University of Nottingham
UK | CHINA | MALAYSIA

Analysis of the *E. coli* CRISPR Cas3 protein and its potential for regulation by a short non-coding RNA

A thesis submitted to the University of Nottingham for the degree of Master of Research

Niamh Kennerdale, BSc (Honours), MSc

March 2023

Abstract

CRISPR-Cas is an RNA-guided adaptive immune system characterised by short repeating DNA sequences, which is an evolutionary adaption in response to survival pressure from invasive genetic elements. Cas3 is a signature protein in type I-E CRISPR-Cas systems, acting as an effector molecule to provide the host cell with immunity from invasion by bacteriophage. Cas3 possesses dual nuclease-helicase activity, acting in conjunction with a large multiprotein complex termed Cascade to degrade invasive DNA for capture and genome integration.

Previous research on the *Escherichia coli* Cas3 indicated the possibility of allosteric regulation via temperature change as a proxy mechanism, hinting at alternative regulatory pathways by non-canonical genetic elements. Here, we signify the antisense promoter antiP*cas,* that overlaps with the *cas3* protein-coding strand and directs a short non-coding RNA transcript. Computational modelling shows the RNA forms a complex secondary structure, which is reminiscent of regulatory RNA molecules such as riboswitches and ribozymes. The potential significance of the anti-*cas* RNA in *cas3* regulation and post-transcriptional modulation has never been investigated.

Genetics and biochemistry were used to examine the relationship between the anti-*cas* RNA on Cas3 activity *in vitro* and the regulatory role of the RNA-encoding gene on *cas3*. Deletion of the antiP*cas* gene was performed to determine the influence on CRISPR function in live cells. Insight was also gained by modelling and sequence analysis of the anti-*cas* RNA using bioinformatics.

Contents

Figures

Chapter 1: Introduction

1.1 CRISPR-Cas systems

Bacteria and archaea are among the oldest and most genetically diverse life forms on earth, and have undergone beneficial adaptations that have enabled survival against predation by viral mobile genetic elements[1]. The adaptation of bacteria and archaea over generations to resist infection is defined as an "arms race", that stimulated the evolution of primitive adaptive immune systems to challenge phage attack[2,3]. The natural selection of prokaryotes with this acquired genomic adaptation enabled species proliferation and consequently species-specific development of this adaptive immune system, which came to be known as the CRISPR-Cas system[1,2].

The CRISPR (clustered-regularly-interspaced-short-palindromic-repeats) locus is a region characterised by short repeating sequences, and functions as a specialised adaptive immune system. It was initially discovered in the *Escherichia coli* strain CSR603, as an anomalous region of short repeats near to the *iap* gene[4]. Subsequent research later identified the sequence in the archaeon *Haloferax mediterranei* and multiple *Escherichia* species[2], before its widespread discovery in prokaryotic genomes, aided largely by advancements in next generation sequencing[2].

The high degree of conservation of CRISPR-Cas systems in both bacterial and archaeal species may indicate that the CRISPR array itself may be interchanged via horizontal gene transfer[6,8], serving as a selfish but beneficial transposon in symbiosis with the host cell to enhance immunity against phage[2, 29]. As a consequence of this selective evolutionary pressure, an

enormous diversity of different CRISPR-Cas systems exist, which are typically categorised based on the proteins involved in enabling CRISPR immunity.

Over 90% of bacterial species and archaea are defined as having class 1 and class 2 CRISPR systems, which are characterised by *cas* (CRISPR-associated) genes that encode Cas enzymatic proteins, previously termed COG (Clusters of Orthologous protein Groups)[2,3]. Cas proteins, or endonucleases, are intrinsically conserved of the CRISPR array, and are key for the cell to mount an effective immune response against phage predation[2,5,7,8].

In both classes, the CRISPR array is situated downstream of the *cas* genes, and consists of repeating DNA sequences 24-48 nucleotides in length, punctuated by short spacer regions of non-coding DNA[2,6] (*Figure* 1.1). Classes 1 and 2 can be further divided into subtypes, ranging from Type I to IV in Class 1, and Type II to VI in Class 2. Essentially, Class 1 systems use multi-protein effector complexes to destroy invasive DNA, while Class 2 systems use a single effector complex[5,8,15]. A notable CRISPR system subtype is Type II in Class 2, characterised by the Cas9 effector endonuclease which was first identified in *Streptococcus pyogenes*[3] and subsequently widely utilised as a versatile gene editing tool. The subclasses can also be defined as Types I -E [15].

The bacteria *Escherichia coli* is classified as having a Type 1 Class 1 CRISPR system, which constitutes the Cas1 and Cas2 proteins involved in CRISPR adaptation, and the signature Cas3 effector protein involved in CRISPR interference[1,11,12]. *E. coli* is a well-established model organism used in molecular biology and biotechnology with a fully sequenced and annotated genome, enabling the robust study of CRISPR-Cas expression and functioning in this species.

The *E. coli* K-12 MG1655 strain is a particularly well-characterised organism that has been used to study *in vivo* CRISPR-Cas activity. The CRISPR locus of this strain contains two CRISPR arrays, the first consisting of thirteen highly conserved repeats, and the second with seven, that flank twelve and six spacer sequences respectively[1,9,39]. The CRISPR-Cas locus is composed of *cas*1 and *cas2* (*ygbT*, *ygbF*) genes encoding the Cas1 and Cas2 proteins, along with the *cas*-A to *cas*-E genes that encode the Cascade complex[37]. The *cas3* gene (*ygcB*) is situated approximately 400bp upstream of the *cas*-A-E array, and is the largest protein-coding gene in the locus. Detailed descriptions of the *cas3* gene and Cas3 endonuclease role in CRISPR-Cas functioning are given in sections 1.1.2, 1.1.3, and 1.1.4.

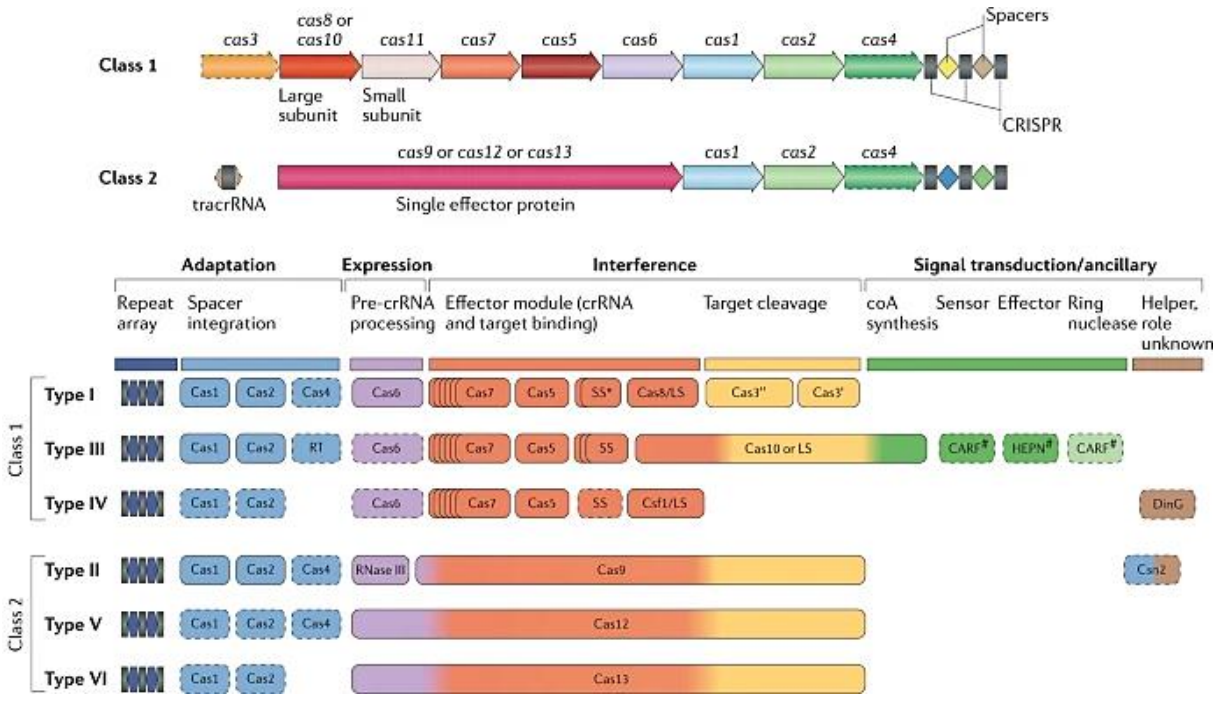


Figure 1.1 **CRISPR-Cas system classification**

An overview of the signature *cas* genes and their functionality in viral defence in each class and subtype found across bacteria and archaea. (a) Illustration of the generic organisation of classes 1 and 2 CRISPR-Cas loci. (b) Functional molecules and complexes involved in the stages of CRISPR immunity across the two main classes. Figure adapted from [1].

1.1.1 CRISPR as a multi-component immune system

CRISPR systems function through an RNA-processing interference mechanism, and behave as a DNA cleavage and sequence integration system to provide adaptive immunity to the host cell[1,2,4]. The integration of short foreign DNA fragments into the genome is performed by Cas1, Cas2 and Cas4 proteins, a process defined as adaptation, that serves as precursor to the initiation of a immune response against future phage invasion[4,5,8,9]. At the expression stage, the CRISPR array is typically transcribed as a single molecule, the pre-CRISPR RNA (pre-crRNA), that is processed into mature CRISPR RNAs (crRNAs), each containing the spacer sequence and parts of the flanking repeats (*Figure* 1.2). At its simplest iteration, crRNA serves as an embedded genetic memory that enables the cell to prime itself against future invasion, thus ensuring survival and proliferation.

The biogenesis and expression of crRNA is crucial for enabling the cell to mount an effective immune response against phage, and primes effector proteins to target viral DNA during CRISPR interference. It is critical that short spacers of non-self MGE-derived DNA be transcribed into small RNA molecules prior to interference. In Class 1 systems transcription is performed by Cas6, whereas in Class 2 it is performed by the RNase III molecule[1]. At the interference stage, the mature crRNA, which typically remains bound to the processing protein complex, serves as a guide to recognize the protospacer, or a highly similar sequence, in invasive phage DNA, which is then cleaved and inactivated by a Cas nuclease that either is part of the effector or is recruited by a multiprotein effector complex[6,7]. Sheared DNA fragments are then captured by Cas1 and Cas2 for genomic integration as new spacers, further extending the CRISPR array and priming the host cell against future infection.

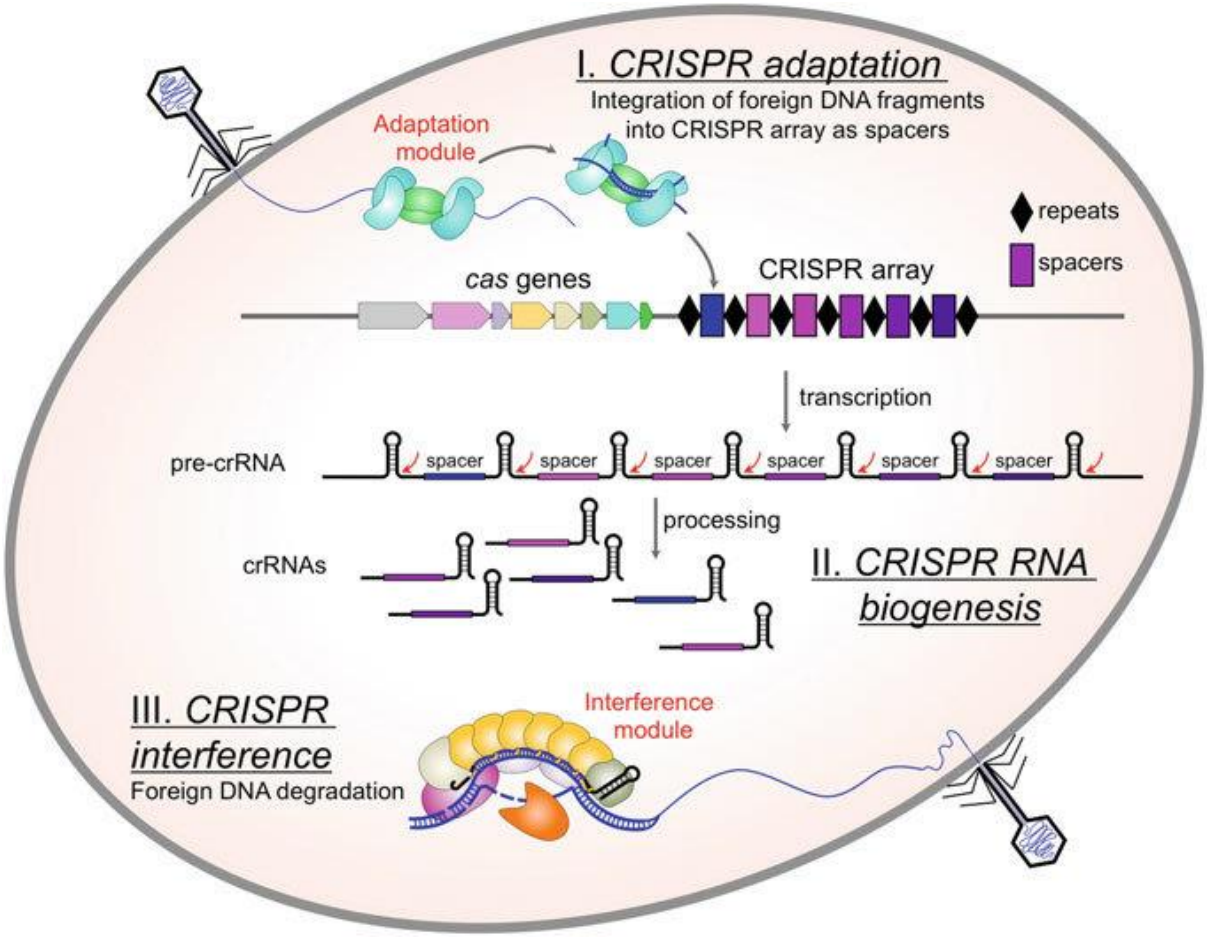


Figure 1.2 **Overview of the CRISPR-Cas immune system in prokaryotes**

Initiation of the CRISPR immune response begins upon the passage of MGE DNA into the cell and subsequent detection and degradation by an interference complex. Cas-enzyme mediated CRISPR interference is guided by spacer crRNA from integrated foreign MGE DNA, enabling the "recognition" of invader DNA from previous infection. Diagram taken and adapted from [9].

CRISPR interference is complex stage in overall CRISPR immunity, as it involves the activity of large protein effector complexes to detect and neutralise MGE DNA. As mentioned previously, the key effector molecule active in class 1 type I-E systems is the Cas3 nuclease-helicase, which I will be focussing on and discussing in the context of this chapter.

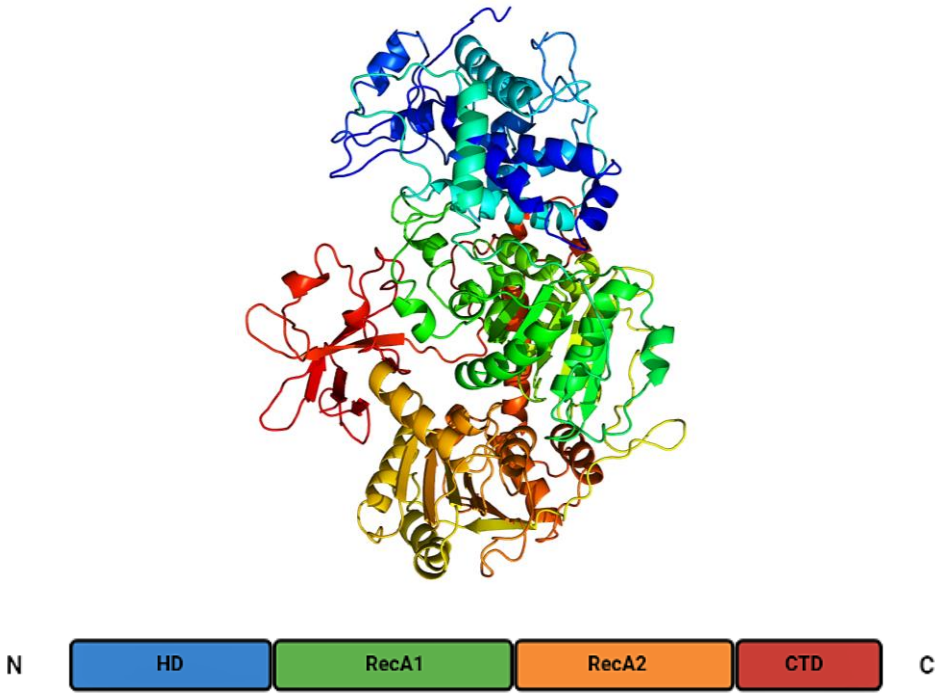
1.1.2 Cas3 as an effector molecule in CRISPR interference

Cas3 is a signature protein in Type 1 Class 1 CRISPR systems, which in *E. coli* species has an integral role in CRISPR immunity. It is a dual-function endonuclease, which cleaves dsDNA through ATP-dependent translocation of DNA strands in a 5` to 3` orientation to a nuclease active site domain[10,59] . Nuclease activity is preferentially dependent on the presence of Mg^{2+} ions, although activity can be supported at a reduced rate by manganese and cobalt ions[11].

