

Analysis of Cas1-2 and Cas3 CRISPR Immunity Proteins in *E. coli* Cell Physiology

By: Ryan Reeves

Supervisor: Dr Ed Bolt

Thesis submitted to the University of Nottingham for the degree of
Master of Research in Molecular Microbiology

January 2019

Abstract

CRISPR-Cas is an adaptive immune system which provides protection from invasive mobile genetic elements in prokaryotes. CRISPR-Cas immunity starts with acquisition: this involves the capture of DNA fragments called 'prespacers', which are then processed to become 'protospacers' and integrated into the host's chromosome as 'spacers'. Acquisition is catalysed by the Cas1-2 complex along with host factors. Although the spacer integration mechanism is well understood, the prespacer capture process remains cryptic. In the *E. coli* CRISPR I-E system, the products of replication fork collisions have been linked to prespacer capture by Cas1-2. By increasing the frequency of replication-fork collisions using an *E. coli* strain with an additional ectopic origin of replication, I show that the level of spacer acquisition increases *in vivo* compared to wild-type; providing evidence that RecBCD-helicase processed fragments from stalled DNA replication forks are captured as prespacers by Cas1-2. Furthermore, I demonstrate that spacer acquisition increases in *E. coli* under nutritional stress. I also show that N-terminal tags on Cas1 have a detrimental effect on acquisition *in vivo*, despite successful spacer integration activity *in vitro*; as well as investigating the role of Cas3 in a cell stress response; and initiating genetic work into the unusual CRISPR-Cas system of *Marinotoga piezophila*.

Abbreviations

Amp	Ampicilin
Apra	Apramycin
APS	Ammonium persulfate
ATP	Adenosine triphosphate
bp	Base Pair(s)
Cas	CRISPR associated
Cascade	CRISPR-associated complex for antiviral defence
CFU	Colony forming units
Chlm	Chloramphenicol
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
dH ₂ O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate(s)
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
H-NS	heat-stable nucleoid-structuring protein
IHF	Integration host factor
IPTG	Isopropyl- β -D-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilodalton
LB	Lysogeny broth
MGE	Mobile genetic element
nt	nucleotide(s)
OD ₆₀₀	Optical density at 600nm

<i>Ori</i>	Origin of replication
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PDB	Protein database
PolA	DNA polymerase I
Rif	Rifampicin
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPM	Revolutions per minute
<i>rnn</i>	Ribosomal RNA operon
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDW	Sterile distilled water
Spect	Spectinomycin
SpIn	Spacer integration
ssDNA	Single-stranded DNA
Strep	Streptomycin
TBE	Tris-Borate-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
<i>Ter</i>	Replication termination site
Tet	Tetracycline
TRIS	Tris(hydroxymethyl)-aminomethane
Tus	terminator utilising substance
UV	Ultraviolet light
v/v	volume/volume
w/v	Weight/volume
wt	wild type
x g	Centrifugal Force (x gravity)
YFP	Yellow Fluorescent Protein

Contents

Abstract	2
Abbreviations	3
Chapter 1: Introduction	7
1.1 CRISPR-Cas, a Prokaryotic Immune System.....	7
1.1.1 The Type I-E CRISPR system of <i>Escherichia coli</i> K12 MG1655...	12
1.2 Acquisition of New Spacers in <i>E. coli</i>	13
1.2.1 The Cas1-2 Adaptive Complex	13
1.2.2 Naïve Acquisition	16
1.2.3 Primed Acquisition	18
1.2.4 Spacer Integration.....	19
1.3 crRNA Processing in <i>E. coli</i>	22
1.4 Interference in <i>E. coli</i>.....	23
1.5 Cas1-2 and DNA Replication	25
1.6 Ectopic Replication Origins.....	25
1.7 The CRISPR System of <i>Marinotoga piezophila</i>.....	28
1.8 The Stringent Response	30
1.9 Aims of the Project	31
Chapter 2: Materials and Methods	33
2.1 Chemicals	33
2.2 Bacterial Strains	34
2.3 Plasmids and Substrates.....	35
2.4 Primers	37

2.5	Buffers, Solutions, Reagents, and Microbiological Media	38
2.6	General Microbiology	41
2.6.1	Making Chemically Competent Cells	41
2.6.2	Transformation of Chemically Competent Cells	41
2.6.3	Isolation of Plasmid DNA.....	41
2.6.4	Agarose Gel Electrophoresis	42
2.6.5	DNA Extraction from Agarose Gel	42
2.6.6	PCR 'Clean-up'	42
2.6.7	Colony PCR	42
2.6.8	Overexpression of Protein	42
2.6.9	SDS PAGE	43
2.7	Genetic Manipulation.....	43
2.7.1	P1 Transduction to Produce EB371/EB372	43
2.7.2	Red Recombineering.....	44
2.7.3	Subcloning of Genes	45
2.7.4	Cloning His-YFP-Cas1	46
2.8	Acquisition Assays.....	46
2.8.1	Passaging of Cells	46
2.8.2	Bacterial Growth Viability ('Spot') Tests	47
2.8.3	Detection of New Spacer(s) by PCR	47
2.9	Protein Purification Work	48
2.9.1	Overexpression of YFP-Cas1 and Cas2	48
2.9.2	Coupled YFP-Cas1 and Cas2 Purification	49
2.10	Genetic Analysis of Interactions Between Cas3 and the Stringent Response	
		51
2.10.1	Bacterial Growth Viability ('Spot') Tests for Rifampicin Resistance	51
Chapter 3:	Results.....	52

3.1	Naïve Acquisition in Wild Type Vs Ectopic Origin <i>E. coli</i> Strains ..	52
3.2	YFP-Tagged Cas1 Acquisition and Purification	65
3.3	Analysis of Acquisition in M9 Media Vs LB Media	73
3.4	Naïve Acquisition Using the <i>M. piezophila</i> CRISPR/Cas system	76
Chapter 4: Cas3 and the Stringent Response		81
Chapter 5: Discussion.....		84
5.1	Prespacer Sources in Naïve Acquisition	84
5.2	Cas1 N-terminal Fusions and Acquisition	87
5.3	The CRISPR System in <i>M. piezophila</i>	88
5.4	Cas3 and the Stringent Response	89
Acknowledgements.....		91
Bibliography		92

Chapter 1: Introduction

1.1 CRISPR-Cas, a Prokaryotic Immune System

Bacteria and Archaea must protect themselves from harmful mobile genetic elements (MGEs). These MGEs can take the form of invasive plasmids and viruses, the latter of which can out-number host cells 10:1 in certain areas, such as oceans¹. Some eukaryotes employ RNAi pathways to protect against the propagation of parasitic invasive MGEs at the single-cell level², however this isn't an option for unicellular prokaryotes. As a result, other defence mechanisms, both innate and adaptive, have been developed. These immune systems are often found within the same host cell, indicating that innate and adaptive immune systems can fulfil different roles: the innate systems can indiscriminately protect the host against parasites; whilst the adaptive systems can allow for MGEs with host benefits, such as antibiotic resistant plasmids, to be accepted by the cell³. This trade-off between different immune systems can sacrifice host immunity in favour of potentially beneficial horizontal gene transfer, which can help to drive evolution⁴. An example of an innate defence mechanism involves modifying outer membrane proteins to block phage adsorption and reduce plasmid conjugation efficiency, such as the OmpA membrane protein in *E. coli* which can interact with the lipoprotein TraT to block MGE binding via surface exclusion⁵. If a virus successfully injects DNA into the cell, restriction-modification (RM) systems can be used to target this 'non-self' DNA: 'self' DNA is methylated by methyltransferase, which protects it from digestion by restriction endonucleases which target non-methylated 'non-self' sequences⁶. Another way in which host cells can degrade non-self viral DNA is via the 'defence island system associated with restriction modification' (DISARM), which can restrict phage replication with an as-yet uncharacterised mechanism⁷. Bacteriophage exclusion (BREX) is another anti-phage mechanism which blocks phage replication⁸. In a similar response to eukaryotic apoptosis, prokaryotes can activate cell death via abortive infection (Abi) systems: here a toxin and anti-toxin are transcribed from a single promoter, but the anti-toxin is easily modified in response to stimuli (such as MGE invasion) resulting in the

toxin destroy the cell, protecting the clonal population⁹. As well as innate responses, bacteria and archaea are also capable of adaptive immunity, such as the 'clustered regularly interspaced short palindromic repeats' (CRISPR) system.

The homologous 28 bp repeat sequences separated by 33 bp spacers that form the CRISPR locus was first noticed in 1987 by Ishino *et al.* when sequencing the *iap* gene in *E. coli*¹⁰. It took a further 20 years for Barrangou *et al.* to demonstrate CRISPR immunity by identifying that *Streptococcus thermophilus* could integrate short sections from phage DNA, forming new CRISPR spacers, thereby obtaining resistance to the phage of origin¹¹. Therefore, these spacers can be traced back to their MGE of origin which will have a corresponding 'protospacer' sequence, acting as a catalogue of previous infections.

CRISPR is an immune system found in roughly 45 % of bacteria and 87 % of archaea¹² utilising MGE-derived spacer sequences, short host-derived palindromic sequences, and CRISPR-associated (Cas) proteins to prevent foreign DNA integrating into the host genome¹³. These are arranged in a repeat-spacer-repeat conformation and form a CRISPR array with an AT-rich leader sequence. In active CRISPR loci, the AT-rich leader sequence is also flanked by cas genes. Whilst CRISPRs share this general common layout, bioinformatic data has revealed that some genomes have over 700 CRISPR spacers (*Haliangium ochraceum*), 52 cas genes (*Rhodospirillum rubrum*), 661 bp spacers (*Anabaena sp.*) and 92 bp repeats (*Shewanella putrefaciens*)¹⁴. *E. coli* strain MG1655 has 2 CRISPR loci, one of which contains the cas genes required for CRISPR immunity, with 28 bp repeats and 33 bp spacers making up the CRISPR arrays¹⁵. A diagram of a generalised CRISPR locus is shown in **Figure 1**.

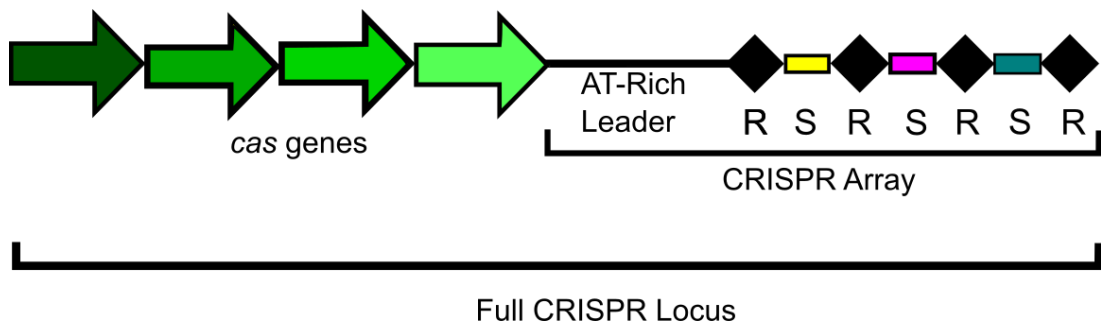


Figure 1 An example of a CRISPR locus. The *cas* genes are shown as green arrows. This is followed by the AT-rich leader sequence, then the repeats (black diamonds labelled 'R') and spacers which are coloured to represent their unique origin (rectangles labelled 'S'). This locus contains 13 repeats, but only 4 have been shown.

The mechanism of CRISPR immunity follows three main steps: acquisition (also called adaptation), crRNA processing (also called expression), and interference. Acquisition involves the detection of an invasive MGE, followed by integration of part of the MGE sequence into the host's CRISPR array to form a 'spacer'. The CRISPR-RNA (crRNA) processing step is where the CRISPR array is transcribed to pre-crRNA, which is then processed to mature crRNA. These mature crRNAs have sequence complementarity to the MGE of origin, as well as secondary RNA structure from the palindromic repeat for Cas-protein binding which is essential for the formation of ribonucleotide-protein effector complexes and subsequent interference. The interference stage is where the crRNA binds to the corresponding MGE to form an effector module, which can either degrade the MGE as a multidomain effector complex when a match is found (type 2) or recruit Cas proteins to destroy the MGE as a multi-subunit effector complex (type 1).¹⁶ A simplified overview of the CRISPR-Cas system is given in **Figure 2**.

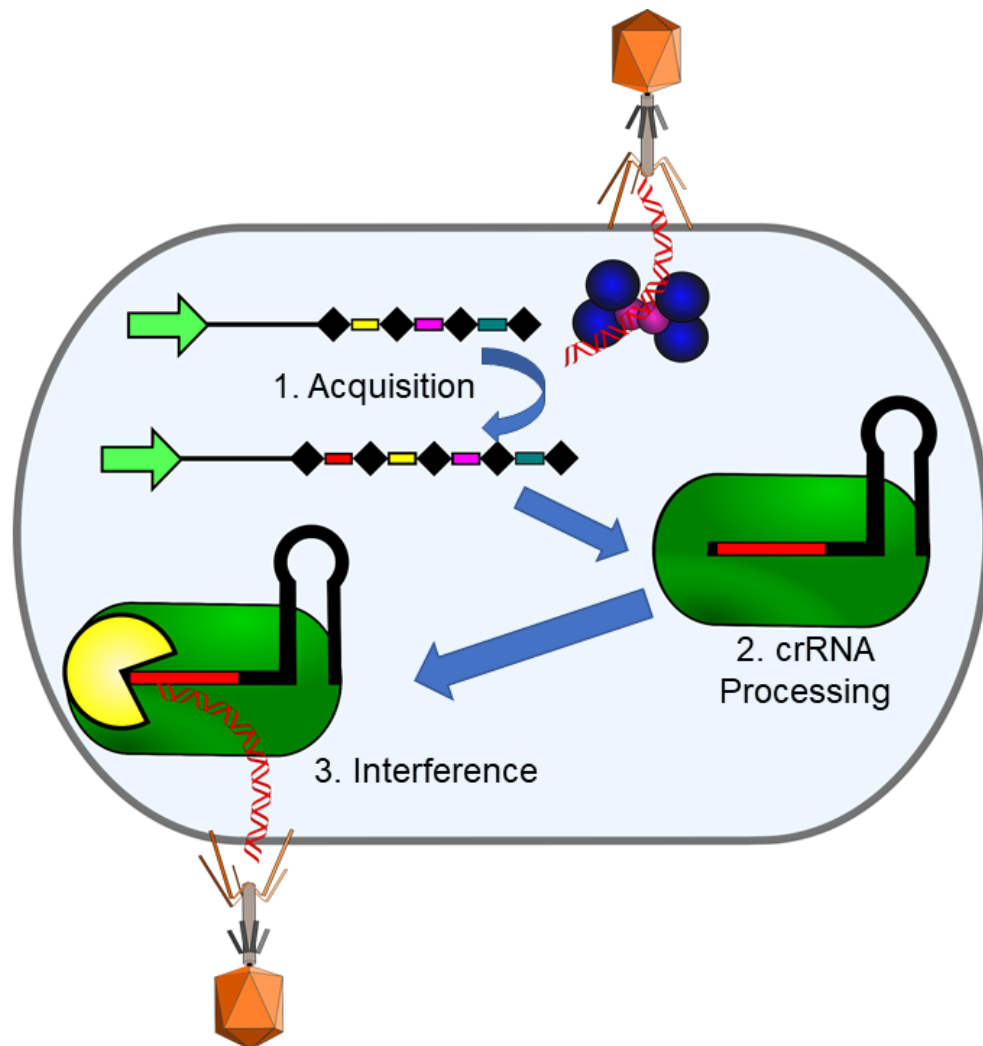


Figure 2 Simplified overview of the CRISPR-Cas immune response. 1. Acquisition: The MGE in this instance is from a phage attacking at the top of the cell. Part of the invading red DNA is integrated into the host CRISPR locus with the help of the blue-pink acquisition Cas proteins. This adds a red spacer to the CRISPR locus. 2. crRNA Processing: the crRNA is transcribed, processed, and bound to the green 'effector' module. 3. Interference: The interference complex recognises an identical phage attacking, so recruits the yellow degradation protein to stop the MGE attack.

Although CRISPR immunity across species is primarily composed of the basic mechanisms of acquisition, expression and interference, the Cas proteins utilised to effect the CRISPR response vary and are used to classify the different CRISPR systems. CRISPR systems are first divided into two separate classes defined by their 'effector' module, which is responsible for binding crRNA, locating the MGE using the crRNA template, and binding to the MGE to

commence the interference stage. Class 1 CRISPR systems have multi-subunit effector complexes, whilst Class 2 CRISPR systems have a single protein as an effector module¹⁷.

The Class 1 CRISPR-Cas system is divided into type I, type III, and type IV. The type I system is further divided into seven subtypes I-A to I-F with an 'unclassified' type-I system also present¹⁸. The hallmarks of the type I system are the effector complex, Cascade, and the translocase-nuclease for interference, Cas3. Cas3 is responsible for degrading invasive DNA, whilst the Cascade complex acts as the effector, binding crRNA and recruiting Cas3 for interference. The type-III system contains the signature Cas10 protein, with two subtypes: III-A contains the Csm complex, whilst III-B contains the Cmr complex. The type IV system is defined by the signature *csf1* gene, as well as a lack of *cas1-2* genes; these systems are functionally uncharacterised¹⁹. The class 2 system is divided into the type II, V, and VI systems. The type II is defined by the presence of Cas9 which targets and cleaves DNA, and can be subdivided into: II-A, which contains the *csn2* gene; II-B which contains the *cas4* gene; and II-C, which only contains *cas1*, *cas2*, and *cas9*. Type V systems contain the effector protein Cas12a (formerly Cpf1)²⁰, whilst type VI systems contain the effector protein Cas13a (formerly C2c2)²¹. A table showing an overview of the different CRISPR types is shown in **Table 1-1**.

Table 1-1 Classification of CRISPR-Cas Systems. Blue highlighted columns are involved in acquisition; green highlighted columns are involved in crRNA processing; orange highlighted columns are involved in interference; grey columns are additional proteins involved in CRISPR-Cas immunity. *proteins are only found in some types. †proteins sometimes exist as multi-subunit proteins, not always monomers. Question mark is for hypothetical function. Adapted from Mohanraju *et al.* 2016¹⁹.

		Spacer Integration	Pre-crRNA Processing	Effector Module	Target Degradation	Ancillary
Class 1	Type I	Cas1 Cas2 Cas4*	Cas6	Cascade	Cas3†	Cas4*(?)
	Type III	Cas1* Cas2*	Cas6*	Cas7 Cas5 Cas10	Cas10	CARF
	Type IV			Cas7 Cas5 Csf1		DinG
Class 2	Type II	Cas1 Cas2 Cas9 Cas4*	Rnase III	Cas9		Csn2
	Type V	Cas1* Cas2* Cas4*		Cas12a		Cas4*
	Type VI	Cas1* Cas2*		Cas13a		Cas4*

1.1.1 The Type I-E CRISPR system of *Escherichia coli* K12 MG1655

The CRISPR system in *E. coli* has the 'CRISPR-associated complex for antiviral defence' (Cascade) as a multi-subunit effector complex, as well as the signature interference protein Cas3²², therefore meeting the criteria of a Class1 Type-I CRISPR system. In K-12 strains, such as MG1655, there are two CRISPR loci. CRISPR-II contains only the leader and repeat-spacer CRISPR array, whilst the CRISPR-I locus (**Figure 3**) contains the CRISPR array and 8 cas genes: *cas1* and *cas2* are the genes responsible for controlling acquisition machinery; *cas6* is involved in crRNA processing; *cse1* (*cas8e*), *cse2* (*cas11*), *cas7*, *cas5*, and *cas6e* (sometimes called *casA-E*) code for the effector complex Cascade; finally *cas3* is responsible for interference²³. The cascade genes as well as the *cas1-2* genes are under the

control of the P_{cas} promoter, whilst the leader sequence also contains a promoter for crRNA processing, P_{CRISPR} . Both promoters are under repression by the heat-stable nucleoid-structuring protein (H-NS) which is a global transcriptional repressor. The *cas3* gene has its own promoter, P_{cas3} , which is not regulated by H-NS²⁴ unless the cells are in stationary phase²⁵. The P_{cas} promoter can be activated by *LeuO*²⁶ and *BaeR*²⁷ to help overcome H-NS silencing, leading to an enhanced CRISPR-Cas response²⁸.

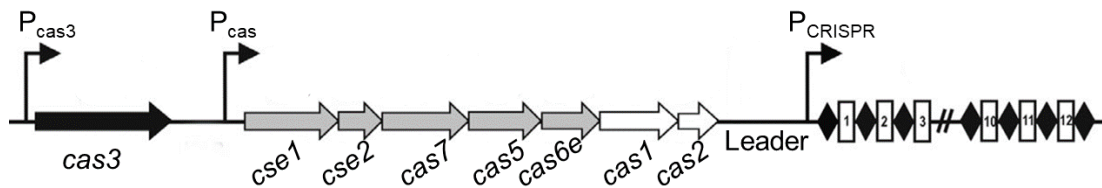


Figure 3 The CRISPR-I locus of *E. coli* MG1655. The *cas* genes are shown as thick arrows, promoters as bent arrows, repeats as diamonds and the 12 spacers as white rectangles. Figure adapted from Westra *et al.* 2010²⁶.

1.2 Acquisition of New Spacers in *E. coli*

Acquisition, which is also known as adaptation, is the first stage of CRISPR immunity. Acquisition involves a new spacer being generated from an invading MGE ‘prespacer’ (spacer capture) which is then processed to a ‘protospacer’ and inserted into the host’s CRISPR locus (spacer integration). There are two different varieties of spacer capture in the type I-E *E. coli* MG1655 CRISPR system: naïve acquisition (**1.2.2**), where a spacer is generated from a newly encountered MGE; or primed acquisition (**1.2.3**), where a previously encountered MGE triggers an enhanced uptake of new spacers²⁹

1.2.1 The Cas1-2 Adaptive Complex

Acquisition is reliant on upon the presence of the proteins Cas1 and Cas2³⁰, as such they are highly conserved amongst active CRISPR-Cas systems. Cas1 proteins form asymmetrical homodimers which are 66 kDa in size. These Cas1 units have divalent-metal nuclease activity on linear ds/ssDNA, branched DNA and ssRNA substrates without target sequence specificity³¹. Mutations in the C-terminal nuclease domain (H208A, D221A)

resulted in loss of spacer acquisition function *in vivo*, indicating Cas1 has a catalytic function in acquisition³². Cas2 proteins form symmetrical homodimers³³ with sequence-specific RNA endonuclease activity; however, loss of catalytic activity in Cas2 does not result in loss of spacer acquisition, indicating a structural role in spacer acquisition³⁴. Cas1 and Cas2 form a complex together (Cas1-2), which is hexameric; two asymmetrical Cas1 homodimers flank a central symmetrical Cas2 homodimer, forming a Cas1₄:Cas2₂ structure³⁵ (**Figure 4**).

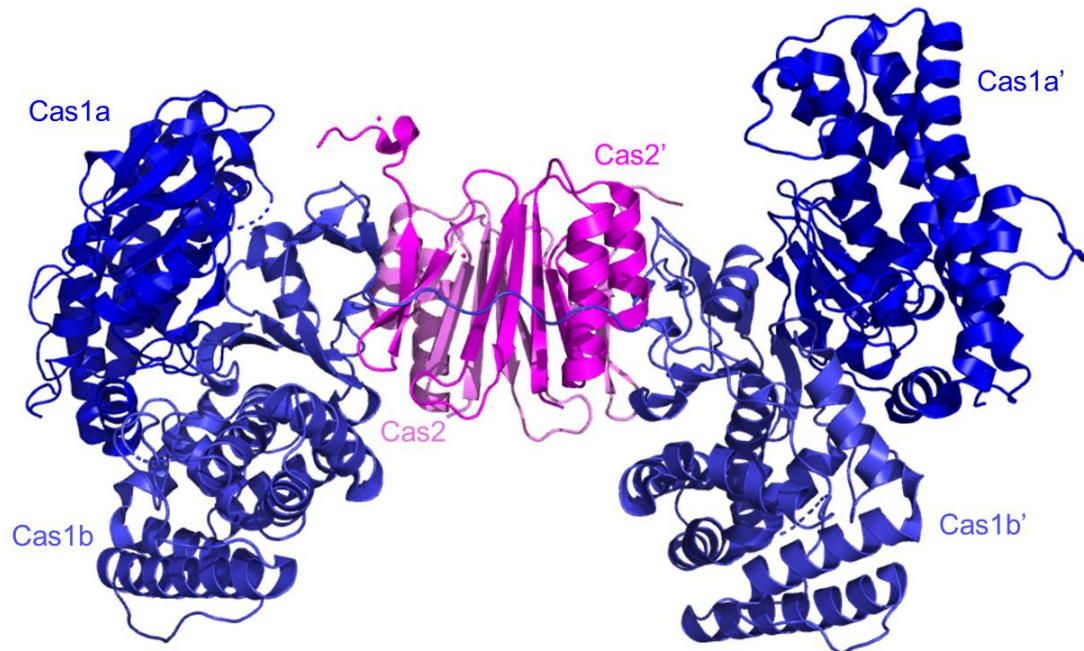


Figure 4 Hexameric *E. coli* Cas1-2. A pair of Cas1 dimers are bound to a Cas2 dimer. Cas1 proteins are shown in blue, Cas2 proteins are shown in pink. Only Cas1b and Cas1b' monomers interact with the Cas2 dimer. Crystal structure from³⁵, Protein Data Bank code 4P6I.

