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

David M. Picton[1], Joshua D. Harling-Lee[1,2], Samuel J. Duffner[1], Sam C. Went[1],
Richard D. Morgan[3], Jay C.D. Hinton[4] and Tim R. Blower [1,*]

[1]Department of Biosciences, Durham University, Stockton Road, Durham DH1 3LE, UK, [2]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Edinburgh EH25 9RG, UK, [3]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA and [4]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool L69 7ZB, UK

ABSTRACT

Bacteria are under constant assault by bacteriophages and other mobile genetic elements. As a result, bacteria have evolved a multitude of systems that protect from attack. Genes encoding bacterial defence mechanisms can be clustered into 'defence islands', providing a potentially synergistic level of protection against a wider range of assailants. However, there is a comparative paucity of information on how expression of these defence systems is controlled. Here, we functionally characterize a transcriptional regulator, BrxR, encoded within a recently described phage defence island from a multidrug resistant plasmid of the emerging pathogen *Escherichia fergusonii*. Using a combination of reporters and electrophoretic mobility shift assays, we discovered that BrxR acts as a repressor. We present the structure of BrxR to 2.15 Å, the first structure of this family of transcription factors, and pinpoint a likely binding site for ligands within the WYL-domain. Bioinformatic analyses demonstrated that BrxR-family homologues are widespread amongst bacteria. About half (48%) of identified BrxR homologues were co-localized with a diverse array of known phage defence systems, either alone or clustered into defence islands. BrxR is a novel regulator that reveals a common mechanism for controlling the expression of the bacterial phage defence arsenal.

INTRODUCTION

Bacteriophages outnumber bacterial prey by about 10-fold (1,2). The estimated $\geq 10^{30}$ bacteriophages (phages) on Earth (1,2) cause infections at a rate of 10^{25} per second (3).

To contend with this extreme selection pressure, bacteria have evolved varied modes of defence against phages and other mobile genetic elements (4–6). Well-established examples of defence systems include restriction–modification (R–M) (7–10), abortive infection (11) and CRISPR-Cas (12) systems. R–M systems have been shown to cluster in 'immigration control regions' (13). Recent comparative genomic analyses have demonstrated how diverse defence systems also commonly cluster into 'defence islands' (14,15). The 'guilt-by-association' approach has allowed gene functions to be inferred from defence islands, and has identified novel defence systems (16). Coupled with renewed interest in technological spin-offs from these systems, and the rise of phage therapy to treat bacterial infections, multiple new systems have been identified, including Bacteriophage Exclusion (BREX) (17), CBASS (18), BstA (19), retrons (20), viperins (21) and pycsar (22). As multiple diverse systems have been assembled into a single locus, expression of the various genes must be meticulously regulated to reduce any impacts on host fitness whilst maximizing the response to phages, and other mobile genetic elements.

It has been postulated that WYL-domain containing proteins act as ligand-binding regulators of phage defence system expression (23). WYL-domains (named after three conserved amino acids), are only found in prokaryotes and are part of the Sm/SH3 superfold family, which is itself subsumed by the larger 'small β-barrel' family (24). Sm proteins are responsible for eukaryotic snRNP complexes and were first discovered as autoantigens in cases of lupus (using sera from a patient named Stephanie Smith) (25), whilst SH3 (Src-homology 3) domains are adaptor domains with diverse roles in eukaryotic cell signalling (26). In prokaryotes, the Sm homologue Hfq uses the Sm/SH3 fold to bind RNAs (27), whilst other WYL-domains bind proteins, peptides, DNA and oligosaccharides (23).

A handful of studies have begun to demonstrate that WYL-domain containing proteins regulate diverse processes in prokaryotes: Sll7009 from *Synechocystis* 6803 represses the CRISPR subtype I-D locus (28); DriD from *Caulobacter crescentus* activates expression of SOS-independent DNA damage response mediators (29); PIF1 helicase from *Thermotoga elfii* has a ssDNA-binding WYL-domain that couples ATPase activity to DNA unwinding (30); the RNA cleavage activity of a Type VI Cas13d protein from *Ruminococcus* is stimulated by a WYL-domain protein named WYL1 (31), which binds ssRNA with high affinity (32); and in Mycobacteria, PafBC is a transcriptional activator of DNA damage response genes (33). Most recently, WYL-domain proteins were found associated with phage defence islands within integrative conjugative elements of *Vibrio cholerae* (34).

We recently characterized a multi-functional phage defence island containing a BREX system (17) and the BrxU GmrSD-family (35) type IV restriction enzyme, encoded on a multidrug resistant plasmid of the emerging animal and human pathogen *Escherichia fergusonii* (36,37). These two systems provide complementary protection against a wide range of environmental coliphages (37). This defence island encodes a WYL-domain containing protein, BrxR, which was hypothesized to act as a transcriptional regulator. Here, we present functional and structural characterization that identifies BrxR as the first member of a large family of transcriptional regulators. BrxR-family homologues are widely associated with diverse phage defence systems and islands. Our findings suggest a possible common thematic approach for the regulation of phage defence systems that may involve a signalling molecule acting as a secondary messenger.

MATERIALS AND METHODS

Bacterial strains and culture conditions

Total genomic DNA (gDNA) was obtained for *E. fergusonii* ATCC 35469 from ATCC. *Escherichia coli* strains DH5α and BL21 (DE3) (ThermoFisher Scientific) were grown at 37°C, either on agar plates or shaking at 220 rpm for liquid cultures. Luria broth (LB) was used as the standard growth media for liquid cultures, and was supplemented with 0.35% (w/v) or 1.5% (w/v) agar for semi-solid and solid agar plates, respectively. Growth was monitored using a spectrophotometer (WPA Biowave C08000) measuring optical density at 600 nm (OD_{600}). When necessary, growth media was supplemented with ampicillin (Ap, 50 μg/ml), tetracycline (Tc, 10 μg/ml), isopropyl-β-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).

Use of environmental coliphages

Escherichia coli phages Pau, Trib and Baz were isolated previously from freshwater sources in Durham, UK (37). 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.

DNA isolation and manipulation

PCR amplicons and plasmids were purified using Monarch DNA kits (NEB). PCR, restriction digests, ligations, transformations and agarose gel electrophoresis were performed using standard molecular biology techniques. Constructed plasmids were confirmed via sequencing with an Abi 3370 DNA sequencer. The pSAT1-LIC-*brxR*$^+$ expression construct adds a cleavable N-terminal His$_6$-SUMO tag. Primers TRB878 and TRB879 were used to amplify *brxR* from pEFER (gene pEFER_0020) for insertion into pSAT1-LIC (38) to produce pTRB446 via Ligation Independent Cloning (LIC) (Supplementary Table S1). Primers TRB876 and TRB877 were used to amplify *brxR* from pEFER which was inserted into pBAD30 (39) to produce pBAD30-*his$_6$-brxR* (Supplementary Table S1). Primers TRB1987 and TRB1988 were used to perform QuikChange (Invitrogen) mutagenesis to produce pBAD30-*his$_6$-brxR*-R17A (Supplementary Table S1).

Efficiency of plating assays

Escherichia coli DH5α were transformed with pBAD30-*his$_6$-brxR* and transformants were used to inoculate overnight cultures. Serial dilutions of phages Pau, Trib and Baz (37) 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. Efficiency of plating (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.

β-Galactosidase assays

Putative promoter regions (R1-12, or mutants thereof) were ligated into the promoterless *lacZ* fusion plasmid, pRW50 (40) (Supplementary Table S1). *Escherichia coli* DH5α was then co-transformed with one of the *lacZ* reporter constructs (or pRW50 as a vector control) and either pBAD30, pBAD30-*his$_6$-brxR* or pBAD30-*his$_6$-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 (41). The plotted data are the normalized mean and standard deviation of three independent replicates.

Protein expression and purification

To make untagged BrxR, *E. coli* BL21 (DE3) was transformed with pSAT1-LIC-*brxR* and single colony was use to inoculate a 25 ml overnight culture of LB, supplemented with Ap and grown overnight. Overnight cultures were used to inoculate 12 l of 2× YT media in 2 l baffled flasks, each containing 1 l of culture. Cultures were grown at 37°C shaking at 180 rpm until an OD_{600} of ~0.6, at which point cultures were supplemented to a concentration of 1 mM IPTG to induce expression. Cultures were incubated overnight at 16°C and cells were pelleted at 4500 × g for 30 min at 4°C. Cell pellets were resuspended in 50 ml of ice-cold A500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 10 mM imidazole and 10% glycerol) and used immediately or flash frozen in liquid nitrogen and stored at −80°C. Pellets were lysed via sonication and centrifuged at 45 000 × g at 4°C for 30 min. All clarified cell lysates were passed over a 5 ml HisTrap HP column (Cytiva), and washed with 50 ml of A500. Bound BrxR was further washed with 50 ml of W500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 40 mM imidazole and 10% glycerol) and eluted from the column in B500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 250 mM imidazole and 10% glycerol). Imidazole was removed via dialysis back into A500 and the sample was treated with hSENP2 SUMO protease overnight at 4°C to remove the fusion tag. The resulting untagged BrxR was loaded on to a second 5 ml HisTrap HP column and the flowthrough was collected and concentrated to 2 ml. BrxR was further purified via size exclusion through a Sephacryl S-300 HR gel filtration column in preparative SEC buffer (20 mM Tris–HCl pH 7.9, 500 mM KCl and 10% glycerol). Fractions were analysed via SDS-PAGE to assess content and purity, and peak fractions were pooled. BrxR was either dialysed into Xtal buffer (20 mM Tris–HCl pH 7.9, 200 mM NaCl and 2.5 mM DTT) for use in crystallization, or was supplemented with glycerol to a final concentration of 30% (w/v) for biochemical assays and stored at −80°C following flash freezing in liquid nitrogen.

To make His_6-BrxR and His_6-BrxR-R17A, *E. coli* DH5α was transformed with the corresponding pBAD30-*his6*-*brxR* or pBAD30-*his6*-*brxR*-R17A construct and a single colony was used to inoculate a 25 ml of overnight culture of LB. Overnight cultures were used to inoculate 1 × 1 l of LB media in a 2 l baffled flask, and grown until an OD_{600} of ~0.6 before induction with L-arabinose to a final concentration of 0.1% (w/v). Cultures were incubated overnight at 16°C and cells were pelleted at 4500 x g for 30 min at 4°C. Cell pellets were resuspended in 50 ml of ice-cold A500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 10 mM imidazole and 10% glycerol). Pellets were lysed via sonication and cen-

trifuged at 45 000 × g at 4°C for 30 min. All clarified cell lysates were passed over a 5 ml HisTrap HP column (Cytiva), and washed with 50 ml of A500. Bound His_6-BrxR and His_6-BrxR-R17A was further washed with a step gradient of imidazole concentration, 50 ml of W500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 40 mM imidazole and 10% glycerol) and eluted from the column in B500 (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 250 mM imidazole and 10% glycerol). Samples were buffer exchanged into storage buffer (50 mM Tris–HCl pH 7.9, 500 mM KCl and 30% glycerol) and flash frozen in liquid nitrogen for storage at −80°C.