The structure of *E. coli* Cas3 is composed of a nuclease domain linked to two central RecA domains, that exhibit similar characteristics to superfamily-2- DNA helicases[10]. In addition, an accessory CTD (C-terminal helicase domain) region is thought to be an additional acquired macromolecule, which exhibits varied morphology in multispecies comparative analysis, and is key for enabling Cas3 to dock onto co-functioning protein subunits in Cascade[14,10].

The *E. coli* Cas3 folded secondary structure and functional domains have been predominately modelled from well-annotated Cas3 crystal structures from species such as *Thermobifida fusca* and *Thermobaculum terrenum* (*Figure* 1.3). Analysis of secondary structures built from atomic residues in Cas3 proteins in a multispecies comparison show a varied morphology of domain size and inter-strand linkage. This variety may reflect the nature of helicase and nuclease domains acquired through differing evolutionary pathways[13]. Cas3 subtypes exhibit structural diversity across multiple species, such as fusion and fission proteins. However, the most common catalytically active form of Cas3 is that of a fused HD-nuclease-translocase, which is characteristic of the *E. coli* Cas3[20,64,68].

A



B

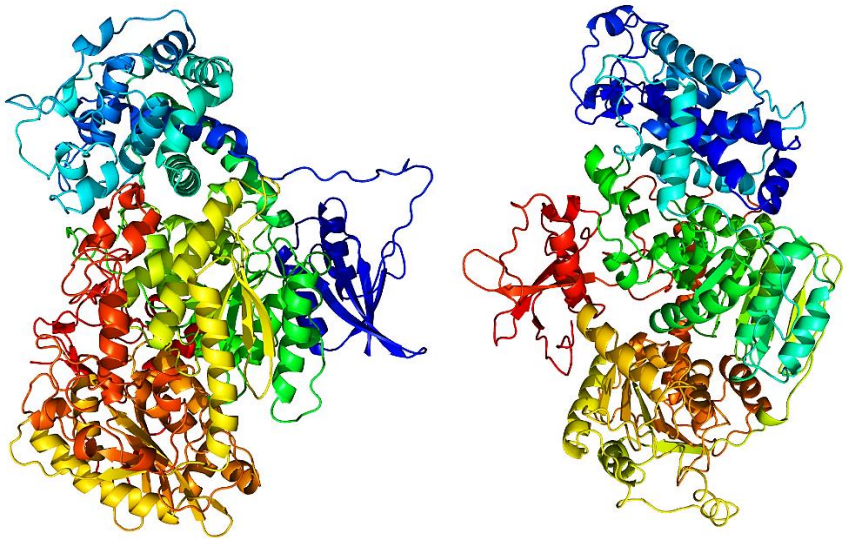


Figure 1.3 **Cas3 structure and domain comparative models**

(A) *Escherichia coli* Cas3 high-resolution model generated by Phyre2 analysis using a FASTA amino acid sequence obtained from SnapGene file of *cas3*/*ygcB*, model constructed using sequence templates from *Thermobifida fusca* (PDB 4QQX) and *Thermobaculum terrenum* (PDB 4Q2D). Model is viewed in a vertical N

terminal to C terminal orientation, with domains highlighted as shown in key (B) Cas3 models from *Pseudomonas aeruginosa* (UniProt Q02ML8) (L) and *Thermus thermophilus* (UniProt Q53VY2) (R).

The CRISPR interference reaction in Type 1 Class 1 systems is distinct from the mode of function in other systems, the effector complex is composed of a priming crRNA strand, the Cas3 endonuclease and the unique ribonucleoprotein Cascade (CRISPR-associated complex for antiviral defence) molecule. The involvement of the Cascade effector complex is crucial in specifically guiding Cas3 to cleave target DNA and prevent non-specific strand degradation, acting as "surveillance" complex to detect for the presence of invasive DNA[11,12,20].

The Cascade complex is comprised of five protein subunits, which work in conjunction to cleave crRNA prior to loading , and enable the docking of Cas3. Cascade is primed by a strand of mature crRNA that fuses at the *Cse6* subunit, and forms a "backbone" intersecting through the protein medial topology[5,8,20]. The structural diversity of the Cascade complex across Type 1 CRISPR systems is made explicit by the seven or eight subtypes, 1-A to F[5,6]. Within these, Cascade complexes vary in composition from three to five protein subunits, however *Cse*5 and *Cse*7 are common to all subtypes, and show variation in catalytic functions of the subunits during loading of the crRNA payload and its DNA targeting [5,8,10,68].

The CRISPR interference stage in Type I-E systems involve the co-functioning of Cas3 with the Cascade-crRNA complex, to target and destroy foreign DNA in a multistage process. The maturation of a pre-crRNA transcript is first mediated by the Cas6e endonuclease, which cleaves pre-crRNA into individual guide molecules[1,63]. Mature crRNA then binds with the *Cse6*, Cas5 and Cas7 Cascade subunits prior to Cas3 recruitment. An evolutionarily conserved region on the outer surface of *Cse1* and positive residues in Cas7 promote facilitated diffusion of the Cascade complex during target searching[8,19].

Cas3 is recruited to the Cascade complex through a docking mechanism, the *Cse1-Cse2* subunits anchoring with the Cas3 alpha helix chain and enabling loading. A key prerequisite for strand targeting and cleavage is the formation of an RNA-DNA hybrid termed an R-loop, caused by the crRNA base-pairing with duplex DNA and displacing the non-target strand in a "bulge". The highly specific protospacer adjacent motif (PAM) recognition codon scans the crRNA backbone of Cascade in a 5` to 3` orientation for a complementary cRNA 3bp base site, typically an NGG codon. This enables complementary base pairing and Cascade binding with the proximal strand of target DNA, at a complementary site in an AT-rich sequence region[5,8,13]. Partial binding of Cascade to target DNA with tolerated mismatches within the spacer sequence, but not the PAM, elicits collateral ssDNA cleavage activity of recruited *Eco*Cas3[12,13]. Conversely, stable binding with complete R-loop formation drives *Eco*Cas3 to nick the non-target strand (NTS) in the bound DNA. Helicase-dependent unwinding then combines with trans ssDNA cleavage of the target strand and repetitive *cis* cleavage of the NTS to degrade the target double-stranded DNA.

High-speed atomic force microscopy demonstrates that *Eco*Cas3 bound to Cascade repeatedly reels in and releases the target DNA, followed by target fragmentation. Cas3 can also rapidly dissociate away from Cascade and translocate along the proximal DNA strand in an ATP-dependent unidirectional movement, causing a prolonged "shearing" of the ssDNA strand (*Figure* 1.4). ssDNA passes into the binding channel and into the nuclease active site to undergo cleavage[18]. Sheared dsDNA fragments are then captured by Cas1-Cas2 for genome integration. The close interplay between interference and adaptation depends on the ability of Cascade to initiate target recognition, and Cas1-Cas2 to effectively integrate new spacers. Primed adaptation is defined as Cascade-Cas3 reinforcing immunity against a recognised phage via base-pairing of a pre-existing spacer crRNA, whereas naïve adaptation occurs in the

absence of interference and is catalysed by Cas1-Cas2 to capture DNA from a MGE not previously encountered.

The central role of Cas3 in neutralising MGE DNA during CRISPR interference may be enhanced by the formation of a primed acquisition complex (PAC), as has been tested in a *T.fusca* model[17,19]. Upon the binding and capture of MGE DNA fragments by Cas1-Cas2 for genome integration, biomolecular fluorescence complementation (BiFC) analysis demonstrated the cohesive binding of Cas1-Cas2 with Cas3-Cascade to form a large 710kDa multi-protein acquisition complex. Cascade forms the keystone of the PAC, as Cas3 and Cas1-Cas2 both require Cascade for stable association with the target DNA. The PAC demonstrated ATP-driven translocation along DNA strands for extended distances, resulting in the integration of MGE spacers at further interspaced loci than seen in non-PAC acquisition by Cas1 and Cas2[19]. *Tfu*Cas3 nuclease activity in the PAC was not found to be significantly impacted, as compared to the levels measured in the typical Cas3-Cascade complex. The close positioning of Cas1-2 bound to Cascade allows for an increased spacer acquisition efficacy, by the direct uptake of cleaved dsDNA fragments from Cas3 strand shearing in a rapid progression movement. The physical and functional coupling of interference with adaptation in a PAC can re-cycle DNA from the phage invader into the CRISPR loci, updating immunity in 'primed' or 'targeted' adaptation[19,68]. The involvement of Cas3 in primed adaptation in a PAC complex is not a typical interference reaction, typically distinct separate stages as previously described. The efficacy of the Cas3-Casacde effector complex in catalysing progressive DNA cleavage has highlighted its potential use in biotechnology and precision genome editing, presenting a complex and highly efficient CRISPR tool comparative with Cas9 and Cas12a nucleases.

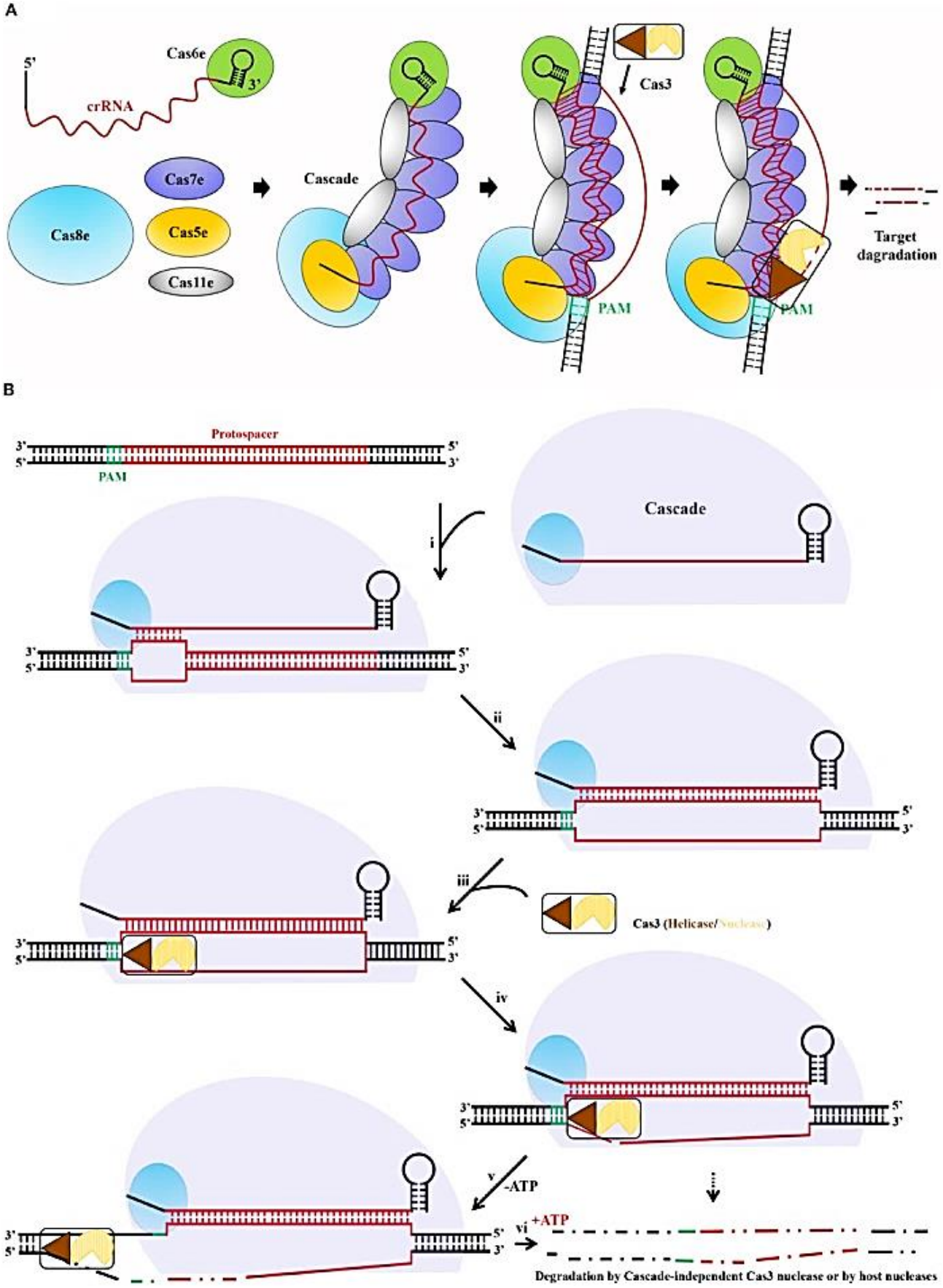


Figure 1.4 *Cont.*

Figure 1.4 **Cas3-Cascade complex and activity in CRISPR interference**

CRISPR interference mechanism in Type I-E systems in *E. coli* K12. Cascade subunits 8, 11, 7, 5 and 6 also termed A-E respectively (A) Mature crRNA acts as a scaffold in the Cascade protein assembly. Upon recognition of PAM in the target DNA, crRNA base pairs with the stand, initiating Cas3 recruitment and docking. (B) Cas8/CasA subunit interacts directly with the PAM, allowing Cascade binding (i). Primary base pairing between crRNA and protospacer is followed by extended base pairing, displacing non-target strand and forming a full R-loop (ii). Conformational changes caused by DNA binding trigger Cas3 to dock (iii). In the absence of ATP, the nuclease domain cuts the displaced strand within the protospacer, leaving an ssDNA gap in the target. In the presence of ATP, Cas3 helicase activity unwinds the dsDNA, and complete degradation of the target DNA is mediated by Cascade-independent Cas3 nuclease activity. Illustration adapted from [15].

1.1.3 Applications of Cas3 in genome editing and biotechnology

CRISPR-Cas genome editing technology primarily utilises single-molecule effector proteins that cleave DNA in conjunction with a mature dual RNA molecule, namely Cas9 and Cas12a endonucleases. Cas9 in particular has been utilised extensively since its suitability for editing was discovered[4,5], with diverse applications in gene therapy, immunology, and most recently in gene drives for the mitigation of insect-borne diseases[21].The revolutionary design of the chimeric sgRNA molecule simplified and expanded the system for widespread usage, particularly in mammalian cells[5]. Cas9-sgRNA complexes have been refined as precision tools arising from their ability to induce blunt double-strand breaks at specific sites, typically 3bp upstream of the PAM determined by the guide RNA[8]. Cas9-mediated cleavage is characterised as a “snipping” mechanism, enabling the facilitation of precise gene knock-ins, which is also useful for studying loss-of-function (LOF) mutations in genes of interest[8,14]. Whilst the tightly controlled excision resulting from a clean DSB caused by Cas9 enables a low risk of polar excision outside of the gene of interest, sequence misreads from the short 20nt binding range of the sgRNA can result in off-target cleavage [12].

As has been discussed, Cas3 DNA cleavage is starkly characterised by a progressive unidirectional “shearing” of the dsDNA strand, in contrast to the singular DSB generated by

Cas9. The ATP-driven “shearing” mechanism of Cas3 enables the degradation of regions up to 200-600 bp in length rapidly, with an upper limit of 100kb[10,11]. This process creates a higher volume of DNA spacer fragments for chromosomal integration in comparison with Cas9 and Cas12a nucleases, thus potentially equipping the host cell with a stronger CRISPR immune response[20]. Studies performed with *Tfu*Cas3 and *Eco*Cas3 demonstrated that large DNA deletions upstream of the PAM occurred without prominent off-target activity; this distinctive feature of Cas3-mediated editing may broaden the application of genome editing by facilitating efficient gene knockouts and knock-ins[20]. Cas3-based editing has been demonstrated to be more efficient at gene knockout than Cas9, based on its affinity for large scale kilobase deletions and strand shearing [13,15,20,]. The ability of Cas3 to generate long-range sequence deletions is ideal for removing sequences of long ncRNA in dysfunctional genes, enabling efficient inactivation[12]. Recent *in vivo* trials using human stem cell models have demonstrated the efficacy of Cas3 in rapid strand degradation, inducing exon skipping in the *DMD* gene, thus presenting a potential novel tool in the treatment of Duchenne’s muscular dystrophy [12,13].

The intrinsic control of Cas3 by the Cascade complex mostly prevents non-specific nucleotide cleavage in non-target DNA sequences, resulting in a lower chance of generating off-target cleavage and polar excision than Cas9-based editing [10]. This enables Cas3 gene editing to be potentially safer than Cas9-based editing, due to the moderately high frequency of off-target mutagenesis generated by the latter. The higher accuracy of PAM complementary base-paring by the longer motif in Cascade also enables a more accurate target recognition rate in comparison to Cas9 [8,10].