The Cas1-2 complex is essential for spacer capture. Spacer capture involves the recognition of the invasive MGE prespacer to form a protospacer which, following capture, becomes integrated in the CRISPR locus to become a spacer. The exact origin of these MGE-derived prespacers and how they are preferentially targeted over 'self' DNA is not yet fully understood. In type I CRISPR interference such as that of *E. coli*, spacer-complementary MGEs are recognised via a Protospacer Adjacent Motif (PAM)³⁶. These PAMs are located to at the 5' end of the prespacer and are usually 3 nucleotides in length. The most common PAM sequence in *E. coli* is 5'-AAG, with the 3'-G being inserted along with the rest of the

spacer sequence, therefore new spacers begin with a 5'-G³⁷. (though the 5'-AA bases are cleaved)³⁸. PAMs are not included in the spacer sequence, helping the host distinguish 'self' (no PAM spacer) from 'non-self' (PAM-containing MGE). A PAM as well as spacer-protospacer match is required for the interference step³⁹ which helps to minimise the occurrence of self-targeting the host's chromosome, a process which can be lethal⁴⁰. Studies involving Cas1-2 binding to artificial protospacers has shown Cas1-2 favours binding to protospacers in the form of dual-forked DNA, though it can bind to ssDNA and dsDNA. The engineered protospacer sequence has a 23 bp duplex region which binds to the flat region of the Cas2 dimers and the inner Cas1 monomers directly bound to the Cas2 (i.e. Cas1b/b' from **Figure 5**). The protospacer fork contains ≥7 nucleotide 3'-overhangs (though 5' overhangs have also been demonstrated to work⁴¹) which interact with the C-terminal domain of the outer Cas1 monomers not bound to Cas2 (i.e. Cas1a/a' from **Figure 5**). The length of the protospacer duplex region is measured by the distance between the tyrosine residues (Y22) in the inner Cas1, thanks to both catalytic Cas1 dimers being bound to the structural Cas2 dimers. The Y22 prevents base pairing, essentially prying the base pairs apart, keeping the duplex at 23 bp length. If the PAM-complimentary sequence (5'-CTT) is located in this overhang, the Cas1-2 complex undergoes a conformational change to ensure strong DNA-protein interaction⁴² which could also be beneficial in Cas1-2 integrating the protospacer into the CRISPR array⁴³. The 3' overhangs are cleaved after 5 nucleotides which is coordinated by Y217 base stacking; this residue positions the 3' DNA to be cut after 5 nt generating the 3'-hydroxyl for nucleophilic attack. An arginine clamp (Cas1 R41 and Cas2 R16, R77, R78- on Cas2) binds non-specifically to the protospacer's 23 bp helical region⁴³, so when processed, the protospacer length is 33 nucleotides long (but as one nucleotide is always part of the PAM and therefore conserved, spacers can be considered 32 bp with 29 bp repeats). When the prespacer is cut, forming a 33 bp protospacer, the DNA sequence produced contains 3'-hydroxyl groups. Cas1-2 processing a prespacer into a protospacer in this way is summarised in **Figure 5**. Recent research by Musharova *et al.* has suggested that Cas1-2 can bind to single-stranded prespacers which then forms a double stranded protospacer, either by re-annealing to the MGE or another undefined mechanism to create the second strand⁴⁴.

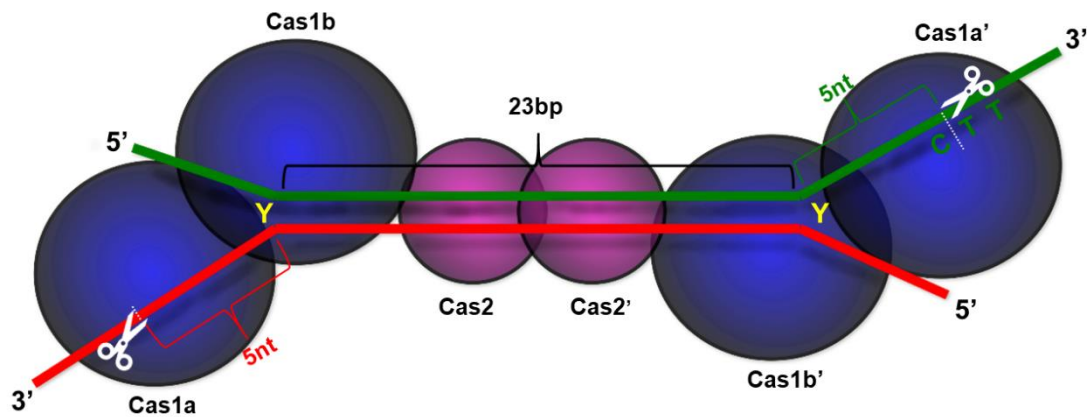


Figure 5 Cas1-2 complex processing a dual-forked protospacer. Cas2 units and Cas1b units bind to the 23 bp duplex DNA nonspecifically using arginine residues. The yellow Y represents with Tyrosine-22 of the Cas1b subunits which help separate the DNA strands. The protospacer 3' overhangs are cut after 5nt, represented by scissors, to leave 3'-hydroxyl groups. The PAM is shown on the green strand, so the processed protospacer will end with a 3'-C residue.

1.2.2 Naïve Acquisition

Naïve acquisition is where a protospacer is selected for acquisition from an MGE which has not been previously encountered. Naïve acquisition is reliant on more than just Cas1-2 to acquire new spacers, with the origin of protospacers which become protospacers (and eventually spacers) still not being fully understood. It has been suggested that RecBCD could play a role in naïve acquisition by generating protospacers for capture. In 2015, sequencing of newly generated spacers by Levy *et al.* revealed that despite PAM sequences appearing evenly across the whole genome, spacers were generated preferentially close to the chromosomal origin (*oriC*), as well as fork-stalling chromosomal replication termination (*Ter*) sites⁴⁵. The *oriC* derived spacers could be due to the relative abundance of this DNA found within the cell, whilst the *Ter* site spacers being due to double strand breaks due to stalled replication forks. These strand breaks would be degraded by RecBCD until a *Chi* site (5'-GCTGGTGG) is reached, where RecBCD stops. Sequencing has determined that spacer sequences can originate from between *Ter* and *Chi* sites, indicating that RecBCD plays a role in generating protospacers⁴⁵. This could explain the bias for non-self over self as high

copy number plasmid MGEs would lead to more replication forks, increasing the chances of stalls/breaks for RecBCD to repair. A high frequency of *Chi* sites on the chromosome prevents RecBCD from degrading the chromosome too much, thereby limiting the prespacers available, limiting host-derived spacers; MGE DNA has fewer *Chi* sites so can be degraded more by RecBCD⁴⁵ (**Figure 6**). Phage DNA is linear and so RecBCD could target it immediately, increasing the chances of it being used as a spacer⁴⁵. However, whilst RecBCD has both helicase and nuclease functions, recent studies have shown that only the helicase-translocase activity of RecBCD is needed for prespacer capture. Radovčić *et al.* showed that nuclease-deficient RecBCD strains can acquire new spacers *in vivo*, and that 3' ssDNA/ 5' ssDNA exonucleases are not required for prespacer capture. This could imply that RecBCD is needed as a helicase/translocase to displace proteins from the potential prespacer DNA with translocase activity, allowing Cas1-2 to bind and process the prespacer with its own nuclease activity⁴¹. In addition to spacers originating from RecBCD helicase with separate nuclease products, Cas3 degradation products have been demonstrated to act as prespacers⁴⁶ and in *Pyrococcus furiosus* spacers were generated from active transposons, CRISPR loci, ribosomal RNA (rRNA) genes and rolling circle origins of replication⁴⁷.

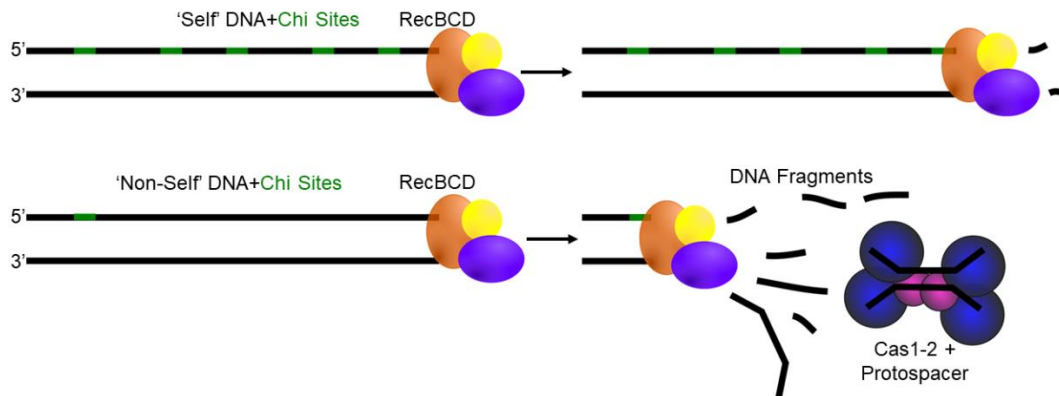


Figure 6 RecBCD on 'Self' Vs 'Non-Self' DNA. The 'Self' DNA, ie chromosomal, has several Chi sites so the DNA is only partially processed with RecBCD helicase/transolcase activity, leaving little for Cas1-2 to process into protospacers with its nuclease activity. The 'Non-Self' DNA, ie MGE, contains 14x fewer Chi sites, so RecBCD is able to process more DNA, providing Cas1-2 with a source of prespacers to process into potospacers.

1.2.3 Primed Acquisition

As well as being able to acquire new spacers via naïve acquisition, the type I-E CRISPR system of *E. coli* MG1655 can create additional spacers from previously encountered MGEs, this is called 'primed acquisition'. This 'Priming' uses the interference machinery (1.4) with Cas1-2 to target previously encountered MGEs as a way to update its own immunity⁴⁸. This means that the CRISPR-Cas system can overcome 'escape mutations' in the MGE's PAM sequence or protospacer 'seed' sequence which would otherwise render the effector complex incapable of adequately recognising the protospacer, therefore not able to illicit the interference step, ergo not able to protect the host against the MGE which it has previously encountered³⁹. Though these altered PAM/protospacer sequences prevent the interference stage, they can trigger primed acquisition with the aid of Cas1-2, the effector complex (Cascade, 1.3) and Cas3, which form the primed acquisition complex (PAC)⁴⁹. The PAC complex works with other host factors, such as PriA and RecG⁵⁰, to produce multiple new MGE-derived spacers²⁹. As with the interference stage, Cascade forms an R-loop, which blocks MGE replication and recruits Cas3. As Cas3 translocates along the MGE and degrades the DNA as part of the PAC, the degraded DNA fragments are captured by the

Cas1-2 complex (forming new 33 nt protospacers) which contain new T-rich PAM-compliment sequences at the 3' end (5'-CTT being the most frequent)^{51 52}. Primed acquisition of spacers may not be due to Cse1 of Cascade being unable to stimulate interference⁵³, but rather by Cascade and Cas3 binding to Cas1-2, forming the PAC⁴⁹. This is demonstrated by primed acquisition being stimulated even when protospacers fully match crRNA, but they are degraded in interference reactions more quickly than priming reactions can occur⁵⁴.

1.2.4 Spacer Integration

Once the 33 bp protospacer has been generated, it is inserted into the CRISPR locus proximal to the AT-rich leader sequence. In *E. coli* MG1655, both CRISPR loci can integrate spacers, though naïve acquisition experiments have demonstrated a preference for the CRISPR-II locus over the CRISPR-I locus⁵⁵. Cas1-2 is able to bind to the CRISPR locus at the leader sequence, which is where new spacers are integrated³⁵.

Although *E. coli* Cas1 alone can integrate a low-level of protospacers into the CRISPR locus⁵⁶, Cas1-2 as a complex achieves the most efficient spacer integration. At this stage, the Cas1-2 complex can integrate larger sequences than the 33 bp protospacer created during the capture stage, so it is capture, not integration that determines spacer length. Furthermore, a PAM sequence is not required for integration, but the 3'-hydroxyl cytosine (3'-C) end of the PAM-compliment sequence is preferentially integrated into the minus strand of the leader-repeat section, explaining with a 5'-Guanine (5'-G) forms the start of most spacers⁵⁶. As well as the Cas1-2 complex, Integration Host Factor (IHF) is required in Type I-E CRISPR systems to integrate protospacers. IHF is a 22 kDa heterodimer made up of an α and a β subunit, and is a histone-like protein which binds to the minor groove of DNA⁵⁷. In spacer integration, IHF induces a $\sim 160^\circ$ bend in the AT-rich leader sequence. This bend in the DNA allows Cas1 to interact with the scissile DNA phosphodiester bond where the leader and first repeat meet⁵⁸; in a similar model to retroviral integration⁵⁹. Without IHF, integration cannot occur unless the target DNA is supercoiled⁵⁶. The protospacer is inserted via two separate nucleophilic attack reactions with Cas1-2 catalysed transesterification (TES). The first reaction takes place at the 5'-end of the first repeat (where the leader meets

repeat-1) and a second reaction on the negative strand 28 nt downstream; this results in 61 new base pairs added to the CRISPR locus⁶⁰. The spacer length is conserved due to two short inverted repeat (IR) sequences found in the CRISPR repeat sequence which help anchor integration machinery (though not confirmed Cas1-2). They act as rulers after the first nucleophilic reaction to coordinate the second nucleophilic attack, which occurs between the leader-spacer junction, keeping the repeats at 28 nt long⁶¹. Firstly, the 3'-hydroxyl of a protospacer (containing the PAM-derived 3'-C base) attacks the leader-repeat-1 boundary, which is followed by a TES reaction catalysed by Cas1. This joins the protospacer to the 5' of the first repeat, resulting in a half-site intermediate. The second nucleophilic attack joins the remaining 3'-hydroxyl of the protospacer to the repeat-1-spacer-1 boundary, linking the protospacer to the negative strand of repeat 1. The protospacer has now been integrated and is a full CRISPR spacer⁶². The gaps between the leader and the new spacer on the positive strand, and between the first and second spacers on the negative strand, are repaired by DNA Polymerase I (PolA)⁵⁰. A summary of Cas1-2 mediated spacer integration is shown in **Figure 7**.

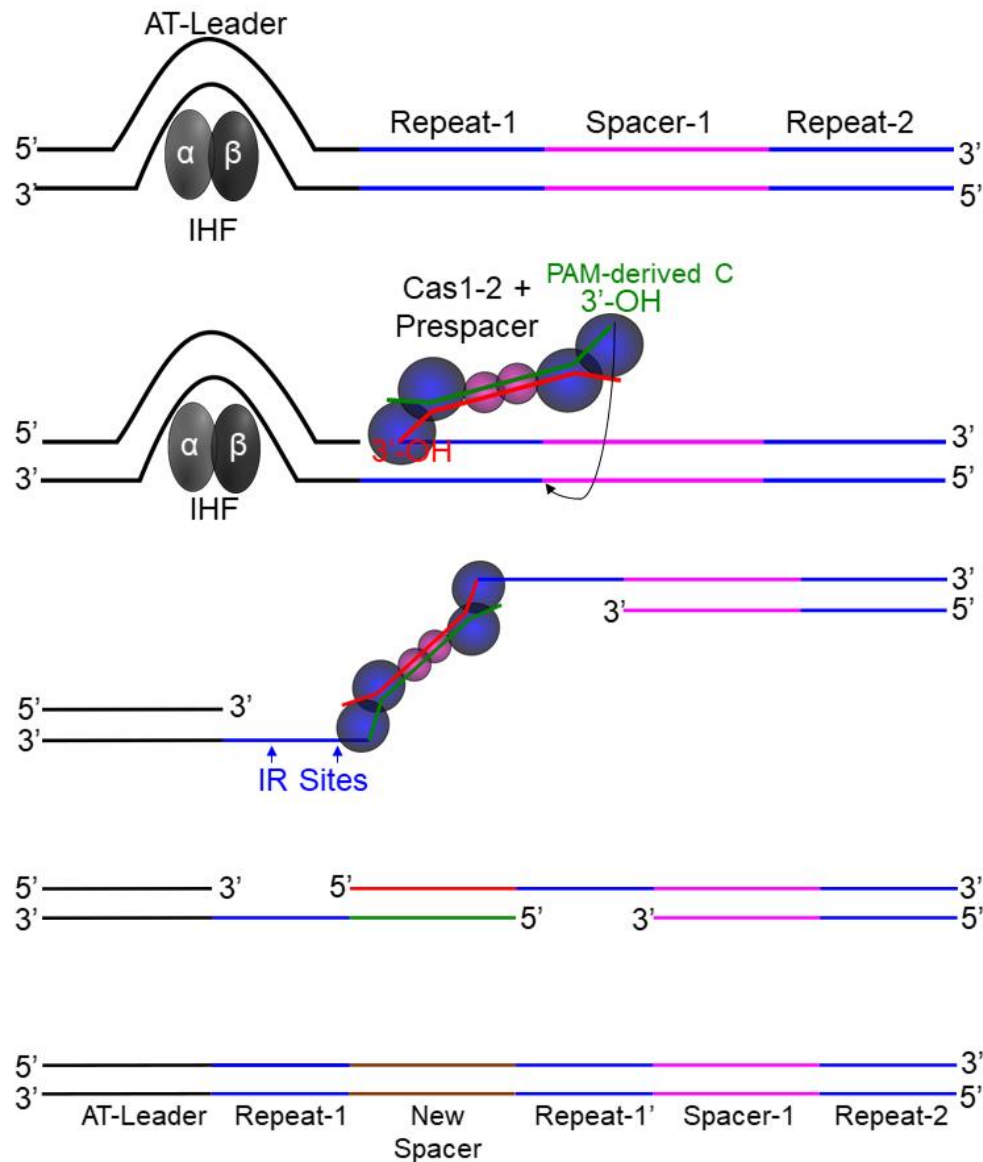


Figure 7 Overview of Spacer Integration. IHF binds to the AT-rich leader sequence, bending it by around 160° . This bent leader sequence allows Cas1-2 to facilitate a nucleophilic attack on the CRISPR locus, first with the 3'-OH of the red protospacer strand which inserts between the leader and first repeat. Next, the green-strand with the PAM-derived 3'-Cytosine is inserted via nucleophilic attach between repeat-1 and the first spacer; this is guided into position by the IR sites, which keep the repeats a consistent length. After nucleophilic attack and transesterification, PolA fills the gaps, fully integrating the new (brown) spacer.

1.3 crRNA Processing in *E. coli*

Following on from the acquisition stage, the next stage in CRISPR-Cas immunity is the crRNA processing. This process begins with the generation of precursor crRNA (pre-crRNA) which is then cleaved to form mature crRNA, bound to the effector complex (also called a surveillance ribonucleoprotein complex), and used as a guide for interference⁶³. The pre-crRNA is generated by transcription of the whole CRISPR array from the leader sequence (P_{CRISPR}) to the final repeat. In type I-E systems Cas6e, an endoribonuclease subunit of Cascade, is then employed to trim the large pre-crRNA down into fully mature crRNA without the need for any further processing. Cas6e recognises the hairpin structures caused by the palindromic repeats and cuts 8 nt upstream from the repeat-spacer boundary⁶⁴. This generates mature crRNAs which are 61 nt in length, containing the unique spacer flanked by 8 nt of the upstream spacer at the 5'-end with a 5'-hydroxyl group (5'-handle), and 3'-end composed of a hairpin structure from the downstream spacer ending with a cyclic 2' 3'-phosphate (3'-handle)⁶⁵.

Once the mature crRNA has been produced, it is attached to the effector complex, Cascade; without the crRNA binding, Cascade cannot be assembled. Cascade is 405 kDa in size and is comprised of eleven subunits: Cse1₁ (CasA), Cse2₂ (CasB), Cas7₆ (CasC), Cas5₁ (CasD), and Cas6e₁ (CasE)⁶⁶ (shown in **Figure 8**). The Cas6e subunit remains bound to the crRNA 3' handle, which is a hairpin structure, from the crRNA processing step. The 5'-handle of the crRNA forms a crook which binds to Cas5, Cse1 and Cas7.⁶⁷ Between the crRNA caps, the Cas7 subunits bind non-specifically to the spacer region of the crRNA via interactions with the phosphate backbone of the crRNA. This forms the spine of the complex, connecting the two caps, holding the crRNA in place, whilst also presenting 5 nt segments on the complex surface for binding crRNA to complementary DNA during the interference stage⁶⁸.

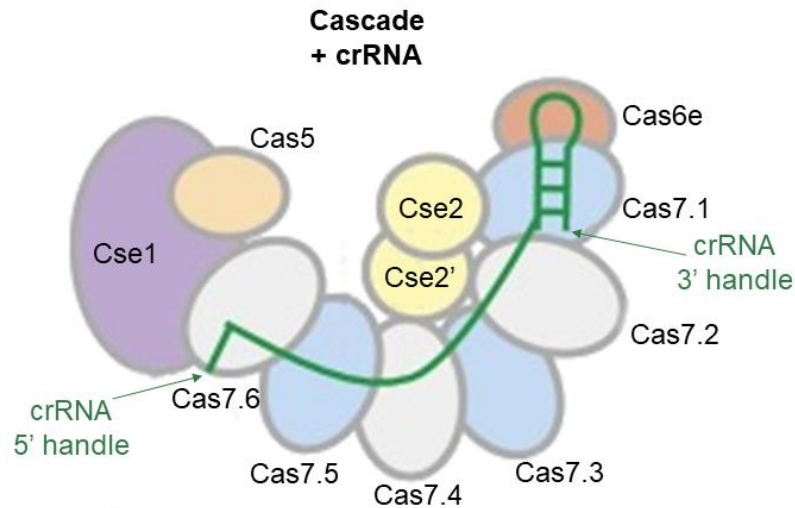


Figure 8 Cartoon showing the Cascade effector complex of *E. coli* bound to crRNA.

The Cas7 subunits have been labelled 7.1-7.6 according to their position in the Cascade complex, though they are all identical. Image modified from Hochstrasser *et al.* 2014⁶⁹.

1.4 Interference in *E. coli*

The final stage of CRISPR-Cas immunity is the interference stage, where a target degrading module is recruited by the effector complex; in the case of *E. coli* Cas3 is employed to destroy MGEs as the Cascade complex has no DNase activity⁶⁵. Cas3 is a 100 kDa protein with N-terminal Histidine-Aspartate (HD) nuclease domains⁷⁰ and C-terminal superfamily 2 (SF2) helicase domains⁷¹. The HD phosphohydrolase/nuclease domain contains HD motifs I-V which are involved in binding metal cofactors and degrading ssDNA⁷². The C-terminal helicase meanwhile unwinds DNA-DNA and DNA-RNA duplexes in a 3'-5' direction in the presence of a Mg^{2+} cofactor and ATP^{73, 74}.

Before Cas3 is recruited, the crRNA-bound Cascade complex must first recognise the MGE which is complementary to its crRNA spacer sequence. Initially, the N-terminal domain of the Cse1 subunit scans dsDNA for PAM sites⁷⁵ (5' AAG). Once the PAM is successfully recognised, the Cas7.5 and Cas7.6 subunits bind to the target DNA⁷⁶ strand close to the PAM to check if the crRNA matches the PAM-targeted DNA sequence⁷⁷. This takes place by matching a 'seed' sequence or 7-8 nt of the crRNA to the target complementary MGE

sequence adjacent to the PAM. If there is a match, one of the Cse2 subunits undergoes a conformational change, which allows the C-terminal domain of Cse1 to pry the MGE DNA strands apart, which allows the crRNA to invade the target strand, forming an R-loop. For successful interference PAM recognition, ≥ 5 nt seed region recognition, and an R-loop of at least 18 bp are required⁷⁸; this DNA/crRNA matching is summarised in **Figure 9**. Cse1 displaces the non-target MGE strand and recruits Cas3 via Cse1 Helix 1 with ATP^{79, 69}. This initial binding of Cas3 has a ~6 second half life and requires a stable R-loop before Cas3 can form a more stably bound formation⁷⁷. Cas3 is then able to nick the ssDNA of the displaced strand, allowing Cas3 to translocate in a 3'-5' direction along the non-target MGE strand, degrading it with the exonucleolytic activity of its HD domain⁸⁰. This is achieved by unwinding 3 bp followed by three individual 1 nt nuclease cuts in a reeling fashion, providing the nuclease domain with ssDNA substrates⁸¹. Cascade can remain bound to the DNA after Cas3 has begun degrading the MGE. In addition to this PAM-dependent interference mechanism, a PAM-independent pathway can degrade an MGE if the protospacer matches the spacer. Cas1-2 then help to load Cas3 to the Cascade-protospacer site, which can translocate in either direction (though still 3'-5')⁷⁷.

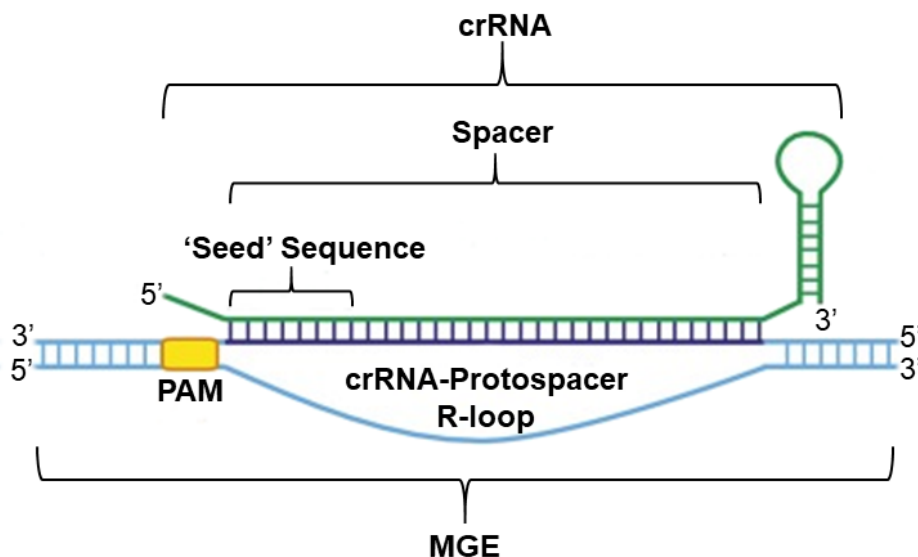


Figure 9 crRNA bound to complimentary MGE. The crRNA spacer sequence has bound successfully to the corresponding 32 nt of the MGE, so an R-loop is formed. The 8 nt 'seed' and PAM sequences which are vital for Cascade binding and Cas3 recruitment are shown.

The Cascade complex and Cas3 interference protein are not shown. Modified from Hochstrasser *et al.* 2014⁶⁹.

1.5 Cas1-2 and DNA Replication

The CRISPR-Cas immune system is reliant on more than just Cas proteins, various host factors also play a role. H-NS is a 15 kDa protein which represses the expression of cas genes in *E. coli* MG1655^{82, 26}. Another essential protein is IHF, which manipulates the CRISPR locus so as to allow Cas1-2-mediated spacer integration⁵⁸. As discussed earlier (1.2.2), proteins involved in DNA replication/repair are also vital for CRISPR-Cas function; specifically the RecBCD complex in naïve acquisition. RecBCD helicase⁴⁵, though not nuclease⁴¹, activity is essential for the prespacer production alongside Cas1-2. RecBCD is required for repairing double stranded DNA breaks in *E. coli* and is also essential for aiding the termination of DNA replication⁸³.