Electrophoretic mobility shift assays

An inverted repeat (IR) region was identified within the R7 promoter region upstream of *brxR*. Probes were synthesized using artificial templates (IDT) containing the target region and a 3′ common region corresponding to the start of *lacZ* within pRW50. Templates consisted of either the wild type (WT) sequence, or mutant sequences which replaced one or both of the IRs with polycytosine (Supplementary Table S1). Incorporation of the pRW50-based common region permitted the use of a single fluorescein-tagged reverse primer TRB1068, in conjunction with the respective forward primer, to produce fluorescein-tagged WT and IR mutant probes via PCR. Probes were purified via gel extraction and quantified via Nanodrop. DNA-binding reactions were performed in 10 μl volumes, containing 1 μl of 2500 fmol of labelled probe, 1 μl of 1 μg/μl poly (dI-dC), 2 μl of 5× EMSA buffer (50 mM Tris–HCl pH 7.9, 750 mM KCl, 2.5 mM EDTA pH 8.0, 1 mM DTT, 0.5% Triton X-100, 65% glycerol), 1 μl of His_6-BrxR or His_6-BrxR-R17A, and made up to 10 μl with water. Specific competitor samples used a 20-fold excess of unlabelled probe, and non-specific competitor samples used a 20-fold excess of the *rv2827c* promoter region from *Mycobacterium tuberculosis* (41) (Supplementary Table S1).

Samples were incubated at 25°C for 30 min before being loaded onto 7% native PAGE gels in 0.5× TBE (45 mM Tris-borate pH 8.0, 1 mM EDTA). Gels were pre-ran at 150 V for 120 min for 2 h at 4°C. Gels were imaged using an Amersham Bioscience Typhoon 9400 in fluorescence mode, emission filter 526 SP. Band intensities of the unbound probe were enumerated using ImageJ. Fractional saturation corresponding to the amount of unbound probe, Y, was calculated using $Y = 1 - (I_T/I_C)$, where I_T is the band intensity of the unbound probe in test lanes and I_C is the band intensity probe in the control lane at 0 mM BrxR. Dissociation coefficients (K_d) were calculated from saturation plots using non-linear regression. Data shown are mean values from triplicate experiments and are plotted with standard error of mean.

Analytical gel filtration

A Superdex 200 Increase (S200i) GL 5/150 (Cytiva) was connected to an ÄKTA Pure system (Cytiva) and equilibrated by running through 2 column volumes of filtered MQ water and 5 column volumes of analytical SEC buffer (20 mM Tris–HCl pH 7.9 and 150 mM NaCl) at 0.175

ml/min. It was then calibrated using standard calibration kits (Cytiva). Calibration curves were used to calculate the oligomeric state of untagged BrxR, His$_6$-BrxR and His$_6$-BrxR-R17A according to their elution volume. Fifty μl samples contained 5 μl of 1000 nM untagged BrxR, His$_6$-BrxR or His$_6$-BrxR-R17A, 5 μl of 5× FPLC sample buffer (100 mM Tris–HCl pH 7.9, 750 mM KCl, 20% (w/v) glycerol) and were made up with water. Samples were loaded onto to the S200i via Hamiliton syringe into a 10 μl loop. Samples were injected onto the S200i and two column volumes of analytical SEC buffer were used to elute BrxR proteins at a flow rate of 0.175 ml/min.

Protein crystallization and structure determination

BrxR was concentrated to 10 mg/ml and crystallization trials were set using a Mosquito Xtal3 robot (Labtech) with commercial screens (Molecular Dimensions). Drops were set at both 1:1 and 2:1 (protein : precipitant) ratios at 18°C and crystals appeared overnight in Pact Premier F8 (0.2 M Na$_2$SO$_4$, 0.1 M Bis–Tris propane pH 7.5 and 20% w/v PEG 3350). Crystals were reproduced manually in 2 μl drops and harvested in nylon cryoloops. Crystals were soaked in Cryo solution (20 mM Tris–HCl pH 7.9, 150 mM NaCl, 2.5 mM DTT and 80% glycerol) and stored in liquid nitrogen. Diffraction data were collected on I04 at Diamond Light Source (DLS). Four datasets collected at 0.9795 Å were merged to produce a single dataset using the DIALS pipeline in iSpyB (DLS). Data scaling was performed using AIMLESS (42). Phases were obtained by molecular replacement using an AlphaFold (43) model of BrxR in PHASER (PHENIX) (44), to produce a starting model which was then further built using BUCCANEER (45). Iterative refinement was performed using PHENIX and manually edited in COOT (46). Structure quality was assessed using PHENIX, COOT and the wwPDB validation server, and BrxR was solved to 2.15 Å. Structural figures were produced in Pymol (Schrödinger).

Comparative genomic analyses

The protein sequences and features of 3,828 reference and representative prokaryote sequences of 'complete' or 'chromosome' quality were downloaded from RefSeq using ncbi-genome-download v0.2.9 (https://github.com/kblin/ncbi-genome-download), in September 2021. A BLAST database of 13,499,153 proteins was constructed, and the protein sequence of BrxR (pEFER_0020) queried against the database using BLASTP with default settings. All homologues with E-value < 1e^{-5} were identified. Marker genes were queried against the same database, using BLASTP at default settings, and all homologues with E-value < 1e^{-5} identified. A second query with a less stringent threshold of E-value < 1e^{-3} was also carried out. The location, proximity, and orientation to identified BrxR homologues was then determined (in-house R script). All BLAST analyses were carried out on a CLIMB server (47). Inverted repeats were detected by extracting the DNA sequence up to 200 bp upstream of each *brxR* homologue, and searching these sequences using EMBOSS 'Palindrome' (v6.6.0.0) with a gap of <8 bp, a minimum repeat length of 20 bp and a maximum of two mismatches.

A representative phylogeny of the 3,828 genomes was downloaded from the NCBI Common Taxonomy Tree resource. The ete3 toolkit (48) provided taxonomic information for each genome. Trees were visualized in R using the ggtree package (49). UpSet plots were produced using the UpSetR R package (50).

RESULTS

The pEFER phage defence island is regulated by BrxR

We previously characterized a phage defence island from *E. fergusonii* ATCC 35469 (37). The island is carried by pEFER, a multidrug-resistant, 55.15 kb, plasmid. By subcloning the 18 kb defence island, we demonstrated that the island provides phage defence via the complementary BREX system (17) and a GmrSD-family type IV restriction enzyme (35), named BrxU (37). We predicted that the third open reading frame of the ten-gene island encoded a helix-turn-helix (HTH) domain, using PHYRE 2.0 (51). We hypothesized that this protein bound DNA to act as a transcriptional regulator of the defence island, and named it BrxR (Figure 1A). Bioinformatic analyses predicted a promoter upstream of *brxR* (52). Part of the transcriptional control of other BREX systems is mediated by promoters upstream of *brxA* and *pglZ* (17). We hypothesized that an additional promoter would lie upstream of *brxS* and *brxT*, to permit independent expression of the accessory genes we found to be necessary for BREX-mediated host genome methylation (37) (Figure 1B).

To investigate the function of the hypothetical promoters, and to determine the ability of BrxR to regulate gene expression, regions of the pEFER defence island denoted R1-12 were cloned into pRW50 (40), which encodes a promoterless *lacZ* reporter gene (Figure 1B). Gene *brxR* was also cloned into pBAD30 (39) to permit L-arabinose-inducible expression of His$_6$-BrxR, yielding pBAD30-*his$_6$-brxR*. *E. coli* DH5α was co-transformed with either pRW50 vector control or reporter plasmids, and pBAD30-*his$_6$-brxR*. These dual plasmid-carrying strains containing both a pRW50 reporter and pBAD30-*his$_6$-brxR* were grown either in the presence of D-glucose (D-glu), to repress *his$_6$-brxR* expression, or L-arabinose (L-ara), to induce *his$_6$-brxR* expression, and the resulting levels of β-galactosidase activity were determined (Figure 1C).

Of the four putative promoter regions, strong expression was observed from a promoter upstream of *brxR*, (P$_{brxR}$), with weaker expression being observed from upstream of *brxA* (P$_{brxA}$). Neither regions upstream of *brxS* nor *pglZ* showed measurable levels of transcriptional activity (Figure 1C). The induction of *his$_6$-brxR* reduced the expression from P$_{brxR}$ and P$_{brxA}$ (Figure 1C). Using pRW50-R7, we then confirmed that repression was due to expression of *his$_6$-brxR* when compared to an empty pBAD30 vector control (Figure 1D). Finally, to confirm that His$_6$-BrxR-mediated repression of transcription was specific to the tested DNA regions, rather than reflecting a global activity of His$_6$-BrxR, we tested whether His$_6$-BrxR could repress expression from pRW50-based reporter plasmids carrying other promoters (41) (Figure 1E). His$_6$-BrxR-mediated repression only occurred for the pEFER-derived promoter carried by pRW50-R7 (Figure 1E). Collectively, these data
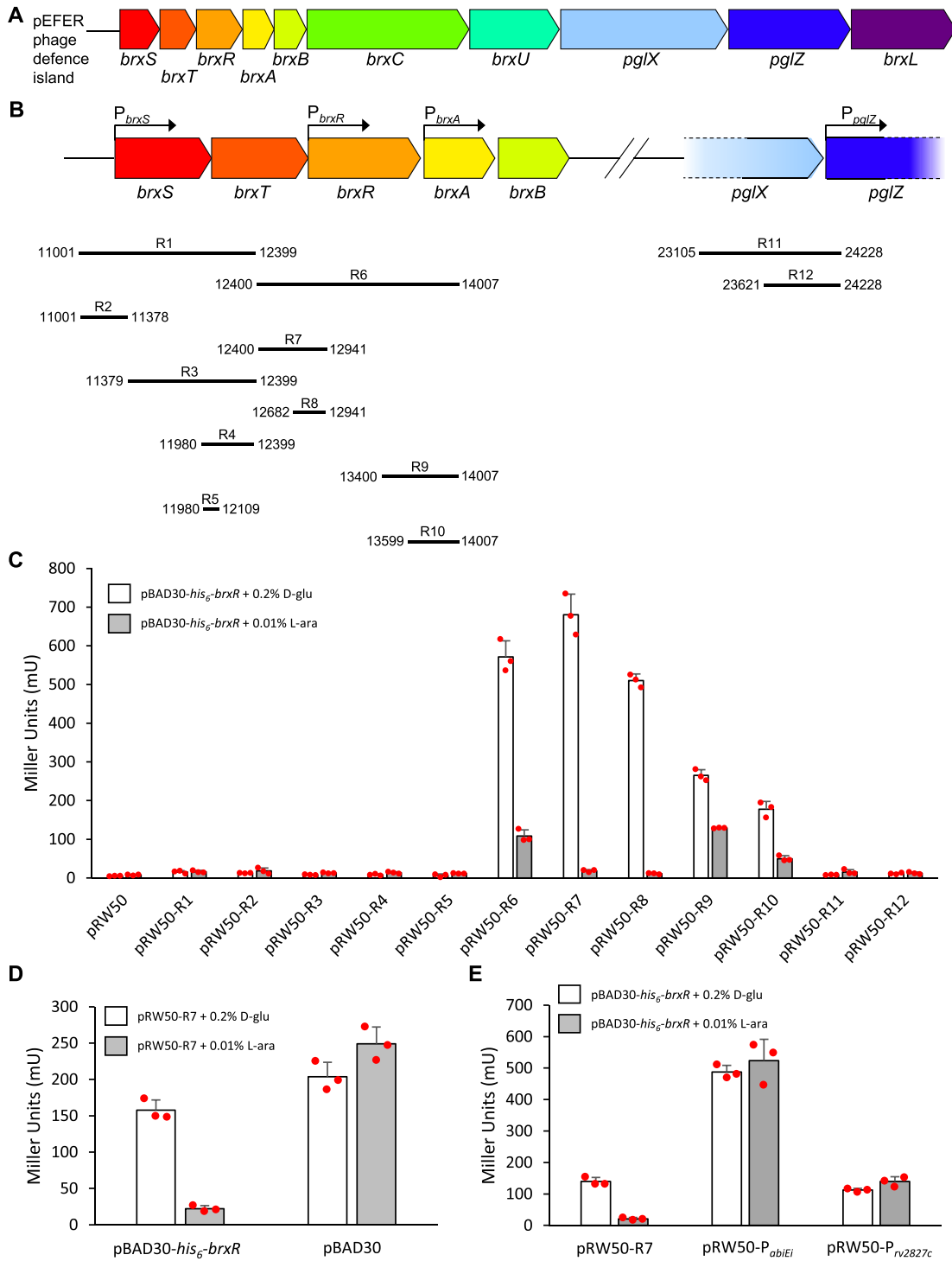
Figure 1. The pEFER phage defence island is regulated by BrxR at the transcriptional level. (**A**) Linear representation of the phage defence island of pEFER. (**B**) Transcriptional organization of the pEFER phage defence island, showing putative promoters P_{brxS}, P_{brxR}, P_{brxA}, and P_{pglZ}, 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} (in bold within (**B**)), 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 triplicate, and error bars represent standard deviation of the mean.

indicate that His$_6$-BrxR is a transcriptional regulator of the pEFER phage defence island that negatively regulates expression.