Currently, challenges remain for the optimisation of Cas3 as a robust gene-editing tool, notably in the delivery of a large multi-protein complex to mammalian cells and potential non-specific gene silencing arising from extensive DNA shearing[20]. Restricting off-target DNA degradation and genotoxicity arising from Cas3 cleavage could be addressed by using anti-CRISPR proteins such as AcrE1 to exert nuclease inhibition, however this method requires thorough optimisation[12,13]. The application of Cas3 as a gene-editing tool would be better harnessed via a deeper understanding of the pathways by which Cas endonuclease function is regulated, both by CRISPR-associated proteins and other regulatory elements in the genome.

1.1.4 Regulation of the CRISPR-Cas system and Cas protein interactions with RNA

The functioning of *cas* genes and subsequently Cas endonuclease activity is largely regulated by promoters situated upstream of *cas* genes and at the head of the CRISPR 1 and 2 arrays[40]. Additional regulatory pathways of the CRISPR-Cas system by host cell proteins have been previously described, one well documented agent of *cas* regulation is the DNA-binding protein H-NS[40]. It has been shown to silence *cas* gene expression via the inhibition of promoters, including the *casA-E* genes encoding Cascade transcription, resulting in the integration of MGE DNA into the host genome without initiating a CRISPR immune response. The H-NS based system of *cas* gene silencing can be described as an inherent innate immune system, in conjunction with the CRISPR-based adaptive immune system acquired through horizontal gene transfer[40].

Cas3 functional stability and enzymatic activity during CRISPR interference also requires the co-expression of a chaperone protein termed HtpG (high temperature protein G) that is widespread in *E. coli* [64,68]. HtpG activity is upregulated in a temperature-dependent manner,

providing phage immunity in conjunction with Cas3 to the host cell at temperatures of 30°C, but displaying phage sensitivity at 37°C [72]. Both HtpG and H-NS function in tandem to regulate Cas3 expression levels and functioning in the host cell, enabling phage immunity and a robust CRISPR immune response.

A wider role of Cas3 in cellular regulation outside of the CRISPR system may suggest potential close interactions with non-CRISPR proteins and genomic elements. In a study conducted on the ectopic overexpression of *Eco*Cas3, it was found to stimulate the uncontrolled replication of ColE1 *ori* plasmids [25].

Plasmids use R-loops at their origin of replication (*ori*) to initiate DNA replication. This is performed by plasmid-encoded RNA molecules, one that primes replication (RNAII) and the other that prevents priming (RNAI)[25]. Cas3 stimulates the priming of replication, which is dependent on its ATPase and translocase activity independent of Cascade. It was theorised that Cas3 can specifically target RNAI-RNAII molecules for dissociation, liberating RNAII to pair with DNA at *ori* and enabling replication via R-loop stabilisation(*Figure 1.5* A).

As discussed previously, Cas3 utilises RNA as a priming basis for dsDNA cleavage, in the form of R-loops, demonstrating its capability to interact with a range of RNA molecules. A putative mechanism of RNA regulation via molecule pairing as given in[19,25] is a tantalising model which leads to more questions about the capability of Cas3 to process and interact with nonCRISPR RNA, potentially via allosteric pathways.

1.1.5 Allosteric protein regulation

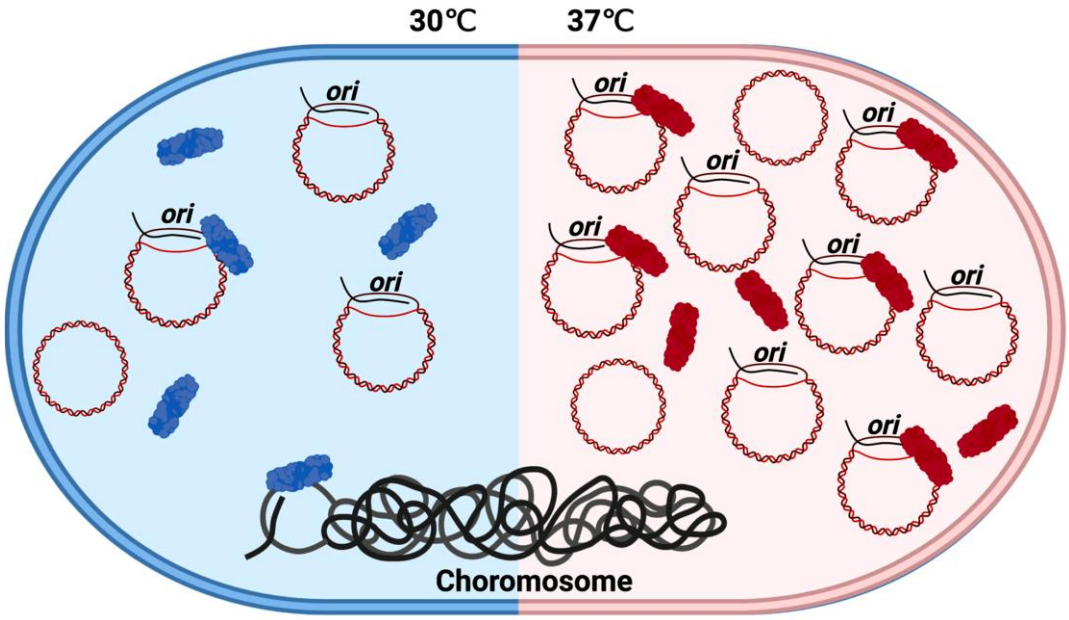
Allostery is defined as where a binding reaction at a specific protein domain induces a conformational change at a distal site in the same protein[84]. Allosteric activity is encoded in

single molecules such as nucleic acids in addition to proteins, and enables key biological processes such as cell signalling and post-translational modifications[84]. Many proteins are allosterically regulated by ligands that bind at a location distinct from the active site of an enzyme, which provides an efficient mechanism for regulating ion transport and catalysis[33]. Similarly, RNA molecules with regulatory functions can fold to form active sites that promote chemical reactions or selective binding sites for ligands and cellular metabolites.

There is limited research into the allosteric control of Cas protein function by regulatory RNA molecules, however there is an increasing interest in this field[40]. The allosteric affinity of the *S.pyogenes cas9* was trialled in a recent study using synthetic aptamer and ribozyme motif sequences, to determine resulting Cas9 expression levels and dsDNA cleavage[90].It was reported that ligand-activated ribozymes were able to trigger the Cas9 nuclease function on target DNA and release the sgRNA from a blocking sequence via self-cleavage. In addition, inactivated ribozymes rendered Cas9 non-functional[28,90] (*Figure 1.5*).

The current understanding of how proteins undergo allosteric regulation by regulatory RNA molecules is limited to studies using synthetic constructs, the activity of naturally occurring regulatory RNAs on local proteins therefore presents a novel research gap.

A



B

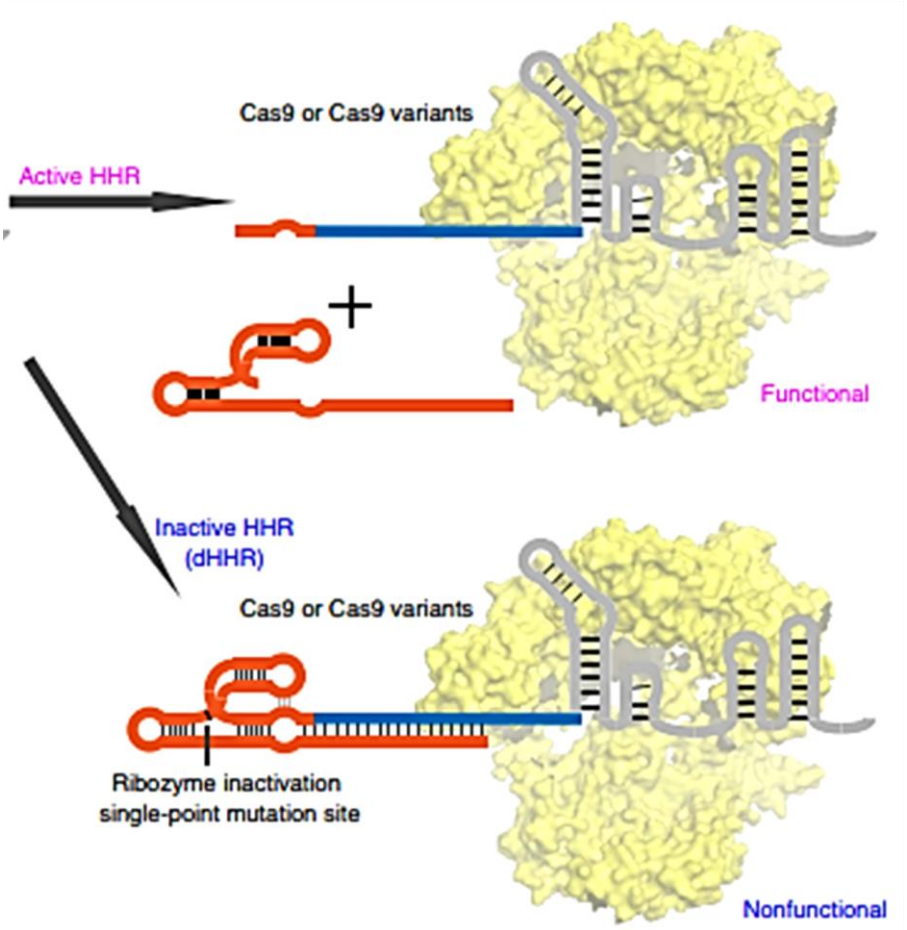


Fig 1.5 *cont.*

***Figure 1.5* Cas endonuclease interaction with nonCRISPR RNA molecules**

(A) *Eco*Cas3 interacts with RNA polymerase II via R-loop formation to override control over the *ori* of ColE1 plasmids, dissociating the polymerase from the DNA strand in a temperature-dependent phenomenon (B) Engineered allosteric control of the *S.pyogenes* Cas9 is imparted by synthetic hammerhead ribozymes (HHR), to control Cas9 expression and sgRNA activity. Diagram adapted from [23] and [90].

1.2 Small regulatory RNA

The RNA world theory proposes that RNA was the first nucleotide to form on earth and was the initial carrier of genetic information, which over time diversified into catalytic molecules for controlling gene expression and transcription[30]. This encompasses the hypothesis that complex RNA molecules such as ribozymes and riboswitches were the primary drivers of metabolic processes in ancient organisms[32]. Several types of catalytic RNAs and many classes of ligand-sensing RNA switches still exist in modern cells. RNA molecules can be broadly grouped into two categories, non-regulatory and protein coding, and regulatory and catalytically active[33]. Regulatory RNA molecules are typically formed from short sequences of non-coding RNA, and are highly prevalent in bacterial genomes.

1.2.1 Riboswitches

Riboswitches are RNA-based regulatory switches, also termed RNA sensors, which function as powerful ligand-responsive mechanisms to control gene expression[26,27]. They have an evolutionary basis as an ancient mechanism of cellular and protein modification, and are defined as cis-acting genetic elements[26,27].They occur within the 5`UTR region of mRNA sequences and possess a high degree of conservation, as detected through genomic multiple sequence alignment analyses[28]. The most common class of riboswitch in *E. coli* are TPP (thiamine pyrophosphate) sensors that can regulate the folding of the aptamer domain to initiate different cellular processes, enabling them to be one of the most versatile of gene regulators [28,29].

Riboswitches possess complex secondary structures that enable them to exert control over and regulate the initial transcription and downstream expression of juxtapose genes. They undergo a conformational change in the presence of cellular metabolites and ions, notably Mg^{+}, which exposes an expression platform, or termination loop, that either enables or inhibits gene transcription (*Figure* 1.6). The expression platform acts as a switching sequence and possesses a high degree of structural plasticity, thus it is variable across multiple species. In contrast, the aptamer domain is integral for a functioning riboswitch and is thus highly conserved across multiple species, allowing for the identification of novel riboswitches through sequence detection software [28].

The structural plasticity and responsive on/off states induced by ligands enables the development of synthetic riboswitches for more versatile uses. Recent applications of regulatory RNA include priming mechanisms for new gene-editing tools in bacteria. A notable example is RiboCas, which functions to tightly control *Spy*Cas9 expression and reduce off-target toxicity, thereby paving the way for a higher editing efficiency[39].

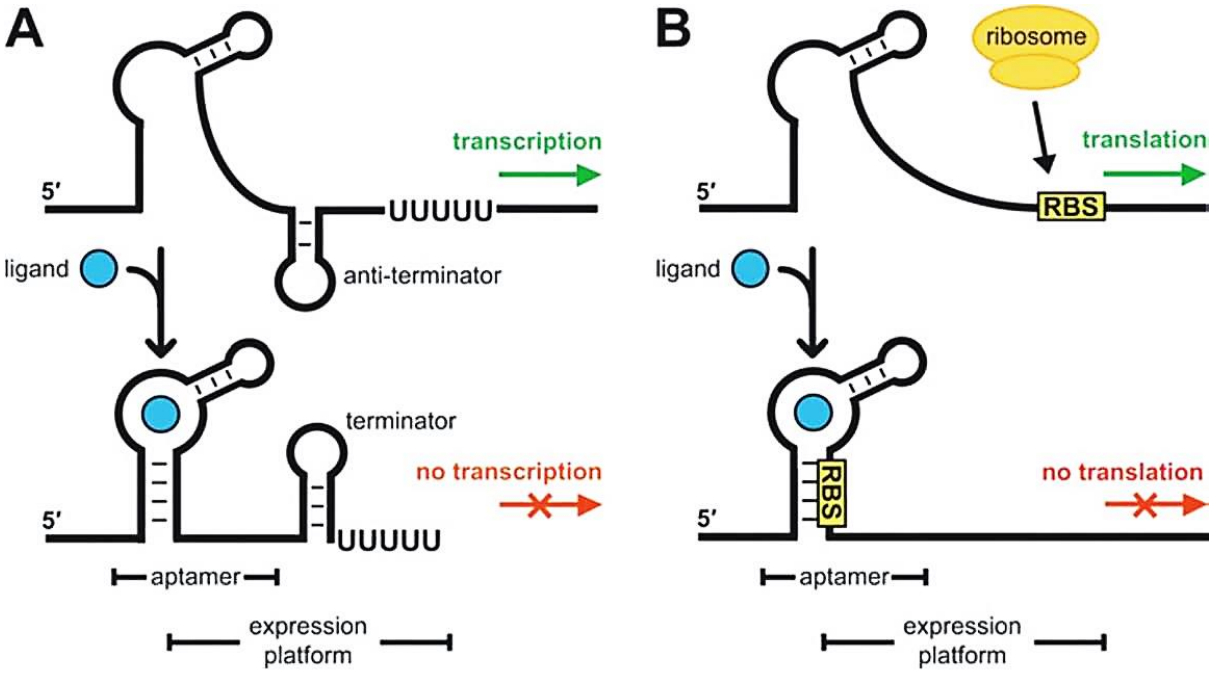


Figure 1.6 **Riboswitches as gene regulators**

Riboswitches can exert control of gene expression at the point of transcription, in response to ligands and cellular metabolites (A) In the absence of cellular ligands, gene transcription is uninhibited by conformational changes in the upstream riboswitch sequence. Upon ligand binding, the aptamer domain undergoes a conformational change into a branched hairpin loop structure, and a stalling terminator loop forms from the expression platform, inhibiting gene transcription (B) A similar process occurs for gene translation, with unimpeded mRNA translation in the absence of ligand binding to the aptamer domain. Diagram adapted from [30].

1.2.2 Small self-cleaving ribozymes

Ribozymes are catalytically active molecules that arise from a ncRNA sequence, and are found ubiquitously in both prokaryotes and eukaryotes[29]. Their key functions involve regulating gene expression via mRNA splicing[34,35].Ribozymes can be grouped into splicing ribozymes and cleaving ribozymes, with the latter divided further into trans-cleaving ribonucleases, and small self-cleaving or 'nucleolytic' ribozymes[29,30]. As nucleolytic ribozymes are more commonly found in *Escherichia* species, I will be focusing solely on this group. To date, nine distinct self-cleaving ribozyme classes have been described, including hairpin, hammerhead, twister, and hatchet [29,31]. In bacteria, hammerhead and twister ribozymes have been found to have roles in mRNA processing, and have the potential to influence gene expression in the 3' UTR[29].

One aspect that has not been a key focus of ribozyme research is the ubiquitous formation of RNA-protein complexes (RNPs) in cellular systems[31]. Their nature and characterization *in vivo* or *in vitro* could shed light on the long-standing question of the biological function of innate encoded ribozymes[29]. The natural formation of RNA:Protein complexes are key for understanding the interaction of a functionally active RNA molecule with a protein encoded from a local gene. The binding sites of the RNA with the protein can also indicate any local activity with the protein domain, and wider role in protein functioning.

Apart from catalysing RNA splicing and indirectly regulating genes, there is limited knowledge about the wider role of ribozymes in cellular functioning or any putative capability in allostery of local proteins[30,33].