1.6 Ectopic Replication Origins

DNA replication is a tightly regulated process in which two identical copies of the chromosome are synthesised by the 'replisome' complex; this process is essential for cells to reproduce⁸⁴. In *E. coli*, DNA replication commences at the origin of replication, *oriC*, and continues in a bi-directional manner creating two 'replichores'. Replication starts with the protein DnaA which melts base pairs, 'opening up' the origin to load the multi-subunit replisome complexes, forming two replication forks^{85, 86}. The replisome is made up of two Pol-III cores with β -clamps for processivity (for leading and lagging strand synthesis), which are attached to a central multiprotein clamp loader, which is also bound to the helicase DnaB; primase and SSB proteins aid the replication of the lagging strand⁸⁷. DNA replication continues in a 5'-3' direction until the two forks meet at termination (*Ter*) regions. These *Ter* sites are found on each replichore, so have directional specificity with the replication fork: *TerA*, *TerD*, and *TerE* are positioned on the right-hand replichore to arrest clockwise replication fork; whilst *TerB*, *TerC*, and *TerF* arrest the anticlockwise replication fork on the left-hand replichore⁸⁸. There are four additional *Ter* sites found on the *E. coli* chromosome

(*TerK*, *TerL*, *TerY* and *TerZ*), but these have weak fork-arresting abilities⁸⁹. As the *TerA/C* sites are the furthest from the origin, therefore of equal replication distance, most forks meet at these sites. At these *Ter* sites, terminator utilising substance (Tus) protein binds. When forks approach Tus bound to *Ter* sites from the permissive direction (ie *TerE* for the clockwise fork), the replication machinery dislodges Tus and proceeds. When the fork reaches the opposite facing Tus (i.e., has moved over halfway around the chromosome into the opposite replicore) Tus stalls the fork so it cannot continue, allowing the slower fork to meet it⁹⁰. When the forks meet, they briefly pass beyond each-other, creating overlapping sections of replicated DNA. SbcCD and ExoI cleave the over-replicated regions⁹¹ and the RecBCD complex is then employed to degrade these cleaved intermediates and join the strands together, completing the replication of DNA⁹². These degradation products may be used as a source of prespacers by Cas1-2. Termination of replication is summarised in **Figure 10**.

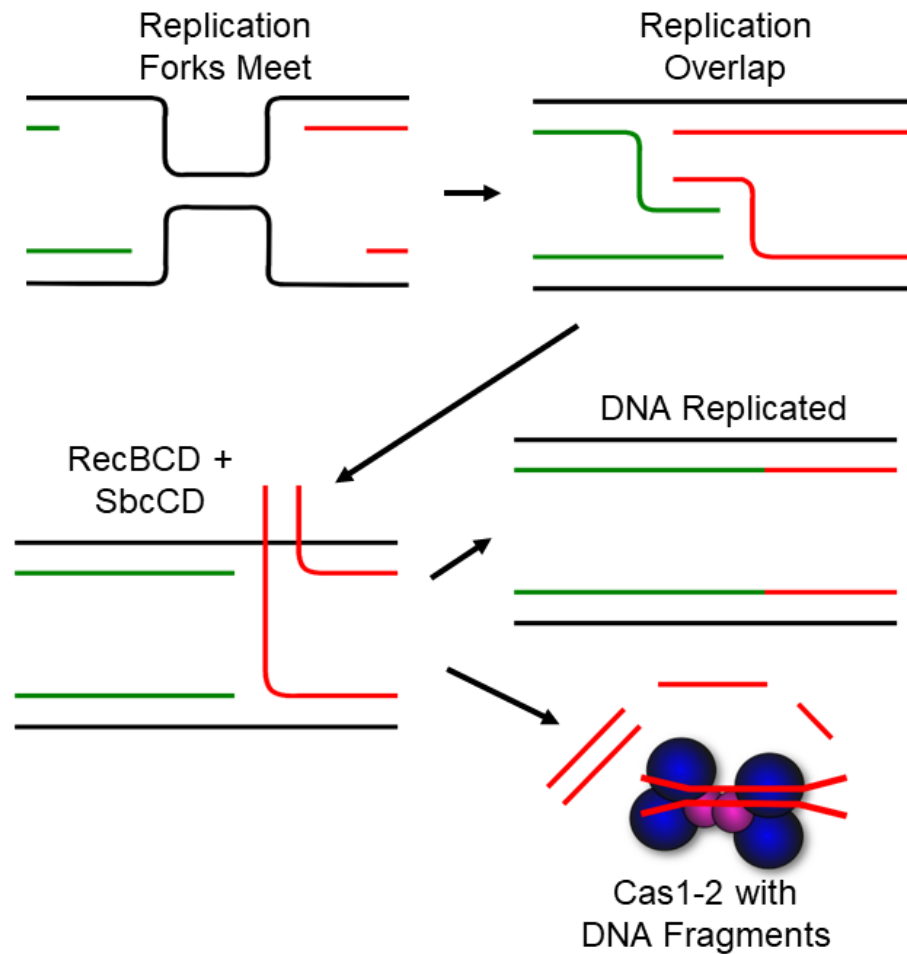


Figure 10 Model of replication forks meeting, terminating DNA replication. Initially, the two replisomes meet. They over-replicate slightly, duplicating part of the chromosome and leading to strands overlapping. The red strands are displaced, then cut with SbcCD and processed by RecBCD. This produces two complete non-overlapping strands of DNA and fragments of the duplicated DNA, which may act as a pool of pre-spacers for Cas1-2 to process into protospacers. Adapted from Courcelle *et al.* 2015⁹².

When ectopic origins are added, such as with the *oriC oriZ* strain RCe504⁹³, the chromosome is replicated by two sets of replication machinery (**Figure 11**). This addition of an additional origin on the right-hand replicore of *E. coli* MG1655 has several effects on replication. As there are only between 10-20 replisomes per cell⁹⁴, the time taken to replicate is slightly increased due to fewer available replication complexes available at DnaA-melted origins. The extra set of forks also means there is another location for forks to collide with one another, where the clockwise *oriC* fork meets the anticlockwise *oriZ* fork; the location of

this fork-fusion is not determined by fork-trap *Ter* sites. Furthermore, the anticlockwise *oriZ* fork travels in the opposite orientation to the wild-type *oriC* fork which it meets, meaning it can collide with RNA-polymerase complexes⁹³ which are traveling in the opposite orientation, such as *rrnH* (a ribosomal DNA operon) which then required RecBCD processing to re-start the replication fork^{95, 96}.

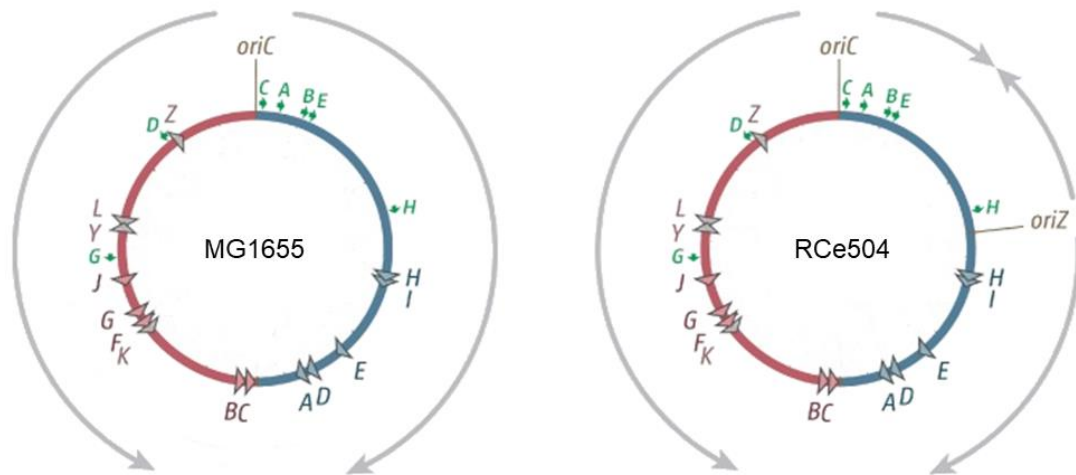


Figure 11 Replichore arrangement of *E. coli* K-12 strains MG1655 and Rce504. The MG1655 has one origin, *oriC* and the directions of replication are shown as grey arrows. Rce504 has an extra origin in the right (blue) replichore, *oriZ* which means that two additional replication forks meet, shown by the grey arrows. *Ter* sites are shown in red/blue, depending on the corresponding replichore, with arrows denoting which direction the replisome can pass through. Highly-transcribed ribosomal operons (*rrnA-E/G-H*) with direction of transcription shown as green arrows. Adapted From Dimude *et al* 2016⁹⁶

1.7 The CRISPR System of *Marinitoga piezophila*

A secondary part of this MRes work was to investigate acquisition in *Marinitoga piezophila*. *M. piezophila*, discovered in 2002, is a rod-shaped thermophilic bacterium which thrives at high pressure; hence the name 'piezophila' which is Greek for 'press'⁹⁷. As with around 45 % of all known bacteria, *M. piezophila* has a CRISPR-Cas immune system. The presence of three CRISPR arrays and Cmr complex mean the CRISPR system could be defined as type III-B, however there are also primase-polymerase (for mPrim-Pol protein) and argonaute (for

mAgo protein) genes present alongside the cas genes⁹⁸, shown in **Figure 12**. The importance of these genes and if/how they interact with CRISPR-Cas is currently unknown, so investigating their possible roles in naïve spacer acquisition was undertaken.

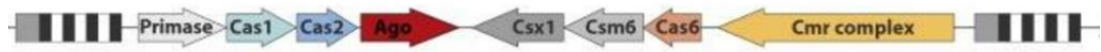


Figure 12 The hybrid mAgo/mPrimPol CRISPR-Cas Type III system of *Marinitoga piezophila*. cas/additional genes are shown as arrows, CRISPR arrays shown as striped boxes with grey-box leaders. An additional CRISPR array has been removed from the figure. From Kaya *et al.* 2016⁹⁸.

Primase proteins, such as DnaG, are involved in generating short RNA chains to act as primers for DNA synthesis to begin or to recommence after a replication fork has stalled⁹⁹. PrimPol nucleotidyltransferase proteins are part of the archeo-eukaryotic primase (AEP) superfamily which, like the DnaG primases, are used for commencing DNA replication by providing a double-stranded primed template. They have also been linked to non-homologous end joining (NHEJ) in bacteria¹⁰⁰ and to initiation of plasmid replication in archaea¹⁰¹. In addition to this, archaea PrimPol proteins have been implicated in translesion synthesis, where PrimPols temporarily replace the native polymerase to synthesise DNA over the lesion, filling the gap¹⁰². This is a similar role to PolA in spacer integration, so mPrimPol could be playing this CRISPR-Cas spacer acquisition role in *M. piezophila*.

Argonautes are proteins found in all domains of life. In eukaryotes Argonaute proteins have a P element-induced wimpy testis (PIWI) domain which has RNase H like catalytic activity; an N-terminal domain for cleaving and dissociating targets; a PIWI-Argonaute-Zwille (PAZ) domain for 3' RNA guide binding; and a middle (MID) domain for 5' RNA guide binding¹⁰³. These eukaryotic Argonautes are associated with RNA interference (RNAi) pathways, where they bind to 5'-phosphorylated small interfering RNAs (siRNA, created by Dicer proteins) and micro RNA (miRNA) to form RNA induced silencing complexes (RISC). These target specific mRNA sequences for degradation, providing additional translation regulation¹⁰⁴. Prokaryotes lack these RNAi pathways, so their Argonaute functions remain elusive. As the

argonaute gene in *M. piezophila* is located amongst the *cas* genes, these Argonaute proteins may have host defence functions by binding specific DNA/RNA sequences and using them as guides to target MGEs¹⁰⁵. Unlike Argonaute proteins in eukaryotes, the *M. piezophila* Argonaute protein has a preference for small 5'-hydroxylated RNA guides⁹⁸, similar to the crRNA produced by Cas6. Additionally, mAgo uses these guides to target ssDNA and ssRNA, but not dsDNA like CRISPR effector modules¹⁰⁶.

1.8 The Stringent Response

An additional aspect of this MRes involved a genetic interaction between CRISPR-Cas and a host stress response. This genetic analysis followed-on from work started a decade ago in this lab, but was left incomplete, which hinted at a possible link between Cas3 and the stringent response. This stress response, like several other stress responses, initiates transcriptome reprogramming which help the host survive adverse conditions. The stringent response is triggered by starvation in *E. coli* and is controlled by the small 'alarmone' (p)ppGpp. Guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) were first implicated in a starvation response in 1973 when *E. coli* cells were starved of essential amino acids; ppGpp and pppGpp accumulated in the presence of an un-charged tRNA bound to the ribosome. Uncharged tRNAs accumulate at the ribosomal A site when amino acid levels are low, meaning they are unbound to their respective amino acid¹⁰⁷. Furthermore, it was noticed that no (p)ppGpp did not accumulate in a $\Delta relA$ strain, indicating that (p)ppGpp is synthesised by RelA¹⁰⁸.

In *E. coli* (p)ppGpp is synthesised by RelA, an 84 kDa ribosome-associated synthase enzyme with a Mn^{2+} dependent HD-nuclease domain; and broken down by SpoT, a 79 kDa cytosolic hydrolase. RelA/SpoT are often part of the same protein, Rel-Spo-Homologue RSH proteins¹⁰⁹. These regulate the amount of the regulatory molecule (p)ppGpp, called an 'alarmone', in a reversible reaction between GTP and GDP with an ATP-phosphate-donor. Whilst most often triggered in response to amino acid starvation, (p)ppGpp can also be increased in response to antibiotics, heat shock and fatty acid limitation¹¹⁰.

The alarmone (p)ppGpp globally alters cell physiology in several ways. One way ppGpp helps the host survive harsh conditions is by binding to the RNA-polymerase (RNAP) between the RNAP 'core' and 'shelf' modules¹¹¹. This affects the opening of the RNAP 'cleft' which means that UvrD can mediate RNAP backtracking for nucleotide excision DNA repair more easily as the energy required to move the RNAP is lowered. This allows ppGpp to play a role in maintaining genome integrity¹¹². ppGpp can also work with DskA to help reprogram the transcriptome by binding to the RNAP to decrease the lifetime of RNAP open-complexes. One such re-programming involves inhibiting RNAP recognition of ribosomal RNA and tRNA promoters; this is achieved by reducing the stability of the RNAP-promoter open complex. Overall this reduces the rate of ribosome production, reducing protein synthesis, helping the bacteria to conserve energy¹¹³. Whilst ppGpp-DskA can cause RNAP to inhibit transcription of ribosomes, they have also been shown to help increase RNAP transcription of amino acids via biosynthetic promoters, though this is less well understood¹¹⁴. As well as conserving energy by altering protein expression, ppGpp can negatively regulate DNA replication initiation by interacting with SeqA, helping SeqA remain bound to origin sequences, which can prevent DnaA from initiating DNA replication¹¹⁵. Further targets for ppGpp include purine nucleotide homeostasis pathways via YdgH, metabolism regulation via HypB, and ppGpp homeostasis via MutT¹¹⁶.

1.9 Aims of the Project

This MRes was primarily made up of genetic analyses of CRISPR-Cas using *E. coli* as the host organism. One of the aims of this project was to investigate the effect ectopic origins of replication have on naïve acquisition *in vivo* in *E. coli*. As additional origins of replication mean having additional replication forks, it also means additional replication fork collisions, as well as chances for replication machinery to collide with transcription machinery. This could lead to additional substrates for RecBCD/Cas1-2 to capture as pre-spacers, which could increase the level of naïve acquisition seen in additional origin strains compared to wild-type. To test this hypothesis, naïve acquisition assays in MG1655 and RCe504 were set up to determine if more colliding replication forks means more naïve acquisition seen. An

additional aim involving naïve acquisition in *E. coli* was to investigate the effect fluorescent tags can have on the acquisition functions of Cas1 *in vivo*.

Aside from type I-E CRISPR acquisition investigations, naïve acquisition with *M. piezophila* Cas/mAgo/mPrmpol proteins was also tested *in vivo*. The aim of this was to see if the mAgo and mPrmpol are required for acquisition by performing acquisition assays using mCas1-2 only, mCas1-2+mPrmpol+mAgo, and mPrmpol+mAgo only. This could determine if these additional non-Cas proteins are located within the CRISPR locus for normal CRISPR function, or perhaps another reason.

The final aim of the project was to investigate the possible role Cas3 could have in the stringent response. As a protein with an HD-nuclease domain like RelA, Cas3 could be able to synthesise ppGpp, or perhaps enhance ppGpp formation with RelA. To test this, strains with combinations of *relA* deletions, *spoT* deletions and *cas3* deletions were tested *in vivo* for their ability to produce ppGpp.

Chapter 2: Materials and Methods

2.1 Chemicals

All of the chemicals used throughout were supplied by Fisher Scientific or Sigma unless stated otherwise.

Table 2-1 Antibiotics used in this study. All dissolved in sterile distilled water (SDW) apart from: * dissolved in Ethanol; † dissolved in Methanol.

Antibiotic	Stock Concentration	Working Concentration
Ampicillin (Invitrogen)	100 mg/ml	100 µg/ml
Apramycin	50 mg/ml	50 µg/ml
Chloramphenicol*	35 mg/ml	35 µg/ml
Kanamycin	50 mg/ml	50 µg/ml
Rifampicin†	10 mg/ml	10-30 µg/ml
Spectinomycin	100 mg/ml	100 µg/ml
Streptomycin	100 mg/ml	100 µg/ml
Tetracycline* (PanReac AppliChem)	15 mg/ml	15 µg/ml

2.2 Bacterial Strains

Table 2-2 *E. coli* strains used in this study

<i>E. coli</i> Strain	Genotype	Source or Reference
BL21-AI	<i>F⁻ ompT hsdS_B (r_B⁻ m_B⁻) gal dcm araB::T7RNAP-tetA</i>	Invitrogen
BL21-CodonPlus (C+)	<i>F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm⁺ Tet^r gal endA Hte [argU ileY leuW Cam^r]</i>	Agilent Technology
DH5α	<i>F⁻ ϕ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(r_k⁻, m_k⁺) phoA supE44 thi-1 gyrA96 relA1 λ⁻</i>	Invitrogen
MG1655	<i>F⁻ λ⁻ rph⁻¹</i>	Bachmann 1996 ¹¹⁷
RCe504	MG1655 <i>oriZ-cat-frt</i>	From Dr Christian Rudolph ⁹³
EB304	MG1655 <i>Δcas3::apra</i>	From Dr Ed Bolt
EB305	MG1655 <i>Δcas3::apra ΔrelA::kan</i>	From Dr Ed Bolt
EB306	MG1655 <i>ΔrelA::kan</i>	From Dr Ed Bolt
EB309	MG1655 <i>ΔspoT</i>	From Dr Ed Bolt
EB310	MG1655 <i>Δcas3::apra ΔspoT</i>	From Dr Ed Bolt
EB371	MG1655 <i>araB::T7RNAP-tetA</i>	P1.BL21-AI x MG1655
EB372	RCe504 ⁹³ <i>araB::T7RNAP-tetA</i>	P1.BL21-AI x RCe504 ⁹³

2.3 Plasmids and Substrates

Table 2-3 Plasmids used in this study

Plasmid Name	Description
pAH4	<i>E. coli</i> cas3 subcloned into pBad via XhoI and EcoRI. From Dr Alexandra Hughes.
pEB505	<i>E. coli</i> cas1 cloned into pET14b via NdeI and EcoRI. N-terminal His-tagged. From Dr Ed Bolt.
pEB625	<i>E. coli</i> cas1-2 cloned into pET-Duet: cas1 via NdeI and XhoI; cas2 via NcoI and NotI. From Dr Ed Bolt.
pEB628	<i>E. coli</i> cas1-2 cloned into pBad via NcoI and XhoI. From Dr Ed Bolt.
pEB680	<i>E. coli</i> cas2 cloned into pACYC-Duet via NcoI and BamHI. From Dr Ed Bolt.
pEB681	<i>E. coli</i> cas2 from pEB680 subcloned into pLau18 YFP-Cas1. From Dr Ed Bolt.
pKD46	Red Recombinase expression plasmid. Datsenko & Wanner, 2000 ¹¹⁸
pLau18 YFP-Cas1	<i>E. coli</i> cas1 N-terminally fused to Yellow Fluorescent Protein. From Dr Christian Rudolph.
pPK250	<i>M. piezophila</i> CRISPR locus in pET-28a. From Prof. Aiden Doherty.
pPK273	His-tagged <i>M. piezophila</i> cas1/2, in pACYC-Duet. From Prof. Aiden Doherty.
pPK276	His+SUMO-tagged <i>M. piezophila</i> primase-polymerase, with His-tagged <i>M. piezophila</i> argonaute, cloned into pET-28a. From Prof. Aiden Doherty.
pRPR1	<i>M. piezophila</i> CRISPR locus from pPK250 subcloned into pCDF-1b via NcoI and BamHI.

pTK7	N-terminal flag-tagged <i>E. coli cas1</i> + wt <i>cas2</i> cloned into pBad. From Dr Tom Killelea.
pTK10	pET19b with <i>E. coli cas2</i> , which contains a C-terminal His-tag. From Dr Tom Killelea.
pACYC-Duet	T7 promoter, Chlm ^R , N-His ₆ tag, p15A origin.
pBad-HisA	<i>araBAD</i> promoter, Amp ^R , N-His ₆ tag, pBR322 origin.
pCDF-1b	T7 promoter, Spect ^R , N-His ₆ tag, CloDF13 origin
pET-28a	T7 promoter, Kan ^R , N-His ₆ tag, ColE1 replicon.
pET-Duet	T7 promoter, Amp ^R , N-His ₆ tag, ColE1 replicon.

2.4 Primers

Table 2-4 DNA oligonucleotide Primers used in this study

Name	Sequence (5'-3')	Reference	Description
BL21-AlacquiF2	AGTTGGTAGATTGTGACTGGCT	Dr Ed Bolt	Detection of new spacer acquisition into <i>E. coli</i> BL21AI CRISPR locus.
BL21-AlacquiR2	TTTGCGTTTCGTTTCAGGTCG		
Marintoga F	ATCTACGAGGTTCCGACAGG	This study	Detection of new spacer acquisition into <i>M. piezophila</i> CRISPR locus from pRPR1 plasmid.
Marintoga R	TGTCACCAGTCATGCTAGCC		
SW1	AGCTGATCITTTAATAAAGGAAT	Stephen Wearne	Detection of new spacer acquisition into <i>E. coli</i> MG1655 CRISPR-I locus.
SW2	CGGATTATTAAGCTGACGGTT		
MG cas1/2 KO F	CCTTAATGTAAACATTTCTTATTATTAAAGATCAGCTAAT TCITTTGTTTCAACAGGATACAAGGGGTGTATGAGC	This study	Used for deleting <i>cas1-2</i> genes from <i>E. coli</i> MG1655 and replacing with a kanamycin resistance cassette, via Red Recombineering.
MG cas1/2 KO R	CTTGCTATCTTTGGCTCCACTGTGATTGAGGTGAATAA AAAGACCACGTTTACAAATTCAGGTGG		
EB628 In MG F	TTGACTCGGCTCATTTATTGCCCGCTTGAGACATTTGTTT CCATATGTACGCGAGTATACACTCCGCTATCG	This study	Used for inserting arabinose inducible <i>cas1-2</i> from pEB628 plus ampicillin resistance, into <i>E. coli</i> MG1655 via Red Recombineering.
EB628 In MG R	TTCATGGGGATGCTGAAAAGAGTAGTAATTGCTGTAA TGACTCCAACAACITGGTCTGACCCAGTTACC		
T7 RNAP FWD NcoI	CATGCCATGGATGAACACGATTAAACATCGCTAAG CATGCCATGGATGAACACGATTACATCGCTAAG	Jack Braithwaite	For detection of T7-RNAP gene, used to test P1-transduction strains for successful integration of T7-polymerase.
T7 RNAP REV HindIII	CATGAAGCTTTTACGCGCGGAACGCGAAGTC		
HYFPCas1F-NcoI	ATTACCATGGATGCATCATCACCATCATCAGTGTAATCT CGAGITGGTGAGGCAAGGGCGGAGGA	This Study	Used to add an N-terminal His-tag to YFP-Cas1 from plau18 YFP-Cas1. NcoI and XhoI cut-sites added for use with pET/pACYC-Duet.
HYFPCas1R XhoI	AGTCCTCGAGTGGCTCTAGATCAGCTACTCCGATGGCC TGCAATC		
Cas1-2 F NcoI	ATGCACCATGGATGACCTGGCTTCCCTTAATCC	Dr Tom Killelea	Used to detect <i>cas1-2</i> genes to ensure recombineering had knocked them out.
Cas1-2 R BamHI	ATGGCGAATTCTCAACACAGGTAAAAAAGACACCAACCTT AAACC		

2.5 Buffers, Solutions, Reagents, and Microbiological Media

Table 2-5 Buffers, solutions, reagents, and microbiological media used in this study

Media Name	Composition (in dH ₂ O unless stated otherwise)
LB Media	10 g/L tryptone (BD Biosciences) 5 g/L (w/v) yeast extract (BD Biosciences) 10 g/L NaCl NaOH to pH 7
5xM9 Minimal Salts	64 g/L Na ₂ HPO ₄ ·7H ₂ O (BDH AnalaR) 15 g/L KH ₂ PO ₄ 2.5 g/L NaCl 5 g/L NH ₄ Cl (BDH AnalaR)
M9 Media	20 % (v/v) 5xM9 Minimal Salts 2 mM MgSO ₄ 100 nM CaCl ₂
1.5 % Agar Plates	1.5 g Agar (VWR International)/100 ml LB/M9 Media
SOB Media	5 g/L yeast extract 20 g/L tryptone 10 mM NaCl 2.5 mM KCl 20 mM MgSO ₄ NaOH to pH 7.5
SOC Media	5 g/L yeast extract 20 g/L tryptone 10 mM NaCl 2.5 mM KCl

SOC Media (continued)	20 mM MgSO ₄ 20 mM D-glucose NaOH to pH 7.5
MC Buffer	100 mM MgSO ₄ 5 mM CaCl ₂
P1 Agar Plates	1.5 % LB agar 5 mM CaCl ₂ 0.15 % w/v glucose
HEPES Buffer	1 mM HEPES NaOH to pH 7.5
10X Tris Boric Acid EDTA (TBE) Buffer	121.1 g/L Trizma base 61.8 g/L boric acid 7.4 g/L EDTA
4x SDS Loading Dye	200 mM Tris-HCl pH 6.8 8 % (w/v) SDS 0.4 % (w/v) Bromophenol Blue 40 % glycerol
10xSDS PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) Running Buffer	30 g/L Trizma base 144 g/L glycine 10 g/L SDS
SDS Coomassie Stain	40 % (v/v) methanol 10 % (v/v) glacial acetic acid 0.5 g/L Coomassie Brilliant Blue R-250
SDS Destain	20 % ethanol 10 % acetic acid
Resuspension Buffer	20 mM Tris pH 7.5 500 mM NaCl 20 mM imidazole

Resuspension Buffer (continued)	10 % (v/v) glycerol 0.5 mM PMSF
Ni ²⁺ Column Charge Buffer	200 mM NiCl ₂
Ni ²⁺ Column Buffer A	20 mM Tris pH 7.5 500 mM NaCl 20 mM imidazole 10 % glycerol.
Ni ²⁺ Column Buffer B	20 mM Tris pH 7.5 500 mM NaCl 500 mM imidazole 10 % glycerol.
Ni ²⁺ - Heparin Column Dialysis Buffer	20 mM Tris pH7.5 150 mM NaCl 1 mM DTT 10 % glycerol
Heparin Column Buffer A	20 mM Tris pH 7.5 150 mM NaCl 10 % glycerol.
Heparin Column Buffer B	20 mM Tris pH 7.5 1 M NaCl 10 % glycerol.