We additionally tested whether the pBAD30-*his$_6$-brxR* plasmid provided any protection from phages that were previously shown to be susceptible to the pEFER defence island (37) (Supplementary Figure S1). His$_6$-BrxR alone had no impact on the ability of the tested phages to form plaques confirming His$_6$-BrxR to be a regulator of, but not a participant within, phage defence (Supplementary Figure S1). We have previously sub-cloned the defence island from pEFER, generating plasmid pBrxXL that demonstrated complementary defence through BREX and BrxU (37). Next, we aimed to test the impact of ablating *brxR* expression in the context of the pBrxXL plasmid. However, the putative *brxR* knockout transformants obtained by either golden gate assembly (53), or Gibson assembly (54), all contained extensive mutations in other parts of the defence island. Furthermore, when commissioned, the *brxR* knockout plasmid could not be generated commercially. Collectively, our findings imply that the repression provided by BrxR in the context of the pEFER defence island may both regulate phage defence and limit inherent toxicity associated with uncontrolled expression of the island.

BrxR binds inverted DNA repeats

Our previously studied HTH transcriptional regulators were shown to bind inverted DNA repeats (41,55). We examined regions R7 and R9 (Figure 1B) and found an 11 bp imperfect inverted repeat (containing a single base difference at the second position), with a 5 bp spacer between repeats, located between P$_{brxR}$ and *brxR* at positions 12,820–12,846 bp (Figure 2A).

We tested the ability of His$_6$-BrxR to bind the inverted repeats downstream of P$_{brxR}$ by electrophoretic mobility shift assay (EMSA). We used a labelled probe containing 70 bp of pEFER (12,801–12,870 bp), which included inverted repeat 1 (IR1), inverted repeat 2 (IR2), and a transcriptional −10 box (predicted by BPROM (52)) upstream of IR1 (Figure 2A). For context, the start codon for *brxR* is 51 bp further downstream of the probe, at 12,921 bp. His$_6$-BrxR bound this DNA probe in a concentration-dependent manner (Figure 2B). The specific, S, control that contained a 20-fold excess of unlabelled probe, and the non-specific, NS, control, that contained a 20-fold excess of unlabelled probe from an unrelated *Mycobacterium tuberculosis rv2827c*-derived promoter (41), confirmed that the His$_6$-BrxR-DNA interaction was DNA-sequence specific (Figure 2B). His$_6$-BrxR-DNA binding generated a single shift of the labelled probe (Figure 2B), implying a single binding event. The presence of the two inverted repeats suggests that His$_6$-BrxR likely forms a stable dimer in solution that binds both IR1 and IR2 simultaneously. Replacing IR1 with a polyC tract still yielded a single binding event to IR2, albeit requiring greater concentrations of His$_6$-BrxR (Figure 2C). Similarly, replacing IR2 with a polyC tract had the same effect (Figure 2D). Replacing both IR1 and IR2 with polyC tracts prevented His$_6$-BrxR binding, unless at such high concentrations to allow non-specific DNA interactions (Figure 2E).

Quantification of complex formation in comparison to unbound probe generated K_d values for each binding event (Figure 2F–I). His$_6$-BrxR bound to the WT IR1-IR2 probe most tightly (K_d of 13.0 ± 2.7 nM), followed by IR1c-IR2 (K_d of 24.0 ± 6.5 nM), then IR1-IR2c (K_d 85.5 ± 18.7 nM) (Figure 2F–I), suggesting that the different base in IR1 reduced the affinity of His$_6$-BrxR binding.

### There are two BrxR binding sites encoded within the pEFER phage defence island

Next, we chose to investigate whether one inverted repeat would be sufficient to provide transcriptional repression in our LacZ reporter assays. Construct pRW50-R7-IR1c-IR2 was generated, containing a polyC tract in place of IR1. This construct was still transcriptionally active, and could be repressed in the presence of His$_6$-BrxR (Figure 3A). These data demonstrate that BrxR is a transcriptional regulator that binds either inverted repeats, or a single cognate binding sequence, to negatively regulate expression of phage defence genes.

Based on the veracity of BrxR binding to IR2, we searched regions R9 and R10 (Figure 1B) for sequences related to IR2, in an attempt to explain the observed BrxR-dependent repression of transcription from these regions (Figure 1C). To our surprise, a near-match was found that made up the equivalent IR2 latter repeat of a pair of 11 bp inverted repeats containing differences at five positions, with a 5 bp spacer between repeats (Figure 3B). Our earlier set of inverted repeats (Figure 2) was re-named R-BOX1, and this second set of inverted repeats, from 13,612–13,638 bp, was named R-BOX2 (Figure 3B). Aligning R-BOX1 and R-BOX2 it is clear that the IR1 repeats poorly align, whereas the IR2 repeats have a single difference, and there is also a single difference in the 5 bp spacer (Figure 3C). We hypothesize that R-BOX2 IR2 is sufficiently conserved with R-BOX1 IR2 (as tested in Figure 3A), to allow repression of P$_{brxA}$.

BrxR represents a family of multi-domain dimeric transcriptional regulators

Untagged BrxR was expressed and purified as described (Materials and Methods). Monomeric, untagged, BrxR has a predicted weight of 33.73 kDa. The size of BrxR in solution was determined by analytical size exclusion chromatography (SEC) (Figure 4A). The elution volume, in comparison to calibration controls, indicated that BrxR forms a dimer in solution (Figure 4A), which was consistent with the observation of single binding events by EMSA, independent of there being one or two IRs (Figure 2).