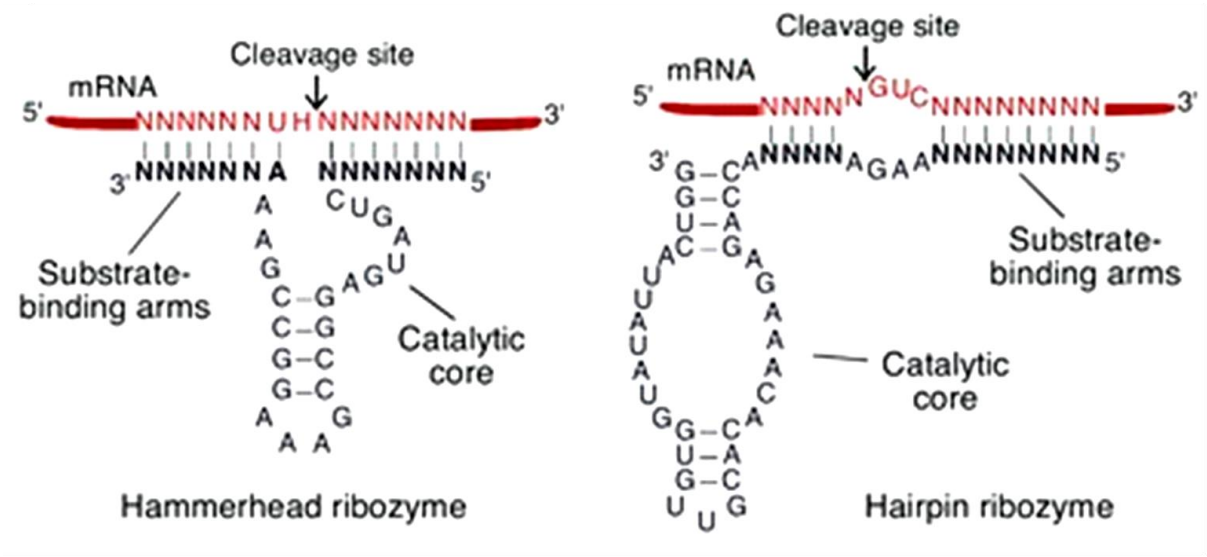


Figure 1.7 **Ribozyme secondary structures exhibit key functional domains**

Secondary structures of Hammerhead (L) and Hairpin (R) ribozymes exhibit common features among sub-types, notably conserved hairpin loops, and have been identified in *Escherichia* species. Diagram taken and adapted from [41]

1.3 Short ncRNA in bacteria

The significance of ncRNA as candidates for gene expression regulators and other key roles in cellular functioning cannot be understated, as what was erroneously termed "junk" DNA has yielded significant insights into hidden aspects of the genome[41]. As discussed previously, ncRNA can form the structural basis for riboswitches and ribozymes, of which a number of classes are prevalent in bacterial genera including *Escherichia*.

Previous research has shown that short non-coding RNA sequences may have a potential role in the regulation of *cas* gene expression by indirect allosteric control [29,31,33]. Largely uncharacterised ncRNA molecules in intergenic regions, for example between *cas-A* and *cas3*, may also have integral roles in gene expression and Cas protein function.

1.3.2 P*cas* and AntiP*cas*

P*cas* and antiP*cas* are promoters situated in the IGLB (Intergenic region between *ygcL* and *ygcB*) upstream of the CRISPR-1 locus, and were initially discovered in the *E. coli* wild-type strain MG1655[40]. The P*cas* sequence directs transcription of *casA* and the CRISPR array in a "sense" 5` to 3` prime direction, whereas antiP*cas* is transcribed in an anti-sense orientation and overlaps with the 3` terminal end of the *cas3* protein-coding strand. P*cas* and antiP*cas* expression levels are regulated by the DNA-binding protein H-NS, their functionality *in vivo* has been demonstrated via binding to spacer DNA in the CRISPR locus [33,40].

The antiP*cas* promoter directs a transcript of short ncRNA 150-200 nt in length, which forms a complex folded secondary structure that exhibits similarity to those characteristic of functionally active RNA molecules, such as riboswitches and ribozymes. The greater significance of a ncRNA molecule with complex secondary structure is unknown, as this may indicate the presence of an active regulatory RNA molecule in the CRISPR locus. An implication of this may be the exertion of control over *cas3* expression and subsequently Cas3 functioning, thus presenting a novel regulatory pathway.

Any modulation of *cas3* post-transcriptional function by the anti-*cas* RNA may manifest in the activity of the Cas3 nuclease-helicase, for example in impaired or enhanced nuclease activity during CRISPR immunity, however this remains to be elucidated.

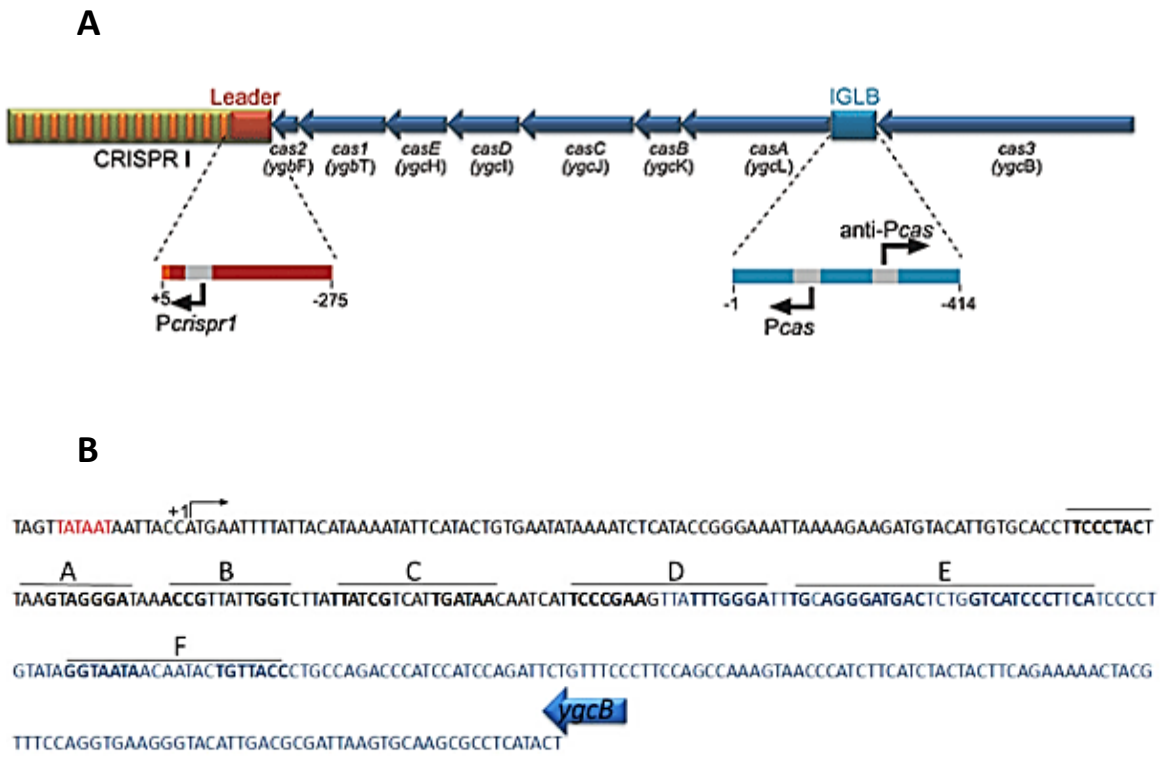


Figure 1.8 **P*cas* and AntiP*cas* promoters and transcriptional orientation**

(A)The P*cas* and AntiP*cas* promoters are situated intergenically within the CRISPR 1 locus, in a sense orientation towards the CRISPR array and anti-sense towards *cas3* respectively. P*cas* controls expression of the casA-cas2 operon, antiP*cas* controls anti-*cas* RNA expression(B) Full-length sequence of the antiP*cas* gene. Promoter region and TSS are denoted by an arrow, the sequence overlap with the 3`protein-coding region of *cas3* is highlighted in blue. Diagram taken and adapted from [40].

1.3.3 Implications

Active Cas3 expression is understood to be predominately controlled and mediated by CRISPR associated proteins such as H-NS and the chaperone HtpG, as has been previously discussed. The presence of a regulatory RNA molecule in the CRISPR locus may have implications for CRISPR immunity, as *cas3* gene expression and subsequently Cas3 may be under transcriptional termination control by the stable anti-*cas* RNA transcript. By uncovering the functional activity of the ncRNA transcript and RNA-encoding gene in relation to *cas3*, this may present a new pathway for *cas* gene regulation by non-coding regulatory RNA.

The efficacy of the Cas3-Cascade effector complex in gene-editing applications has been demonstrated in numerous studies; the technology, though still in its infancy, may become as widely used as Cas9 and Cas12a endonucleases. A greater understanding of the regulatory mechanisms underpinning CRISPR immunity and Cas3 mediated interference in Type 1 Class 1 systems can enable further research into how to optimise Cas3-Cascade mediated DNA cleavage, for future use in therapeutic applications and biotechnology.

1.4 Research Aims and Objectives

Cas3 protein functionality is regulated at the transcriptional and post-translational level, however the possibility of Cas3 to interact with non-coding RNA molecules may present an intriguing alternative pathway. A key hypothesis was formulated based on the functional response of Cas3 to temperature, potentially by allosteric regulation by unknown genomic elements. Previous research on the *in vivo* activity of the antiP*cas* promoter and the ncRNA anti-*cas* molecule indicated a regulatory role based on the complex secondary structure. We propose that the Cas3 protein may undergo allosteric regulation by the short non-coding RNA named anti-*cas*. This hypothesis was tested by performing biochemical and genetic analyses to uncover the activity of the antiP*cas* gene and RNA molecule through *in vitro* interactions with the Cas3 nuclease-helicase, and *in vivo* modulation within the host cell with both overexpressed and absent antiP*cas*.

Key objectives were to determine if the anti-*cas* RNA could form a bound complex with the Cas3 endonuclease, whether it could impact Cas3 nuclease activity on target DNA, and if the antiP*cas* gene modulated CRISPR immunity. This project was developed further by the use of additional cell strains from a collaboration with prominent researchers from the University of Zagreb to study primed adaptation.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Reagents

All commercial and industrial reagents used in this study were supplied by Sigma Aldritch (Merck KGaA) and ThermoFisher Scientific unless stated otherwise.

2.1.2 Commercial enzymes

All enzymes used in this study were supplied by New England Biolabs (NEB) unless stated otherwise.

2.1.3 Antibiotics

Table 1. Antibiotic reagents used for *in vitro* analyses of anti-*cas* RNA and in general microbiology and cell culture preparation

Antibiotic	**Industrial supplier**	**Stock concentration**	**Solvent**	**Working concentration**
Ampicillin (Amp)	Sigma Aldritch	100 mg/mL	SDW	100 µg/mL
Kanamycin (Km)	ThermoFisher	50mg/mL	SDW	50µg/mL
Chloramphenicol (Chlm)	Sigma Aldritch	35mg/mL	Ethanol	35µg/mL

2.1.4 Bacterial strains

Table 2. Description of cell strains used as referenced in methodology subsection and results chapter

Escherichia coli	**Industrial supplier or source**	**Genotype**
MG1655 *st*. K12	Invitrogen/Blattner lab	F- lambda- ilvG- rfb-50 rph-1
EB408/ IIB969	BW40114 x BW39651 strains, Ivančić Baće lab	+λT3 *lac*UV5-*cas3 cat*::*araB*p8-*casA*
DH5α	Invitrogen	F- ϕ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hdR17 (rk-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1
EB409/ NK001, MG1655ΔantiP*cas*	Bolt lab, generated in this work	F- lambda- ilvG- rfb-50 rph-1 ΔantiP*cas*::kan

2.1.5 Cas3 proteins

Table 3. Description of His_{x6} -MBP-tagged WT and mutant Cas3 proteins used in this study

***Escherichia coli* Cas3**	**Molecular weight and molar concentration**	**Source or literature reference**	**AA sequence base change**
*Eco*Cas3 WT	143.5 kDa 1µM monomer, 4.5µM polymer	Bolt lab, Howard *et al.,* 2011 [15]	N/A
*Eco*Cas3 ^{W406A}	143.5 kDa 1.67µM	He *et al*., 2021 [37]	WT Trp (W) to Ala (A)
*Eco*Cas3 ^{D75G}	143.5 kDa 5.1µM	He *et al.,* 2021 [37]	WT Trp (W) to Gly (G)

2.1.6 Plasmids

Tables 4-5. Description and source of plasmids used for biochemical analyses of antiP*cas* activity including plasmid construct used for *in vivo* spacer detection analyses

Plasmid	**Description**	**Plasmid size and *ori* type**	**Literature reference or source**
PUC19/18	Empty backbone vector containing *lacZ* gene and the *lac* promoter, MCS situated within *lacZ* gene. PUC18 is identical but with MCS in antisense orientation. Amp^R	2686 bp ColE1 *ori*	Messing lab, Norrander, Kemp and Messing, 1983
pDM13	pBadHisA-cloned plasmid containing antiP*cas* DNA sequence and *cas3* protein coding 3` terminal end.	4531 bp ColE1 *ori*	Ivančić-Bače lab, University of Zagreb
pT7-7	Backbone vector conferring resistance to ampicillin and containing the T7 promoter. High copy number plasmid.	2470 bp ColE1 *ori*	Tabor and Richardson, 1985
pKD4	Plasmid containing a gene encoding aminoglycoside phosphotransferase, originally from Tn5 transposase. Gene insertion confers resistance to	3267 bp R6K *ori*	Datsenko and Wanner, 2000

	kanamycin, neomycin and geneticin.		
pSim6	pBadHisA-cloned plasmid containing a temperature sensitive variant of the phage *λ* repressor gene. Red recombinase expression controlled by native *λ* phage promoter, induced by heat-shock using the temperature-sensitive *λcI857* repressor Also contains Shine-Dalgarno sequence. Has a low copy number *ori*	6716 bp PSC101 *ori*	Sharan *et al.,* 2009,
pTK43	pBadHisA-cloned plasmid containing lambda phage R gene, used as donor plasmid for subcloning and generating plasmid construct	4471 bp ColE1 *ori*	Bolt lab, University of Nottingham, for collaboration with Ivančić-Bače lab, University of Zagreb
pTK133/ pRSF-1b	Plasmid containing aminoglycoside phosphotransferase gene cassette, used as a control during RED recombineering procedure.	3669 bp RSF *ori*	Novagen
pBadHisA	Backbone vector for pDM13. Contains amp^R antibiotic cassette sequence	4102 bp ColE1 *ori*	Invitrogen
pACYCDuet 1	Backbone vector with two MCS and lacI repressor gene. Confers resistance to Chloramphenicol in donor cell. Medium copy number plasmid	4008 bp P15A *ori*	Novagen
pJWR2	pBluescript-cloned plasmid containing the *E. coli* CRISPR-1 locus, amp^R	1000 bp ColE1 *ori*	Howard *et al.,* 2011

pJWR5	*E. coli* Cascade genes (*cas1-6*) inserted into pBadHisA Kan^R	9421 bp CoLE1 *ori*	Ivančić-Baće *et al.,* 2013
pAH4	*E. coli cas3* ygcB gene inserted into pBadHisA ColE1 backbone vector, amp^R	6769 bp ColE1 *ori*	Ivančić-Baće *et al.,* 2013 Construct from 2007 in Bolt lab

Construct name	Gene or insert	Vector	Donor
pNK01	R, target of *λ*T3 spacer and antiP*cas*	pACYDuet1	pTK43

2.1.7 Media and Solutions

Table 6. Description of nutrient medias used for bacterial culture

Media	Composition
LB/Mu Broth	10g/L Tryptone (BD) 10g/L NaCl 5g/L Yeast Extract (BD) Adjusted to pH 7 using NaOH
M9 minimal media (1X)	5X Minimal salts { 15 g/L KH_2PO_4 64 g/L $NA_2PO_4 \cdot 7\ H_20$ 2.5 g/L NaCl 5.0 g/L NH_4Cl } LB/Mu Broth
SOB/SOC	2% (w/v) Tryptone (BD) 10mM NaCl 0.5% (w/v) Yeast extract (BD) 2.5mM KCl 20mM Glucose

LB agar (1.5%)	3g per 200mL LB broth
Soft agar (0.4 , 0.7%)	0.4-0.7g per 100mL LB broth
TBM	10g/L Tryptone 4g/L NaCl 0.02% (v/v) Maltose

2.1.7.1 Gel compositions

Table 7. 8% SDS-PAGE gel composition and 3mL stacking gel used for analysis of protein-RNA binding

Resolving gel (10mL) **Solution name**	**Volume**
SDW	5.8 mL
30% Acrylamide	2.7 mL
3M Tris pH8.8	1.3 mL
10% SDS	100 µL
10% APS	100 µL
TEMED	10 µL

Stacking gel (3mL) **Solution name**	**Volume**
SDW	1.75 mL
30% Acrylamide	500 µL
3M Tris pH8.8	750 µL
10% SDS	30 µL
10% APS	30 µL
TEMED	3 µL