2.6 General Microbiology

2.6.1 Making Chemically Competent Cells

A single colony from a plate of the desired cell strain was used to inoculate 5 ml of LB media; this was then grown overnight at 37°C with rotation (New Brunswick-TC7). 500 µl of the overnight culture was used to inoculate 50 ml of fresh LB media in conical flask. This was grown at 37°C in a shaking water bath (220 RPM, Fischer Scientific) until the OD₆₀₀ reached 0.6 (OD₆₀₀ was measured using a Beckman Coulter DU 530). Once at the correct OD₆₀₀, the cells were transferred to a Falcon tube (Sarstedt) and pelleted at 2,000 x g, 4°C for 10 minutes (using an Eppendorf Centrifuge 5430 R). After this, the supernatant was discarded and 6.25 ml chilled 0.1 M CaCl₂ was used to gently re-suspend the pellet; this was then left on ice for 2 hours. After this incubation, the cells were pelleted again as before and resuspended in 6.25 ml chilled 0.1 M CaCl₂, with Glycerol added to a final concentration of 30 %. Finally, the cells were aliquoted (100 µl-400 µl), flash frozen with dry-ice, and stored at -80°C.

2.6.2 Transformation of Chemically Competent Cells

A 100 µl aliquot of competent cells was thawed on ice. To this, 5-100 ng of the desired plasmid was added, and mixed by gently tapping the tube; this mixture was then left to incubate on ice for 30 minutes. After this initial incubation, the cells were subjected to a 90 second heat-shock at 42°C in a water bath (Grant Sub). The cells were then left on ice for a further 5 minutes. 900 µl LB media was then added to the cells, which were then incubated for 1 hour at 37°C in a shaking water bath. The cells were pelleted at for 1 minute at 16,000 x g (using an Eppendorf Centrifuge 5424), had 800 µl supernatant discarded, and were re-suspended in the remaining 200 µl. Between 50-200 µl of the cells were spread onto LB-agar plates (1.5 % w/v agar) containing the appropriate antibiotic to select for the desired plasmid (see 'Table 2-1 Antibiotics used in this study' for details of antibiotics used).

2.6.3 Isolation of Plasmid DNA

Plasmid DNA was extracted from cells (*E coli* DH5α cells unless otherwise stated) and purified using QIAprep Spin Miniprep Kit. The protocol used can be found here ¹¹⁹. The

concentration of the plasmid DNA was measured using a Thermo NanoDrop 2000 Spectrophotometer.

2.6.4 Agarose Gel Electrophoresis

Before loading onto a gel, 6xPurple Loading Dye (NEB) was added to the DNA samples and gently mixed. Agarose powder was dissolved with heating in 1xTBE buffer, the % w/v added depending on the DNA being analysed: 1 % w/v Agarose-TBE gels were used for DNA sequences greater than 2000 bp; 2 % Agarose-TBE gels were used to analyse DNA sequences below 2000 bp. 0.2 µg/ml Ethidium Bromide was added to the gel before setting. 12 µl of the DNA sample was loaded into the wells, with 4 µl pre-stained ladder as a guide (NEB). Gels were run at 100 V for 110 minutes using a PowerPac Basic power supply (BioRad), and imaged using a U:Genius 3 (Syngene International) UV transilluminator

2.6.5 DNA Extraction from Agarose Gel

DNA bands were extracted from agarose gels using the QIAquick Gel Extraction Kit, the protocol for which can be found here¹²⁰. A Stratagene Transiluminator 2040 EV was used to view the bands for excision, keeping UV exposure to ≤10 seconds to minimise DNA damage.

2.6.6 PCR 'Clean-up'

DNA products from PCR reactions were purified using the QIAquick PCR Purification Kit, the protocol for which can be found here¹²¹.

2.6.7 Colony PCR

A Single colony was added to 50 µl SDW, re-suspended, and boiled at 95°C for 10 minutes. 2 µl of this sample was used for PCR.

2.6.8 Overexpression of Protein

Overnight cultures were set up in LB and subsequently used to inoculate 50 ml of fresh LB at a 1:100 dilution. These cultures were grown to OD₆₀₀ 0.6, at which point they were induced with 0.2 % L-arabinose and 0.1 mM IPTG. The proteins were left to express for 4

hours; then samples were taken, re-suspended in SDW with 4xSDS loading buffer + 30 mM DTT, boiled for 10 minutes,

2.6.9 SDS PAGE

Protein samples had 4xSDS Loading Dye and 30 mM DTT added, which were subsequently boiled at 95°C for 10 minutes in a Techne Dryblock DB100/2. The samples were then loaded onto an SDS gel with a 5 % acrylamide stacking layer, and a 15 % resolving layer unless otherwise stated. Broad-Range Unstained Ladder (NEB) was loaded to measure protein size. The SDS gels were run using the Mini Protean II system (BioRad), with 120 V supplied by a PowerPac Basic (BioRad) for 90 minutes. Once the protein samples had fully migrated down the gel, the gels were stained using Coomassie Brilliant Blue stain for 30 minutes, before being left in de-stain buffer overnight.

2.7 Genetic Manipulation

2.7.1 P1 Transduction to Produce EB371/EB372

P1 transduction of *E. coli* MG1655/RCe504 was required to express *cas1-2* from the pEB628 plasmid; as BL21-AI successfully expressed these proteins using this plasmid, the arabinose-inducible T7-polymerase was transduced from this strain into the MG1655/RCe504 strains. P1 transduction method was modified from¹²². Overnight culture of BL21-AI was used to inoculate fresh LB containing 5 mM CaCl₂ at a 1:50 dilution. Cultures were grown to OD₆₀₀ 0.8, then added to 3 overlay tubes containing 3 ml of molten 0.4 % LB agar, at a 30-fold dilution. P1 phage stock (roughly 10⁸ pfu/ml), diluted 10-fold in MC buffer, was then added to the culture-containing soft-agar tubes; 0.1 ml phage stock added to two of the plates to form P1 lysates, the third left as a no-P1 control. The molten agar-culture-P1 mix was then poured on top of P1 agar plates and left for 16 hours at 37°C. Lysed cultures were scraped from the P1-containing plates into a 15 ml Falcon tube, whereby 1 ml of MC buffer and 0.5 ml of chloroform was added. The tubes were vortexed for 30 seconds using a WhirliMixer (FSA Laboratory Supplies) then centrifuged at 6,500 x *g* for 30 minutes at 4°C Eppendorf Centrifuge 5430 R. The supernatant was extracted and used as a BL21-AI.P1 lysate.

The recipient strains, MG1655 and Rce504, were grown overnight in 5 ml LB media. The overnight cultures were spun-down at 2,000 x *g* for 10 minutes at 4°C. The pelleted cells were then re-suspended in 1 ml of chilled MC buffer and left to incubate at room temperature for 10 minutes. 250 µl of cells were then added to 2 Eppendorf tubes per strain. Two of the tubes had 50 µl MC buffer added and were used as MG1655/Rce504 no-P1 controls, the other two tubes had 50 µl of the BL21-ΔI.P1 lysate added; the four tubes were incubated at 37°C for 30 minutes in a sealed beaker within a Carbolite incubator to allow phage infection. Infection was stopped by the addition of 200 mM Sodium Citrate and leaving at room temperature for 5 minutes. The infected cells were then mixed with 2.5 ml of 42°C molten LB agar (0.6 % w/v agar) and poured onto the top of standard LB agar plates containing 7.5 µg/ml tetracycline (50 % of the usual tetracycline due to the sodium citrate). The plates were incubated for 24 hours at 37°C.

Successful transductants were re-streaked onto plates containing 7.5 µg/ml tetracycline and 5 mM Sodium Citrate, and incubated at 37°C for 12 hours. This process was repeated to ensure no P1 was remaining in the transduced colonies. After this second re-streaking step, a single colony from each strain was verified by colony PCR using T7-RNAP primers.

2.7.2 Red Recombineering

Red recombineering was utilised to remove the need for plasmids (pEB628) from the naïve acquisition assays. This would remove the plasmid-instability issues as well as removing a source of MGE prespacers, leaving only the chromosomal sources of prespacers.

Chromosomal *cas1-2* was to be replaced with a kanamycin resistance cassette (MG *cas1/2* KO primers), then replaced with an arabinose-inducible *cas1-2* with ampicillin resistance from pEB628 (EB628 In MG primers) Red recombineering method for gene insertions/deletions based on¹¹⁸.

PCR primers were designed with 20 bp of homology to the gene-of-interest to be inserted, and 50 bp of homology to the *E. coli* genome region of interest for insertion/replacement of gene. PCR was used to amplify genes of interest for insertion, and PCR 'clean-up' to purify the fragments.

EB371 and EB372 strains were transformed by pKD46 and overnight cultures grown in LB with ampicillin at 30°C. Overnight cultures were inoculated into flasks containing 40 ml SOB media (1:100 dilution) with 0.2 % L-arabinose; these were grown at 30°C in a shaking water bath until OD₆₀₀ reached 0.6. Once the required OD₆₀₀ was reached, the flasks were removed from the incubator and placed on ice for 20 minutes. The cells were poured into Falcon tubes and centrifuged at 4,000 x *g* for 5 minutes at 4°C. The cells were re-suspended in 1 ml ice-cold 1 mM HEPES pH 7.5, using chilled pipette tips, and transferred to an Eppendorf tube on ice. The 1 ml resuspended cells were then pelleted at 4,000 x *g* for 2 minutes at 4°C and re-suspended in 1 ml ice-cold 1 mM HEPES pH 7.5. This process was repeated twice, using chilled HEPES, pipette tips, and centrifuge. Finally, the cells were spun at 4,000 x *g* for 2 minutes at 4°C and re-suspended in 100 µl chilled 1 mM HEPES pH 7.5.

65 µl of the chilled cells were added to a pre-chilled electroporation cuvette with 500 ng (10-20 µl) cleaned PCR product added. These were briefly mixed by pipetting on ice, before being electroporated with a 1.7 mV pulse (BioRad Gene Pulser and Pulse Controller). The electroporated cells were immediately added to 700 µl SOC media in an Eppendorf tube, grown in a shaking water bath at 30°C for 2 hours, and spread onto LB agar plates containing antibiotics selective for the inserted/replacement gene (Kanamycin for *cas1-2* genetic knockout, Ampicillin for pEB628 *cas1/2* insertion).

Successful colonies were re-streaked onto antibiotic-containing LB agar plates and grown at 42°C to purge the pKD46 plasmid. Colonies grown after this stage were streaked onto new plates as a 'grid test', so multiple colonies could be tested for the successful genetic insertion/replacement via colony PCR (Cas1-2 F NcoI and Cas1-2 R BamHI primers) using fewer plates (10 colonies streaked per plate).

2.7.3 Subcloning of Genes

All subcloning was performed using NEB restriction enzymes, NEB restriction buffers and NEB T4 ligase according to manufacturer's instructions and NEBcloner software¹²³.

2.7.4 Cloning His-YFP-Cas1

A His₆ tag was added to the YFP-Cas1 from pLau18 YFP-Cas1 using the primers 'HYFPCas1R XhoI' and 'HYFPCas1F-NcoI'. pLau18 YFP-Cas1 was used as the template for the PCR. PCR product was 'cleaned-up' and cloned into pET-Duet (2.7.3). Plasmids produced were digested with NcoI and XhoI to ensure insert was present and run on a 1 % agarose TBE gel. Plasmids with the correct insert were transformed into BL21-AI and overexpression of protein induced (2.6.8) followed by SDS-PAGE (2.6.9).

2.8 Acquisition Assays

These experiments were used to detect naïve spacer acquisition into the CRISPR loci of BL21-AI, MG1655, RCe504, EB371, and EB372. Assays for detecting naïve spacer acquisition adapted from³⁰.

2.8.1 Passaging of Cells

Overnight samples were set up by inoculating a single colony into a test tube containing 3 ml LB with 0.1 % L-arabinose (pBad based plasmids) and 0.1 mM IPTG (pET/pACYC/pCDF based plasmids and BL21-AI strain). These cultures were incubated overnight at 37°C with rotation at around 60 RPM (New Brunswick TC7). After 12-16 hours, 150 µl of the overnight cultures ('Passage 1') were used to inoculate a baffled flask containing 45 ml of LB containing 0.1 % L-arabinose (+0.1 mM IPTG). These flasks containing the new cultures ('Passage 2') were grown to an OD₆₀₀ of 0.6, at which point 0.1 % L-arabinose (+0.1 mM IPTG) was added to induce extra expression of the Cas1/2 proteins. These cultures were incubated in a shaking water bath for 8-12 hours to allow time for acquisition to occur. After 8-12 hours, these cultures were used to inoculate 45 ml fresh LB with 0.1 % L-arabinose in a baffled flask as before (to produce 'Passage 3'). This process was repeated for up to 7 passages. 100 µl of each passage was also taken for PCR analysis at the same time as subculturing occurred, i.e. after 12 hours of growth when acquisition could be analysed.

2.8.2 Bacterial Growth Viability ('Spot') Tests

Samples taken from cultures during passaging were serially diluted in M9 media from 10^{-1} to 10^{-7} . 10 μ l of the diluted samples were spotted in pentaplicate onto LB agar plates containing appropriate antibiotics. These were incubated at 37°C for 16 hours (Carbolite incubator) and colonies counted to calculate colony forming units of the culture.

2.8.3 Detection of New Spacer(s) by PCR

Samples were taken for PCR analysis when passaging cell cultures (2.8.1). These 100 μ l passage samples were diluted 1:4 in sterile distilled water (SDW) and incubated at 95°C for 10 minutes. Samples were then frozen until ready for PCR. The composition and conditions of the PCR reactions are shown in **Table 2-6** and **Table 2-7** respectively. PCR reactions took place using a Veriti 96 well thermal cycler (Applied Biosystems). PCR products were run on a 2 % TBE agarose gel as described in section 2.6.4. Intensity between parental DNA bands and expanded DNA bands was measured using ImageJ software to calculate the percentage of acquisition observed.

Table 2-6 PCR reaction mix for new spacer detection.

Component	Final Concentration
Dream Taq Green 2x Master Mix	1x
Forward Primer	150 nM
Reverse Primer	150 nM
Sample	10 % v/v
Up to 20 μ l with SDW	

Table 2-7 PCR conditions for detection of new spacers acquired.

Temperature/°C	Time/Seconds	Cycles
95	180	1
95	30	30
44 (SW1/2), 56 (BL21Alacqui1/2, Marinitoga F/R)	30	
72	35	
72	300	1

2.9 Protein Purification Work

2.9.1 Overexpression of YFP-Cas1 and Cas2

Before proceeding with a full-scale overexpression followed by purification, individual pilot overexpressions of YFP-Cas1 and *E. coli* Cas2 were set up.

BL21-AI was transformed with the plasmid pLau18 YFP-Cas1 in order to overexpress YFP-Cas1; whilst BL21-AI transformed with pTK10 was used to overexpress Cas2. A positive control of BL21-AI transformed with pEB505 (His₆-tagged Cas1) in place of YFP-Cas1 was set up to ensure the purification method worked, which was purified in the same way as described below, with Cas2 biomass from pTK10. Cultures of these strains were separately set up in 5 ml LB with ampicillin in test tubes which were grown overnight at 37°C with rotation. These overnight cultures were used to inoculate 50 ml LB with ampicillin in a flask at a 1:100 dilution. These new cultures were grown in a shaking water bath at 220 RPM, 37°C, until the OD₆₀₀ reached 0.6. Once at the desired OD₆₀₀, protein overexpression was induced by the addition of L-arabinose to a final concentration of 0.2 %, followed 10 minutes later by the addition of IPTG to a final concentration of 1mM. The induced cultures were placed back in the shaking water bath, and YFP-Cas1 or Cas2 was left to express for 3.5 hours.

A 1 ml sample of each culture was taken, centrifuged at 16,000 x *g* in an Eppendorf Centrifuge 5242 and re-suspended in 150 µl SDW. These protein samples were tested for

successful overexpression by running an SDS-PAGE gel as previously described (2.6.9). If the overexpression was successful, the experiment was scaled-up and repeated.

For the scaled-up overexpression, overnight cultures were set up as before but in 50 ml flasks in a shaking water bath at 220 RPM, 37°C. The YFP-Cas1 overnight was inoculated at a 1:100 dilution into two 5 L flasks, each containing 2 L of LB with ampicillin. The Cas2 overnight culture was used to inoculate a single 5 L flask containing 2 L of LB with ampicillin, again at a 1:100 dilution. These different ratios reflect the stoichiometry of the Cas1-2 complex: Cas1₄:Cas2₂. These were grown with shaking at 37°C as before, the only difference being the incubator used (Kuhner Shaker ISF1-X). The cultures were induced at the same OD₆₀₀ as before, with the same concentrations of L-arabinose/IPTG, and left for the same amount of time to express.

The cell biomass was harvested by centrifugation at 2,000 x *g* for 10 minutes using an Avanti J-26 XP centrifuge (Beckman Coulter) with a JLA 10.500 rotor. Cells were re-suspended in 7 ml resuspension buffer per 400 ml centrifuged. Each of the YFP-Cas1 overexpressed cells were pooled together, added to two 50 ml Falcon tubes, and flash frozen for storage at -80°C. The Cas2 cells were also pooled together into a single Falcon tube, flash frozen, and stored at -80°C.

2.9.2 Coupled YFP-Cas1 and Cas2 Purification

Stored biomass was thawed on ice for 2 hours, then separated into 7 ml aliquots for lysis by sonication. Sonication occurred on ice using a Vibra Cell sonicator (Jencons). Sonication at 60-80 % amplitude was used for 10 seconds, with 10 seconds off, repeated for a total of 3 minutes per 7 ml aliquot. The lysed samples were then clarified by centrifugation at 21,000 x *g*, 4°C for 1 hour, using an Avanti J-26 XP centrifuge (Beckman Coulter) equipped with a JA 25.50 rotor. The supernatant of the YFP-Cas1 and Cas2 lysates were combined and gently mixed for 2 hours at 4°C with 15 RPM rotation (New Brunswick TC7).

The combined clarified lysates were loaded through a 5 ml HiTrap Ni-Chelating column (GE Healthcare) which was charged with Ni²⁺ Column Charge Buffer and pre-equilibrated with Ni²⁺ Column Buffer A, on an ÄKTA-Start FPLC system (Amersham Pharmacia). Elution of

bound proteins was achieved by using a 0-100 % Ni^{2+} Column Buffer B gradient over 5 column volumes. Fractions which contained eluted proteins, indicated by the elution trace given, were run on a 15 % SDS-PAGE gel to determine which fractions contained YFP-Cas1 and Cas2. Column flow-through and starting sample were also analysed on the 15 % SDS-PAGE gel. The YFP-Cas1 and Cas2 containing fractions were then pooled together into dialysis tubing (BioDesign). Dialysis took place overnight at 6°C with magnetic stirring in 3 L of Ni^{2+} - Heparin Column Dialysis Buffer.

Following the overnight dialysis, the samples were loaded onto a 1 ml HiTrap Heparin column (GE Healthcare) pre-equilibrated with Heparin Column Buffer A. Proteins were eluted using a 0-100 % Heparin Column Buffer B gradient over 20 column volumes. Again, fractions with eluted proteins were analysed on a 15 % SDS-PAGE gel with flow-through and starting sample lanes too.

2.10 Genetic Analysis of Interactions Between Cas3 and the Stringent Response

2.10.1 Bacterial Growth Viability ('Spot') Tests for Rifampicin

Resistance

Plate cultures of *E. coli* strains MG1655 (wt), EB304 ($\Delta cas3$), EB305 ($\Delta cas3 \Delta relA$), EB306 ($\Delta relA$), EB309 ($\Delta spoT$), and EB310 ($\Delta cas3 \Delta spoT$) were used to inoculate 5 ml LB with appropriate antibiotics. These cultures were grown overnight at 37°C with rotation. Samples were collected and diluted in M9 media, spotted onto LB agar plates with differing concentrations of rifampicin (10-30 µg/ml). The plates were incubated at 37°C and growth was compared. For complementation tests, EB304 and EB305 were transformed with pAH4 to re-introduce the *cas3* gene. Rifampicin spot tests using these strains took place using the same method.

Chapter 3: Results

3.1 Naïve Acquisition in Wild Type Vs Ectopic Origin

E. coli Strains

As stated previously (1.9) one of the aims of this study was to explore the roles Cas1 and Cas2 play in cell physiology, specifically with regards to prespacer capture. One aspect which was investigated, was that of Cas1-2 being linked to DNA replication, more specifically the roles that colliding forks and *Ter* sites may play in providing a source of prespacers for capture (1.6). It has been shown previously that spacers can be generated where DNA-replication forks stall⁴⁵. Therefore, if the frequency of fork collisions is increased, so too may be the spacers generated, increasing Cas1-2 mediated spacer acquisition. To test this, wild-type *E. coli* K-strain MG1655 was compared to RCe504 (a gift from Dr Christian Rudolph, Brunel University), a strain of *E. coli* which contains an additional origin of replication: as well as the *oriC* found in the wild-type chromosome, RCe504 also contains an ectopic '*oriZ*' which leads to additional colliding forks of DNA replication machinery⁹³.

Both strains were transformed with the plasmid pEB628 (pCas1-2), which contains *E. coli* *cas1* and *cas2* genes (Figure 14A). As both strains still have *cas* genes in their genomes, albeit under repression by H-NS, an empty vector control was also transformed into each strain to ensure any acquisition is due to the *cas1-2* from the plasmid, not the strain. Naïve acquisition experiments were carried out as described in 2.8. Strains were passaged four times, with each passage being left to grow/acquire spacers for 12 hours. L-arabinose was used to induce Cas1-2, thereby stimulating naïve acquisition.

Acquisition was detected via PCR (2.8.3): primers were used to amplify the region between the AT-rich leader sequence and the CRISPR spacers as new spacers are inserted proximal to the leader. This generated a 440 bp 'Parental' strand, which is an un-extended section of the CRISPR-I locus of *E. coli* MG1655 between the AT-rich leader sequence and the sixth CRISPR repeat. When acquisition occurs, a spacer of 33 bp along with a new 28 bp repeat

are integrated into the CRISPR locus adjacent to the AT-rich leader sequence. This increases the length of the CRISPR locus by 61 bp, which can be identified when run on a 2 % agarose-TBE gel stained with Ethidium Bromide. An overview of acquisition detection by PCR is summarised in **Figure 13**.

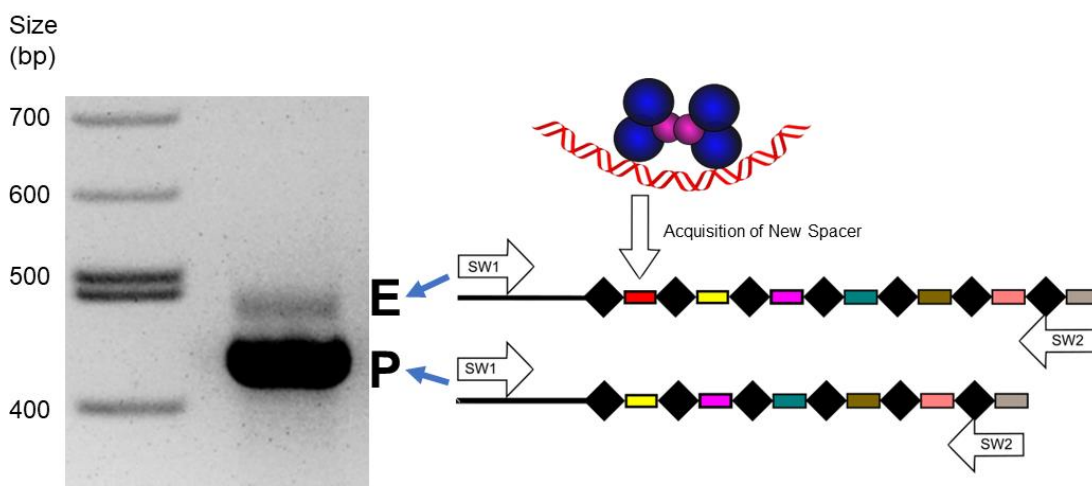


Figure 13 A diagram summarising the detection of naïve acquisition using PCR. A 2 % agarose-TBE gel showing an un-extended 440 bp ‘parental’ band, and an extended 500 bp band which is the result of acquisition. A cartoon of the Cas1-2 complex is shown integrating a new red-coloured DNA spacer into the CRISPR locus, leading to a new spacer (red DNA) and a new repeat (black diamond) being added to the CRISPR locus. The SW1 and SW2 primers are shown bound to the CRISPR locus, and their PCR amplification products are shown on the adjacent gel. ‘P’=parental locus band, ‘E’=expanded locus band.

This initial experiment demonstrated both MG1655 and RCe504 being unable to integrate new spacers using the pEB628 plasmid, summarised in **Figure 14B**. However, during parallel experiments investigating acquisition using a YFP-tagged Cas1 (**3.2**), with un-tagged Cas1 from pEB628 as a control, the BL21-AI with pEB628 successfully demonstrated acquisition (**Figure 21**). This was also shown previously by¹²⁴.

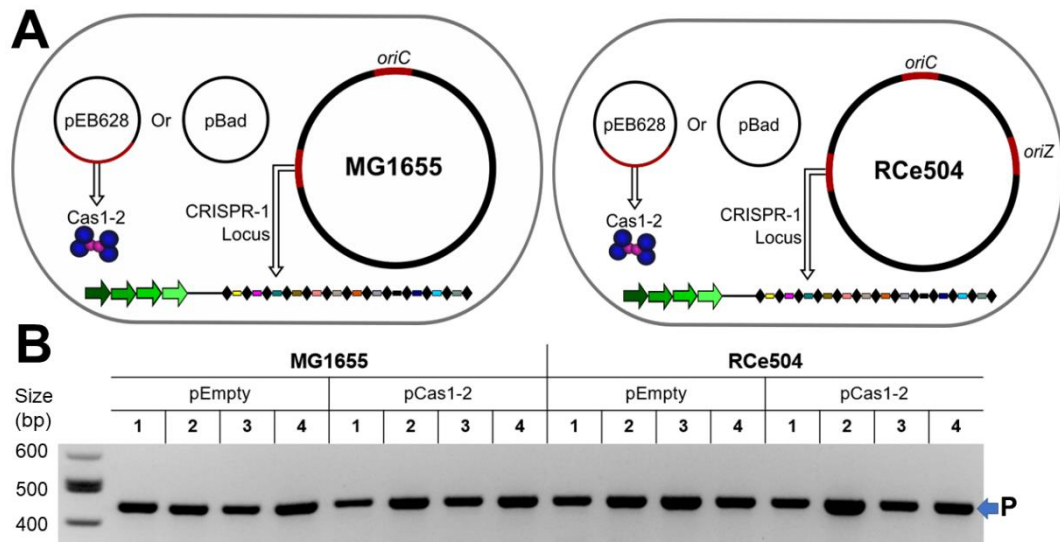


Figure 14 Naïve adaptation not being stimulated in MG1655 or RCe504 using pEB628.

(A) A diagram showing how naïve acquisition was investigated in MG1655 and RCe504.