We solved the structure of BrxR by X-ray crystallography, to 2.15 Å (Figure 4, Table 1). The asymmetric unit contained four BrxR dimers, supporting our previous SEC data that indicated BrxR is a dimer in solution. Each BrxR protomer consists of three domains (Figure 4B). BrxR comprises an N-terminal winged-HTH domain (residues 1–94) (56), followed by a WYL-domain (so called due to a previous analysis of conserved amino acids), which has been implicated as a potential ligand-binding domain with a role in phage defence (residues 120–190) (23), and a C-terminal dimerization domain (residues 200–295) (Figure 4B).
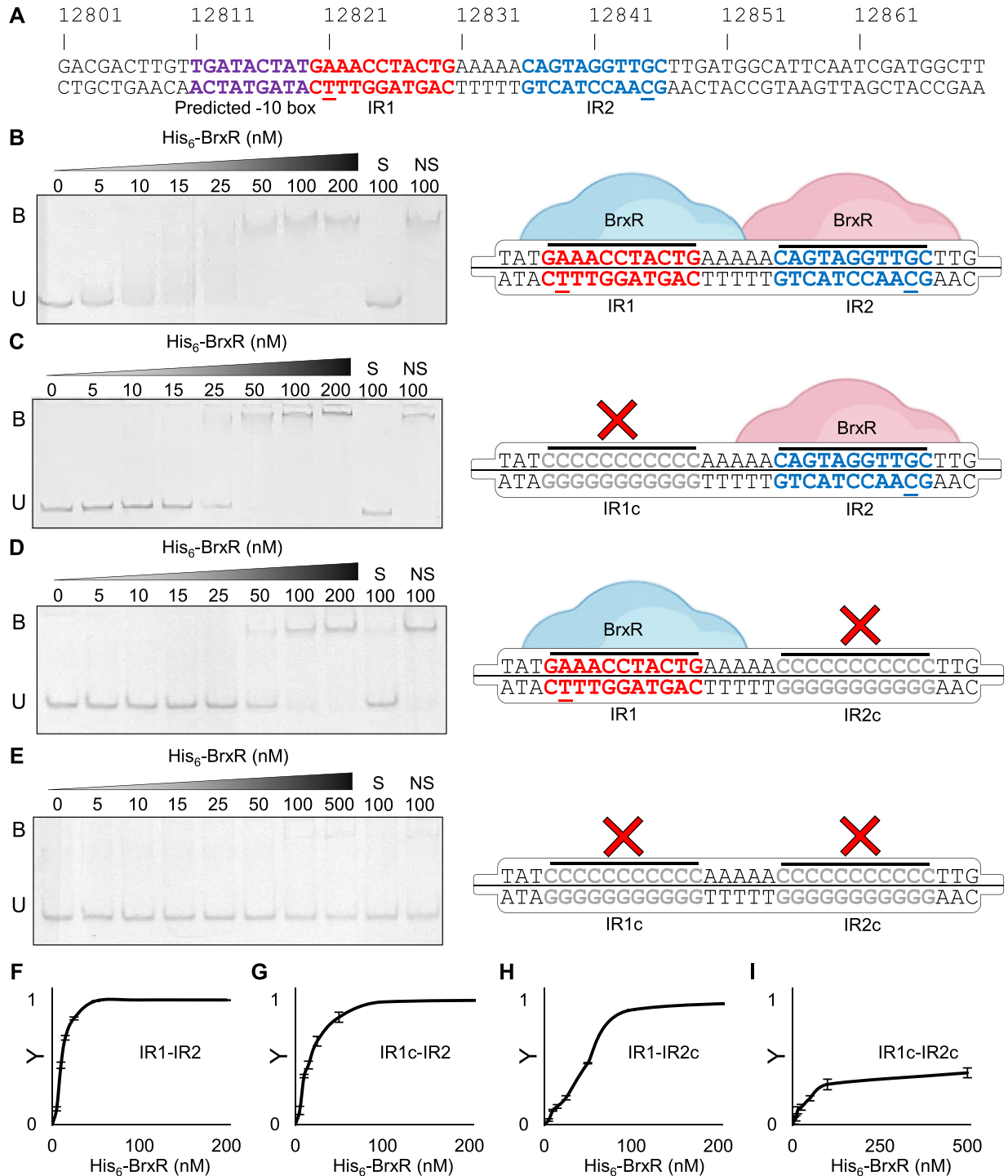
Figure 2. His$_6$-BrxR binds DNA inverted repeats *in vitro*. (**A**) Sequence of pEFER 12,801–12,870 bp included in EMSA probe. The predicted transcriptional −10 is indicated in purple, and inverted repeats 1 and (IR1, IR2), are shown in red and blue, respectively, with the 1 bp difference in sequence underlined. (**B–E**), Electrophoretic mobility shift assays (EMSAs) of titrated His$_6$-BrxR protein with dsDNA probes spanning pEFER nucleotide locations 12,801–12,870. Target probes were amplified to incorporate fluorescein and contain either the native WT promoter region IR1-IR2 (**B**) or substituted regions where either IR1 (**C**), IR2 (**D**) or both IRs (**E**) were replaced by polycytosine residues. Protein concentration is shown above each lane together with binding events (**B** – bound, **U** – unbound). Control lanes correspond to samples prepared with a 20-fold excess of unlabelled specific DNA (S) or non-specific DNA (NS), respectively. Experiments were run in triplicate and a representative gel from each experiment is shown. Each EMSA is accompanied with a schematic of the binding capacity of His$_6$-BrxR relative to the presence/absence of its target motif. IR mutations are shown in grey. Probe sequence diagrams are truncated to show only the IR regions of dsDNA probes. (**F–I**), Saturation curves were plotted using EMSA band intensity of unbound probe to determine Y values. Y values were calculated using $Y = 1 - (I_T/I_C)$, where I_T is the band intensity of the unbound probe in test lanes, and I_C is the band intensity probe in the control lane at 0 nM His$_6$-BrxR. Points plotted are mean values from triplicate data and error bars correspond to standard error of the mean. (**F**) Native promoter (**G–I**), Mutated promoter regions with polycytosine substitution of IR1 (**G**), IR2 (**H**) or both (**I**).
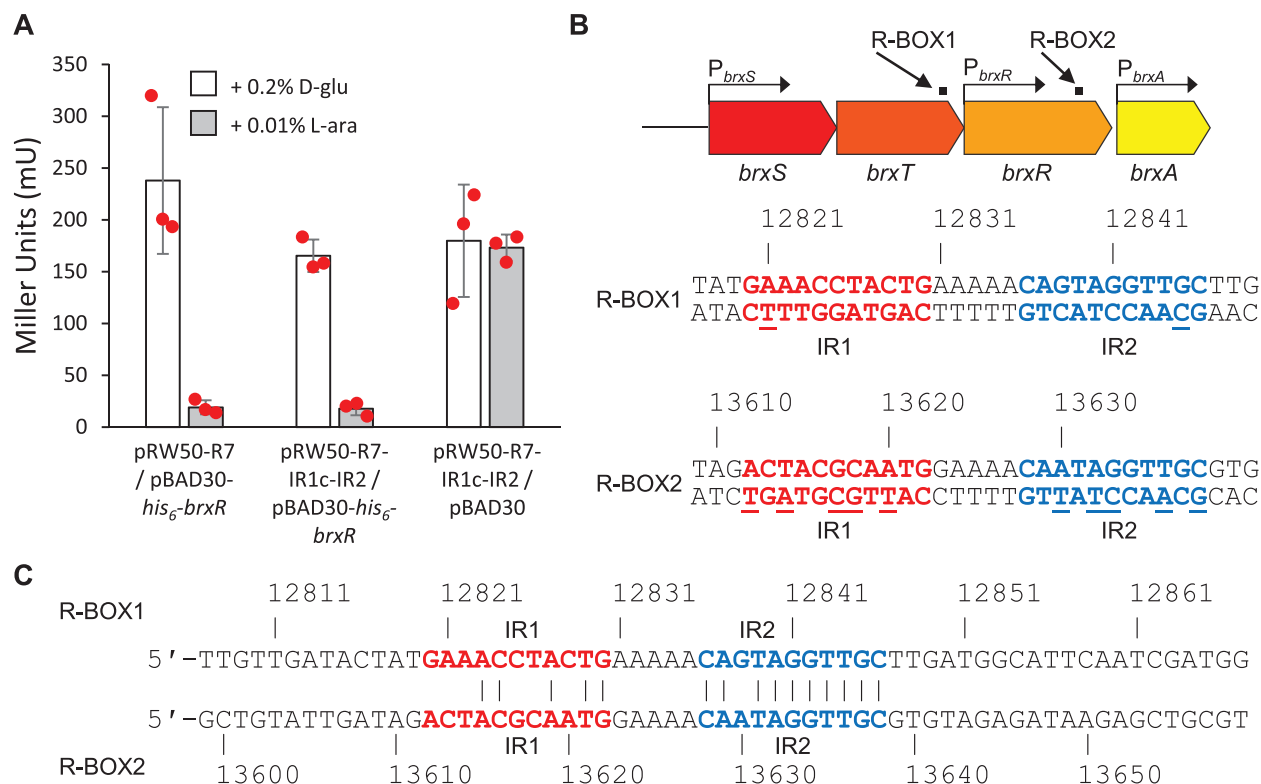
Figure 3. Plasmid pEFER encodes a second, putative, BrxR binding site. (**A**) LacZ-reporter assays using pRW50-R7 and pRW50-R7-IR1c-IR2 with and without induction of pBAD30-*his₆-brxR* or a pBAD30 vector control. Data are shown in triplicate, and error bars represent standard deviation of the mean. (**B**) Positions of R-BOX1 and R-BOX2 within the pEFER defence island, and sequences of the inverted repeats. Underlined bases indicate differences between inverted repeats within an R-BOX. (**C**) Alignment of R-BOX1 and R-BOX2. Conserved residues are only indicated if they are within the IR1 and IR2 repeats.

The wHTH domains are spaced ∼25 Å apart in the BrxR dimer, indicating additional movement is required to optimize the positions for interaction with the major grooves of target DNA (Figure 4C). The wHTH domains are exchanged between the protomers, such that upon exiting the wHTH the protein fold crosses to the other side of the dimer, to the WYL-domain. This cross-over begins with an α-helix aligned in parallel with that of the opposing protomer (residues 84–94), before entering a long loop section (residues 95–119) that circles round either end of both the central helices, interacting with all three domains of the opposing protomer around the circumference, before forming the WYL-domain. The first α-helix of each WYL also lines-up in parallel either side of the two central cross-over helices, to form a row of four parallel helices, alternating between protomers. The WYL-domains do not appear to directly interact, but the C-terminal dimerization domains extend across like two left hands shaking, interacting through the opposing C-terminal domain through the palms, and with the opposing WYL-domain through an α-helix at the utmost tip of the protomeric dimerization domain (residues 202–232) (Figure 4C). Two sulphate molecules are bound within each WYL-domain (Figure 4C). As the crystals were formed in conditions containing 0.2 M sodium sulphate, it is expected that the abundance of sulphate in the crystallization condition allowed these ions to be resolved in the structure. Nevertheless, the position of the two sulphates cor-

responds to a solvent-exposed basic patch formed by each BrxR protomer (Figure 4D).

Protein sequences homologous to BrxR were selected with Consurf (57), and used for multiple sequence alignment and subsequent calculation of residue conservation. The conservation output was then mapped onto the BrxR surface (Supplementary Figure S2A). Interestingly, conservation showed a similar distribution to the electrostatic potential (Figure 4D), with greatest conservation in the DNA-contacting helices of the wHTH domain, the sulphate-binding residues of the WYL-domain, the central line of interfacing α-helices, and the protomer interface residues of the C-terminal dimerization domain (Supplementary Figure S3A). The DALI server (58) was used to search the PDB for structural homologues of BrxR (Supplementary Table S2). The highest scoring hit reached a Z-score of only 11.3, indicating that there was no clear match to BrxR within the PDB. Of the obtained hits, each was shown to overlay either with the wHTH domain (a common DNA-binding motif) (56), or the WYL-domain itself (33). No hits matched the BrxR C-terminal dimerization domain. We concluded that BrxR is the first solved structure of a new family of WYL-domain containing transcriptional regulators.

One other structure of a BrxR homologue that appears to regulate a phage defence island from *Acinetobacter* sp. NEB394 (BrxR*Acin*) has been solved by Luyten et al., in a study co-submitted with this article (59). We exchanged
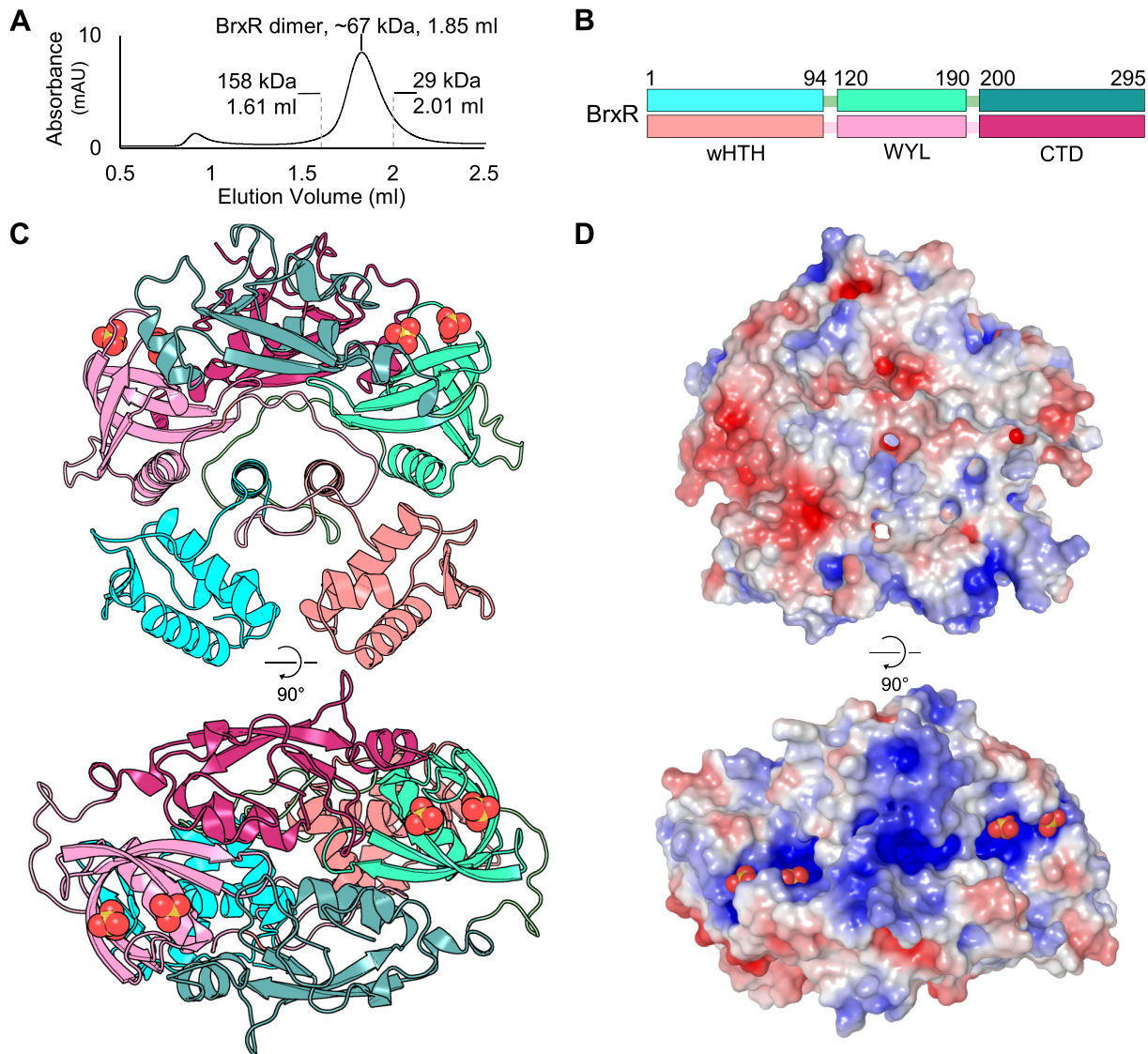
Figure 4. BrxR forms a dimer and exhibits significant surface electropositivity. (**A**) Size exclusion chromatography of untagged BrxR resolved via a Superdex 200 increase GL 5/150 gel filtration column. BrxR elutes at an elution volume of 1.85 ml, corresponding to a mass twice its M_r, indicating dimer formation. No additional peak is observed for residual monomers. Calibration standards are indicated. (**B**) Organization of the three domains within BrxR, separated by linker regions. Each protomer is coloured in either shades of cyan or pink, with domains indicated by the amino acid residue numbers shown. (**C**) Cartoon overview of the BrxR dimer coloured corresponding to (A), presented in orthogonal views (PDB: 7QFZ). Sulphate ions are represented as yellow and red spheres. (**D**) Electrostatic representation of surface BrxR charges in orthogonal views. The blue electropositive patches around the helices of the wHTH domains, and within the WYL domain surrounding the bound sulphate ions, are shown.