Table 8. SDS-PAGE reagents

Solution	**Composition**
4x Sodium dodecyl sulphate (SDS) Polyacrylamide gel electrophoresis (PAGE) loading buffer	0.5mL 1M Tris-HCl pH6.8 2mL 10% SDS 10mg Bromophenol blue 1.25mL 80% (w/v) glycerol SDW
Coomassie Brilliant blue stain	40% methanol 10% Glacial acetic acid 0.05% Coomassie Brilliant Blue R-250
Destain	20% methanol 10% Acetic Acid made up to 1L with distilled water

Table 9. 10% Native-PAGE gel and corresponding dye used for nuclease assays

Solution	**Volume**
30% Acrylamide (37:5:1)	13.3mL
10X TBE	4mL
SDW	22.7mL
10% APS	200µL
TEMED	50µL

Solution	**Composition**
Orange G loading dye 1	80% Glycerol (*v/v*) SDW Orange G dye

Table 10. 5% Denaturing PAGE gel and corresponding dye used for RNA substrate analysis and *in vitro* analyses

Solution	Volume
40% Acrylamide (19:1)	5mL
10X TBE	4mL
7M Urea	16.8g
Formamide	2mL
SDW	7mL
10% APS	200µL
TEMED	50µL

Solution name	Composition
Orange G loading dye 2	79% formamide (*v/v*) 20% glycerol (*v/v*) SDW Orange G dye

Tables 11 – 18. Descriptions of buffer compositions and oligonucleotide sequences used in *in vitro* and *in vivo* analyses

2.1.7.2 Buffers

Running Buffer Stock solution	Composition
10X Tris-borate-EDTA (TBE) buffer	1M Tris 1M Boric acid 20mM EDTA
10X SDS PAGE running buffer	250mM Tris 1.92M Glycine 1% (v/v) SDS
10X Tris-acetate-EDTA (TAE) buffer	1M Tris 17.4 M Glacial acetic acid

	20mM EDTA
10X Tris-borate (TB) buffer	1M Tris 1M Boric acid

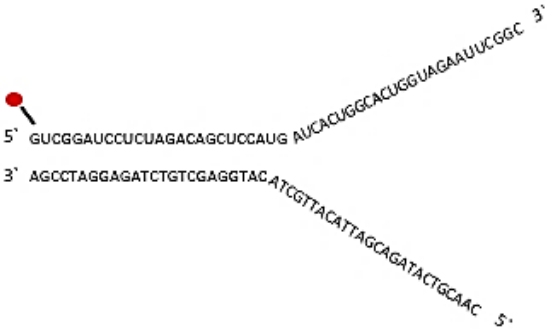
Assay Buffer **Buffer name**	**Composition**
10X Annealing buffer	100 mM Tris pH 7.5 500 mM NaCl 10 mM EDTA
Replication Fork DNA elution buffer	4 mM Tris pH8.0 10 mM NaCl
5X Buffer O	50mM Tris-HCl pH7.5 10mM $MgCl_2$ 100mM NaCl 1mg/mL BSA 20mM DTT 0.2% Tween20
Buffer E	50mM Tris-HCl pH7.5 100mM NaCl 1mg/mL BSA 20mM DTT 0.2% Tween20
5X Stop buffer	50mM Tris pH8.0 100mM EDTA 5mg/mL proteinase K (Invitrogen) 1% SDS

5X HB buffer	100mM Tris-HCl pH7.5 25% Glycerol 500µg/mL BSA
5X HB Stop buffer	2mg/ml proteinase K (Invitrogen) 2.5% SDS 200mM EDTA
2X TEA buffer	40mM TEA-HCl pH8.5 2mM EDTA 300mM NaCl 1mM DTT

2.1.8 Oligonucleotides used for substrate synthesis

Protein-Nucleic acid assays **Substrate name**	**Oligonucleotide**	**Oligonucleotide sequence**	**Extinction coefficient L/ (mole.cm)**
Fork 2B DNA	MW12 5' Cy5	Cy5-GTCGGATCCTCTAGACAGCTC CATGATCACTGGCACTGGTAGAAT TCGGC	471,700
	MW14	CAACGTCATAGACGATTACATTGCTAC ATGGAGCTGTCTAGAGGATCCGA	495,600
Fork 2B dsRNA	MW12 5' Cy5	Cy5-GUCGGAUCCUCUAGACAGCU CCAUGAUCACUGGCACUGGUAGAAU UCGGC	471,700
	MW14	CAACGTCATAGACGATTACATTGCTAC ATGGAGCTGTCTAGAGGATCCGA	495,600

Substrates generated by *in vitro* transcription **Substrate name**	**Substrate type, molecular weight and master stock molar concentration**	**Sequence 5` to 3`**	**Sequence template**
Anti-*cas*	ssRNA 51.138 kDa 97.75 µM	(Cy5)-UAAUACGACUCACUAUAGGUC CCUACUUAAGUAGGGAUAAACCGU UAUUGGUCUUAUUAUCGUCAUUG AUAACAAUCAUUCCCGAAGUUAU UUGGGAUUUGCAGGGAUGACUCU GGUCAUCCCUUCAUCCCCUGUAU AGGUAAUAACAAUACUGUUACC -	Purified pDM13 plasmid, transcribed from top strand of RNA-encoding gene in sense orientation 159 nt in length
Cy5 Anti-*cas*	ssRNA 51.138 kDa 21.31 µM		

Replication Fork structures	
	

2.1.9 Primers for *in vitro* analyses

Anti-*cas* RNA **Primer names**	**Oligonucleotide sequence 5` to 3`**	**Tm**	**Length**
anti-cas A_gg	TAATACGACTCACTATAGGTCCCTACTTAAGTAGGGATAAACCG	56°C	44
anti-cas B	GGTAACAGTATTGTTATTACC		21
Anti-*cas* forward primer 1	TAATACGACTCACTATAGGGGGTACCAGTATGAGGCGCTTG CACTTAAT	52°C	49
Anti-*cas* reverse primer 2	GCGGAGCTCTAGTTATAATAATTACCATG		29
CRISPR-1F	GAGATGCAGGCCATCGGA	69°C	18
CRISPR-1R	GCGACCGCTCAGAAATTCCAGACCCGATCCAAA		33

2.1.10 Primers for Lambda RED recombineering

Primer names	**Oligonucleotide sequence 5` to 3`**	**Tm**	**Length**
1st set RED forward	ACCGGGAAATTAAAAGAAGATGTACATTGTGCACCTTCCCTACTT AAGTAATTGTGTAGGCTGGAGCTGCTTC	55°C	73
1st set RED reverse	TACAGGGGATGAAGGGATGACCAGAGTCATCCCTGCAAATCCCA AATAACATGAATATCCTCCTTAGTTC		70
2nd set RED forward	GACAATAATGCATGCATTTCAAATCTGCAA	67°C	30
2nd set RED reverse	TGGGTCTGGCAGGGTAACAG		20

2.2 Methods

2.2.1 General Microbiology

2.2.1.1 Chemically competent cell preparation

Escherichia coli IIB994, IIB969, DH5α and NK001 cells were streaked on selective antibiotic LB plates and incubated overnight at 37°C. Colonies were selected and inoculated into fresh LB selective media for o/n incubation prior to transferral at a 1:100 dilution into 50ml of non-selective LB media in a 250ml baffled flask. Cells were grown at 37°C in a shaking water bath until OD_{600} = 0.4. Cells were then pelleted at 4000 RPM at 4°C using an AG Eppendorf Centrifuge 5430 R, and resuspended in filtered 0.1M $CaCl_2$ that was pre-chilled. Cells were incubated on ice for 90 minutes and subsequently pelleted at 4000 RPM at 4°C before resuspension in filtered and chilled 0.1M $CaCl_2$ with 30% glycerol. 400µL aliquots were prepared in 1.5ml sterile Eppendorf tubes and flash frozen on dry ice for storage at -80°C.

NK001 and IIB969 cells were prepared in an identical manner. Cells were pelleted for final resuspension in chilled 0.1M $CaCl_2$ with 10% glycerol and flash frozen on dry ice for storage at -80°C.

2.2.1.2 Bacterial transformation

Competent cells were thawed on ice. 1-2µL of vector plasmid was introduced per 100µl of competent cells and mixed by gentle pipetting. Cultures were incubated on ice for 30 minutes before being heat shocked at 42°C for 90 seconds in a SUB water bath (Grant) and placed back on ice. 900µL of LB broth was added under sterile conditions (under blue roaring Bunsen burner flame) and incubated in a shaking water bath for 1 hour at 37°C. Cells were pelleted at 13000 RPM at RT using a Heraeus PICO 17 centrifuge (Thermo Scientific) and concentrated by resuspension in $1/8^{th}$ volume LB broth. All cells were pipetted and evenly distributed onto

LB agar plates inoculated with commercial antibiotics using a spread plate technique. Plates were incubated overnight at 37°C in a Carbolite dry air incubator. Antibiotics used at recommended working concentrations, see table 1 for reference.

2.2.1.3 Cloning and subcloning

Cloning of the amplified *spf* SPOT 42 sequence was performed using purified PCR DNA and the high copy number backbone vectors PUC19 and PUC18. Restriction digestion, dephosphorylation, ligation, and transformation for downstream applications were performed using commercial NEB (New England Biolabs) reagents and adhered to recommended NEB protocol. Preparation of the DNA insert for ligation into PUC19/18 was performed using an identical procedure as described below;

Subcloning was performed using the ColE1 *ori* donor plasmids pTK43 and pDM13 and backbone vector pACYDuet1, which contains a P15A *ori* and has a medium copy number propagation. pTK43 and pDM13 were cut respectively in a double RE digest using Ncol and Xhol high fidelity enzymes (New England Biolabs) and single RE digest using high fidelity Hind-III (New England Biolabs). pACYDuet1 was digested using Ncol and Xhol in an identical manner. Digestion was confirmed on a 1% TBE agarose gel, with 100ng of cut and uncut plasmid loaded as a control in an adjacent lane. pACYDuet1 was dephosphorylated with Quick Cip (NEB) prior to genetic manipulation to allow cohesive end joining with the desired DNA inserts.

Ligation of the desired inserts into the pACYDuet1 vector was subsequently performed at RT using T4 DNA ligase (NEB) and incubated for a duration of 10 minutes. The genotype of the vector was verified on an 0.8% TBE agarose gel with empty pACYDuet1 as a control. 5µL of plasmid constructs were transformed into 50µL of competent DH5α cells for o/n incubation.

Viable plasmid DNA was extracted and purified for downstream applications using the Promega Wizard® Plus SV Minipreps DNA purification kit.

2.2.2 DNA substrate preparation

The DNA fork substrate used in protein-nucleic acid assays to evaluate anti-*cas* effect on Cas3 nuclease-helicase activity (shown in section 2.1.7) was constructed by annealing 5µM of the 5' labelled oligonucleotide MW12 and 6µM of the unlabelled oligonucleotide MW14 in a buffer composed of 100mM Tris pH7.5, 500mM NaCl, and 10mM EDTA. The DNA was incubated at 95°C for 10 minutes and left to cool to RT on bench. The annealed DNA fork was separated from free oligonucleotides by migrating the sample for 2 hours on a 10% acrylamide Native-PAGE 1xTBE gel. Gel band containing the substrate was excised, and the substrate was subsequently eluted in buffer (4 mM Tris pH8.0, 10 mM NaCl) and left to diffuse for 48 hours at 4°C.

The fork substrate concentration was quantified using a DeNovix NanoDrop 2000 spectrophotometer (ThermoFisher Scientific). The A_{260nm} absorption reading and extinction coefficient value within the Beer-Lambert law were used to calculate the substrate µM concentration. The Oligoanalyser 3.1 (IDT) tool was used to calculate the extinction coefficient of each oligonucleotide individually and a total value.

2.2.3 Agarose gel electrophoresis

Native TBE gel

Samples were mixed with 6x Purple Gel Loading dye (NEB B7025) and loaded onto 1xTBE agarose gels premade with 0.2µg/ml ethidium bromide (EthBr) (Sigma Aldrich). Gels were run for electrophoresis using a PowerPac Basic power supply (BioRad) in 1xTBE buffer. Gels were imaged using a U:Genius3 Bio-imaging system (Syngene) with UV exposure.

Denaturing TAE gel

Samples were mixed with 79% formamide loading dye (60% *v/v*) and heat denatured prior to loading onto 1xTAE agarose gels. Gels were not pre-stained with ethidium bromide, 1µL of EthBr (10mg/ml solution) was mixed with all samples prior to gel loading. Gels were run using a PowerPac Basic power supply (BioRad) in 1xTAE buffer. Gels were imaged using a U:Genius3 Bio-imaging system (Syngene) with UV exposure.

Rapid migration TB gel

Samples were mixed with 5µL Orange G dye 1 and loaded onto a 0.5 X TB agarose gel premade without ethidium bromide. Electrophoresis was performed in 0.5 X TB buffer using a PowerPac Basic power supply (BioRad) for a run duration of 20 minutes at 12.5 V/cm.

2.2.4 RNA synthesis and fluorescent labelling

A partial sequence from the antiP*cas* gene in the pDM13 plasmid was amplified using primers containing the T7 promoter sequence of 5` TAATACGACTCACTATAGGG 3` and termination codon of 3` TTA 5`. This was in preparation for subsequent *in vitro* synthesis of the RNA sequence. The PCR product was 159bp in length, and PCR cleanup was performed using Promega Wizard™ PCR and SV Clean-up Kit, and eluted with 40µL of nuclease-free water.

For RNA synthesis, the NEB HighScribe™ T7 High Yield RNA kit was used. The reaction was prepared in a total volume of 20µL and incubated overnight at 37°C in a Carbolite dry air incubator. The RNA was purified and concentrated using the NEB Monarch® RNA Cleanup Kit (50µg), and eluted in a final volume of 30µL. The ng/µL concentration of the RNA substrate was quantified using the DeNovix NanoDrop 2000 spectrophotometer (ThermoFisher

Scientific). The A_{260} absorption reading, Beer-Lambert Law and kDa of the RNA sequence were used to calculate the final µM concentration of the substrate.

In vitro fluorescent labelling of the anti-*cas* substrate was performed using the NEB HighScribe™ T7 High Yield RNA kit and a Cy5 conjugated oligonucleotide (Jena Biosciences). The reaction was prepared in a total volume of 20µL and incubated overnight at 37°C in a dry air incubator. Labelled Cy5 anti-*cas* RNA was purified and concentrated using the NEB Monarch® RNA Cleanup Kit (50µg), and eluted with nuclease free water to a final volume of 40µL.

2.2.5 Protein-Nucleic acid assays

2.2.5.1 EMSA

Cy5 end-labelled anti-*cas* RNA at 40nM and a concentration titration of monomeric Cas3 WT were prepared in nuclease free water to a final reaction volume of 20µL. A control reaction contained only RNA. Reactions were incubated at 30°C for 1 hour in Buffer E (50mM Tris-HCl pH 7.5, 100mM NaCl, 0.1mg/mL BSA, 20mM DTT) and then transferred to room temperature. Reactions were mixed with 25µL Orange G loading dye 1 (80% *v/v* glycerol and Orange G) for electrophoresis in a 1xTBE 5% acrylamide native PAGE gel and subsequently in a 2.5% 0.5 X TB agarose gel (Section 2.2.3). Results were imaged using the Amersham Typhoon™ 5 Biomolecular imager (GE healthcare) and a U:Genius Bio-imager with UV exposure (Syngene). Agarose gels were post-stained with 1:10000 dilution of SYBR Gold nucleic acid stain in 100ml 0.5X TB buffer, and destained with 100ml 0.5X TB buffer prior to UV imaging.

2.2.5.2 Nuclease assays

Preliminary assays were performed with Cas3 WT, Cas3^{W406A} and Cas3^{D75G} mutant proteins to measure DNA cleavage with Cy5 Fork2B DNA, Cy5 Fork2B dsRNA and Cy5 labelled anti-*cas*

RNA substrates to determine comparative nucleotide degradation and optimal assay conditions. Experiments were performed according to published protocol in [35].

*Eco*Cas3 WT was pre-incubated with 20nM of anti-*cas* RNA for 10 minutes at 30 or 37 °C. Reactions were initiated upon the addition of Buffer O (see Table 12.) and 20nM Fork DNA (see Table 11) and incubated for 4 hours at 30 or 37 °C. Reactions were stopped by the addition of Proteinase K (Invitrogen) followed by further incubation for 15 minutes. RNase T1 (ThermoFisher) was added to neutralise the anti-*cas* RNA. Reactions were mixed with Orange G loading dye 1 for electrophoresis in a 10% acrylamide 1x TBE Native-PAGE gel, run at 160 volts for 120 minutes. Results were visualised using an Amersham Typhoon™ 5 Biomolecular imager (GE Healthcare) and U:Genius Bio-imager (Syngene).