The diagrams highlight the plasmids used, and the key difference between the chromosomes of both strains. **(B)** The outcome of the four-passages acquisition test. Only one DNA band, the parental 'P' band, was seen for each sample at around 440 bp. This corresponds to an un-extended CRISPR locus.

As acquisition had not been stimulated in MG1655 or RCe504, but had been stimulated in BL21-AI, P1 transduction was used to add an arabinose-inducible T7-polymerase to the two strains. The addition of the T7 polymerase would allow for better expression of *cas1-2* using T7-based vectors, leading to more Cas1-2 protein, leading to successful acquisition. The transduction of the arabinose inducible T7 polymerase with tetracycline resistance cassette from BL21-AI resulted in two new strains: EB371, based on MG1655; and EB372, based on RCe504.

After transduction, both strains were once again transformed with pEB628 and pBad, as well as the pET-Duet based pEB625 with the empty vector as a control. After transformations, acquisition assays were conducted exactly as with the MG1655/RCe504 strains described previously, but with only three 12-hour passages; the number of passages was reduced as acquisition was seen in BL21-AI clearly in passages 2 and 3, less so in passage 4. Despite being under the control of a T7 promoter, the acquisition stimulated

using the pEB625 (**Figure 15**) was far weaker than the arabinose-operon based pEB628 (**Figure 16**). Therefore, the comparison of acquisition between EB371 and EB372 was made using the pBad-based pEB628 (pCas1-2) with pBad (pEmpty) as the negative control. As can be seen in **Figure 16B**, the acquisition observed in the additional-origin EB372 was stronger than in EB371. Another key point is the pBad empty-vector controls showing no acquisition, so despite having *cas1-2* genes in their chromosomes, EB371 and EB372 use the pEB628 Cas1-2 for acquisition.

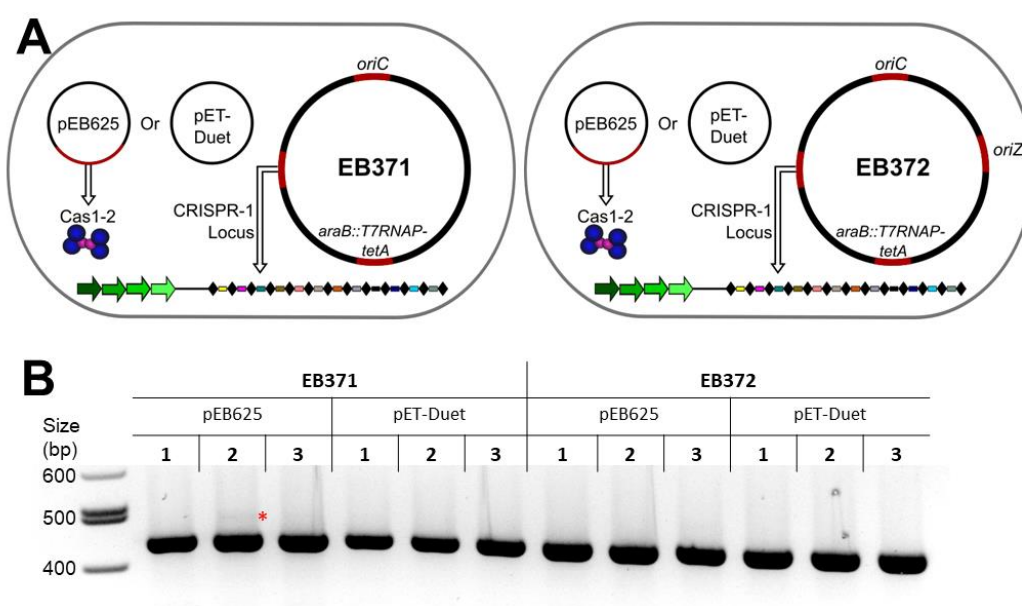


Figure 15 Naïve acquisition in EB371 and EB372 strains with pEB625. (A) A diagram showing how naïve acquisition using the EB371/EB371 strains and the pET-Duet based pEB625 (*E. coli cas1-2*). **(B)** A 2 % agarose-TBE gel showing CRISPR arrays amplified by PCR. PCR samples were taken at each passage, with passages every 12 hours. The red asterisk marks a faint 500 bp expanded CRISPR array. Expression of Cas1-2 induced with 0.2 % arabinose and 0.2 mM IPTG.

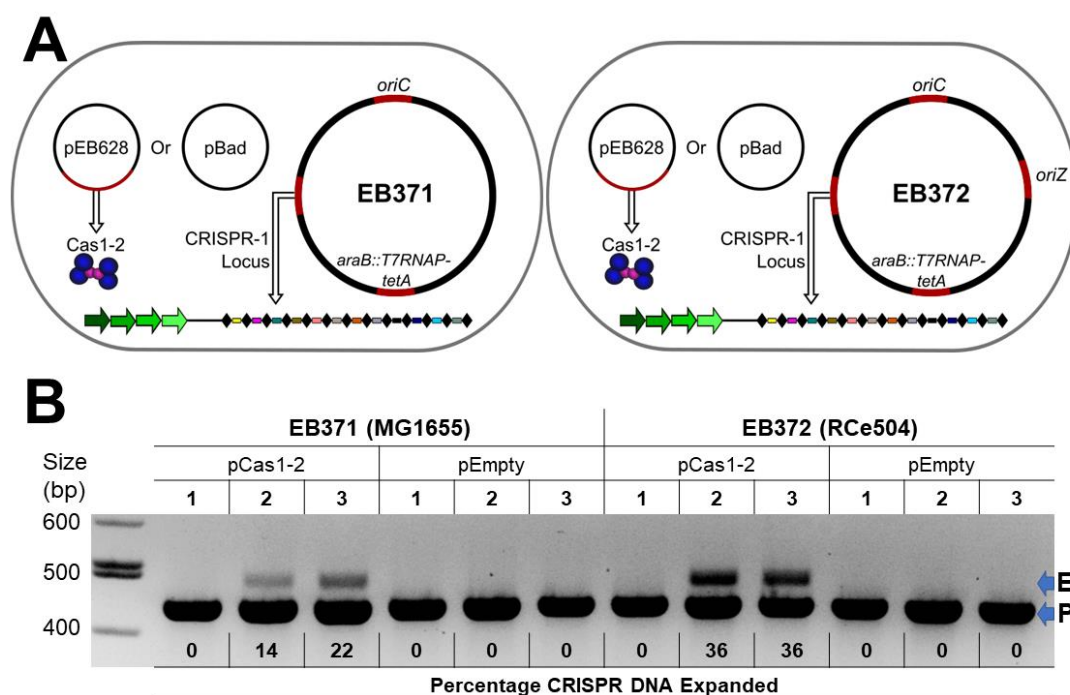


Figure 16 Naïve acquisition in EB371 and EB372 strains with pEB628. (A) A diagram showing how naïve acquisition was investigated in EB371 (MG1655) and EB372 (RCe504). The diagrams highlight the plasmids used, and the replication origin(s) on the chromosomes of both strains. **(B)** A 2 % agarose-TBE gel showing CRISPR arrays amplified by PCR. Samples were taken at each passage, with passages every 12 hours. The difference between the 440 bp parental band, 'P', and the 500 bp expanded band, 'E', is shown as 'Percentage CRISPR DNA Expanded' at the bottom of the gel.

After the initial 3-passage tests, additional passages were added to see if acquisition is consistently more active in the double-origin EB372 than in the single-origin EB371, or if they display similar levels of acquisition but at different rates. The cultures were passaged every 12 hours and carried out exactly as before, but for up to seven passages. In addition to samples being taken for PCR and passaging, extra samples were taken for viable-cell-count spot tests (2.8.2); this was to quantify the number of cells to ensure the adaptation observed was due to additional acquisition events rather than simply a greater number of cells, as well as ensuring the plasmids were not cured.

A 100 µl sample from each passage was taken and diluted 10-fold in M9 media. From this diluted sample, additional 10-fold dilutions were carried out, down to a 10^7 dilution. The 10^3 -

10⁷ dilutions were added to LB-agar plates supplemented with antibiotics as 10 µl spots, with five spots per dilution (i.e. in pentaplicate). They were 'spotted' onto plates containing tetracycline as a measure on total cell count as EB371 and EB372 are both chromosomally tetracycline resistant (Tet^R) due to the *araB::T7RNAP-tetA* genotype being present. The samples were also plated as spots onto ampicillin-containing plates to see if the pBad-based ampicillin resistant (Amp^R) plasmids were still present in the cells, as plasmid loss could explain poor acquisition levels.

Loss of plasmids proved to be an issue with both EB371 as well as EB372. The seven-passage acquisition assay of EB371 (**Figure 17A**) shows the levels of acquisition when transformed with pEB628 slowly rise from 8 % at passage 2, consistently increasing with each passage to 49 % multiple-band acquisition at passage 7. The cell-count spot test for this assay (**Figure 17C**) shows that around 10⁹ CFU/ml of EB371 was in each passage culture when plated on ampicillin or tetracycline. However, this was not the case for EB372 during this seven-passage acquisition assay. When transformed with pEB628, EB372 started out with a higher percentage of acquisition than EB371, 24 %, which increased to 43 % at passage 5 (**Figure 17B**). Then in passages 6 and 7, the levels of acquisition dropped to lower levels than EB371, culminating in 18 % acquisition at passage 7. This decrease in acquisition levels could be due to the plasmid instability, shown in in **Figure 17D**. The CFU/ml of EB372 on both ampicillin and tetracycline is around 10⁹ for the first 4 passages, albeit slightly lower than EB371. After passage 5 however, the CFU/ml plated on ampicillin decreased, indicating that the plasmid was either being lost due to instability, or was being used as a source of spacers for acquisition (plasmids are MGEs after all). With fewer EB372 cells containing the pEB628 plasmid, less Cas1-2 could be expressed which lead to weaker acquisition.

A summary of the 12-hour passage acquisition assays is shown in **Figure 17E**. When passaged up to 5 times, EB372 seems to more-readily acquire spacers than the 'wild-type' EB371. However, the levels of acquisition could be being skewed by the pEB628 (pCas1-2) plasmids not being retained adequately, hindering Cas1-2 expression, ergo acquisition

levels. The non-repeated seven-passage acquisition assay has been left in to show how plasmid instability could be skewing the data.

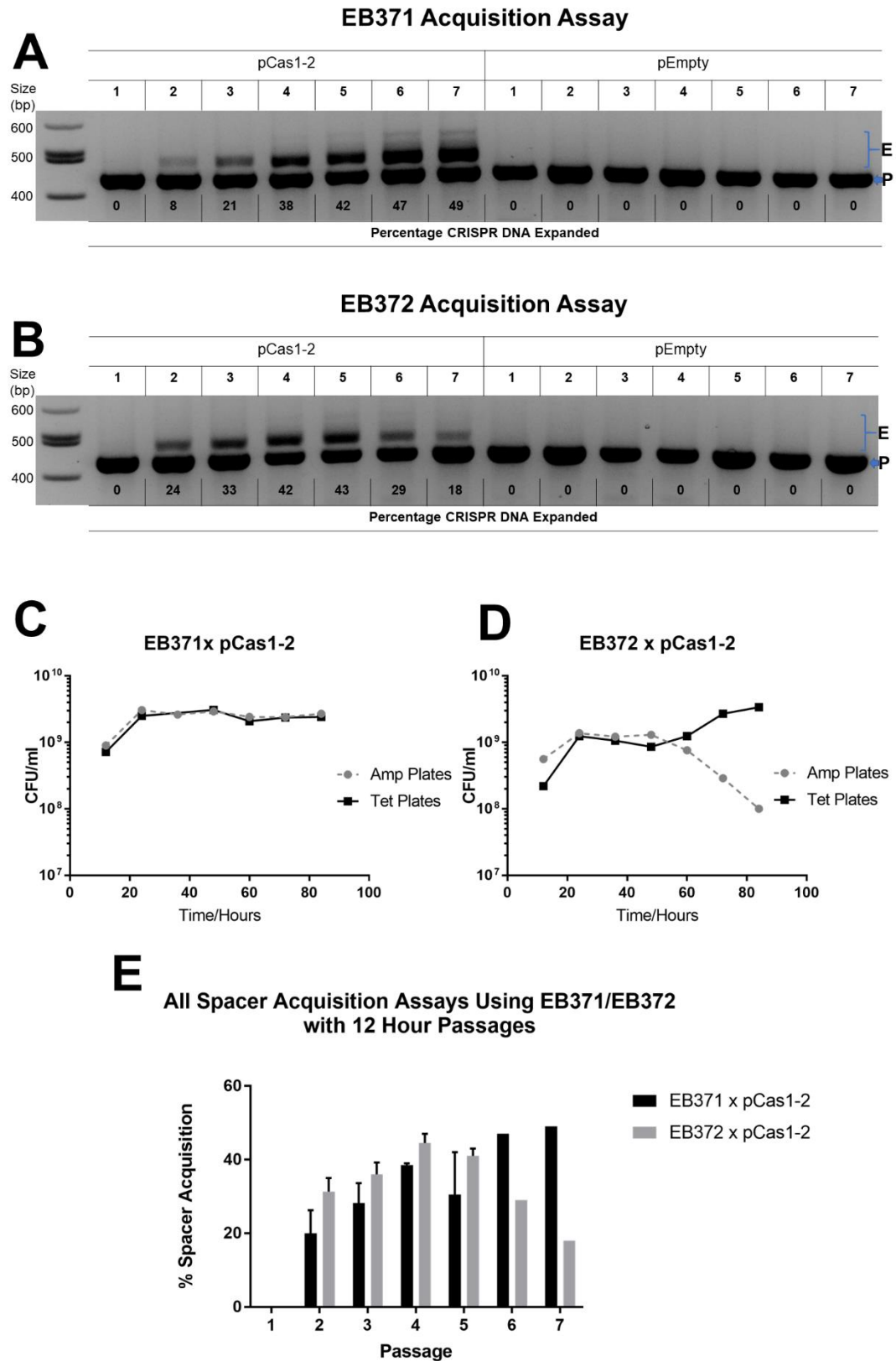


Figure 17 Summary of every 12-hour passage naïve acquisition tests with EB371 and EB372 strains. (A) and (B) 2 % agarose-TBE gels showing PCR-amplified CRISPR loci of

EB371 (A) and EB372 (B) after seven 12-hour passages, with samples taken at each passage. pCas1-2 = pEB628, pEmpty = pBad. Parental bands ('P') at 440 bp, with expanded arrays ('E') at around 500 bp, 550 bp, and 600 bp. The difference between the parental and expanded arrays is shown as 'Percentage CRISPR DNA Expanded' at the bottom of the gel. **(C) and (D)** viable cell count spot tests for EB371 (C) and EB372 (D). Samples were taken at each passage, serially diluted in M9 media, and plated on separate tetracycline (chromosomal resistance) and ampicillin (plasmid resistance) plates. These only show the spot tests for the seven-passage test, therefore n=1. **(E)** Percentage acquisition from all 12 hour acquisition tests comparing EB371 and EB372 strains represented in a bar graph. For passages 1-3, N=4; for passages 4-5, N=2; for passages 6-7, N=1. Mean and standard error calculated in Prism 7 (GraphPad).

To try to overcome the plasmid-stability issues seen in EB371 and EB372 over several passages, two solutions arose. The first involved Red-Recombineering (**2.7.2**) the pEB628 plasmid into a EB371/EB372, via first using recombineering to remove the existing *cas1-2* genes; this way Cas1-2 could be expressed chromosomally without plasmid instability being an issue. However, colony PCR on kanamycin-resistant colonies revealed that the *cas1-2* genes were still present and had not been replaced by the Kanamycin resistance (Kan^R) cassette as intended (**Figure 18**).

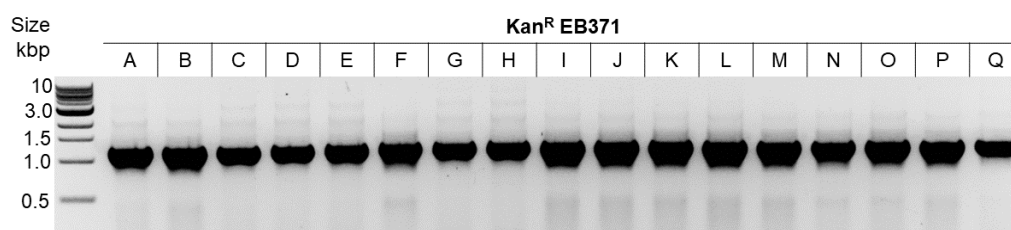


Figure 18 Colony PCR of Kan^R EB371 colonies after recombineering on a 2 % agarose TBE gel. Bands from each colony (A-Q) correspond to *cas1-2* genes, showing *cas1-2* genes weren't replaced by the Kan^R cassette as intended.

The second solution involved more acquisition assays with only two passages. These assays involved setting up a 12-hour passage in LB with a single addition of 0.2 % L-arabinose, instead of two separate additions of 0.1 % L-arabinose. As before, these starter

cultures were used to inoculate a second culture, this time containing 0.2 % L-arabinose with no extra arabinose added at OD₆₀₀ 0.6. These cultures were then incubated with shaking for 24 hours to express the Cas1-2 proteins and therefore acquire new spacers. Samples were taken for PCR analysis of acquisition every hour for 12-hours, then every 4-hours up to 24-hours (gels shown in **Figure 19**). Samples were taken for viable-cell-number spot tests at hours 6, 12 and 24. As shown in **Figure 20A**, the levels of acquisition seen in EB372 was consistently higher than EB371. Another outcome of this experiment was that the previous additions of 0.1 % L-arabinose at OD₆₀₀ 0.6 were not necessary to stimulate acquisition; 0.2 % L-arabinose added to the start of the culture stimulated acquisition. The cell count spot tests (**Figure 20B**) showed that plasmids were retained by both strains, and that the number of EB372 cells in culture was always slightly lower than EB371.

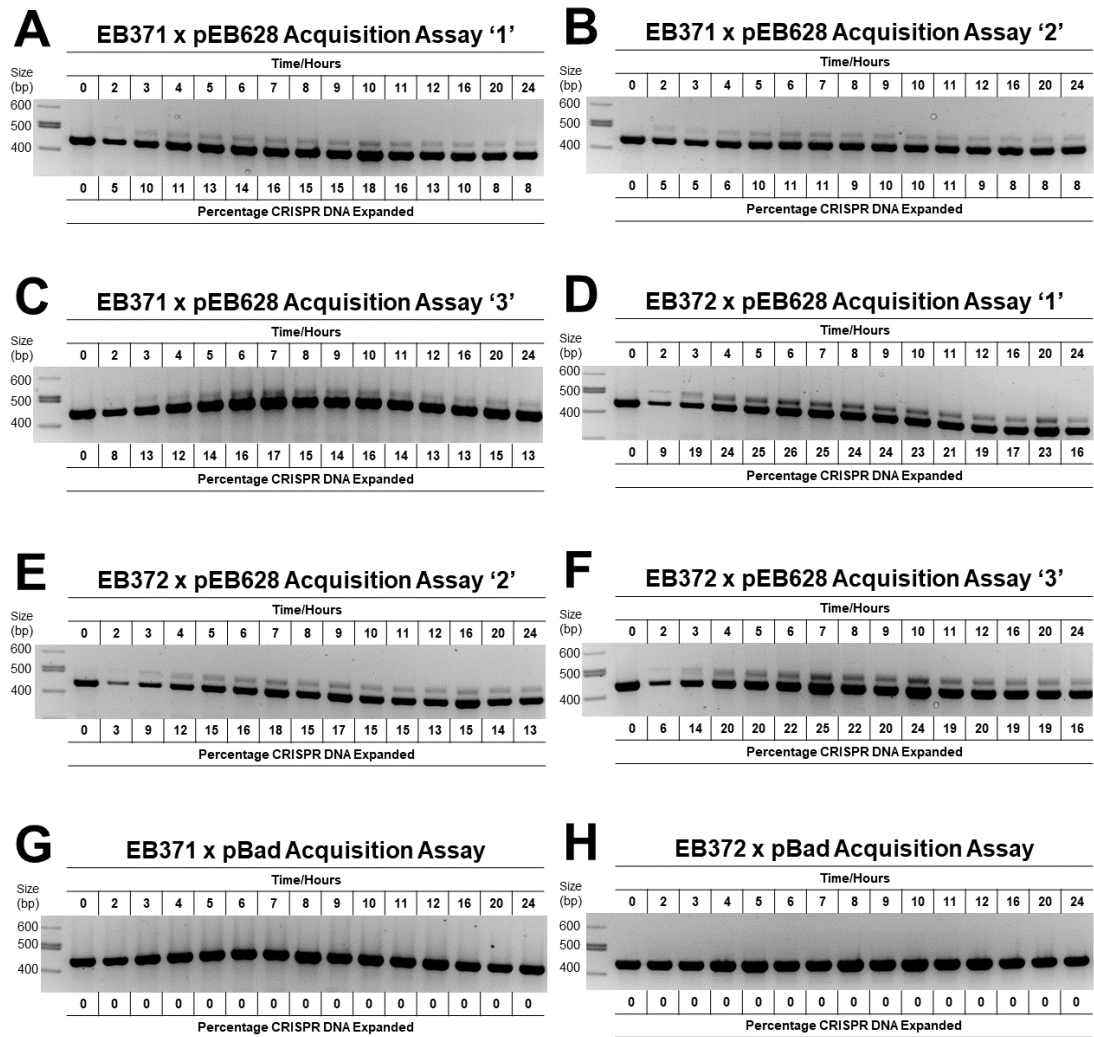
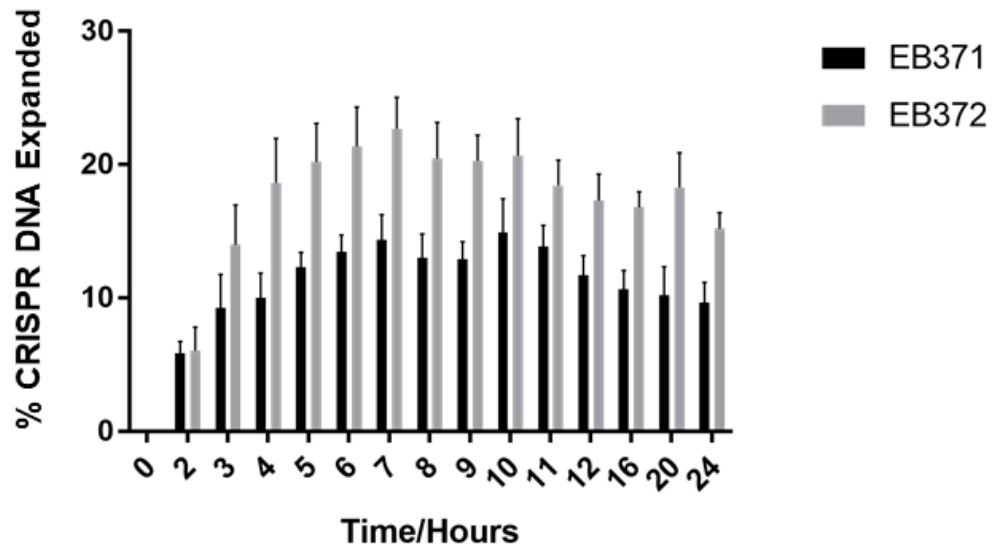


Figure 19 2 % agarose-TBE gels showing PCR amplified CRISPR arrays from the 24 hour acquisition assay. EB371 x pEB628 (A-C), EB372 x pEB628 (D-F), and empty vector controls (G-H). Cultures were passaged once, then induced for 24 hours, with samples being taken for PCR as shown. Un-expanded parental CRISPR arrays shown around 440 bp, with expanded arrays at around 500 bp. The difference between the un-expanded and expanded bands is shown as 'Percentage CRISPR DNA Expanded' underneath the gels.

A

Spacer Acquisition Assay with EB371 and EB372 Over 24 Hours

**B**

Viable Cell Count Spot Tests

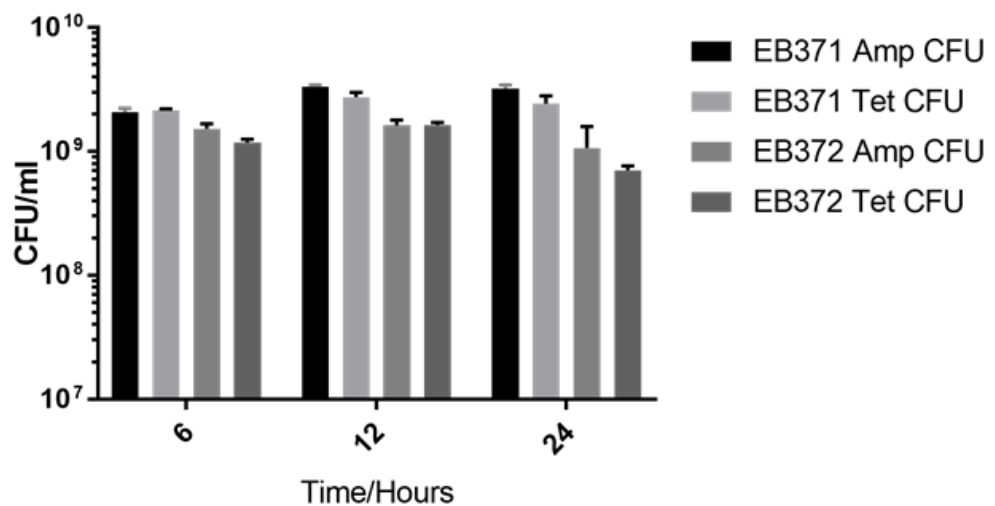


Figure 20 New spacer acquisition from a 24 hour acquisition test **(A)** comparing EB371 and EB372 strains, represented in a bar graph. pEB628 plasmid used to generate Cas1-2. Mean and standard error calculated in Prism 7 (GraphPad). Assay performed in triplicate, N=3. **(B)** Viable cell count from spot tests using EB371/EB372 transformed with pEB628.

Samples taken, serially diluted and spotted onto tetracycline (chromosomal resistance) and

ampicilin (plasmid resistance) to estimate cell count and plasmid retention respectively.

Mean and standard error calculated in Prizm 7 (GraphPad). Assay performed in triplicate,

N=3

3.2 YFP-Tagged Cas1 Acquisition and Purification

Fluorescent tags are often fused to proteins as reporters for protein expression and monitoring localisation within a cell. However, sometimes the addition of large tags onto proteins can affect their normal function. As a collaborator (Dr Christian Rudolph at Brunel University) is using an N-terminal fluorescent tag to study localisation of Cas1 with regards to replication fork stalls/collisions, it was of utmost importance to see if Cas1 can perform its normal function, spacer acquisition, when bound to the yellow-fluorescent-protein (YFP) from *Aequorea victoria*.