BrxR homologue structural models for comparison. A sequence-independent superposition of the two structures generated a Root Square Mean Deviation (RMSD) of 3.37–3.63 Å, depending on which of our modelled dimers of BrxR from *E. fergusonii* (from now on referred to as BrxR$_{Efer}$) was used (Supplementary Figure S2B). Although the relatively low RMSD value suggests poor overall structural homology, both homologues have a similar arrangement of the same three domains, with variations in the relative positioning of each domain and secondary structure elements (Supplementary Figure S2B). For instance, whilst the wHTH domains remain ~25 Å apart in the BrxR$_{Acin}$ structure, they are tilted further out along the short axis

of the dimer compared to BrxR$_{Efer}$. Furthermore, the central parallel helices within BrxR$_{Efer}$ are tilted in BrxR$_{Acin}$, and the loop extending around and towards the protomeric WYL-domain of BrxR$_{Acin}$ donates a β-strand to form an extended β-sheet with the opposing WYL domain as it passes.

Luyten et al. (59) also obtained a structure of BrxR$_{Acin}$ in complex with DNA, having identified a similar set of target DNA inverted repeats, each of 10 bp (single difference at position 5) and separated by 5 bp. This DNA-bound structure shows that BrxR$_{Acin}$ bends the target DNA, which allows interactions with the major grooves despite the spacing of only ~25 Å between wHTH domains. When superposed

Table 1. Crystallographic data collection and refinement statistics

	BrxR Native
PDB ID code	7QFZ
Number of crystals	1
Beamline	Diamond I04
Wavelength, Å	0.9795
Resolution range, Å	82.35–2.15 (2.15–2.19)
Space group	$P\,4_1\,22$
Unit cell	
$a\,b\,c$ (Å)	131.13 131.13 358.369
$\alpha\,\beta\,\gamma$ (°)	90.000 90.000 90.000
Total reflections	324,782
Unique reflections	169,928
Multiplicity	1.9
Completeness (%)	100
Mean I/sigma(I)	9.5
R_{merge}	0.04
R_{meas}	0.057
$CC_{1/2}$	0.998
R_{work}	0.216
R_{free}	0.241
No. of non-hydrogen	
atoms	19,401
Macromolecules	80
Ligands	16
Solvent	884
Protein Residues	2,276
RMSD (bonds, Å)	0.008
RMSD (angles, °)	1.16
Ramachandran favoured (%)	96.85
Ramachandran allowed (%)	3.15
Ramachandran outliers (%)	0
Average B-factor	59
Macromolecules	60.05
Ligands	63.06
Solvent	53.13

against the BrxR$_{Acin}$-DNA structure, the recognition helices of the wHTH domains from BrxR$_{Efer}$ also fit into the major grooves (Supplementary Figure S2C), but the winged β-sheet clearly clashes with the DNA backbone (Supplementary Figure S2C, *inset*). This suggests that upon BrxR$_{Efer}$ binding DNA, a conformational shift will be needed.

Comparative modelling of BrxR$_{Efer}$ against the BrxR$_{Acin}$-DNA structure identified BrxR$_{Efer}$ R17 as a residue predicted to be important for DNA binding (Figure 5A). This residue is equivalent to BrxR$_{Acin}$ R11, which makes a bidentate interaction with the DNA phosphate backbone (Figure 5A). A mutant pBAD30-his_6-$brxR$-R17A construct was generated in order to test the ability to repress transcription in cells. Whilst the pBAD30-his_6-$brxR$ construct provided strong repression of pRW50-R7, pBAD30-his_6-$brxR$-R17A had no impact, comparable to a vector-only pBAD30 control, when induced (Figure 5B). Monomeric His$_6$-BrxR has a predicted weight of 34.61 kDa, whilst monomeric His$_6$-BrxR-R17A has a predicted weight of 34.52 kDa. His$_6$-BrxR-R17A was expressed and purified, then examined by analytical SEC alongside His$_6$-BrxR (Figure 5C). As found with untagged BrxR (Figure 4A), both His$_6$-BrxR and His$_6$-BrxR-R17A eluted as dimers (Figure 5C). This elution profile suggests that His$_6$-BrxR-R17A is correctly folded, and that the observed phenotypes are due to loss of activity through mutation, rather than misfolding (Figure 5C). By EMSA, His$_6$-BrxR-R17A was not able to bind the WT IR1-IR2 probe, unless at such high concentrations as to

allow non-specific DNA interactions (Figure 5D). These data show that R17 is essential for both DNA binding and transcriptional repression by BrxR$_{Efer}$ in cells. This builds a stronger model for DNA interactions by BrxR homologues.

WYL-domain of BrxR as a potential ligand sensor

WYL-domains have been proposed as ligand-binding domains that could act as sensors of phage infection to regulate phage defence systems (23). The fold of the WYL-domain from BrxR$_{Efer}$ corresponds exactly with the expected features of the superfold Sm/SH3 family, itself a subset of the larger and pervasive small β-barrel (SBB) protein domain urfold family (24). The BrxR$_{Efer}$ WYL-domain folds as an N-terminal α-helix, followed by five β-sheets (Figure 6A). The RT loop links sheets β1-β2 (numbered within this domain, not across the entire BrxR$_{Efer}$ protein), the n-Src loop links sheets β2-β3, the distal loop links sheets β3-β4, and the short 3$_{10}$ helix links sheets β4–β5 (Figure 6A). The top DALI hit (Supplementary Table S2) was the WYL-domain from PafBC (PDB: 6SJ9) (33). The WYL-domains of BrxR$_{Efer}$ and PafBC superpose with an RMSD of 0.662 Å, which is an overall good match, but there are distinct movements in the RT loop used by BrxR$_{Efer}$ to bind sulphates (Figure 6B). We noted that the overall arrangement of domains differs between PafBC and BrxR$_{Efer}$, and in contrast to the WYL-domains, the C-terminal domains superpose poorly with an RMSD of 4.38 Å.

Sm/SH3 domains are known to bind diverse polymeric ligands such as DNA, RNA, oligosaccharides, proteins and peptides (24). Detailed examination of the sulphate binding site in BrxR$_{Efer}$ shows an abundance of polar groups that have captured the sulphate ions and could theoretically recognize other small molecule ligands (Figure 6C). These residues are found within the core β-strands but also the conserved loops, such as S143, S145 and S146 on the RT loop; K173 and H174 on the distal loop; and R184 on the 3$_{10}$ α-helix (Figure 6C). We propose that ligand-binding at the WYL-domain basic patch could induce conformational changes to BrxR$_{Efer}$ that relieve transcriptional repression.

BrxR-family homologues are predominantly found in Proteobacteria

We wanted to investigate the extent of this newly identified BrxR family. Homologues were identified through bioinformatic searches of a protein database constructed from representative RefSeq genomes (see Materials and Methods), using BrxR$_{Efer}$ as a search sequence with a conservative threshold of E-value $< 1e^{-5}$. This threshold was chosen to exclude false positives associated with the prevalence of both wHTH and WYL-domains, the numbers of regulatory proteins in general, and the relative size of BrxR$_{Efer}$. Our search identified 347 homologues within 281 genomes, including 59 genomes (57 proteobacteria, 1 firmicute, 1 planctomycete) encoding more than one BrxR-family protein. This corresponds to BrxR-family homologues in 7.79% of the 3828 genomes in our representative dataset. All homologues were found in bacterial genomes, with no homologues identified in the 222 archaeal genomes.
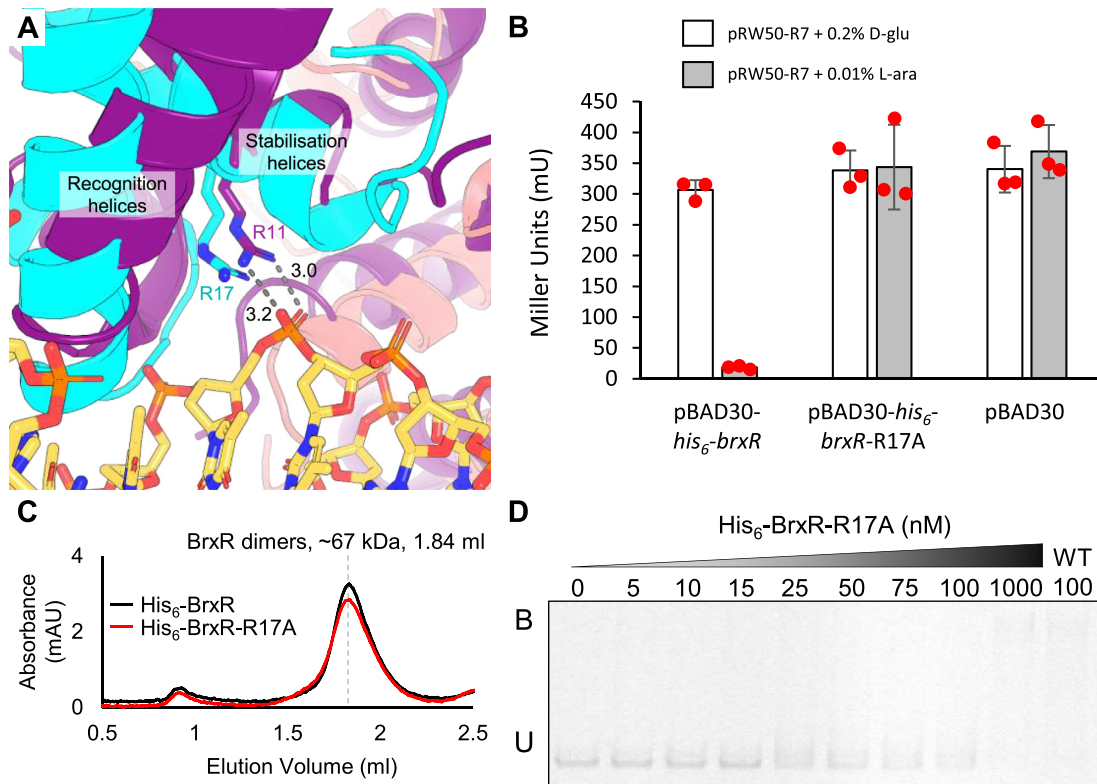
Figure 5. BrxR$_{Efer}$ R17 is essential for transcriptional repression and DNA binding. (**A**) Close-up alignment of wHTH domains from the BrxR$_{Efer}$ *apo* structure (cyan and salmon cartoons, PDB: 7QFZ), and BrxR$_{Acin}$-DNA structure (purple cartoon and DNA as sticks, PDB: 7T8K). Stabilization and recognition helices roughly overlay, and there is co-localization of BrxR$_{Efer}$ R17 with BrxR$_{Acin}$ R11. BrxR$_{Acin}$ R11 makes bidentate hydrogen bonds with the DNA phosphate backbone. Distances shown are in angstroms. (**B**) LacZ-reporter assays using active pRW50-R7 construct with and without induction of His$_6$-BrxR$_{Efer}$ or His$_6$-BrxR$_{Efer}$-R17A from pBAD30. Data are shown in triplicate, and error bars represent standard deviation of the mean. (**C**) Size exclusion chromatography of His$_6$-BrxR$_{Efer}$ and His$_6$-BrxR$_{Efer}$-R17A resolved via a Superdex 200 increase GL 5/150 gel filtration column. Both proteins eluted at 1.84 ml, corresponding to a mass approximately twice their respective M$_r$, indicating correct folding into dimer formation. No additional peak was observed for residual monomers. (**D**) Electrophoretic mobility shift assay (EMSA) of titrated His$_6$-BrxR$_{Efer}$-R17A protein with WT dsDNA probe IR1-IR2 spanning pEFER nucleotide locations 12,801–12,870. Target probe was amplified to incorporate fluorescein and contains the native promoter region. Protein concentration is shown above each lane together with binding events (B – bound, U – unbound). Control lane of His$_6$-BrxR (WT) is included for comparison. Experiment was run in triplicate and a representative gel is shown.