2.2.5.3 Helicase assay

Control reactions were prepared with *Eco*Cas3 WT protein in a molar gradient to measure dsDNA strand unwinding with 25nM Fork 2B DNA and 2.5µM MW12 cold trap DNA. Assays were performed by adapting experimental protocol from [15].

2.2.6 *In vitro* transcription assay

The assay was performed using the HiScribe T7 High yield RNA synthesis kit (NEB). Preliminary controls were performed with a PCR DNA template amplified from the pDM13 anti*cas* plasmid. The DNA template was 407bp in length and contained a truncated 204 bp segment of the *ygcB cas3* gene that contained both 5` and 3` terminal end regions. The template contained flanking regions in proximity to *cas3,* which acted as overhangs beyond the region that anti-*cas* RNA bound to. Prolonged incubation with T7 polymerase can result in RNA substrate displacement, therefore control reactions were prepared to determine an optimal

incubation period for yielding nascent RNA substrates. Control reactions were incubated at 37°C without shaking in a SUB water bath (Grant) for durations of 30 to 120 minutes.

Stable Cy5 labelled anti-*cas* RNA was premixed at RT with 25nM of DNA template in a total reaction volume of 20µL in nuclease-free water. Anti-*cas* RNA was titrated in at a range of 25 to 250nM, with the molar ratio to template increasing by a factor of 2 per reaction. Cy5 labelled cRNA2 was premixed with the DNA template at a final concentration of 250nM. Reactions were initiated upon incubation at 37°C and continued for a duration of 45 minutes. The incubation period was kept minimal to prevent displacement of nascent transcribed RNA by the polymerase molecule, as described in [25].

Reactions were terminated upon removal from the incubation temperature and placed immediately on ice. Prior to gel loading, reactions were mixed with 8µL of Orange G dye 2 and spun down in a tabletop vortexer (ThermoFisher). Reactions were loaded for electrophoresis through a 1xTBE 5% denaturing PAGE gel. Results were imaged using the Amersham Typhoon™ 5 Biomolecular imager (GE Healthcare). The gel was post-stained with 1:10000 dilution SYBR™ Gold nucleic acid stain (ThermoFisher) in 1XTBE prior to scanning with the U:Genius3 Bio-imaging system (Syngene).

2.2.8 Lambda RED recombineering

The knockout of the antiP*cas* gene was performed according to published protocol in [33, 34,35]. The methodology is described as follows. Oligonucleotide sequences were designed to be homologous to 5’ and 3’ orientated ends of a gene encoding aminoglycoside phosphotransferase, which confers resistance to neomycin and kanamycin. The antibiotic cassette gene was a 1577 bp sequence from the pKD4 donor plasmid. Primers were designed to incorporate 50 nt homologous “anchors” which anneal to flanking regions of a short

segment of the antiP*cas* gene in the MG1655 CRISPR loci. A second set of primers were designed to flank the region excised in the primary set, in order to confirm successful antiP*cas* gene deletion via PCR. Primers used are shown in table 11. PCR was performed with the first set of primers and linearised pKD4 DNA template, and verified on a 1.3% agarose gel. The sample was purified using the Promega Wizard® PCR Cleanup and Gel extraction kit, and eluted in a final volume of 40µL.

Competent MG1655 cells were transformed by pSim6 (amp^r, λ*ts* repressor) on ampicillin inoculated agar plates o/n at 30°C. Colonies from the successful RED expressing strain were then grown at 30°C o/n in 5ml of LB in shaking waterbath. An aliquot of 0.5ml of cells were then inoculated into a 250ml baffled flask containing 25ml of fresh LB media and grown in a shaking water bath at 30°C until OD_{600} = 0.5. Once optimal growth had been reached, cells were transferred to 42°C for 15 minutes to induce RED gene expression. Immediately after, the culture was chilled in ice slurry and swilled by hand. After heat induction, recombination activity is lost rapidly at temperatures above 37°C.

Cells were resuspended in 5ml of 10% glycerol solution in 50ml Falcon tubes and mixed by gentle pipetting. This was done in order to remove residual salts that may degrade PCR DNA in later stages. Cells were washed and pelleted 3 times at 4000rpm for 3 minutes at 0°C using an AG Eppendorf Centrifuge 5430 R. After the final wash, cells were resuspended in 1ml of 10% glycerol in a 1.5ml sterile Eppendorf on ice and kept on ice for 20 minutes.

Aliquots of cells and PCR DNA were prepared on ice in 1.5ml sterile Eppendorfs prior to transferral into electroporation cuvettes. 100µL of cells were mixed with 1.5µL of DNA before transferral into Perspex cuvettes. DNA from the pTK133 plasmid was used as a control. Cells were then electroporated at 1.8kV for approximately 5 msec (BioRad Coli Gene Pulser) to

enable effective DNA passage through the membrane. 100µL of SOC was immediately added to the cells to promote survival, samples were then outgrown at 30°C for 2 hours in 1.5 ml sterile Eppendorf tubes.

After incubation, viable cells were purified by streaking on LB agar demarcated plates inoculated with 25µg/ml of kanamycin stock stored at room temperature and incubated o/n at 37°C in a dry air incubator (Carbolite). Viable colonies were then re-steaked the following day onto LB agar plates inoculated with a standard kanamycin concentration of 50µg/ml and incubated o/n at 37°C. The genotype of the deletion strain was confirmed using denatured viable cells in a colony PCR with a second set of primers that annealed to a flanking region outside the excision site, and visualised using gel electrophoresis on a 1.2% agarose gel. pSim6 was purged from the cells via inoculation of a cell culture into 5ml of LB with 100µg/ml of Ampicillin and o/n incubation at 42°C.

Sanger sequencing was performed externally by GENEWIZ® (Azenta Life Sciences) on a prepared PCR sample with specific primers to confirm the sequence read match to the kanamycin gene cassette in pKD4. Parallel alignments of the NK001 sequence with the MG1655 WT locus were subsequently performed using the Clustal Omega software ([31]) .

2.2.10 Cell growth assays

The physiology of the anti-*cas* deletion strain (named NK001, see table 2, section 2.1.3) was first examined by comparing cell growth over 16 hours with MG1655 WT in a range of nutrient medias with different carbon sources.

Overnight cultures of NK001 and MG1655 were first aliquoted into 1.5mL Eppendorf tubes under a roaring blue Bunsen flame and placed on ice to prevent further cell growth at RT.

Fresh LB broth and 1XM9 liquid medias containing maltose, glucose, and glucose plus casamino acids were aliquoted in sterile conditions in 200µL volumes into a clear flat bottom Nuncler 96 well plate (ThermoFisher). Cells were diluted in a factor of 1:100 into nutrient medias under a blue Bunsen flame in duplicate lanes at RT before transferral to an incubation unit preheated to 37°C. Cells were incubated for 16 hours with orbital well shaking, with OD_{600} readings taken every 30 minutes between shaking cycles.

Assays were subsequently repeated in duplicate with NK001 cells transformed with pDM13 to restore antiP*cas* and the WT genotype, and NK001 transformed with empty pBadHisA as a control. Cells were grown o/n in medias containing 100µg/mL ampicillin and 0.02% L arabinose to induce gene expression and overcome bacteriolysis.

NK001 and MG1655 cell growth was visualised by plotting OD_{600} data points against time in minutes for each nutrient media, with error bars representing the means and SD applied.

2.2.11 Anti-*cas* overexpression and transformation

This procedure was performed as described in [37] and using general microbiology protocol. The methodology is summarised as follows: MG1655 IIB969 cells were streaked out onto LB agar plates without appropriate selective antibiotics, and incubated at 37°C o/n in a dry air incubator. Viable cells were then made chemically competent and stored at -80°C until transformation with the anti-*cas* plasmid pDM13 and the backbone vector pBadHisA as a control.

Transformation of competent cells and target plasmids was performed, and cells were spread onto selective LB agar containing 100µg/mL ampicillin prior to o/n incubation at 37°C. Viable colonies were then selected for further use in performing *in vivo* assays to simulate primed adaptation.

2.2.12 Phage lysate stock preparation and phage titre

Master stocks of λ*vir* C and phage P1 WT lysates were used for the preparation of lysate aliquots for use in host cell infectivity to determine phage titre and transduction with target cell strains for later application in performing primed adaptation.

A lysate stock of λ*vir* was prepared using protocol in [44] and is described as follows. DH5α cells were grown o/n in 5mL LB broth in preparation for infection with phage stock. 1M $CaCl_2$ and 1M $MgCl_2$ were sterilised by autoclave and added in 10µL volumes to 10mL of fresh LB broth in a Corex glass culture tube. 100µL of o/n bacterial culture was inoculated into the broth and incubated at 37°C with shaking for 60 minutes. The culture was transferred to RT and infected with 100µL of high titer λ*vir* stock lysate under sterile conditions before incubation at 37°C with shaking for approximately 5 hours, until lysate cleared.

Lysate was decanted into a new 15mL falcon tube under sterile conditions and centrifuged at 4000 rpm at RT for 25 minutes. Cleared phage supernatant was sterilised using a 0.22µM filter (Sartorius) to yield a bacteria-free pure lysate and stored at 4°C.

Phage titre of the lysate was quantified using a plaque assay on a phage susceptible bacterial strain. DH5α cells were streaked onto 25mL LB agar plates and incubated o/n at 37°C. Viable colonies were then inoculated into 10mL of fresh LB and incubated at 37° with shaking until the OD_{600} reached 2-3.

Simultaneously, λ*vir* lysate was serially diluted in 1xM9 media from a factor of 10^{-1} to 10^{-7} and kept at RT. When DH5α cells reached the required OD, 200µL aliquots were transferred into sterile 1.5mL Eppendorf tubes and subsequently mixed with phage dilutions in a final volume of 300µL. Mixed cultures were incubated at RT for 10 minutes prior to inoculation into 3mL of molten 0.7% overlay LB agar at 42°C.

Bacterial overlay was mixed by swirling by hand and poured onto 25mL LB plates preheated at 37°C. The cell-phage mixture was incubated at RT for 15 minutes before o/n incubation at 37°C. Following incubation, the number of plaques were quantified on each plate corresponding to the phage dilution factor, and the λ*vir* stock lysate concentration was determined using the following formula:

$$\text{PFU (Plaque Forming Units)/mL} = N \times 1/DF \times 1/V$$

Where *N* is the number of plaques of lysis counted on the plate (expressed as PFU); DF is the dilution factor and V is the volume of phage dilution poured on the plate.

A P1 phage lysate stock was prepared using established laboratory protocol, and is summarised as follows. An o/n culture of IIB994 cells were diluted 1:16 into 8mL of LB broth containing 100µL of 0.5M $CaCl_2$ and incubated at 37°C with shaking until OD_{650} reached 1.0.

Simultaneously, a wild-type stock of phage P1 grown on DH5α was diluted 10-fold in MC buffer to give a phage titre of 10^8 PFU/mL. P1 agar plates were also prepared with a final concentration of 0.13% glucose and 5mM $CaCl_2$. LB soft agar was prepared to a concentration of 0.4% and when molten was dispensed into four 15ml Corex glass culture tubes at 42°C.

When cells reached OD_{650} 1.0, 100µL aliquots were aseptically transferred into culture tubes containing soft overlay agar at 42°C in a SUB water bath. Immediately after, aliquots of the P1 master stock were added to the cells in volumes of 200µL, 100µL and 50µL respectively, with one cell culture as a no phage control. After phage introduction, cell cultures were removed from the water bath and briefly swirled by hand before pouring onto P1 plates. Plates were incubated at RT for overlay agar to set evenly before o/n incubation at 37°C.

After incubation, bacterial overlay from plates with the highest rates of phage plaques and cell lysis was aseptically transferred into a 15mL Falcon tube containing 1mL MC buffer and 0.5mL chloroform and vortexed thoroughly at RT. When the lysate had cleared, tubes were centrifuged at 10,000 rpm at 4°C for 20 minutes using an AG Eppendorf Centrifuge 5430 R. Supernatant was decanted into a new 15mL Falcon tube and mixed by inversion with 0.5mL chloroform. Phage lysate was stored at 4°C.

2.2.13 Phage sensitivity assay

Plaque assays were performed to determine the physiological and phenotypic response of NK001 cells compared to WT cells when challenged by phage. Assays were performed in triplicate and incubated at 30°C and 37°C o/n in optimal conditions. Experimental design was adapted from publications [35,40].

Isolated colonies of NK001 and MG1655 WT cells were grown in 5mL TBM medium o/n at 37°C with shaking, and aliquoted into 1.5mL Eppendorf tubes under a blue Bunsen flame before being placed on ice. Subsequently, molten 0.6% overlay agar was aliquoted in 3mL volumes into 15 mL glass Corex tubes at 42°C under sterile conditions. Cell cultures were inoculated in 200µL volumes into the molten overlay agar in appropriately labelled tubes under a blue flame. The mixtures were quickly swilled by hand at RT before being evenly poured onto set non- selective LB agar plates pre-incubated at 37°C in a dry air incubator (Carbolite). Overlay agar was allowed to set at RT before phage was applied.

Under a roaring blue Bunsen flame, λ*vir* phage was serially diluted in 10µL volumes in 10mM $MgSO_4$, from a factor of 10^{-1} to 10^{-5}. The phage dilutions were applied to the set agar in spots of 3 x 3.3µL volumes beginning with 10^{-1}concentrations, through to 10^{-5}. Control lawn plates with no phage were prepared in addition. Phage was allowed to soak in at RT before plates

were incubated o/n at both 30 and 37°C. Cell sensitivity was compared between NK001 and WT strains by calculating the PFU/mL and number of plaques on cell lawns for each phage dilution factor.

2.2.14 CRISPR adaptation and spacer acquisition assays

The primed adaptation and acquisition of DNA spacers into the CRISPR locus of MG1655 WT, NK001 transductants and inducible IIB969 cells with overexpressed antiPc*as* was performed according to protocol in [37].

For determining the effect of overexpressed anti-*cas* RNA on spacer acquisition, the strain IIB969 was first transformed with the antiP*cas* plasmid pDM13 and backbone vector pBadHisA as a control. IIB969 contains an IPTG-inducible *cas3* and arabinose inducible Cascade, with the addition of an crRNA anti-lambda spacer in the CRISPR array.

Infectivity efficacy depends on the ratio of viral phage particles per cell in a media, measured in an equation. The multiplicity of infection (MOI) was then calculated by dividing the number of plated viable cells by phage titre, expressed below as a short formula:

$$\text{MOI} = \frac{pfu/mL}{cfu/mL}$$

Where, for example if $2x10^5$ cells are infected by 40µL of virus with titre of 10^7,

$$\therefore 0.04*10^7/2*10^5 = 2.0 \text{ MOI}$$

Isolated colonies were selected from antibiotic selective plates and inoculated into 5mL LB broth containing 10µL of $MgCl^2$ and $CaCl^2$, and inducers 0.2% L-arabinose and 1mM isopropyl-β-D-thiogalactoside (IPTG). Control media lacked inducers. Cultures were incubated at 37°C with shaking until OD_{600} reached 0.4-0.6. At this point phage λ*vir* lysate was added to an MOI of 1.0, ($0.05*10^5$ PFU/mL) before phage adsorption for 15 minutes at 37°C without shaking.

Cells were then inoculated 1:10 into fresh LB media containing inducers as before, and incubated o/n at 37°C with shaking. Colony PCR was performed after incubation to amplify the CRISPR-1 locus.

Assays were subsequently repeated with the pNK01 plasmid construct containing the viral phage R gene to stimulate acquisition *in vivo*, due to the low efficacy and adsorption of phage λ*vir.* At OD_{600} 0.4-0.6, cells were transformed with 40ng of plasmid and incubated without shaking at 37°C for 5 minutes and then 20 minutes with shaking. Infected cells were inoculated in a 1:10 dilution into fresh LB broth containing inducers plus control LB media lacking inducers. Cultures were incubated o/n at 37°C with shaking. Colony PCR was subsequently performed after incubation using 2µL of cells denatured at 95°C and primers to amplify the CRISPR-1 locus of MG1655 (CRISPR-1F and CRISPR-1R, described in Table 16), reactions were then analysed on a 2% 1XTBE agarose gel.

Table 19. **PCR thermocycling method for DNA spacer detection** PCR was performed using Vent® DNA polymerase (NEB #M0254)

Step	Temperature (°C)	Time (seconds)	Cycles
Initial denaturation	95	180	1
Denaturation	95	45	
Annealing	58	30	30
Extension	72	60	
Final extension	72	300	1

2.3 Bioinformatics analysis

2.3.1 Sequence analysis and Modelling

To generate an initial model of the anti-*cas* secondary structure, the RNA fold hacker software was used [46]. The secondary structure model was constructed using a MFE fold algorithm to calculate the minimum distance between nucleotide bases needed for a complex and functional structure, based on the anti-*cas* RNA sequence. A colour key indicated the statistical likelihood of fold and lobe domains from highlighted regions of the RNA sequence. Models were also generated from the RNA sequences of small self-cleaving ribozyme classes found in *Escherichia* species, notably RNase P, Hammerhead, Hairpin, Twister, Twister-sister, Hatchet, and Pistol classes. Structural motif mapping of the anti-*cas* secondary structure was performed using the RNA Analyzer software programme [87], to identify conserved motifs common to known regulatory RNA molecules.

