssDNA, as in PIF1[21], or ssRNA as in WYL1[32]. Another nucleic acid molecule could also be the cognate BrxR ligand, acting as a messenger of DNA damage due to phage infection, as has been previously observed[33].

Much further study is necessary to identify the ligand bound by BrxR, whatever it may be. This is a challenging prospect in many ways, including the documented promiscuity of WYL domain ligand recognition and the significant quantity of ligand candidates. Speculatively, 2′-3′ phosphate cyclic nucleotide or other nucleic acid polymers have been identified as primary candidates, due to their emergence as important signalling factors for other phage defence systems[5,34]. Aside from trialling a library of candidate ligands to investigate BrxR binding, attempts could be made to isolate BrxR protein in its ligand-bound form

Other questions remaining to be answered include the potential toxicity of BREX components, as indicated by the non-viability of *brxR*-knockout mutations which allow constitutive $BREX_{Efer}$ expression, the binding relationship between BrxR protein and R-BOX2 inverted repeats found upstream of *brxA* and the P_{BrxA} promoter, and the location of potential as yet unidentified non-inverted repeat BrxR binding sites. Finally, RNA-seq or qPCR studies can be performed before and after exposure to phage, to investigate how pEFER defence island expression varies during infection.

5. Bibliography

1. Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L. & Brüssow, H. Phage-Host Interaction: an Ecological Perspective. *J Bacteriol* **186**, 3677–3686 (2004).

2. Loenen, W. A. M. & Raleigh, E. A. The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res* **42**, 56–69 (2014).

3. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67–83 (2020).

4. Owen, S. v. *et al.* Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe* **29**, 1620-1633.e8 (2021).

5. Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Mol Cell* **77**, 723-733.e6 (2020).

6. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* **34**, 169–183 (2015).

7. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. v. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J Bacteriol* **193**, 6039–6056 (2011).

8. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, (2018).

9. Koonin, E. v, Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* **71**, 233–261 (2017).

10. Chinenova, T. A., Mkrtumian, N. M. & Lomovskaia, N. D. Genetic characteristics of a new phage resistance trait in Streptomyces coelicolor A3(2). *Genetika* **18**, 1945–52 (1982).

11. Beck, I. N., Picton, D. M. & Blower, T. R. Crystal structure of the BREX phage defence protein BrxA. *Curr Res Struct Biol* **4**, 211–219 (2022).

12. Gordeeva, J. *et al.* BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res* **47**, 253–265 (2019).

13. Hui, W. *et al.* A Novel Bacteriophage Exclusion (BREX) System Encoded by the *pglX* Gene in Lactobacillus casei Zhang. *Appl Environ Microbiol* **85**, (2019).

14. Zaworski, J. *et al.* Reassembling a cannon in the DNA defense arsenal: Genetics of StySA, a BREX phage exclusion system in Salmonella lab strains. *PLoS Genet* **18**, e1009943 (2022).

15. Picton, D. M. *et al.* The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses. *Nucleic Acids Res* **49**, 11257–11273 (2021).

16. Picton, D. M. *et al.* A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. *Nucleic Acids Res* **50**, 5191–5207 (2022).

17. Brennan, R. G. & Matthews, B. W. The helix-turn-helix DNA binding motif. *Journal of Biological Chemistry* **264**, 1903–1906 (1989).

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

19. Modell, J. W., Kambara, T. K., Perchuk, B. S. & Laub, M. T. A DNA Damage-Induced, SOS-Independent Checkpoint Regulates Cell Division in Caulobacter crescentus. *PLoS Biol* **12**, e1001977 (2014).

20. Hein, S., Scholz, I., Voß, B. & Hess, W. R. Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol* **10**, 852–864 (2013).

21. Andis, N. M., Sausen, C. W., Alladin, A. & Bochman, M. L. The WYL Domain of the PIF1 Helicase from the Thermophilic Bacterium *Thermotoga elfii* is an Accessory Single-Stranded DNA Binding Module. *Biochemistry* **57**, 1108–1118 (2018).

22. LeGault, K. N. *et al.* Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* **373**, (2021).

23. Berger, S. & Kimmel, A. Guide to molecular cloning techniques. *Methods Enzymol* **152**, 1–812 (1987).

24. Im, H. The Inoue Method for Preparation and Transformation of Competent E. coli: Ultra Competent Cells. *Bio Protoc* **1**, (2011).

25. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121–4130 (1995).

26. Lodge, J., Fear, J., Busby, S., Gunasekaran, P. & Kamini, N. R. Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett* **95**, 271–276 (1992).

27. Beck, I. N., Usher, B., Hampton, H. G., Fineran, P. C. & Blower, T. R. Antitoxin autoregulation of *M. tuberculosis* toxin-antitoxin expression through negative cooperativity arising from multiple inverted repeat sequences. *Biochemical Journal* **477**, 2401–2419 (2020).

28. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858 (2015).

29. Hampton, H. G. *et al.* 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* **430**, 1141–1156 (2018).

30. Luyten, Y. A. *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* **50**, 5171–5190 (2022).

31. Blankenchip, C. L. *et al.* Control of bacterial immune signaling by a WYL domain transcription factor. *Nucleic Acids Res* **50**, 5239–5250 (2022).

32. Zhang, H., Dong, C., Li, L., Wasney, G. A. & Min, J. Structural insights into the modulatory role of the accessory protein WYL1 in the Type VI-D CRISPR-Cas system. *Nucleic Acids Res* **47**, 5420–5428 (2019).

33. Müller, A. U., Leibundgut, M., Ban, N. & Weber-Ban, E. Structure and functional implications of WYL domain-containing bacterial DNA damage response regulator PafBC. *Nat Commun* **10**, 4653 (2019).

34. Tal, N. *et al.* Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728-5739.e16 (2021).