We then considered the taxonomic distribution of the BrxR-family, and noted that 338/347 BrxR homologues were found throughout Proteobacteria (97.41% of total homologues), most commonly in *Pseudomonas* (24/347; 6.92%), *Shewanella* (18/347; 5.19%) and *Vibrio* (15/347; 4.32%) (Figure 7). Though widespread, no homologues were found within Deltaproteobacteria (Figure 7). BrxR homologues were found in 271 of 1,589 proteobacter genomes within the dataset (Figure 7). Hits were additionally found in firmicutes (6 homologues in 607 genomes), spirochaetes (1 homologue in 49 genomes), planctomycetes (1 homologue in 62 genomes) and verrucomicrobia (1 homologue in 110 genomes). Collectively, these data show that the BrxR-family is widespread amongst Proteobacteria.

BrxR-family homologues are associated with diverse phage defence systems and islands in bacteria

Having identified a list of 347 BrxR homologues, we wanted to know how many phage defence systems could potentially be regulated. We compiled a list of 110 reference protein sequences, comprised of key genes from diverse known phage defence systems and sub-types (Supplementary Table S3), which was used to identify phage defence homologues within our database, using a BLASTP threshold of *E*-value $< 1e^{-5}$. For each BrxR homologue, we tested for the presence of one or more phage defence homologues within 50 kb downstream of *brxR*, identifying 382 phage defence protein homologues. These 382 protein homologues were within 210 phage defence systems, and these were co-localized with 164 of 347 BrxR homologues (48.41%) (Supplementary Table S4). A less stringent threshold of *E*-value $< 1e^{-3}$ was also tested, but this did not increase the number of BrxR homologues associated with known phage defence systems. We also examined the 50 kb upstream of each BrxR homologue, identifying a further 77/347 BrxR homologues including 29/347 with phage defence systems both upstream and downstream (Supplementary Table S5). This equates to an additional 48/347 BrxR homologues that were co-localized with at least one phage defence system, taking the total of associated BrxR homologues to 212/347 (61.10%). As BrxR$_{Efer}$ controls phage defence systems downstream, we chose to be conservative and focussed only on the downstream matches for further analysis.
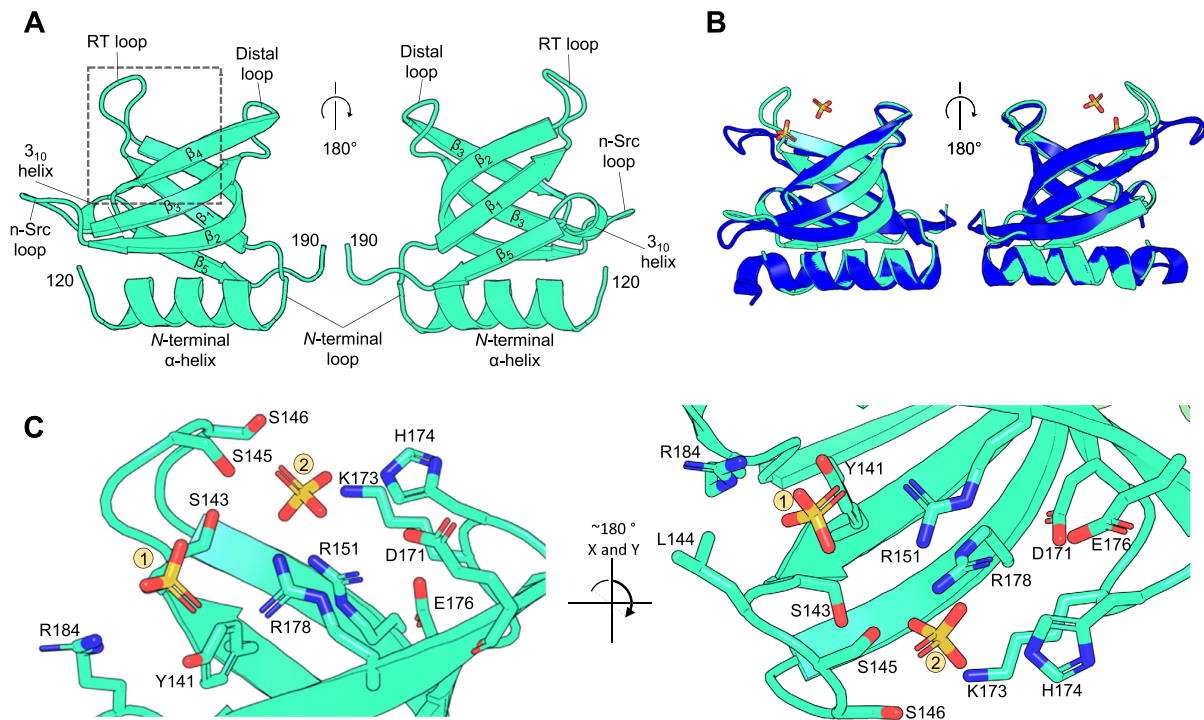
Figure 6. The BrxR WYL-domain shows ligand binding potential via extensive sidechain coordination. Orthogonal views are shown for each panel as indicated. (**A**) Close-up of the WYL domain of BrxR. Terminal residues are numbered, and secondary structural elements and loops for this domain are labelled. (**B**) Structural superposition of BrxR$_{Efer}$ with the WYL-domain of PafBC (RMSD 0.662 Å; PDB: 6SJ9) shows clear structural similarity. Differences are observed at the RT loop of BrxR$_{Efer}$, which has moved inwards to bind the two sulphate ions. BrxR$_{Efer}$ is shown in cyan and PafBC is shown in blue. (**C**) A close-up view of the dashed boxed area of (A) shows the hydrogen bond coordination of two sulphate ions bound within the WYL domain of BrxR$_{Efer}$. Interacting sidechains extend from the core β-strands and intervening loops. Nitrogen atoms are shown in blue and oxygen atoms in red. Sulphate ions are shown as yellow (sulphur) and red (oxygen) sticks.

Sorting the BrxR-associated phage defence systems by class showed BrxR homologues are predominantly co-localized with BREX systems (70/210 BrxR-associated phage defence systems, 33.33%; Figure 8A). Next, they are co-localized with type IV and type I restriction enzymes, 37/210 (17.61%) and 34/210 (16.19%), respectively. CRISPR-Cas systems were similarly well represented, comprising 21 of 210 BrxR-associated phage defence systems (10.00%). Whilst not all toxin-antitoxin system families have been shown to abort phage infections, there are multiple examples where toxin-antitoxin system types I-IV cause abortive infection (60–62). BrxR was predominantly co-localized with type II and IV toxin-antitoxin systems (Figure 8A). More recently defined phage defence systems such as Wadjet, Zorya, Thoeris (16), Pycsar (22) and CBASS (18) were also co-localized with BrxR homologues (Figure 8A). We visualized the BrxR-associated systems in relation to the host phylogeny (Supplementary Figure S3).

Not only did our analysis identify individual phage defence systems that were associated with a BrxR homologue, but we also found examples of multiple systems, clustered into phage defence islands, that were associated with a BrxR homologue (Figure 8B). The most common clustering was between BREX and type IV restriction, as seen for pEFER and previously noted to be the most common pairing of phage defence systems (14,15,37). The next most common clusters included type III and type IV restriction systems,

then BREX and type III restriction systems (Figure 8B). There were also individual examples of multiple forms of diverse clusters, including islands containing homologues from three different systems (Figure 8B). BrxR-associated phage defence systems were also further divided by sub-type of phage defence system (Supplementary Figure S4).

Finally, we took the 347 BrxR homologues and examined the upstream 200 bp of DNA sequence for inverted repeats that might indicate likely BrxR binding sites. This identified 330 inverted repeats with a minimum length of 20 bp, and a gap of <8 bp, which were associated with 193 BrxR homologues (Supplementary Table S6). As not all homologues had inverted repeats, they may rely on single site binding events (as was observed to be possible for BrxR$_{Efer}$, Figure 3A), more distant binding sites, or other means of regulation. Collectively, these data show how BrxR homologues contribute to the regulation of a wide range of phage defence islands that can be highly mosaic in phage defence system content, and within diverse hosts. Our findings confirm that BrxR is an archetypal member of a new family of transcriptional regulators involved in protecting bacteria from phages and mobile genetic elements.