Models of Cas3 from *E. coli* and homologs from *T.fusca* and *T.terrenum* were generated using the Phyre2 software from FASTA files sourced from PDB and UniProt [12], as is shown in Chapter 1. Multiple sequence alignment (MSA) analyses were performed with the anti-*cas* RNA sequence, and a pool of ribozyme sequences found in the *Escherichia* genus using the R-Coffee and Clustal Omega software [28,37]. Controls used were randomised short (100-200nt) RNA sequences. Sequence similarity, alignment scores and the number of conserved sites were indicated by highlighted regions and asterisks denoting top hits. Sequences with a high degree of shared conservation and alignment were then analysed further by generating secondary structure models for comparing domains.

The phylogeny of the anti-*cas* RNA to bacterial ribozymes was performed using the MEGA11 software programme, to generate a tree of maximum likelihood showing the homology and divergence between the anti-*cas* RNA and next common species.

Chapter 3: The AntiP*cas* gene is integral in cellular physiology and adaptive immunity against phage

3.1 Introduction

As has been previously theorised, the orientation of the antiP*cas* promoter and sequence overlap between the RNA-encoding gene with *cas3* poses an intriguing question about the regulation of *cas3* expression and functional activity. This may signify a regulatory role of the gene in *cas3* gene expression and potential post-translational modulation of Cas3-nuclease-helicase activity.

This led to a research question being formulated about the role of the RNA-encoding gene in CRISPR-Cas regulation, which would be answered through the use of genetic methods to uncover more about the role of the RNA gene *in vivo.* The innate function of the RNA-encoding gene and any putative regulatory effect is unknown, an elegant method to answer this was to generate an *E. coli* strain with the RNA-encoding gene deleted, and determine the resulting cell phenotype and physiological effect.

3.2 RED recombineering

The antiP*cas* gene underwent a partial deletion using the Lambda RED recombineering methodology to insert a kanamycin antibiotic resistance cassette, as described in Chapter 2. The excision site lay in the middle of the gene, the scar codon of TAG generated from the inserted cassette sequence was situated 7bp downstream of the 3` protein coding strand of *cas3*, therefore avoiding a deleterious effect on *cas3* expression via polar excision. The successful insertion of the kanamycin antibiotic cassette was confirmed via colony PCR of the

CRISPR locus in the deletion strain, named NK001, with the inclusion of WT MG1655 cells as a control.

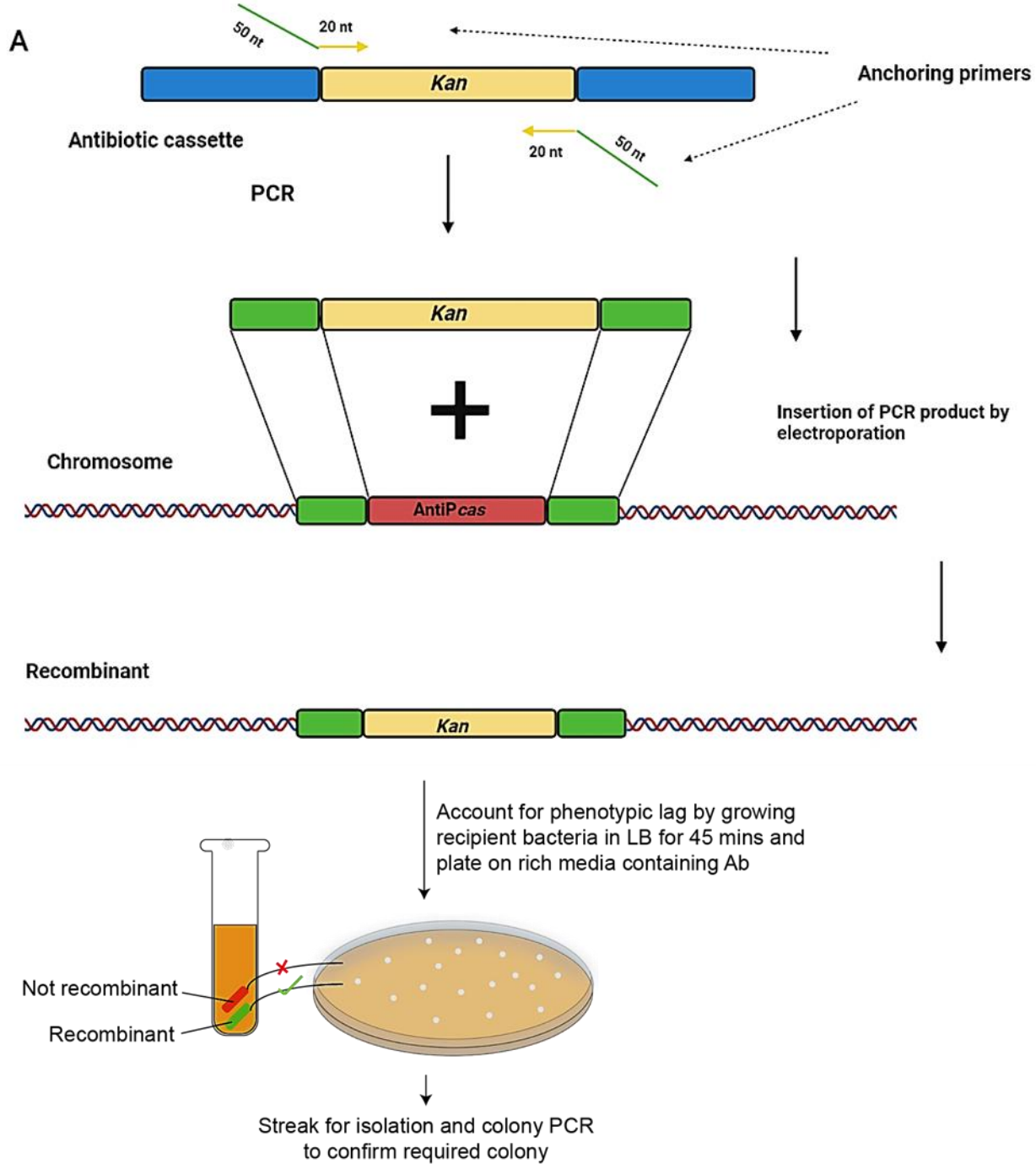
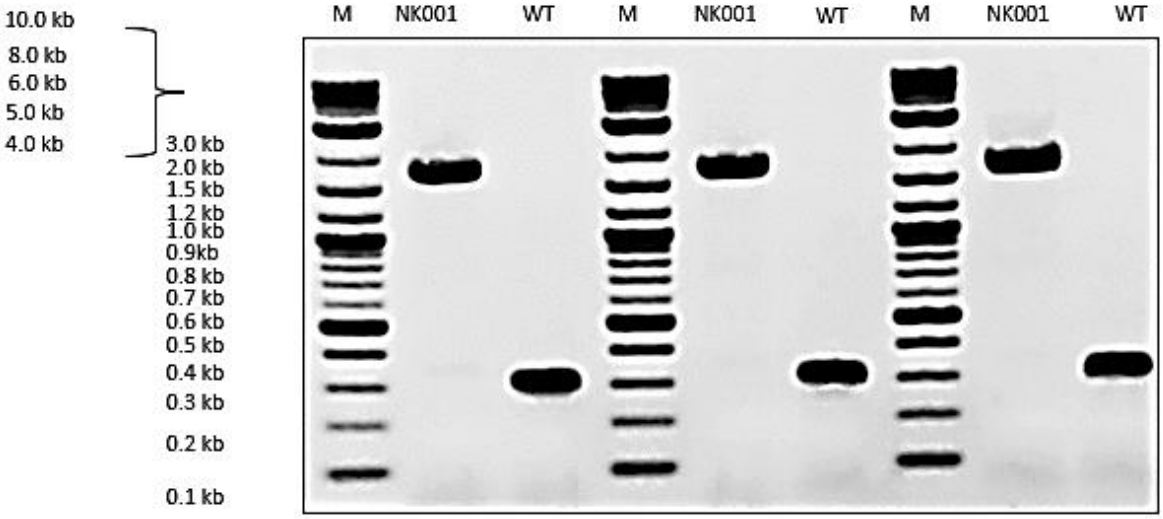
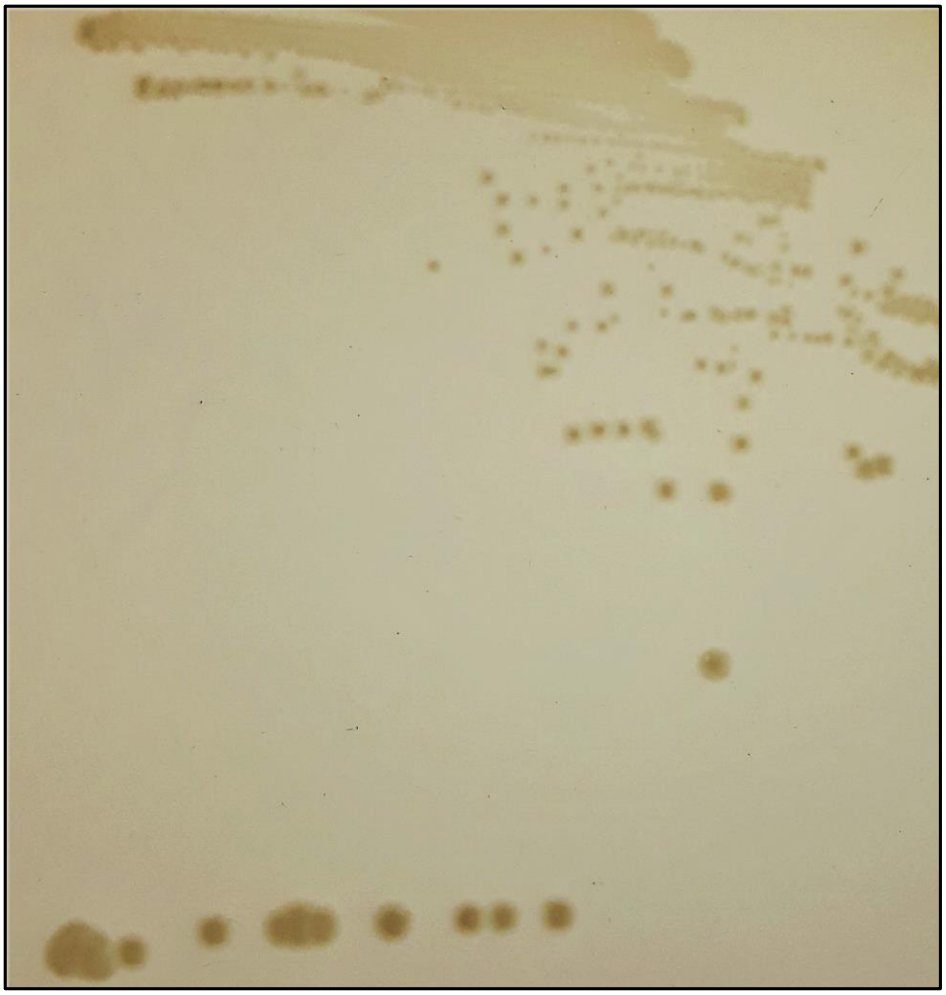


Fig 3.1 *cont*

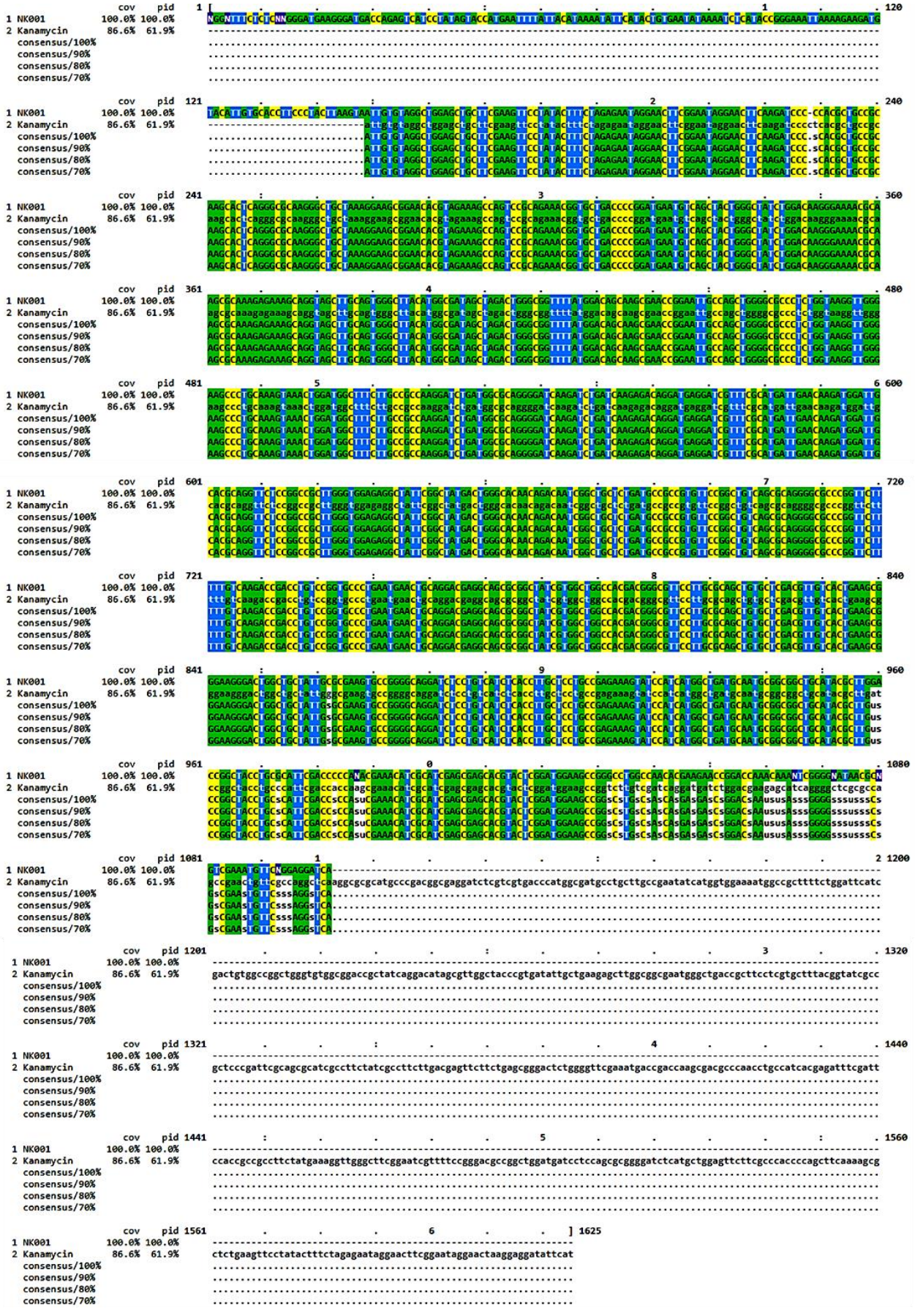
B


10.0 kb
8.0 kb
6.0 kb
5.0 kb
4.0 kb
3.0 kb
2.0 kb
1.5 kb
1.2 kb
1.0 kb
0.9kb
0.8 kb
0.7 kb
0.6 kb
0.5 kb
0.4 kb
0.3 kb
0.2 kb
0.1 kb
M
NK001
WT
M
NK001
WT
M
NK001
WT