Initially, pLau18 YFP-Cas1 (pYFP-Cas1, this plasmid was a gift from Dr Christian Rudolph's lab at Brunel University) was transformed into BL21-AI with pEB680 (pCas2) to express YFP-Cas1 and Cas2 respectively. pEB628 was used to express wt Cas1-2 (pCas1-2), along with the empty vector pACYC-Duet so two plasmids were used throughout. pBad and pACYC-Duet were used as empty vector controls (**Figure 21A**). As BL21-AI was used with T7-expression//*lac* operator vectors, protein expression was stimulated with the addition of L-arabinose and IPTG.

Cultures were passaged every 12 hours for five passages, with samples being taken for PCR at every passage. The outcome of these BL21-AI double-plasmid acquisition assays (**Figure 21B**) showed that expression of N-terminally-fused YFP-Cas1 could not facilitate naïve acquisition. Only the parental bands were visible for the YFP-Cas1 + Cas2 samples, as well as the empty vector controls. The positive control, wt Cas1-2 from pEB628 + empty vector pACYC-Duet, was able to integrate new spacers. As spacer integration was most prominent in the first two passages, the overall number of passages was dropped for the second phase of YFP-Cas1 naïve acquisition testing.

As these experiments were being run in parallel with the ectopic-origin naïve acquisition study, the MG1655-based EB371 strain was adopted for further naïve acquisition tests; this was due to the higher levels of acquisition seen when transformed with pEB628, when compared to BL21-AI transformed with pEB628. Another change made was to the plasmids used. Rather than using two plasmids per cell, creating the potential for plasmid instability,

the Cas2 was subcloned from pEB680 into pLau18 YFP-Cas1, so that all the acquisition machinery was on a single plasmid, pEB681 (pYFP-Cas1+Cas2) just like pEB628 (**Figure 21C**). Acquisition assays were carried out as before, but this time without IPTG additions and with only 3 passages. The outcome (**Figure 21D**) was the same as before, the YFP-Cas1 cannot acquire new spacers *in vivo*.

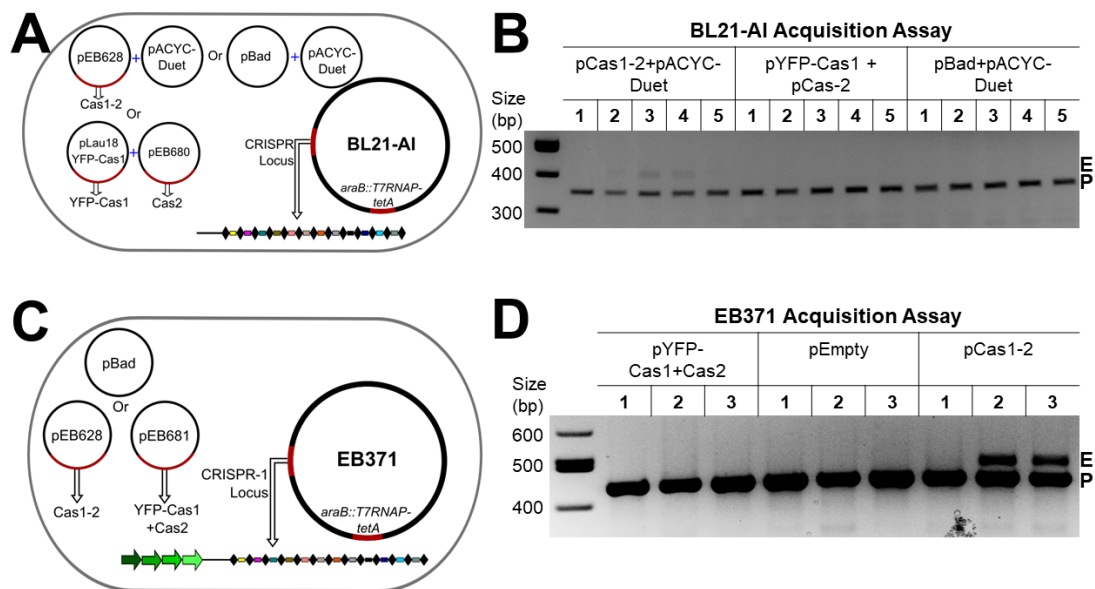


Figure 21 Naïve acquisition experiments using Cas1 with an N-terminal YFP-tag (A) A diagram showing the different plasmids transformed into BL21-AI to investigate the function of YFP-Cas1 *in vivo*. **(B)** The results of a PCR reaction amplifying the CRISPR locus from BL21-AI run on a 2 % agarose-TBE gel. Each sample is from a 12-hour passage. Un-expanded parental bands ('P') are around 350 bp, with expanded arrays ('E') visible at around 410 bp. pCas1-2 = pEB628, pYFP-Cas1 = pLau18 YFP-Cas1, pCas2 = pEB680. **(C)** A diagram showing the different plasmids introduced into EB371 for investigating YFP-Cas1's acquisition ability when co-expressed with Cas2 from the same plasmid. **(D)** A 2 % agarose-TBE gel showing PCR products from a 3-passages acquisition assay, with each passage lasting 12 hours. Parental arrays shown around 440 bp, expanded arrays shown around 500 bp. pYFP-Cas1+Cas2 = pEB681, pEmpty = pBad, pCas1-2 = pEB628.

As the YFP tag is a large tag (28 kDa), a smaller Flag-tag (7 residues) Cas1 was tested to see whether it could facilitate the acquisition of new spacers *in vivo*. The plasmid pTK7 (from

Dr Tom Killelea) was used for the flag-tagged Cas1 with Cas2 (pFlag Cas1-2), alongside a positive control of pEB628 (pCas1-2). Acquisition with flag tagged Cas1 was tested in EB371 and EB372, but no acquisition was seen with the N-terminal flag tagged Cas1.

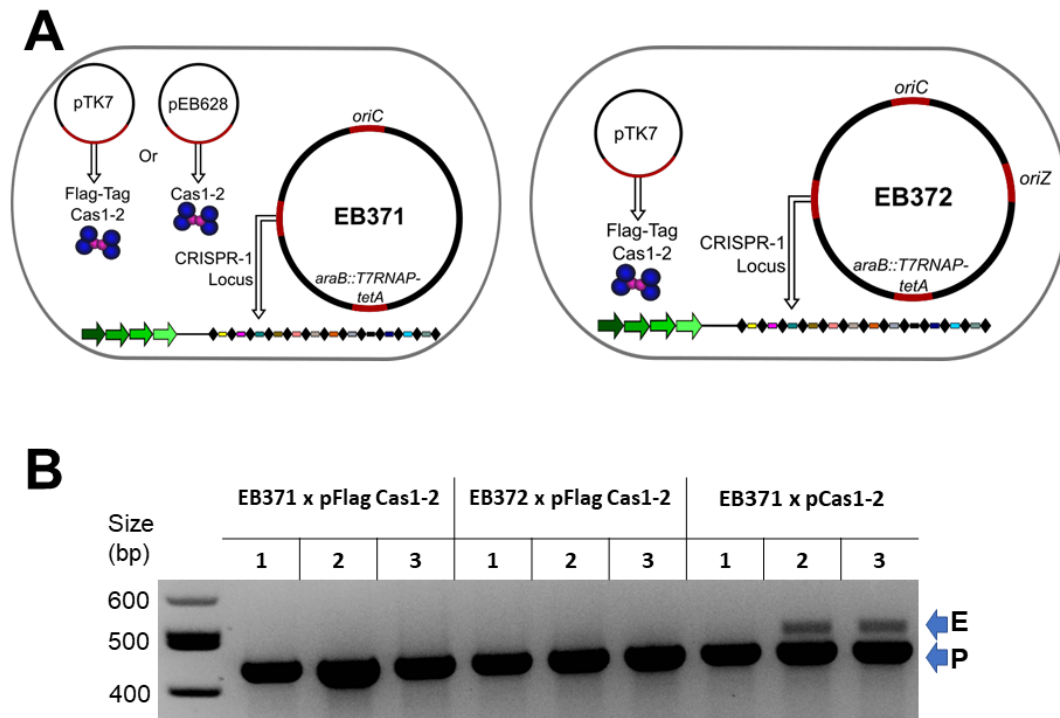


Figure 22 Naïve acquisition experiments using Cas1 with an N-terminal Flag-tag. (A) A diagram showing the different plasmids transformed into EB371/EB372 for N-terminal Flag-tag Cas1 acquisition assays. **(B)** A 2 % agarose-TBE gel showing PCR products from a 3-pass acquisition assay, with each passage lasting 12 hours. Parental arrays shown around 440 bp, expanded arrays shown around 500 bp. pFlag Cas1-2 = pTK7, pCas1-2 = pEB628.

Naïve acquisition could not be stimulated *in vivo* using N-terminal YFP or Flag tags, meaning that either the prespacer capture function, spacer integration function, or both were being compromised by the YFP tag. N-terminally His₆ tagged Cas1 has been shown to integrate spacers *in vitro*⁴¹, leaving the possibility that the N-terminal tags could be affecting prespacer capture, but not spacer integration (SpIn), therefore the YFP-Cas1 could still be semi-functional. This SpIn assay would require purifying YFP-Cas1, which was attempted, though ultimately unsuccessful.

The biomasses for wt Cas1, wt Cas2 and YFP-tagged Cas1 were prepared for purification as described earlier (2.9.1) with overexpressed products checked on a 15 % SDS-PAGE gel (Figure 23A). The wt Cas1-Cas2 would be purified together to act as a control to ensure the purification was successful, whilst YFP-Cas1 and wt Cas2 were also purified together to see if YFP-Cas1 has spacer-integration functions. Biomasses were then lysed and clarified (2.9.2). The clarified lysates of the wt-Cas1 Cas2 and YFP-Cas1 wt Cas2 biomasses were mixed for 2 hours at 6°C with slow rotation, this was to allow YFP/wt-Cas1 to interact with Cas2 for combined purification (ie purifying wt Cas1-Cas2 together and YFP-Cas1 wt Cas2 together).

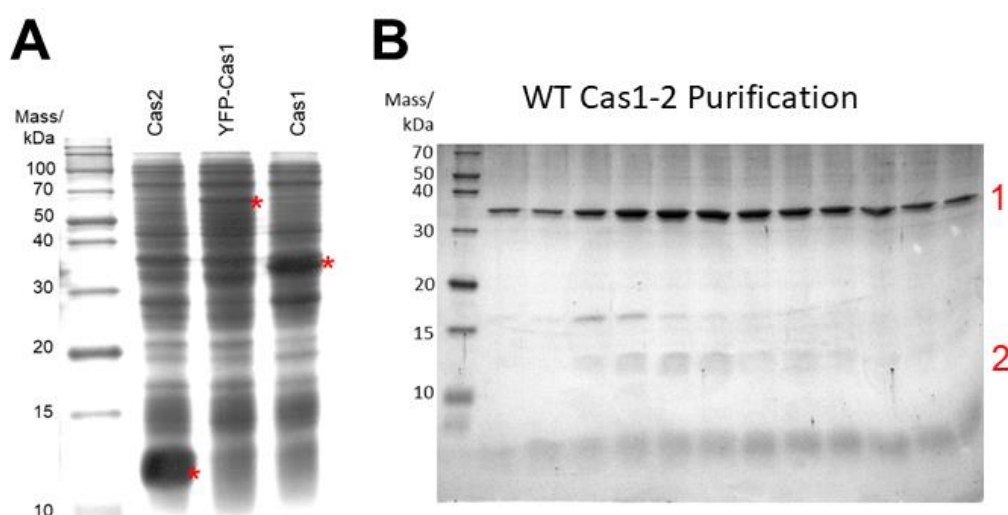


Figure 23 Overexpression and purification of YFP/wt-Cas1-2 (A) Pilot overexpressions of Cas2 (pTK10), YFP-Cas1 (pLau18 YFP-Cas1), and Cas1 (pEB505); run on a 15 % SDS-PAGE gel, stained with coomassie brilliant blue. Overexpressed proteins marked with red asterisks. Overexpression performed by Dr Tom Killelea. **(B)** The 15 % SDS PAGE gel after purifying wt-Cas1-2 using a Ni^{2+} chelating column and a heparin column. The red '1' at 36 kDa represents where Cas1 is present, the red '2' at 11 kDa represents where Cas2 is present. This was run by Dr Tom Killelea in tandem alongside the YFP-Cas1 Cas2 purification.

The first stage of the purification was to load the combined lysates onto a 5 ml Ni^{2+} chelating column to bind the histidine residues from the His-tagged proteins (Cas2 and Cas1).

Increasing concentrations of imidazole were used to out-compete the His₆-tagged proteins, with protein elution being visible on the chromatographic trace. The fractions which corresponded to UV absorbance peaks at 280 nm (A_{280}) were run on a 15 % SDS PAGE gel (**Figure 24A**). Cas2 was visible in the sample, so the fraction samples were then pooled together, dialysed to remove imidazole and lower the NaCl concentration, then loaded onto a 1 ml heparin column (**Figure 24D**). As it was hard to tell if YFP-Cas1 was in the sample with the Cas2, the flow-through of the column was collected, dialysed and loaded onto a 1 ml heparin column (**Figure 24B**). This column was used to separate DNA-binding proteins, such as the Cas1-2 complex, from other proteins. Again, the fractions corresponding to A_{280} trace peaks were run on 15 % SDS PAGE gels to see if YFP-Cas1 with Cas2 were present (**Figure 24E**). The YFP-Cas1 Cas2 Ni²⁺ column fractions run on the heparin column showed that Cas2 had been purified, but no YFP-Cas1. YFP-Cas1 was also not present in the Ni²⁺ flow-through heparin-column elution (**Figure 24C**). The fact that YFP-Cas1 alone wouldn't be able to be purified in this wasn't surprising, but when co-purified with a His₆-tagged Cas2 could indicate that the proteins are not interacting; if Cas2 could be bound to a Ni²⁺ column, YFP-Cas1 may also be bound if it was interacting with Cas2. The wt-Cas1-Cas2 Ni²⁺-elution heparin-elution showed the presence of Cas1-2, demonstrating the method does work for wild-type Cas1-2 (**Figure 23**).

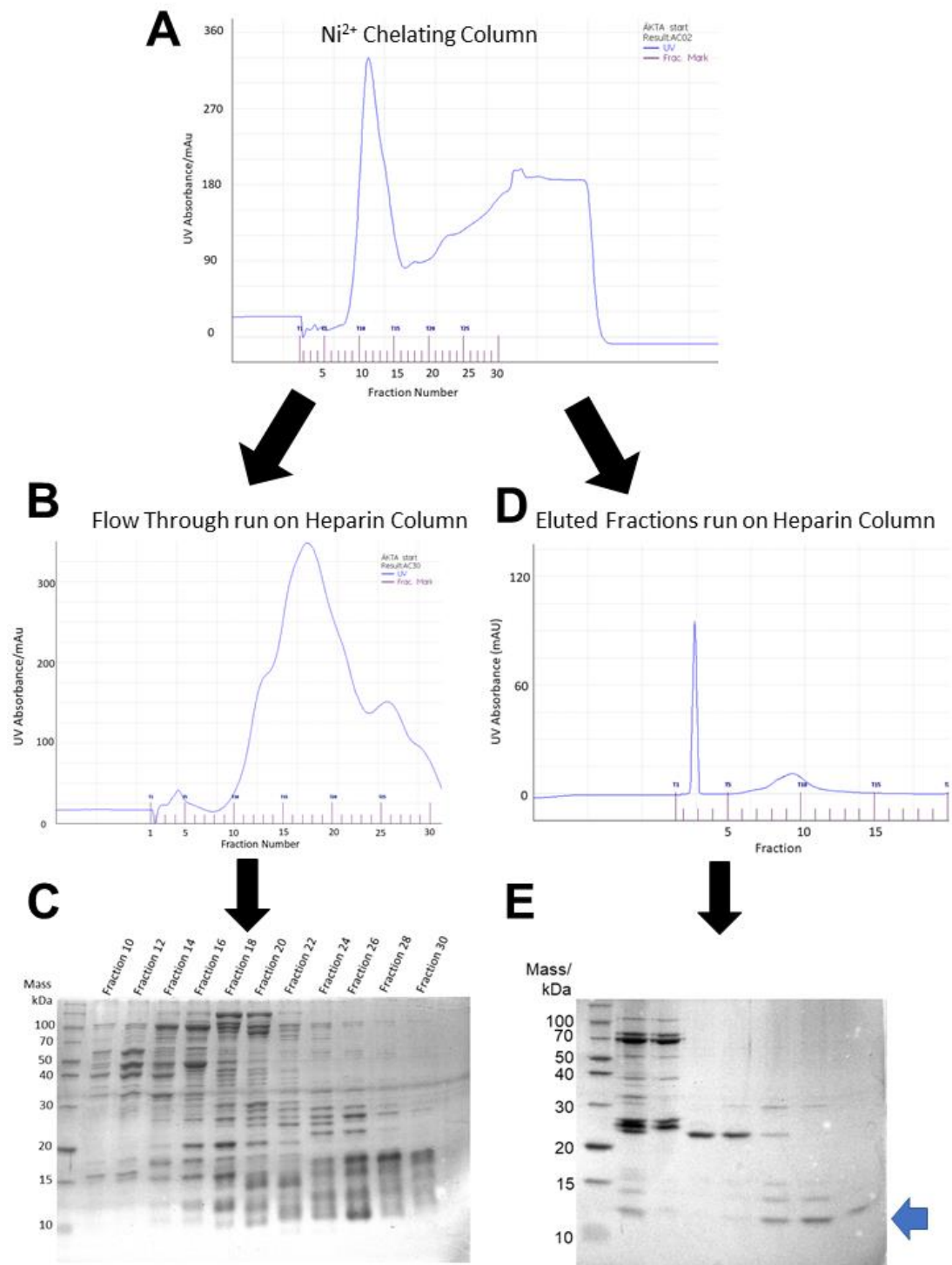


Figure 24 A flow diagram summarising the attempted purification of YFP-Cas1 (A) Chromatographic trace of YFP-Cas1 and Cas2 being run through a Ni^{2+} chelating column. UV absorbance is measured at 280nm (A_{280}). (B) Samples which flowed-through the Ni^{2+}

chelating column were run on a Heparin column. Chromatogram trace of these samples being eluted from the heparin column at A_{280} . **(C)** A 15 % SDS PAGE gel loaded with fractions eluted from the heparin column which were likely to contain protein, as indicated by the trace shown in (C). **(D)** Chromatographic trace of fraction samples previously eluted from the Ni^{2+} chelating column run on a heparin column. Fractions which were bound and eluted from this fraction were run on a 15 % SDS-PAGE gel **(E)**. The blue arrow at around 11 kDa indicates the presence of Cas2. There is no band around 65 kDa to indicate the presence of YFP-Cas1. 'D' and 'E' run by Dr Tom Killelea.

To advance this work further, a His₆-tag could be added to the YFP-Cas1 to purify this protein separately using the same purification protocol. Then a 'pull-down' assay could be attempted, by immobilising the His-YFP-Cas1 on a Ni^{2+} column, running Cas2 clarified lysate through the column, and seeing if the Cas2 interacts with the His-YFP-Cas1. The addition of a His₆ tag onto the YFP-Cas1 is described here **2.7.4**. Despite test digests of the new His-YFP-Cas1 plasmids (pHis-YFP-Cas1A-C) showing the correct-sized inserts **(Figure 25A)**, a pilot overexpression showed the protein to be around 35 kDa, not 65 kDa which corresponds to the mass of His-YFP-Cas1 **(Figure 25B)**.

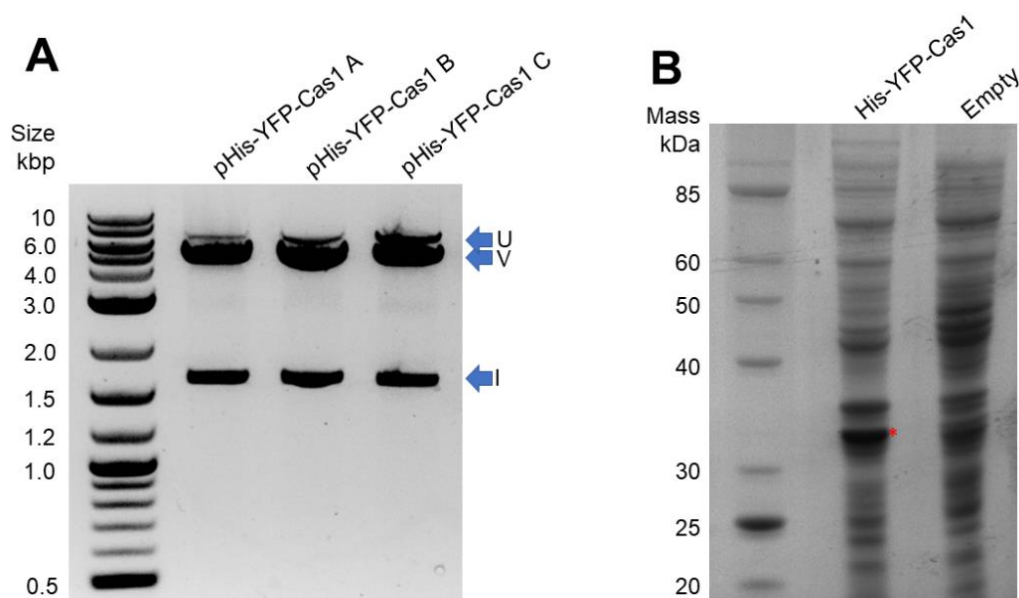


Figure 25 Adding a His₆ Tag to YFP-Cas1. **(A)** The products of the new His-YFP-Cas1 plasmids digested with NcoI and XhoI run on a 1 % agarose-TBE gel. 'U' corresponds to uncut plasmid; 'V' corresponds to pET-Duet vector backbone; 'I' corresponds to the ~1600

bp insert containing the His₆ tagged YFP-Cas1. **(B)** Pilot overexpression of His-YFP-Cas1 in BL21-AI run on a 10 % SDS-PAGE gel. Red asterisk shows overexpressed product at ~35 kDa, with no product at 65 kDa where His-YFP-Cas1 should be. Empty=pET-Duet in BL21-AI.

3.3 Analysis of Acquisition in M9 Media Vs LB Media

As well as investigating the effects of additional origins and protein-tags on Cas1-2 acquisition, nutrient deprivation was also studied. Nutrient deprivation can lead to RNA-Polymerase backtracking, which can then collide with the DNA replication machinery¹²⁵, perhaps having a similar effect seen when ectopic origins are used (**3.1**). The EB371 strain was employed to investigate the effects that nutritional deprivation could have on naïve acquisition in comparison to the routinely-used LB media.

EB371, as before, was transformed with pEB628 (pCas1-2) and pBad (pEmpty). The acquisition assays were set up as described earlier (**3.1**), with three 12-hour passages. 0.2 % L-arabinose was added to the growth media at to induce Cas1-2 protein expression, thereby stimulating spacer acquisition. LB media was used as the positive control; M9 media supplemented with 0.4 % glycerol as a carbon source (so as not to interfere with arabinose-based expression) and 1% casamino acids as a protein-nitrogen source was used as a minimal media. The experiment was performed in triplicate. As shown in **Figure 26A**, the levels of acquisition from EB371 x pCas1-2 when in supplemented M9 media is rapidly stimulated, with acquisition evident in each first-passage of the three M9 samples. The percentages of acquisition (**Figure 26B**) in the M9 media samples also far exceeds that of the LB media samples, with over 50 % of CRISPR-I loci being extended by the third passage.

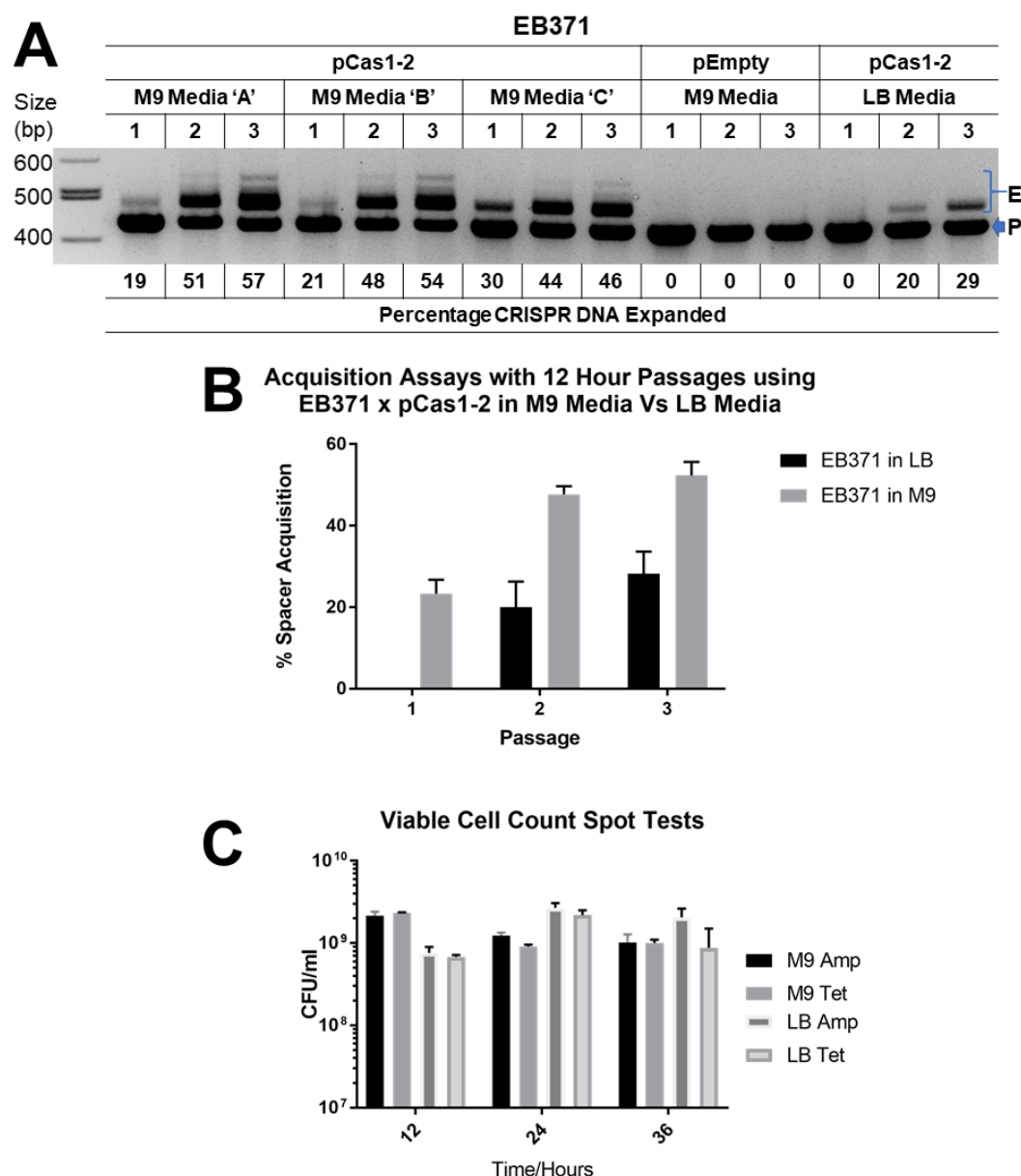


Figure 26 Acquisition assays in M9 media Vs LB media using the *E. coli* strain EB371.