DISCUSSION

Plasmid pEFER was known to encode a phage defence island encoding complementary BREX and type IV re-
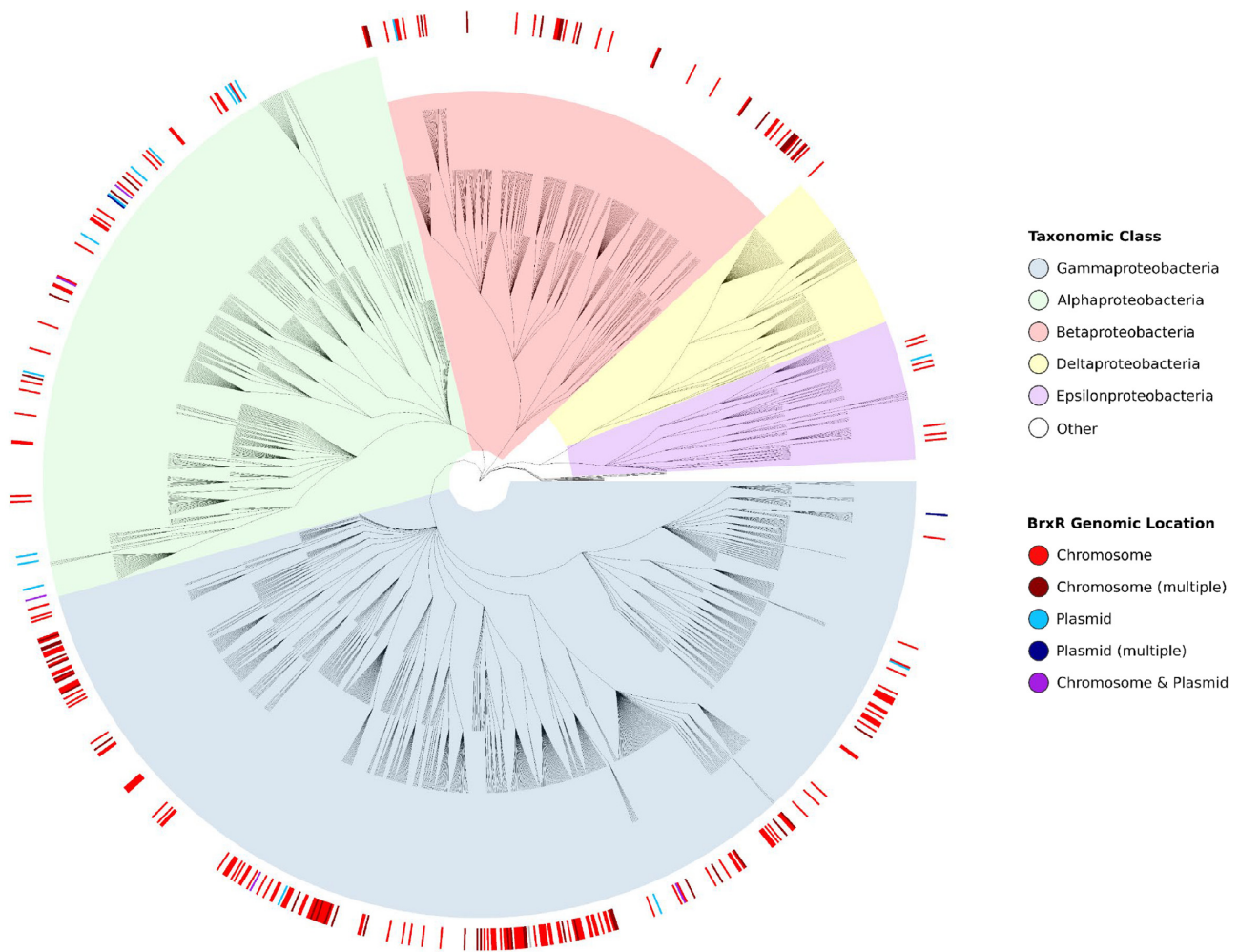
Figure 7. BrxR is widely distributed in the phylum proteobacter. Phylogenetic tree of 1,589 proteobacterial genomes, as downloaded from the NCBI Taxonomy resource and background highlighted according to taxonomic class. BrxR hits are indicated in the exterior circle as a heatmap, coloured to show whether *brxR* is located on the chromosome or a plasmid.

striction systems (37). Here, we have shown that this defence island is regulated by a WYL-domain containing protein, BrxR$_{Efer}$. These data corroborate similar findings co-published for other homologues from Gram-negative strains; BrxR$_{Acin}$, found upstream of a BREX system in *Acinetobacter* (59), and CapW, discovered within a CBASS system of *Stenotrophomonas maltophilia* (63).

BrxR$_{Efer}$ acts as a transcriptional repressor, blocking transcription from a promoter upstream of *brxR* that controls the canonical first BREX gene, *brxA* (Figure 1). Whilst the BREX loci from *Bacillus cereus* contained another promoter upstream of *pglZ* (17), we were unable to detect a promoter in the comparable region of pEFER (Figure 1). EMSA studies demonstrated that BrxR$_{Efer}$ bound as a stable dimer to inverted DNA repeats, termed R-BOX1, in a sequence-dependent manner (Figure 2). The R-BOX1 repeats were positioned immediately downstream of the predicted P$_{brxR}$ promoter sequence, and so repression is likely due to sterically blocking the RNA polymerase. Our data also indicate that a single repeat is sufficient for DNA binding and repression (Figure 3A), perhaps because BrxR$_{Efer}$

both exists in solution, and binds DNA, as a dimer. This suggests how BrxR-family homologues without obvious associated inverted repeats might mediate transcriptional control. Based on these data we also went on to find a second set of inverted repeats, R-BOX2, upstream of P$_{brxA}$ (Figure 3). We hypothesize that R-BOX2 allows BrxR to perform the transcriptional repression observed for regions R9 and R10 (Figure 1). Further searches of the pEFER sequence did not find any further likely R-BOX sequences.

It is clear that not all BREX loci require a BrxR homologue (17,64), so it is worth considering why transcriptional regulation of phage defence genes might be required. We suggest that some BREX homologues can be toxic, which could be exacerbated by the genomic context (chromosomal or plasmid-based) or the methylation status of the host. We have noted that PglX$_{Efer}$ product is toxic when overexpressed and our repeated inability to make a *brxR*$_{Efer}$ knockout mutant supports the hypothesis that the repression of BREX genes is required to reduce fitness costs to the host prior to phage infection. A further hypothesis could be that BrxR control provides temporal regu-
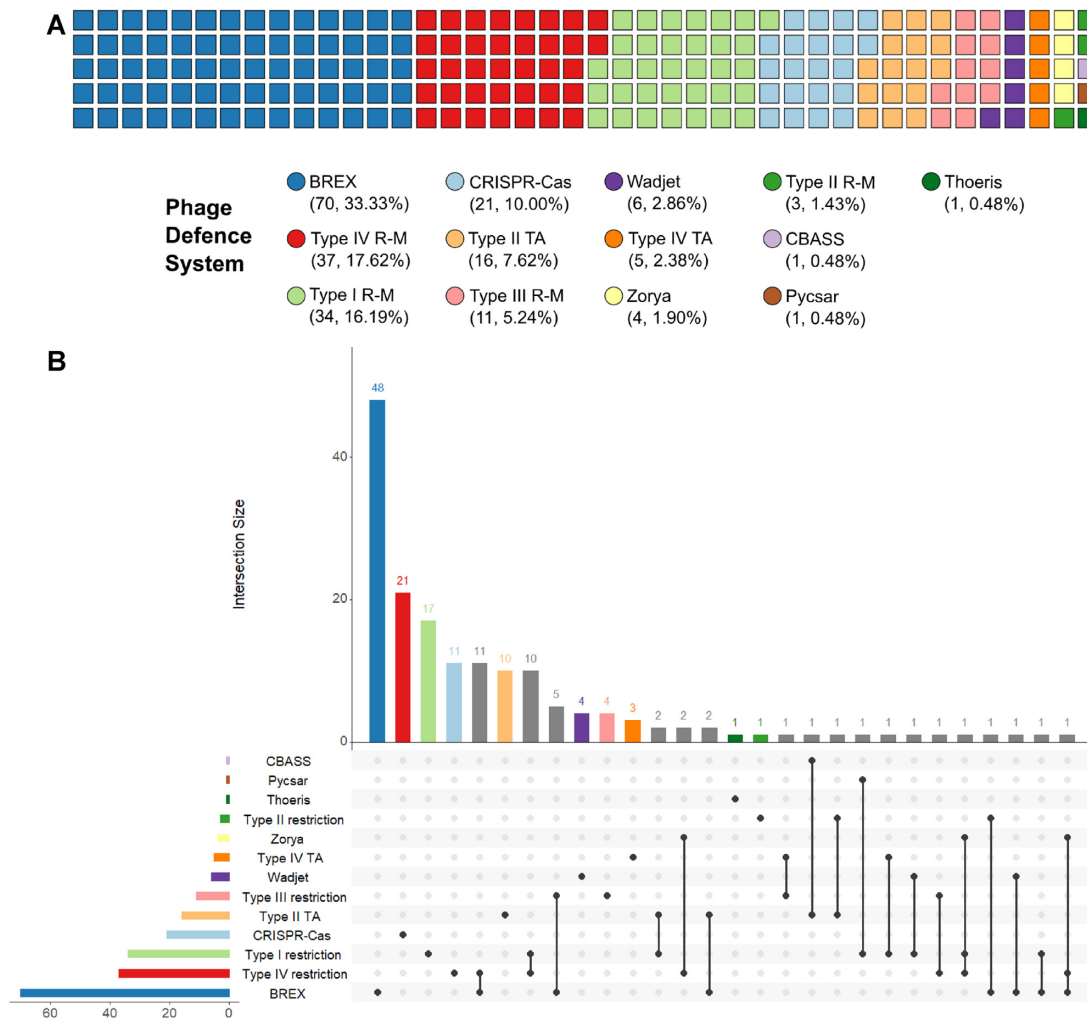
Figure 8. BrxR is associated with a variety of different phage defence systems. (**A**) Waffle chart showing the distribution of defence systems identified downstream of *brxR*, coloured to show the general class of system identified. (**B**) UpSet plot showing the co-occurrence of phage defence systems downstream of *brxR*. Set and intersection bars are coloured by single defence systems, with overlapping intersections in grey. The matrix indicates overlapping set intersections, with set size as the horizontal bar chart, and the vertical bar chart showing intersection size (the number of times any systems are found in combination).

lation of phage defence and this remains to be tested. It should also be considered that though not all BREX loci (or indeed other phage defence systems) have obvious BrxR-family homologues, this does not rule out transcriptional control by other, perhaps currently unknown, means of regulation.

BrxR-family homologues contain an N-terminal wHTH domain, a WYL-domain and a C-terminal dimerization domain (Figure 4), and DNA-binding and repression is dependent on key residues within the wHTH domain (Figure 5) (59). The structures of BrxR$_{Efer}$, BrxR$_{Acin}$ (59) and CapW (63), are the first for this family, but increasing numbers of WYL-domain proteins have recently been characterized. The HTH-WYL protein Sll7009 has previously been shown to negatively regulate a CRISPR locus in *Synechocystis* (28). In this case, however, BrxR$_{Efer}$ and Sll7009 share no significant sequence similarity. Further WYL-domain proteins hypothesized to be transcriptional regulators have also been identified through computational analyses of phage

defence islands associated with integrative conjugative elements (34). WYL-domain containing proteins can also act as transcriptional regulators in contexts other than phage defence. For instance the HTH-WYL-WCX protein PafBC, which has a very different overall domain arrangement to BrxR$_{Efer}$, is a transcriptional activator in response to DNA damage in mycobacteria (33). It is worth noting that Luyten et al. examined the C-terminal domain of BrxR$_{Acin}$ and observed the same core fold as the WYL C-terminal extension domain (WCX) of PafBC (59), implying BrxR and PafBC may well be distant homologues. As a further example of a WYL-domain transcriptional regulator, the much larger DriD protein (914 amino acids to the 295 amino acids of BrxR$_{Efer}$), contains HTH-WYL domains and is involved in upregulation of the DNA damage response in *C. crescentus* (29). WYL-domains can also play a role in regulating catalysis, with PIF1 helicase activity dependent on the WYL-domain (30), and Cas13d activity enhanced by the accessory WYL1 protein (31,32).

It has previously been predicted that WYL domains could function as regulatory domains, either as switches to alter the activity of enzymes, or for transcriptional regulation, as part of phage defence (23). To perform such biological roles, the WYL-domains likely bind ligands; in PIF1, the WYL-domain binds ssDNA (30), whereas in WYL1 the domain binds ssRNA (32). To respond to DNA damage, it has been postulated that the WYL-domains bind ssRNA, ssDNA, or some other nucleic acid molecule or secondary messenger (29,33). The reported promiscuity of WYL-domain ligand recognition will make it a challenge to identify the specific ligands experimentally. We speculate that a suitable candidate for recognizing a phage infection might be a cyclic 2′-3′ phosphate, a cyclic nucleotide as in the Pycsar system (22), or other nucleic acid polymers.