(A) PCR-amplified CRISPR loci run on a 2 % agarose-TBE gel. pCas1-2 = pEB628, pEmpty = pBad. Three 12-hour passages were used in these acquisition assays. Parental bands ('P') at 440 bp, with expanded arrays ('E') between 500 bp and 600 bp. The difference between un-expanded and expanded CRISPR loci given as 'Percentage CRISPR DNA Expanded' at the bottom of the gel. The experiment was performed in triplicate ('A', 'B' and 'C'). **(B)** A graph showing graphically the differences between % spacer acquisition of EB371 x pCas1-2 (pEB628) in M9 media and LB media. Mean and standard error calculated in

Prizm 7 (GraphPad), N=3. **(C)** Viable cell count from spot tests using EB371/EB372 transformed with pEB628. Samples taken, serially diluted and spotted onto tetracycline (chromosomal) and ampicilin (plasmid) to estimate cell count and plasmid retention. Mean and standard error calculated in Prizm 7 (GraphPad). Assay performed in triplicate, N=3

3.4 Naïve Acquisition Using the *M. piezophila*

CRISPR/Cas system

In addition to testing naïve acquisition in MG1655 CRISPR systems, the CRISPR-Cas system of the bacteria *Marinitoga piezophila* was investigated (1.7). The specific reason for studying naïve acquisition in this bacterial species is due to its unusual CRISPR locus, which contains a primase-polymerase (prim-pol) gene and an argonaute gene, as well as the usual cas genes (Figure 27). Argonaute proteins are usually used in RNAi pathways, with Prim-Pol proteins being used in DNA repair; these genes could indicate that *M. piezophila* uses these factors for acquiring spacers in place of usual factors (such as IHF). Naïve acquisition assays were designed to see if acquisition was possible with only the *M. piezophila* cas genes present; with only the prim-pol and argonaute genes present; or whether a combination of both are required for CRISPR-Cas function. This could offer an explanation as to why the prim-pol/argonaute genes are located within the CRISPR locus.

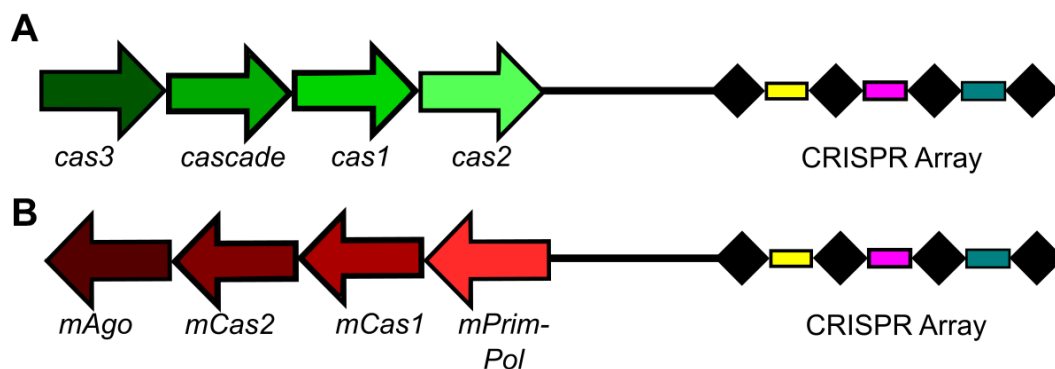


Figure 27 A simplified diagram comparing the CRISPR loci of *E. coli* and *Marinitoga piezophila*. The *E. coli* CRISPR locus has green arrows representing cas genes, the *M. piezophila* has red arrows for cas/other genes.

Before acquisition assays could commence, part of *M. piezophila*'s CRISPR locus (mCRISPR) was subcloned from the pET-28a-based pPK250 (a gift from Professor Aiden Doherty, University of Sussex), into pCDF-1b, forming the plasmid pRPR1. This was so three separate plasmids could be introduced into the same strain without encountering issues with competing plasmid origins or antibiotic resistances. As each of the plasmids

contained T7 promoters as well as lac operators, BL21-AI strain was used with L-arabinose and IPTG to induce protein expression. Along with pRPR1, two more plasmids were transformed into the *E. coli* BL21-AI strain: pPK273, which encoded *M. piezophila* *cas1-2* genes (for mCas1-2 proteins) in pACYC-Duet; and pPK276, which encoded the *M. piezophila* primase-polymerase and argonaute (for mPrim-Pol and mAgo proteins) in pET-28a. Empty vector controls were used to potentially isolate the roles of the mCas1-2 proteins and mPrim-Pol/mAgo proteins in acquisition. The plasmid combinations for investigating *M. piezophila* naïve acquisition is shown in **Figure 28A**.

Naïve acquisition was tested in a similar way to **3.1**: four 12-hour passages in LB media, with induction of protein expression courtesy of 0.1 % L-arabinose and 0.1 mM IPTG from the starting culture and again at OD₆₀₀ 0.6. Samples were taken for PCR analysis of spacer acquisition at the same time as the cultures were passaged. The results of the acquisition assays (**Figure 28B**) revealed that no new spacers were integrated into the mCRISPR locus; only un-expanded CRISPR loci were detected despite passaging cells in the presence of mCas1-2, mPrim-Pol/mAgo, or both mCas1-2 with mPrim-Pol/mAgo.

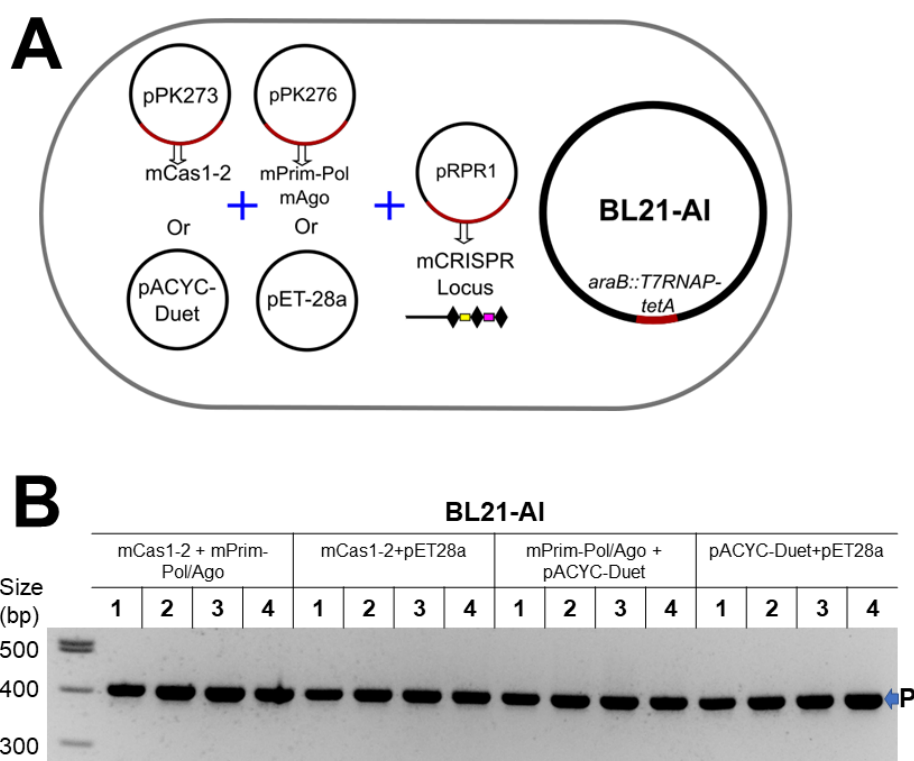


Figure 28 Naïve acquisition assays using the *Marinitoga piezophila* CRISPR Cas system. (A) A diagram showing the different plasmids transformed into BL21-AI to investigate the CRISPR Cas system of *M. piezophila*. **(B)** The results of a PCR reaction amplifying the CRISPR locus from *M. piezophila* (from the plasmid pRPR1) run on a 2 % agarose-TBE gel. mCas1-2 were expressed using pPK273, mPrim-Pol/mAgo were expressed using pPK276, with pACYC-Duet and pET28a as empty vectors. Each sample is from a 12-hour passage. Only the un-extended parental bands ('P') are present at around 410 bp.

As no spacer acquisition was apparent, a pilot overexpression of the BL21-AI with mCas1-2, mPrim-Pol and mAgo was performed to ensure the mCas1-2 and mPrim-Pol/Ago proteins were being expressed. As a negative control, the BL21-AI containing empty vectors, pACYC-Duet/pET-2a, was also overexpressed. The overexpression protocol was carried out as described in 2.6.8 and samples were run on a 10 % SDS PAGE gel. The results of the overexpression (**Figure 29**) showed than only some of the *M. piezophila* proteins were expressed.

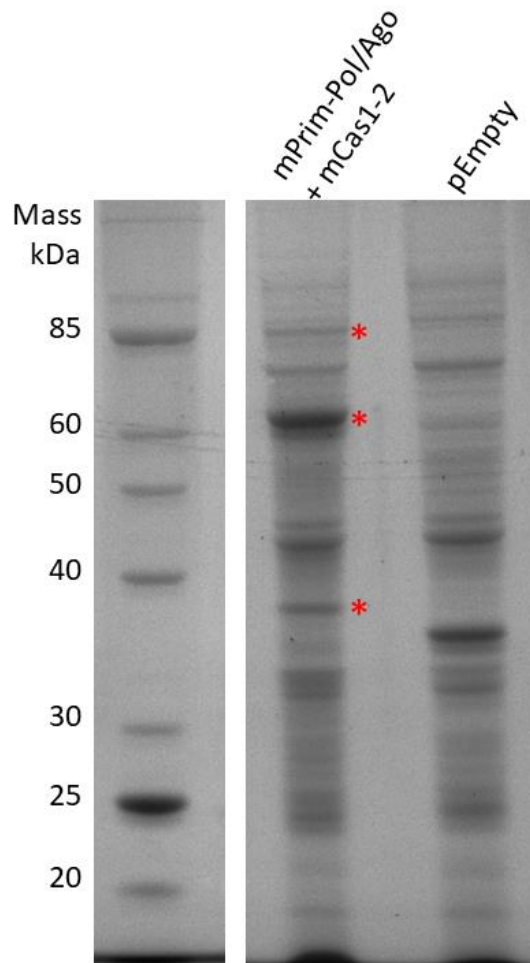


Figure 29 10 % SDS PAGE gel showing the products of overexpression of BL21 AI expressing *M. piezophila* proteins. mCas1-2 expressed from pPK273, mPrim-Pol/Ago expressed from pPK276, pACYC-Duet/pET-28a as empty vectors. Asterisks mark where overexpressed proteins are: mAgo around 85 kDa, mPrim-Pol around 64 kDa, mCas1 around 39 kDa. mCas2 not seen on gel. Irrelevant lanes have been cropped from the gel, hence the gap shown.

As *M. piezophila* proteins from each plasmid were expressed, it couldn't have been an issue with plasmid instability. However, an analysis of codon bias revealed that some of the genes used rare codons to produce the proteins. The online tool 'GenScript Rare Codon Analysis Report' was used to analyse the *M. piezophila* protein sequences to check for rare codon usage, mCas1 was found to have a Codon Adaptation index of 0.65, below the ideal 0.8-1 for good protein expression¹²⁶. To try to overcome this, the BL21-AI strain was swapped for BL21-C+, which contains additional rare tRNAs for expression of heterologous proteins with

differing codon bias to B-strain *E. coli*; this helps to reduce stalling of translation due to a lack of available tRNAs. However, this required sub-cloning of the *M. piezophila* genes from existing plasmids into new plasmids so as not to interfere with the chloramphenicol-resistant BL21-C+. Unfortunately, sub-cloning three plasmids proved unsuccessful, though this could be overcome in future using PCR-amplification sub-cloning methods (such as those described in **2.7.4**).

Chapter 4: Cas3 and the Stringent Response

As well as investigating the CRISPR-associated Cas1-2 proteins, the Type-I interference protein Cas3 was also studied (1.4). *E. coli* Cas3 is a 100 kDa monomeric protein which has an N-terminal HD-nuclease domain and C-terminal SF2 helicase domains. It is essential for the interference stage of type-I CRISPR systems where it degrades foreign DNA which matched CRISPR spacers. As well as its role in interference, Cas3 has been implicated to have a wider role within host cells, such as suppressing biofilm production¹²⁷. It is a potential role outside of CRISPR which was investigated in this study, specifically linking Cas3 to the Stringent Response. (1.8).

The Stringent Response is a physiological response to low amino acid/carbon/phosphate as well as heat shocks. It helps to reduce non-essential transcription/translation via the small molecule 'alarmone' ppGpp. ppGpp (also pppGpp) is generally produced by the RelA protein and degraded by SpoT. Therefore, to investigate if Cas3 has a role in the stringent response, *E. coli* strain MG1655 with deletions in *relA* (EB306); *relA* and *cas3* (EB305); *cas3* (EB304); *spoT* (EB309); *spoT* and *cas3* (EB310) were employed. To test the stringent response, rifampicin was used; this binds to the RNA-polymerase β site to inhibit RNA synthesis thereby preventing transcription. ppGpp has been shown to bind at the same site as rifampicin¹²⁸ and so when ppGpp is produced, rifampicin cannot inhibit transcription, meaning that the stringent response can allow a bacteria to grow on the antibiotic.

Overnight cultures with the desired *relA/spoT/cas3* deletions were set up in LB with appropriate antibiotics, alongside a wild-type MG1655 control. After 16 hours of growth, the cultures were serially diluted (10^0 - 10^7 fold dilutions) and spotted as 10 μ l spots onto LB plates supplemented with 10 μ g/ml rifampicin (**Figure 30A**), as well as LB-only control plates to count the viable cells in the culture. The plates were then incubated and the dilution at which the strains could no longer survive on rifampicin was noted (rifampicin resistance, **Figure 30B**). When *spoT* is deleted (EB309), meaning ppGpp can accumulate, the cells become much more resistant to rifampicin than wild-type. When *cas3* is also deleted with *spoT* (EB310), the same rifampicin-resistant phenotype is seen. When *cas3* (EB304) and

relA (EB306) are deleted individually, the bacteria have marginally improved rifampicin resistance when compared to wild type. When *relA* and *cas3* are deleted together (EB305), the bacteria were the least resistant (i.e., sensitive) to rifampicin, even more so than wild-type, with the bacteria never able to grow on rifampicin beyond a 10-fold dilution. This could imply that Cas3, along with RelA (and some SpoT activity) could help to produce the alarmone ppGpp required to activate the stringent response.

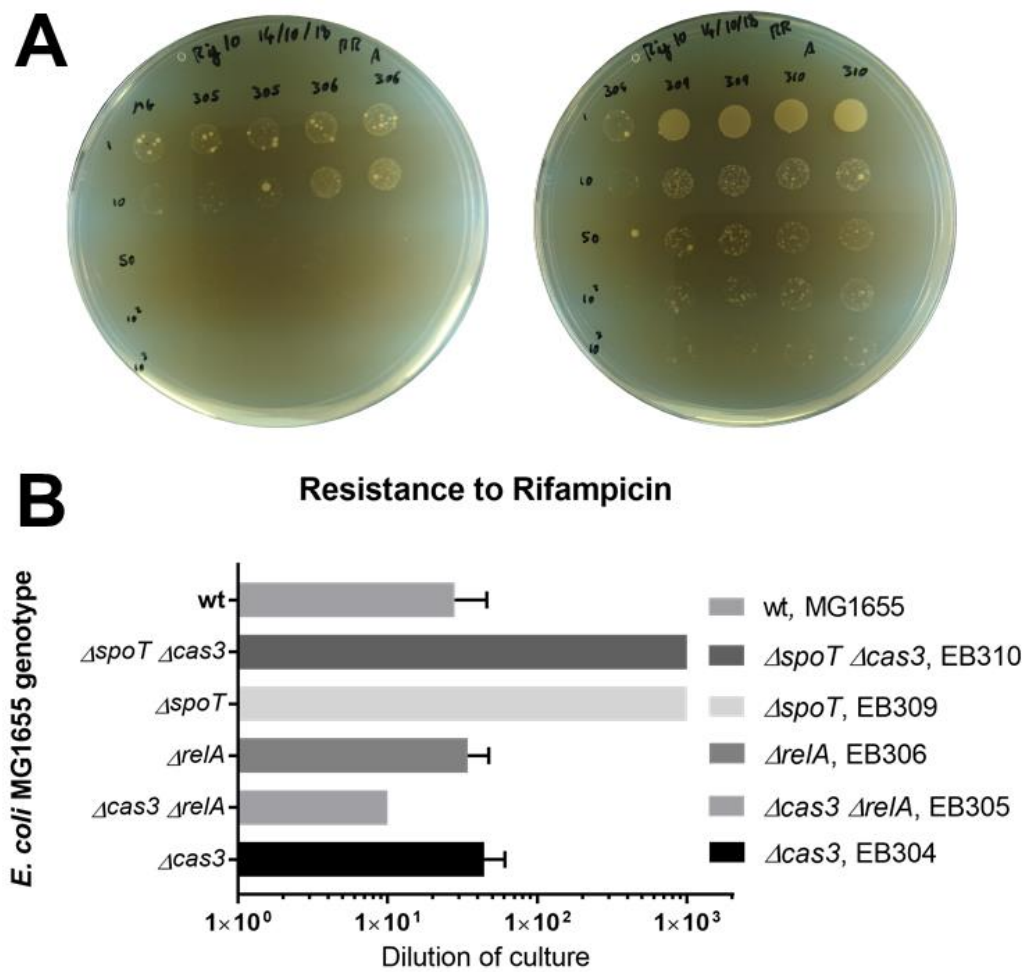


Figure 30 Rfampicin spot test results. (A) Showing 2 LB plates supplemented with 10 µg/ml rifampicin. Overnight culture was serially diluted in M9, 10 µl spotted onto the plate and incubated. The dilutions increase from 10⁰ to 10³ from the top to the bottom of the plate. **(B)** Where colonies grew, the dilution at which they did not grow (became rifampicin sensitive) was noted in Prism7, where the mean and standard error was calculated. For EB305 ($\Delta cas3 \Delta relA$), EB306 ($\Delta relA$), EB309 ($\Delta spoT$) and EB310 ($\Delta spoT \Delta cas3$), N=7; for

wt (MG1655) and EB304 (Δ cas3) N=4. The triangle indicates increasing resistance to rifampicin.

To test this further, a complementation test was carried out. *cas3* was re-introduced into EB305 using the *cas3*-containing plasmid pAH4, so it only had *relA* deleted (in a way to produce the same genotype EB306). Empty vector pBad was also transformed into EB305 and EB306 to ensure the additional cell stress caused by plasmids is not causing phenotypic changes. 0.1 % L-arabinose was used to express the Cas3 protein. However, introducing plasmids to EB305 (pCas3 and pEmpty) as well as EB306 caused decreased resistance to rifampicin (**Figure 31**).

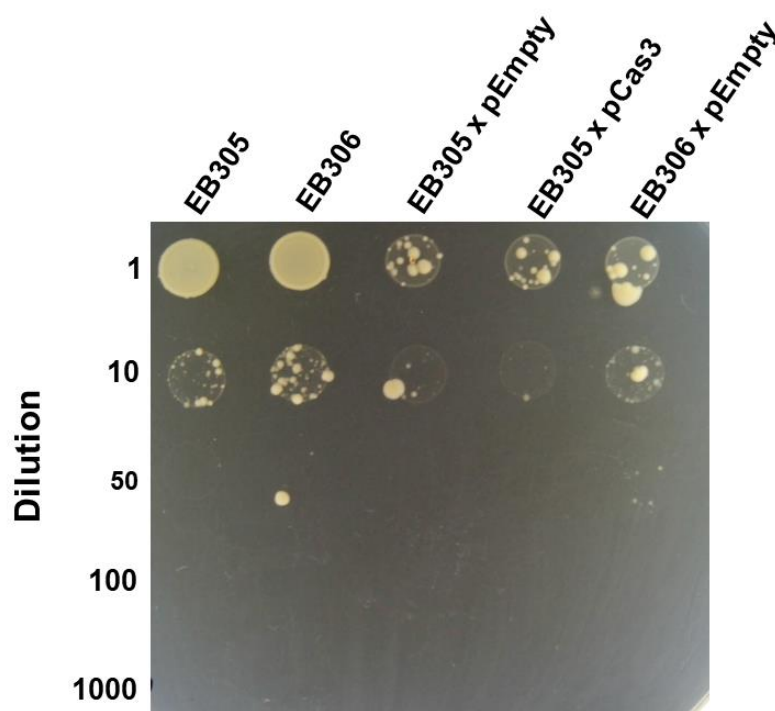


Figure 31 An example of the complementation spot tests, re-introducing *cas3* to **EB305**. Overnight cultures with appropriate transformations (pEmpty=pBad, pCas3=pAH4) were grown in LB with 0.1 % arabinose. Cultures were diluted in M9 as indicated and spotted onto LB agar+10 µg/ml rifampicin.

Chapter 5: Discussion

5.1 Prespacer Sources in Naïve Acquisition

Spacer capture is the first stage of CRISPR immunity, yet it is still poorly understood how these prespacers are selected (**1.2.2**). It was shown by Levy *et al.* in 2015 that new spacers can be acquired from the products of replication termination⁴⁵, with further work showing that these prespacers are processed in some way by RecBCD helicase activity, with additional separate nuclease activity, possibly from Cas1-2⁴¹. As replication fork collision frequency can be increased by the addition of an ectopic origin of replication (**1.6**), it seems logical that more replication fork collisions could provide RecBCD/SbcCD⁹¹ with more DNA to process, in-turn providing Cas1-2 with a greater pool of prespacers with which to implement naïve acquisition.

Despite initial problems with stimulating acquisition in MG1655 and RCe504 (**Figure 14**) and subsequent issues with plasmid instability in both EB371 and EB372 (**Figure 17**), it was eventually shown, over a 24-hour period, that a strain with an additional origin (EB372) is able to acquire more spacers in its CRISPR-I locus than wt (EB371) (**Figure 19** and **Figure 20**). As the only variable between the compared strains is the extra replication origin, it seems likely that additional replication forks are responsible for the extra acquisition demonstrated, backing-up the work of Levy *et al.* that replication termination can act as a prespacer source. The extra collisions between the clockwise *oriC* fork and the anticlockwise *oriZ* replicore on the right hand replicore mean that there is more RecBCD helicase-processed substrate for Cas1-2 to capture, and therefore the levels of acquisition are higher (**Figure 32**). To further clarify this hypothesis, it would be interesting to sequence the CRISPR-I loci to determine where the spacers originate from: if this hypothesis is correct, spacers from EB371 would be derived from the pEB628 plasmid and from the *TerA-C* fork-meeting point on the chromosome; the spacers from EB372 meanwhile would also presumably be derived from pEB628, *TerA-C* and between *oriC-oriZ* where forks collide without the control of Tus. It would also be interesting to repeat naïve acquisition assays as

before, but amplify both CRISPR loci, rather than just CRISPR-I, to ensure that acquisition is genuinely different between strains instead of preferentially using one CRISPR locus. In addition to sequencing and extra PCR, Red Recombineering could be utilised to introduce arabinose-inducible Cas1-2 from pEB628 into the host chromosome. This would eliminate the MGE (plasmid) source of prespacers, so only the fork colliding regions on the chromosome would provide prespacers for acquisition; this was attempted (**Figure 18**) but as of the time of writing has not yet been successful. Moreover, the termination of replication is only one source of fork collisions, replication forks can also collide with transcription machinery. As the highly transcribed *rrnH* operon is usually co-directional with the clockwise replication fork of *oriC*, with the addition of *oriZ* providing an anticlockwise fork over this section of the chromosome, it could lead to head-on collisions between the RNA-polymerase transcription machinery and the DNA replication machinery⁹³ (**Figure 32**). As it would be difficult to determine if collisions in this region of the chromosome are due to RNAP-replication collisions or two replication forks colliding, sequencing may not provide the necessary answers here. As such, it would be interesting to measure how acquisition is affected by using a mutated RNA-polymerase which is less prone to pausing and backtracking, therefore not able to stall replication forks. As demonstrated by Trautinger *et al.* in 2005, RNAP with a mutated β -subunit, *rpoB**35 due to an H1244Q substitution causes the polymerase to form short-lived open complexes, as though ppGpp is bound¹²⁹, this can help to stop polymerase backtracking, which prevents RNAP from forming blocks for the replisome, meaning replication can continue without stalling^{130 131}, limiting prespacer generation.

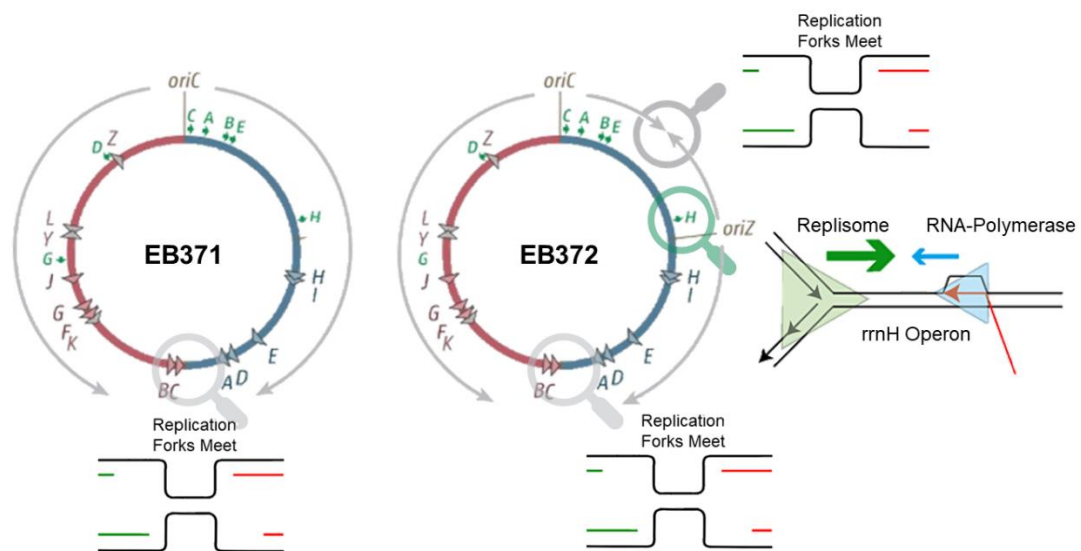


Figure 32 Differing chromosomal sources of prespacers in EB371 and EB372. The differing sources of prespacers generated by EB371 and EB372 are highlighted on the chromosome via magnifying glasses. The single-origin 'wild-type' strain EB371 is able to produce prespacer from where replication forks meet, opposite *oriC*. The dual-origin strain EB372 is able to produce spacers from where replication forks meet opposite *oriC*, where replication forks meet between *oriC* and *oriZ*, and where the anticlockwise *oriZ* replication fork collides with the RNA-Polymerase at the *rrnH* operon. Figure adapted from Dimude *et al.*⁹⁶ and McGlynn *et al.* 2012¹³².

The action of the replisome colliding with the RNAP complex could also explain the high levels of acquisition seen when using minimal-supplemented media compared to rich media (**Figure 26**). When nutrient supply is low, the lack of amino acids can slow the rate of translation at the ribosome. As translation is coupled to transcription in *E. coli*, the slow ribosome is not able to prevent RNAP backtracking¹³³, which can in-turn lead to a stalled RNAP, which can then be an obstruction for replication machinery¹²⁵ despite the highly-transcribed *rrn*-operons being co-directional with the replication forks in this case. This could indicate that RNAP-replisome collisions could be a major source of prespacers, perhaps more-so than colliding replication forks. Again, to test this the *rpoB**35 mutant strains could be tested against wt in supplemented M9 to see if this drops the spacer acquisition levels to wt in LB.

5.2 Cas1 N-terminal Fusions and Acquisition

YFP-fusion tags can be useful for localising proteins within a cell, giving valuable information about where they form foci. However, it is important that these tags do not interfere with the normal protein function, as such it is important to assess their function. A collaborating lab at Brunel University is using YFP-tags to monitor Cas1, therefore it is vital to see if YFP-tagged Cas1 can facilitate spacer acquisition *in vivo*. As shown in **Figure 21**, YFP-tagged Cas1 was not able to acquire spacers *in vivo* regardless of whether YFP-Cas1 was co-expressed with Cas2 on the same plasmid (EB371 x pEB681) or separate plasmids (BL12-AI x pLau18 YFP-Cas1 and pEB680). This could be simply due to having YFP, a 28 kDa protein¹³⁴, fused to the N-terminal of Cas1, a 33 kDa protein, which could cause a major steric hindrance. As shown in **Figure 33**, the N-terminals of Cas1 are located where Cas1 units bind to the central Cas2 units, having YFP here could be affecting complex formation. To test whether the Cas1 N-terminal was important for naïve acquisition, flag-tagged Cas1 (pTK7) was utilised in the EB371 strain. As the flag-tag is only 7 residues long, it should prove less of a steric hindrance than the large YFP tag. However, even with only a short tag at the N-terminal of Cas1, naïve acquisition could not be stimulated (**Figure 22**). N-terminally His₆-tagged Cas1 has been demonstrated integrating spacers into a CRISPR locus *in vitro*⁴¹, so the Flag/YFP-tag could be affecting the Cas1 spacer-capture function *in vivo*, but not necessarily the spacer integration function. To see if YFP-Cas1 could integrate spacers *in vitro*, a purification of YFP-Cas1 was attempted. YFP-Cas1 was co-purified with Cas2, however as YFP-Cas1 had no His-tag, it did not bind to the Ni²⁺ column and therefore only Cas2 was purified (**Figure 24**); this indicates that Cas2 may not form a complex with the YFP-Cas1 as otherwise some YFP-Cas1 would have been co-eluted from the columns with Cas2. If YFP-Cas1 can't form a complex with Cas2, the Y22 residues of (YFP)-Cas1 wouldn't be able to measure the correct 23 bp duplex region for spacer capture, nor would the arginine residues of Cas2 be able to non-specifically bind to the protospacer phosphate backbone, holding it in place. The next step towards testing spacer integration was to add a His₆-tag to the YFP-Cas1, so the YFP-Cas1 would bind to the Ni²⁺ column. However, at the time of writing, this cloning and protein purification has not been successful (**Figure 25**).

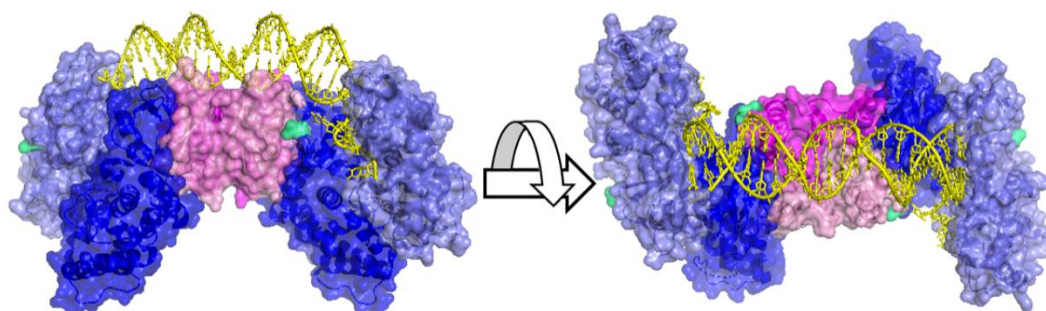


Figure 33 Cas1-2 bound to DNA with Cas1 N-terminals highlighted. The Cas1 subunits are shown in blue, Cas2 as pink, DNA as yellow. The N-terminals of Cas1 are shown in green.

5.3 The CRISPR System in *M. piezophila*

The unusual CRISPR locus of *Marinitoga piezophila* includes argonaute and primpol genes in addition to type III cas genes. Whilst the mAgo protein has garnered attention due to its preference for 5'-hydroxylated RNA structures which bear similarity to crRNA^{98 135}, thereby possibly being functionally similar to a CRISPR effector complex, the fundamental CRISPR mechanism of *M. piezophila* has not been investigated. As a first step towards understanding this CRISPR system, acquisition assays were performed using an *M. piezophila* CRISPR locus (pRPR1) and combinations of mCas1-2 (pPK273) and mAgo/mPrimPol (pPK276) in the *E. coli* strain BL21-AI. These *in vivo* assays showed that spacers could not be captured by mCas1-2 and integrated into the mCRISPR, regardless of whether mPrimPol/mAgo were present (**Figure 28**). As the *M. piezophila* proteins utilised tRNAs which are rare in BL21-AI, the BL21-C+ strain was employed to ensure the proteins were properly expressed; however, the subcloning required for this was not successful. There may be other reasons why these *M. piezophila* proteins could not acquire spacers; differing host proteins in *E. coli* and *M. piezophila* could be important factors. Whilst CRISPR-Cas systems have been successfully transferred between organisms, such as using the *S. thermophilus* CRISPR-Cas system in *E. coli*¹³⁶, the difference between *M. piezophila* and *E. coli* could prevent the CRISPR system working correctly outside of the original host organism. In *E. coli*, IHF is required for binding to the CRISPR leader

sequence and bending the DNA sufficiently to allow Cas1-2 to integrate spacers⁵⁸; *M. piezophila* however does not have known IHF genes¹³⁷ and therefore may require other proteins for guiding Cas1-2 with protospacers to the CRISPR leader. This could be tested by purifying the mCas1-2 and mAgo/mPrimPol and performing an *in vitro* spacer integration (Spln) assays with IHF^{58 41}. If these *in vitro* integration tests worked (with or without the mAgo/mPrimPol) it could highlight that only the spacer capture process of naïve acquisition is not functioning, perhaps due to differences between prespacer-generating host factors such as the full RecBCD complex, which is found in *E. coli* but not *M. piezophila*¹³⁷.

5.4 Cas3 and the Stringent Response

The final aspect of this project involved investigating the role of Cas3 in cell physiology, specifically the stringent response, continuing on from work in our lab which was started a decade ago. As the stringent response is coordinated by RelA and SpoT which synthesise and degrade the alarmone (p)ppGpp respectively, *E. coli* with knock-outs for these genes along with *cas3* were tested for their ability to produce (p)ppGpp. The ability to produce ppGpp was measured using host resistance to rifampicin: ppGpp binds to RNA-Polymerase at the same place as rifampicin¹²⁸, so when ppGpp is produced the host becomes more resistant to rifampicin, allowing transcription to occur, allowing the cell to grow. The deletion of *spoT* with or without *cas3* resulted in increased resistance to rifampicin; this was expected as these deletions would allow ppGpp to accumulate as SpoT is needed to break down ppGpp. When *relA* was deleted, cells became slightly resistant to rifampicin compared to wt, as when *cas3* alone was deleted. When both *relA* and *cas3* were deleted the cells were the least rifampicin resistant (i.e. most rifampicin sensitive) tested; this implies that Cas3 may have some ppGpp synthesising function *in vivo* with RelA. (**Figure 30**). However, many colonies with *relA/cas3* deletions were of abnormally large size implying suppressor mutations were taking place; meaning that the colonies were growing due to mutations which gave resistance to rifampicin, rather than ppGpp being produced and blocking rifampicin binding to the RNAP. These suppressor mutations made it hard to quantify colony forming units of ppGpp producing cultures. These mutations were also seen when

attempting to re-introduce *cas3* to see if the rifampicin sensitivity phenotype is reversed (i.e. increase resistance to rifampicin) (**Figure 31**). Furthermore, the re-addition of *cas3* via plasmids decreased resistance to rifampicin, even with empty vectors. As the rifampicin resistance phenotype change between *relA*, *relA-cas3*, and *cas3* were so small and suppressor mutations were prevalent, it is hard to conclusively say that Cas3 can help produce ppGpp. These *in vivo* tests could be built upon by using DNA helix distorting mutagens, such as UV or nitrofurazone, in place of rifampicin. The DNA lesions produced would be repaired by nucleotide excision repair, which can be regulated by ppGpp¹¹²; if ppGpp is being produced, the cells should be more resistant to the genotoxic stresses, much in the same way as with the rifampicin tests.

Acknowledgements

In alphabetical order, thank you to: Amy, Andy, Ed, Emily S, Emily W, Hannah, He, Jake, Killa-T, Liam, Lola, Lukas, Ronnie, Shivy-B, Staples/Other Ryan, Tabi. This robotic list does little to convey how much you have done to welcome me, support me, and keep me amused; you've made this year more than cromulent. Other thanks to Aiden's lab for the *M. piezophila* plasmids; Christian's lab for the RCe504 strain and YFP-Cas1; Ivana's lab for the λ -vir; Thorsten's lab for the meetings; and the undergraduates who were subjected to my 'teaching' (not as ominous as it sounds). Final sappy thanks to my parents and grandma, obviously.

Bibliography

1. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
2. Westra, E. R. *et al.* The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. *Annu. Rev. Genet.* **46**, 311–339 (2012).
3. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
4. Faure, G., Makarova, K. S. & Koonin, E. V. CRISPR–Cas: Complex Functional Networks and Multiple Roles beyond Adaptive Immunity. *J. Mol. Biol.* **431**, 3–20 (2018).
5. Riede, I. & Eschbach, M. L. Evidence that TraT interacts with OmpA of *Escherichia coli*. *FEBS Lett.* **205**, 241–245 (1986).
6. Vasu, K. & Nagaraja, V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiol. Mol. Biol. Rev.* **77**, 53–72 (2013).
7. Ofir, G., , Sarah Melamed, Hila Sberro, Zohar Mukamed, Shahar Silverman, G. Y. S. D. & Sorek, and R. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).
8. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
9. Blower, T. R. *et al.* A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat. Struct. Mol. Biol.* **18**, 185–191 (2011).
10. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakamura, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
11. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in

- Prokaryotes. *Science* (80-.). **315**, 1709–1712 (2007).
12. Pourcel, C. CRISPRdb. <http://crispr.i2bc.paris-saclay.fr/>; (2017).
 13. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **117**, 119–128 (2015).
 14. Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: A CRISPR interactive database. *Bioinformatics* **25**, 3317–3318 (2009).
 15. Al-Attar, S., Westra, E. R., Van Der Oost, J. & Brouns, S. J. J. Clustered regularly interspaced short palindromic repeats (CRISPRs): The hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392**, 277–289 (2011).
 16. Barrangou, R. & Horvath, P. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* (80-.). **327**, 167–170 (2010).
 17. Koonin, E. V, Makarova, K. S., Zhang, F. & Engineering, B. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67–78 (2017).
 18. Makarova, K. S. *et al.* An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
 19. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* (80-.). **353**, (2016).
 20. Yan, M., Yan, H., Ren, G., Zhao, J. & Guo, X. CRISPR-Cas12a-Assisted Recombineering in Bacteria. **83**, 1–13 (2017).
 21. O'Connell, M. R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *J. Mol. Biol.* **431**, 66–87 (2018).
 22. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* (80-.). **321**, 960–964 (2008).
 23. Hidalgo-Cantabrana, C., Goh, Y. J. & Barrangou, R. Characterization and Repurposing of Type I and Type II CRISPR–Cas Systems in Bacteria. *J. Mol. Biol.* **431**, 21–33 (2018).

24. Pul, Ü. *et al.* Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512 (2010).
25. Majsec, K., Bolt, E. L. & Ivančić Baće, I. Cas3 is a limiting factor for CRISPR-Cas immunity in *Escherichia coli* cells lacking. *BMC Microbiol.* 1–9 (2016).
doi:10.1186/s12866-016-0643-5
26. Westra, E. R. *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393 (2010).
27. Perez-Rodriguez, R. *et al.* Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* **79**, 584–599 (2011).
28. Pougach, K. *et al.* Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* **77**, 1367–1379 (2010).
29. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945–947 (2012).
30. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
31. Wiedenheft, B. *et al.* Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. *Structure* **17**, 904–912 (2009).
32. Babu, M. *et al.* A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502 (2011).
- 33.: Nocek, B., Skarina, T., Brown, G., Yakunin, A., Joachimiak, A. Crystal structure of a putative ssRNA endonuclease Cas2, CRISPR adaptation protein from *E. coli*. *RSCB PDB* (2013).
34. Beloglazova, N. *et al.* A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol.*

Chem. **283**, 20361–20371 (2008).

35. Nuñez, J. Davies, C., Doudna, J., Kranzusch, P. Noeske, J. & Wright, A. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
36. Mojica, Francisco J M, C. Díez-Villasenor, J. G.-M. and C. A. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
37. Yosef, I. *et al.* DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc. Natl. Acad. Sci.* **110**, 14396–14401 (2013).
38. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, 1–7 (2012).
39. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci.* **108**, 10098–10103 (2011).
40. Vercoe, R. B. *et al.* Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genet.* **9**, (2013).
41. Radovčić, M. *et al.* CRISPR–Cas adaptation in *Escherichia coli* requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic Acids Res.* **46**, 10173–10183 (2018).
42. Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840–853 (2015).
43. James K. Nuñez, Lucas B. Harrington, Philip J. Kranzusch, Alan N. Engelman, and J. A. D. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* **527**, 535–538 (2015).

44. Musharova, O. *et al.* Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* **45**, 3297–3307 (2017).
45. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. **520**, 505–510 (2015).
46. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 1–13 (2016).
47. Shiimori, M. *et al.* Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in *pyrococcus furiosus*. *Nucleic Acids Res.* **45**, 11281–11294 (2017).
48. Killelea, T. & Bolt, E. L. CRISPR-cas Adaptive Immunity and the Three Rs. *Biosci. Rep.* **0**, BSR20170297 (2017).
49. Dillard, K. E. *et al.* Article Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. 1–13 (2018).
50. Ivančić-Bace, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and Naïve adaptation in *E. Coli* CRISPR-Cas immunity. *Nucleic Acids Res.* **43**, 10821–10830 (2015).
51. Ramachandran, A. & Bailey, S. Memory Upgrade: Insights into Primed Adaptation by CRISPR-Cas Immune Systems. *Mol. Cell* **64**, 641–642 (2016).
52. Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol. Cell* **63**, 852–864 (2016).
53. Krivoy, A. *et al.* Primed CRISPR adaptation in *Escherichia coli* cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. *Nucleic Acids Res.* **46**, 4087–4098 (2018).
54. Semenova, E. *et al.* Highly efficient primed spacer acquisition from targets destroyed

- by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci.* **113**, 7626–7631 (2016).
55. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S. & Lundgren, M. Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Sci. Rep.* **7**, 1–10 (2017).
 56. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
 57. Aeling, K. A. *et al.* Indirect recognition in sequence-specific DNA binding by *Escherichia coli* integration host factor: The role of DNA deformation energy. *J. Biol. Chem.* **281**, 39236–39248 (2006).
 58. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol. Cell* **62**, 824–833 (2016).
 59. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
 60. Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. M. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol.* **10**, 792–802 (2013).
 61. Goren, M. G. *et al.* Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. *Cell Rep.* **16**, 2811–2818 (2016).
 62. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* **4**, 1–19 (2015).
 63. Charpentier, E., Richter, H., van der Oost, J. & White, M. F. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**, 428–441 (2015).

64. Jason Carte, Ruiying Wang, Hong Li, Rebecca M. Terns, and M. P. T. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **15**, 3489–3496 (2008).
65. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
66. Plagens, A., Richter, H., Charpentier, E. & Randau, L. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol. Rev.* **39**, 442–463 (2015).
67. Zhao, H. *et al.* Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014).
68. Jackson, R. N. Crystal structure of the CRISPR RNA – guided surveillance complex from. **1473**, 1473–1480 (2014).
69. Hochstrasser, M. L. *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci.* **111**, 6618–6623 (2014).
70. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 1–26 (2006).
71. Mulepati, S. & Bailey, S. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J. Biol. Chem.* **286**, 31896–31903 (2011).
72. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342 (2011).
73. Howard, J. A. L., Delmas, S., Ivančić-Baće, I. & Bolt, E. L. Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem. J.* **439**, 85–95 (2011).

74. Mulepati, S. & Bailey, S. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA Target. *J. Biol. Chem.* **288**, 22184–22192 (2013).
75. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **46**, 606–615 (2012).
76. Van Erp, P. B. G. *et al.* Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Res.* **43**, 8381–8391 (2015).
77. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* **163**, 854–865 (2015).
78. Cooper, L. A., Stringer, A. M. & Wade, J. T. Determining the specificity of cascade binding, interference, and primed adaptation In Vivo in the Escherichia coli type I-E CRISPR-cas system. *MBio* **9**, 1–18 (2018).
79. Van Erp, P. B. G. *et al.* Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. *ACS Chem. Biol.* **13**, 481–490 (2018).
80. Westra, E. R. *et al.* CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).
81. Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. *Mol. Cell* **70**, 385–394.e3 (2018).
82. Dorman, C. J. H-NS: A universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* **2**, 391–400 (2004).
83. Taylor, A. F. *et al.* Control of RecBCD Enzyme Activity by DNA Binding- and Chi Hotspot-Dependent Conformational Changes. *J Mol Biol* **426**, 3479–3499 (2014).
84. Leonard, A. C. & Me, M. DNA Replication Origins. *Cold Spring Harb. Perspect. Biol.* **5:a010116**, 1–17 (2013).

85. Skarstad, K. & Katayama, T. Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* **5**, 1–17 (2013).
86. Reyes-Lamothe, R., Nicolas, E. & Sherratt, D. J. Chromosome Replication and Segregation in Bacteria. *Annu. Rev. Genet.* **46**, 121–143 (2012).
87. Georgescu, R. E., Yao, N. Y. & O'Donnell, M. Clamps To Bypass Replication Barriers. *FEBS Lett.* **584**, 2596–2605 (2011).
88. Hill, T. M. Arrest of Bacterial DNA Replication. *Annu. Rev. Microbiol.* **46**, 603–633 (1992).
89. Duggin, I. G. & Bell, S. D. Termination Structures in the Escherichia coli Chromosome Replication Fork Trap. *J. Mol. Biol.* **387**, 532–539 (2009).
90. Berghuis, B. A. *et al.* What is all this fuss about Tus? Comparison of recent findings from biophysical and biochemical experiments. *Crit. Rev. Biochem. Mol. Biol.* **53**, 49–63 (2018).
91. Wendel, B. M., Cole, J. M., Courcelle, C. T. & Courcelle, J. SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. *Proc. Natl. Acad. Sci.* **115**, 349–354 (2018).
92. Courcelle, J., Wendel, B. M., Livingstone, D. D. & Courcelle, C. T. RecBCD is required to complete chromosomal replication: Implications for double-strand break frequencies and repair mechanisms. *DNA Repair (Amst)*. **32**, 86–95 (2015).
93. Ivanova, D. *et al.* Shaping the landscape of the Escherichia coli chromosome: Replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res.* **43**, 7865–7877 (2015).
94. Kelman, Z. DNA Polymerase III Holoenzyme: Structure and Function of a Chromosomal Replicating Machine. *Annu. Rev. Biochem.* **64**, 171–200 (1995).
95. de Septenville, A. L., Duigou, S., Boubakri, H. & Michel, B. Replication fork reversal after replication-transcription collision. *PLoS Genet.* **8**, (2012).

96. Dimude, J. U., Midgley-Smith, S. L., Stein, M. & Rudolph, C. J. Replication termination: Containing fork fusion-mediated pathologies in *Escherichia coli*. *Genes (Basel)*. **7**, (2016).
97. Alain, K. *et al.* Thermo-Piezophilic Bacterium Isolated Under High Hydrostatic Pressure From a Deep-Sea Hydrothermal Vent. *Int. J. Syst. Evol. Microbiol.* **52**, 1331–1339 (2002).
98. Kaya, E. *et al.* A bacterial Argonaute with noncanonical guide RNA specificity. *Proc. Natl. Acad. Sci.* **113**, 4057–4062 (2016).
99. Rowen, L. & Kornberg, A. Primase, the dnaG Protein of *E. coli*. *J. Biol. Chem.* **254**, 758–764 (1978).
100. Della, M. *et al.* Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science (80-.)*. **306**, 683–685 (2004).
101. Guillian, T. A., Keen, B. A., Brissett, N. C. & Doherty, A. J. Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. *Nucleic Acids Res.* **43**, 6651–6664 (2015).
102. Guillian, T. A. & Doherty, A. J. Primpol—prime time to reprime. *Genes (Basel)*. **8**, (2017).
103. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).
104. Azlan, A., Dzaki, N. & Azzam, G. Argonaute: The executor of small RNA function. *J. Genet. Genomics* **43**, 481–494 (2016).
105. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009).
106. Doxzen, K. W. & Doudna, J. A. DNA recognition by an RNA-guided bacterial Argonaute. *PLoS One* **12**, 1–14 (2017).

107. Raina, M. & Ibba, M. TRNAs as regulators of biological processes. *Front. Genet.* **5**, 1–14 (2014).
108. Haseltine, W. A. & Block, R. Synthesis of Guanosine Tetra- and Pentaphosphate Requires the Presence of a Codon-Specific, Uncharged Transfer Ribonucleic Acid in the Acceptor Site of Ribosomes. *Proc. Natl. Acad. Sci.* **70**, 1564–1568 (1973).
109. Hogg, T., Mechold, U., Malke, H., Cashel, M. & Hilgenfeld, R. Conformational Antagonism between Opposing Active Sites in a Bifunctional RelA/SpoT Homolog Modulates (p)ppGpp Metabolism during the Stringent Response. *Cell* **117**, 57–68 (2004).
110. Atkinson, G. C. & Hauryliuk, V. Evolution and Function of the RelA/SpoT Homologue (RSH) Proteins. *eLS* 1–8 (2012). doi:10.1002/9780470015902.a0023959
111. Zuo, Y., Wang, Y. & Steits, T. A. The Mechanism of E. coli RNA Polymerase Regulation by ppGpp Is Suggested by the Structure of their Complex. *Mol Cell* **50**, 430–436 (2013).
112. Kamarthapu, V. *et al.* ppGpp couples transcription to DNA repair in E. Coli. *Science* (80-.). **352**, 993–996 (2016).
113. Rasouly, A., Pani, B. & Nudler, E. A Magic Spot in Genome Maintenance Aviram. *Trends Genet.* **33**, 58–67 (2017).
114. Molodtsov, V. *et al.* Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol. Cell* **69**, 828–839.e5 (2018).
115. Ferullo, D. J. & Lovett, S. T. The stringent response and cell cycle arrest in Escherichia coli. *PLoS Genet.* **4**, 21–23 (2008).
116. Zhang, Y., Rejman, D. & Gerdes, K. Novel (p)ppGpp Binding and Metabolizing Proteins of Escherichia coli. **9**, 1–20 (2018).
117. Bachmann, B. J. Derivations and genotypes of some mutant derivatives of Escherichia coli K-12. *Escherichia coli Salmonella Cell. Mol. Biol.* 2460–2488 (1996).

doi:[http://dx.doi.org/10.1016/0968-0004\(88\)90241-1](http://dx.doi.org/10.1016/0968-0004(88)90241-1)

118. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
119. Gel, Q., Kit, E., Spin, Q. & Kit, M. Quick-Start Protocol: QIAprep® Spin Miniprep Kit. 5–6 (2010).
120. Protocol, Q. QIAquick® Gel Extraction Kit. 6–7 (2015).
121. QIAGEN. QIAquick® PCR Purification Kit. *Qiagen* 5–6 (2010). doi:1063920
122. Thomason, L. C., Costantino, N. & Court, D. L. *E. coli* Genome Manipulation by P1 Transduction. *Curr. Protoc. Mol. Biol.* 1.17.1-1.17.8 (2007).
doi:10.1002/0471142727.mb0117s79
123. NEB. NEBcloner v1.3.8. (2018). Available at: <http://nebcloner.neb.com/#!/>.
124. Wearne, S. J. CRISPR-Cas adaptation: the basis for prokaryotic immunity. A thesis submitted to the University of Nottingham for the degree of Master of Research (2015).
125. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533–543 (2011).
126. GenScript. Rare Codon Analysis. <https://www.genscript.com/tools/rare-codon-analysis> (2018).
127. Cady, K. C. & O'Toole, G. A. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.* **193**, 3433–3445 (2011).
128. Reddy, P. S. & Raghavan, A. Evidence for a ppGpp-binding site on *Escherichia coli* RNA polymerase: proximity relationship with the rifampicin-binding domainEvidence for a ppGpp-binding site on *Escherichia coli* RNA polymerase: proximity relationship with the rifampicin-binding domain. *Mol Microbiol* **15**, 255–265 (1995).

129. Trautinger, B. W., Jaktaji, R. P., Rusakova, E. & Lloyd, R. G. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell* **19**, 247–258 (2005).
130. Mahdi, A. A., Buckman, C., Harris, L. & Lloyd, R. G. Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. *Genes Dev.* **20**, 2135–2147 (2006).
131. Rudolph, C. J., Dhillon, P., Moore, T. & Lloyd, R. G. Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair (Amst)*. **6**, 981–993 (2007).
132. McGlynn, P., Savery, N. J. & Dillingham, M. S. The conflict between DNA replication and transcription. *Mol. Microbiol.* **85**, 12–20 (2012).
133. Proshkin, S., Rahmouni, R., Mironov, A. & Evgeny, N. Polymerase in Transcription Elongation. **328**, 504–508 (2010).
134. UniProt Entry YFP (A0A059PIR9_AEQVI). (2017). Available at: <https://www.uniprot.org/uniprot/A0A059PIR9>.
135. Lapinaite, A., Doudna, J. A. & Cate, J. H. D. Programmable RNA recognition using a CRISPR-associated Argonaute. *Proc. Natl. Acad. Sci.* **115**, 3368–3373 (2018).
136. Sapranaukas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
137. Lucas, S. *et al.* Complete genome sequence of the thermophilic, piezophilic, heterotrophic bacterium *Marinitoga piezophila* KA3. *J. Bacteriol.* **194**, 5974–5975 (2012).