The structure of BrxR$_{Efer}$ showed sulphate ions bound within the WYL-domain (Figures 4 and 6). This highly conserved fold is known to bind a large range of ligands (23,24), and BrxR$_{Efer}$ has an abundance of functional groups located in a conserved basic, solvent-exposed patch at the top of the WYL-domain, which are predicted to recognize the target ligand (Figure 6). We propose that ligand-binding alters the conformational state of BrxR to release the bound DNA, and de-repress transcription of phage defence genes. Interestingly, the structures of Efer$_{Acin}$ (both *apo* and DNA-bound), present a C-terminal strap extending back over the protomeric WYL-domain, perhaps indicating some form of lid mechanism that regulates ligand recognition and binding (59). It is clear that future systematic analysis of potential ligands, combined with extensive mutagenesis studies, are required to identify the molecules that bind BrxR, and determine whether they do cause de-repression.

Comparative genomic analyses identified a larger family of BrxR homologues, widespread within Proteobacteria (Figure 7). Stringent thresholds were used to exclude the many prokaryotic WYL-domain containing proteins. Attempts to match these BrxR-family homologues with known phage defence systems demonstrated that nearly half were associated with a diverse array of single defence systems, or a variety of collections of systems within defence islands (Figure 8). BREX systems and type IV restriction enzymes were most highly represented, consistent with previous studies that show this pairing was the most prevalent in defence islands (14,15).

In addition, BrxR-family homologues were associated with a large array of other systems and islands, suggesting that BrxR-family homologues might not simply function to avoid fitness costs (as hypothesized above), but also to regulate the time-course or stages of phage defence. Such a mechanism would allow each system to provide protection, depending on context of infection and the counter-defence systems present on the invading phage. The possibility of phage-dependent ligands binding to BrxR also provides an opportunity for phage defence systems or islands to respond dynamically to the type of attack, be it via a phage, or a mobile genetic element. In this manner, only selected invading DNAs (or RNAs) might be targeted.

Because almost half of the BrxR-family homologues were associated with known phage defence systems, there is an exciting possibility that other conserved genes associated with BrxR-family homologues represent core genes of yet undiscovered phage defence systems. By using BrxR to hunt for new systems, it may be possible to further expand our knowledge of phage-host interactions and identify novel tools for biotechnology.

DATA AVAILABILITY

The crystal structure of BrxR has been deposited in the Protein Data Bank under accession number 7QFZ. All other data needed to evaluate the conclusions in the paper are present in the paper and/or Supplementary Data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Chibani-Chennoufi,S., Bruttin,A., Dillmann,M.L. and Brussow,H. (2004) Phage-host interaction: an ecological perspective. *J. Bacteriol.*, **186**, 3677–3686.
2. Wommack,K.E. and Colwell,R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.*, **64**, 69–114.
3. Lima-Mendez,G., Toussaint,A. and Leplae,R. (2007) Analysis of the phage sequence space: the benefit of structured information. *Virology*, **365**, 241–249.
4. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
5. Stern,A. and Sorek,R. (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays*, **33**, 43–51.

6. Hampton,H.G., Watson,B.N.J. and Fineran,P.C. (2020) The arms race between bacteria and their phage foes. *Nature*, **577**, 327–336.

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

8. Loenen,W.A.M., Dryden,D.T.F., Raleigh,E.A. and Wilson,G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.*, **42**, 20–44.

9. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.

10. Rao,D.N., Dryden,D.T.F. and Bheemanaik,S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45–55.

11. Lopatina,A., Tal,N. and Sorek,R. (2020) Abortive infection: bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.*, **7**, 371–384.

12. Makarova,K.S., Wolf,Y.I., Iranzo,J., Shmakov,S.A., Alkhnbashi,O.S., Brouns,S.J.J., Charpentier,E., Cheng,D., Haft,D.H., Horvath,P. *et al.* (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.

13. Raleigh,E.A. (1992) Organization and function of the mcrBC genes of *Escherichiacoli* K-12. *Mol. Microbiol.*, **6**, 1079–1086.

14. Makarova,K.S., Wolf,Y.I., Snir,S. and Koonin,E.V (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.*, **193**, 6039–6056.

15. Makarova,K.S., Wolf,Y.I. and Koonin,E.V (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.*, **41**, 4360–4377.

16. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, eaar4120.

17. Goldfarb,T., Sberro,H., Weinstock,E., Cohen,O., Doron,S., Charpak-Amikam,Y., Afik,S., Ofir,G. and Sorek,R. (2015) BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.*, **34**, 169–183.

18. Lau,R.K., Ye,Q., Birkholz,E.A., Berg,K.R., Patel,L., Mathews,I.T., Watrous,J.D., Ego,K., Whiteley,A.T., Lowey,B. *et al.* (2020) Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol. Cell*, **77**, 723–733.

19. Owen,S.V., Wenner,N., Dulberger,C.L., Rodwell,E.V., Bowers-Barnard,A., Quinones-Olvera,N., Rigden,D.J., Rubin,E.J., Garner,E.C., Baym,M. *et al.* (2021) Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe*, **29**, 1620–1633.

20. Millman,A., Bernheim,A., Stokar-Avihail,A., Fedorenko,T., Voichek,M., Leavitt,A., Oppenheimer-Shaanan,Y. and Sorek,R. (2020) Bacterial retrons function in anti-phage defense. *Cell*, **183**, 1551–1561.

21. Bernheim,A., Millman,A., Ofir,G., Meitav,G., Avraham,C., Shomar,H., Rosenberg,M.M., Tal,N., Melamed,S., Amitai,G. *et al.* (2021) Prokaryotic viperins produce diverse antiviral molecules. *Nature*, **589**, 120–124.

22. Tal,N., Morehouse,B.R., Millman,A., Stokar-Avihail,A., Avraham,C., Fedorenko,T., Yirmiya,E., Herbst,E., Brandis,A., Mehlman,T. *et al.* (2021) Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell*, **184**, 5728–5739.

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

24. Youkharibache,P., Veretnik,S., Li,Q., Stanek,K.A., Mura,C. and Bourne,P.E. (2019) The small β-barrel domain: a survey-based structural analysis. *Structure*, **27**, 6–26.

25. Tsokos,G.C. (2006) In the beginning was Sm. *J. Immunol.*, **176**, 1295–1296.

26. Cohen,G.B., Ren,R. and Baltimore,D. (1995) Modular binding domains in signal transduction proteins. *Cell*, **80**, 237–248.

27. Schumacher,M.A., Pearson,R.F., Møller,T., Valentin-Hansen,P. and Brennan,R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.*, **21**, 3546–3556.

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

29. Modell,J.W., Kambara,T.K., Perchuk,B.S. and Laub,M.T. (2014) A DNA damage-Induced, SOS-independent checkpoint regulates cell division in *Caulobacter crescentus*. *PLoS Biol.*, **12**, e1001977.

30. Andis,N.M., Sausen,C.W., Alladin,A. and Bochman,M.L. (2018) The WYL domain of the PIF1 helicase from the thermophilic bacterium *Thermoto gaelfii* is an accessory single-stranded DNA binding module. *Biochemistry*, **57**, 1108–1118.

31. Yan,W.X., Chong,S., Zhang,H., Makarova,K.S., Koonin,E.V, Cheng,D.R. and Scott,D.A. (2018) Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell*, **70**, 327–339.

32. Zhang,H., Dong,C., Li,L., Wasney,G.A. and Min,J. (2019) 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.

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

34. LeGault,K.N., Hays,S.G., Angermeyer,A., McKitterick,A.C., Johura,F.T., Sultana,M., Ahmed,T., Alam,M. and Seed,K.D. (2021) Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science*, **373**, eabg2166.

35. Machnicka,M.A., Kaminska,K.H., Dunin-Horkawicz,S. and Bujnicki,J.M. (2015) Phylogenomics and sequence-structure-function relationships in the GmrSD family of type IV restriction enzymes. *BMC Bioinformatics*, **16**, 336.

36. Gaastra,W., Kusters,J.G., van Duijkeren,E. and Lipman,L.J.A. (2014) *Escherichia fergusonii*. *Vet. Microbiol.*, **172**, 7–12.

37. Picton,D.M., Luyten,Y.A., Morgan,R.D., Nelson,A., Smith,D.L., Dryden,D.T.F., Hinton,J.C.D. and Blower,T.R. (2021) 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.

38. Cai,Y., Usher,B., Gutierrez,C., Tolcan,A., Mansour,M., Fineran,P.C., Condon,C., Neyrolles,O., Genevaux,P. and Blower,T.R. (2020) A nucleotidyltransferase toxin inhibits growth of *Mycobacteriumtuberculosis* through inactivation of tRNA acceptor stems. *Sci. Adv.*, **6**, eabb6651.

39. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P$_{BAD}$ promoter. *J. Bacteriol.*, **177**, 4121–4130.

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

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

42. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 235–242.

43. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

44. Adams,P.D., Afonine,P.V, Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 213–221.

45. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 1002–1011.

46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.

47. Connor,T.R., Loman,N.J., Thompson,S., Smith,A., Southgate,J., Poplawski,R., Bull,M.J., Richardson,E., Ismail,M., Thompson,S.E.- *et al.* (2016) CLIMB (the cloud infrastructure for microbial bioinformatics): an online resource for the medical microbiology community. *Microb. Genomics*, **2**, e000086.

48. Huerta-Cepas,J., Serra,F. and Bork,P. (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.*, **33**, 1635–1638.

49. Yu,G., Lam,T.T.-Y., Zhu,H. and Guan,Y. (2018) Two methods for mapping and visualizing associated data on phylogeny using Ggtree. *Mol. Biol. Evol.*, **35**, 3041–3043.

50. Conway,J.R., Lex,A. and Gehlenborg,N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.

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

52. Solovyev,V. and Salamov,A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. In: Li,R.W. (ed). *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, pp. 62–78.

53. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**, e3647.

54. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.

55. Hampton,H.G., Jackson,S.A., Fagerlund,R.D., Vogel,A.I.M., Dy,R.L., Blower,T.R. and Fineran,P.C. (2018) 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.

56. Gajiwala,K.S. and Burley,S.K. (2000) Winged helix proteins. *Curr. Opin. Struct. Biol.*, **10**, 110–116.

57. Ashkenazy,H., Abadi,S., Martz,E., Chay,O., Mayrose,I., Pupko,T. and Ben-Tal,N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, 344–350.

58. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

59. Luyten,Y., Hausman,D., Young,J.C., Doyle,L.A., Higashi,K.M., Ubilla-Rodriguez,N.C., Lambert,A.R., Arroyo,C., Forsberg,K.J., Morgan,R.D. *et al.* (2022) Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkac311.

60. Fineran,P.C., Blower,T.R., Foulds,I.J., Humphreys,D.P., Lilley,K.S. and Salmond,G.P. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 894–899.

61. Blower,T.R., Fineran,P.C., Johnson,M.J., Toth,I.K., Humphreys,D.P. and Salmond,G.P. (2009) Mutagenesis and functional characterisation of the RNA and protein components of the *toxIN* abortive infection and toxin-antitoxin locus of *Erwinia. J. Bacteriol.*, **191**, 6029–6039.

62. Dy,R.L., Przybilski,R., Semeijn,K., Salmond,G.P.C. and Fineran,P.C. (2014) A widespread bacteriophage abortive infection system functions through a type IV toxin-antitoxin mechanism. *Nucleic Acids Res.*, **42**, 4590–4605.

63. Blankenchip,C.L., Nguyen,J.V., Lau,R.K., Ye,Q., Gu,Y. and Corbett,K.D. (2022) Control of bacterial anti-phage signaling by a WYL domain transcription factor. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkac343.

64. Gordeeva,J., Morozova,N., Sierro,N., Isaev,A., Sinkunas,T., Tsvetkova,K., Matlashov,M., Truncaite,L., Morgan,R.D., Ivanov,N.V *et al.* (2019) BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res.*, **47**, 253–265.