C



D



E

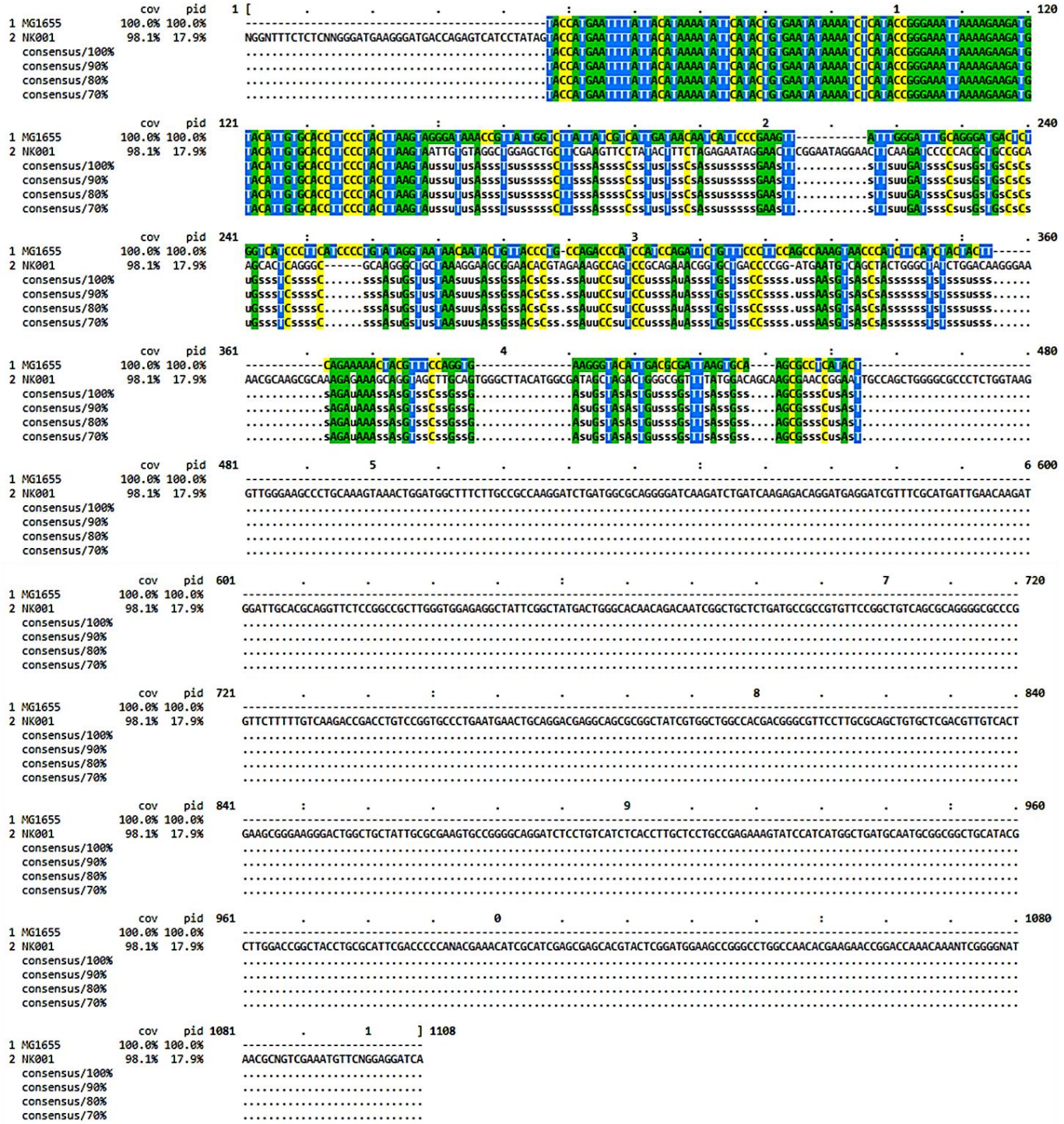


Figure 3.1 *cont.*

F

Figure 3.1 **Generation of the NK001 anti-*cas* deletion strain**

(A) Illustration showing the mechanism of Lambda RED recombineering *in vivo*, as performed in this study (B)Ethidium bromide stained 1.2% 1xTBE agarose gel analysis of the amplified deletion site. PCR was performed with a second set of RED primers designed to flank the antiP*cas* sequence region to confirm a successful deletion (C) NK001 viable colonies after antibiotic selection (D-E) A multiple sequence alignment demonstrates the difference between the CRISPR locus in MG1655 WT compared to NK001, across the anti-*cas* excision site. Sanger sequencing output reads of the deletion site locus in NK001. Aligned reads show an identical match to antibiotic cassette sequence in pKD4 donor plasmid, demonstrating the successful insertion of the kanamycin marker and deletion of the antiP*cas* gene *in vivo*. (F) The Scar codon site of TAG was located 5bp downstream of the *cas3* protein coding strand, thus avoiding a polar excision and delivering a clean deletion of the gene of interest.

3.3 AntiP*cas* overexpression in WT cells

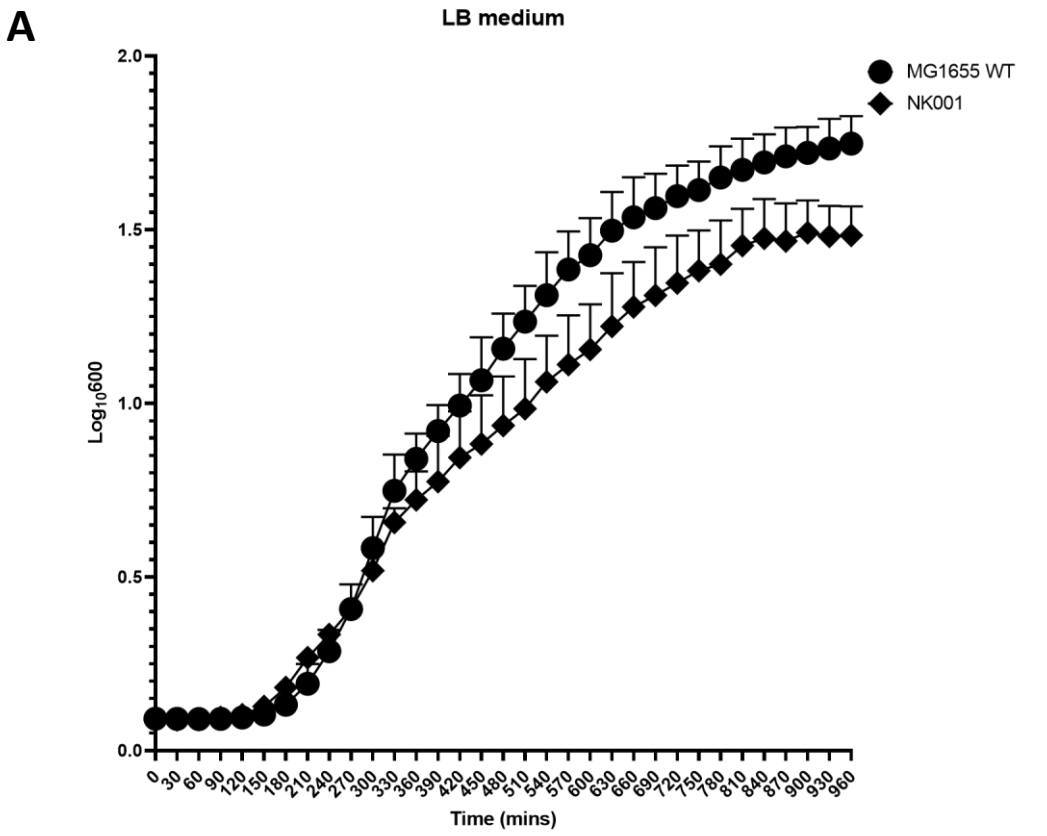
The role of the antiP*cas* gene in normal live *E. coli* cells was also examined by overexpressing the gene via plasmid delivery to the inducible promoter strain named IIB969. The purpose of this was to simulate primed adaptation *in vivo*, and determine any subsequent effect on CRISPR immunity and spacer acquisition into the CRISPR-1 locus. IIB969 contains an araB-inducible *cas3* gene and Cascade gene array, in addition to the crRNA T3 synthetic gene spacer necessary during CRISPR interference.

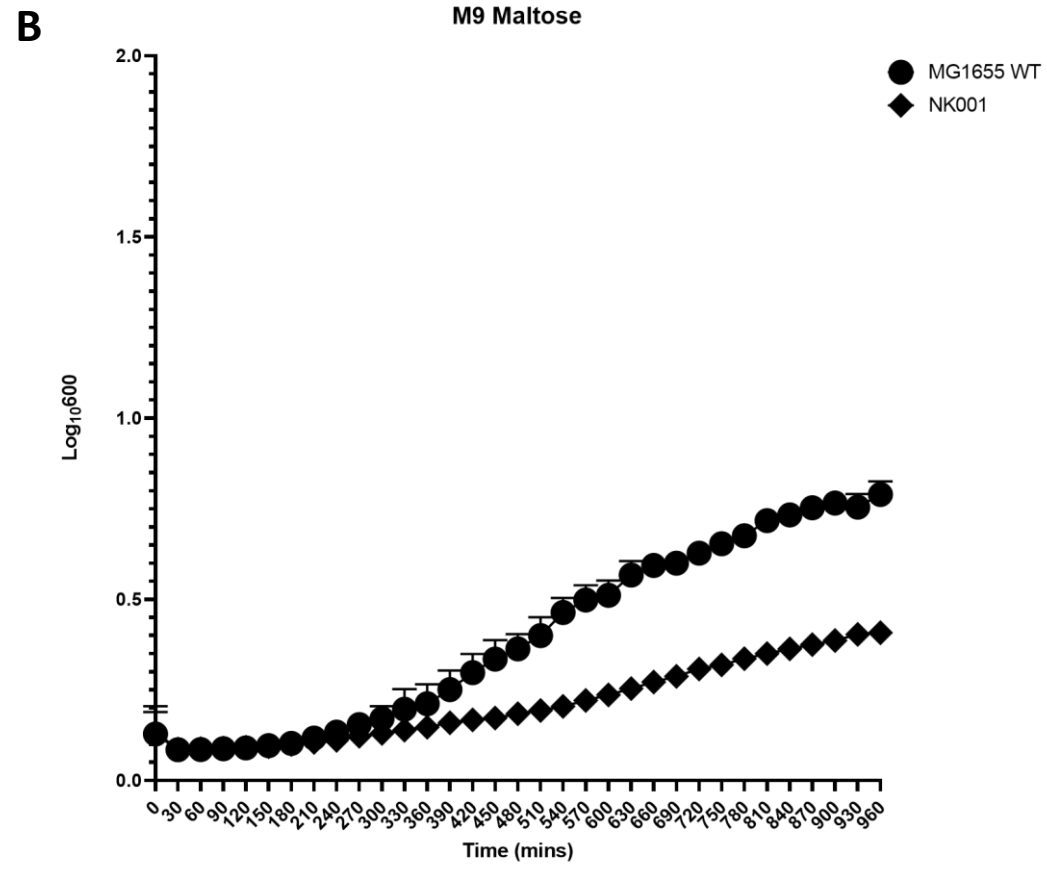
3.4 Genetic analyses

3.4.1 Cell growth and complementation assays

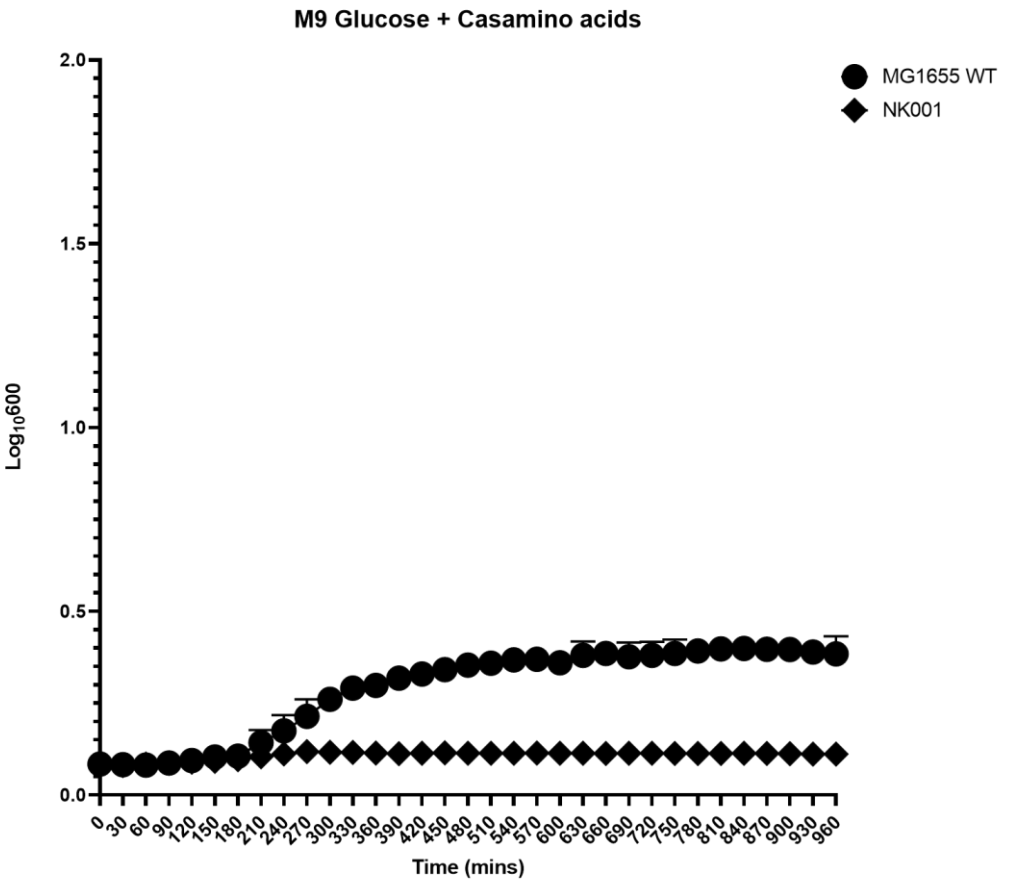
The phenotype of the anti-*cas* deletion strain, termed NK001, was first determined by measuring cell growth rates in nutrient medias with different carbon sources, over the course of 16-hour periods. NK001 cell growth rates were compared to WT MG1655 cells in identical media for the same time period. Results were represented as growth curves of cellular optical density readings, with means and SD applied to all data points. Optical density readings were taken after every 30 minutes for the duration of the assay.

NK001 cells exhibited markedly reduced growth rates in nutrient M9 medias supplemented with glucose and casamino acids, but showed no impaired growth in LB media in comparison with MG1655 cells (*Figure* 3.2 panel A). In M9 media supplemented with maltose, NK001 cells exhibited a stronger growth phenotype in comparison to the minimal growth curves recorded for media with glucose and casamino acid nutrients (*Figure* 3.2 panel C). WT cell growth was markedly reduced in all M9 medias other than LB. This is anomalous as it possesses the metabolic capability to synthesis various compounds, even though glucose and maltose sugars serve as poor carbon sources for energy metabolism. Complementation with the MG1655 was tested using NK001 cells with the anti-*cas* gene restored via transformation with pDM13, and empty pBadHisA as a control. Restoration of the RNA-encoding gene did not complement the mutant phenotype of deleted anti-*cas* to WT cells. This was apparent in M9 medias (*Figure* 3.2 panels E-F); however growth was not impaired in LB media (*Figure* 3.2 panel D). Both NK001 cell cultures exhibited a slower growth in M9 supplemented with maltose compared to NK001 cells without a plasmid vector (*Figure* 3.2 panel E).

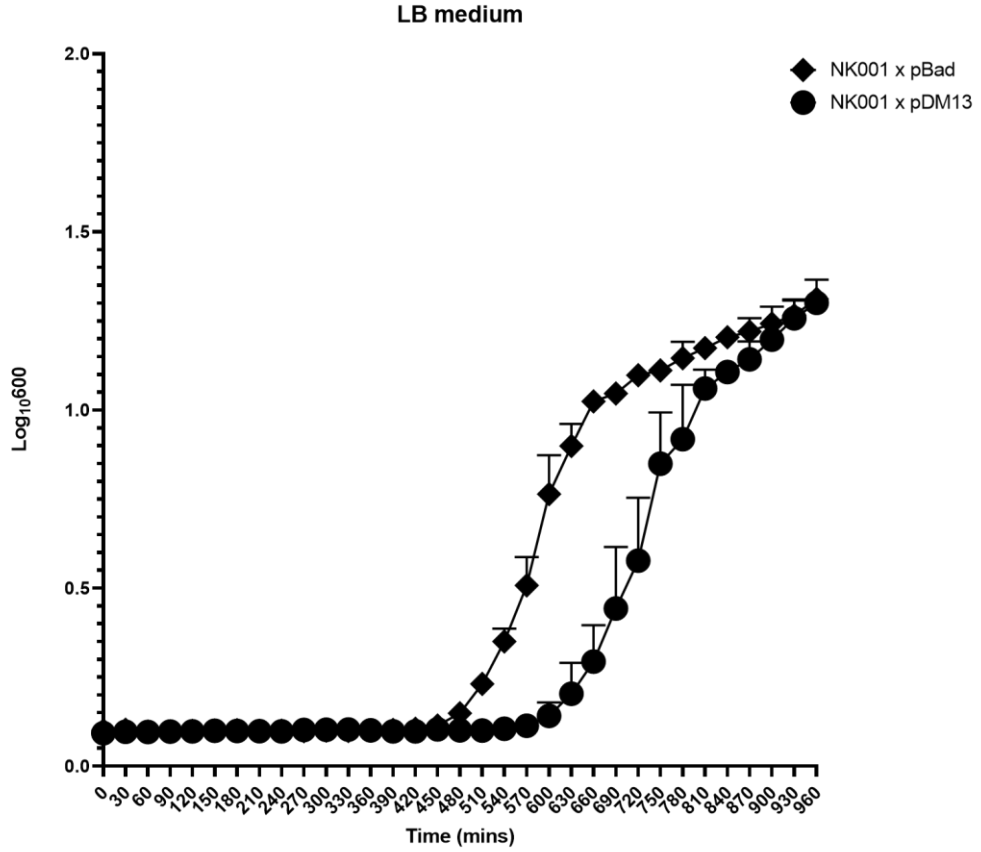

A
LB medium
MG1655 WT
NK001
Log10600
Time (mins)


B
M9 Maltose
MG1655 WT
NK001
Log10600
Time (mins)

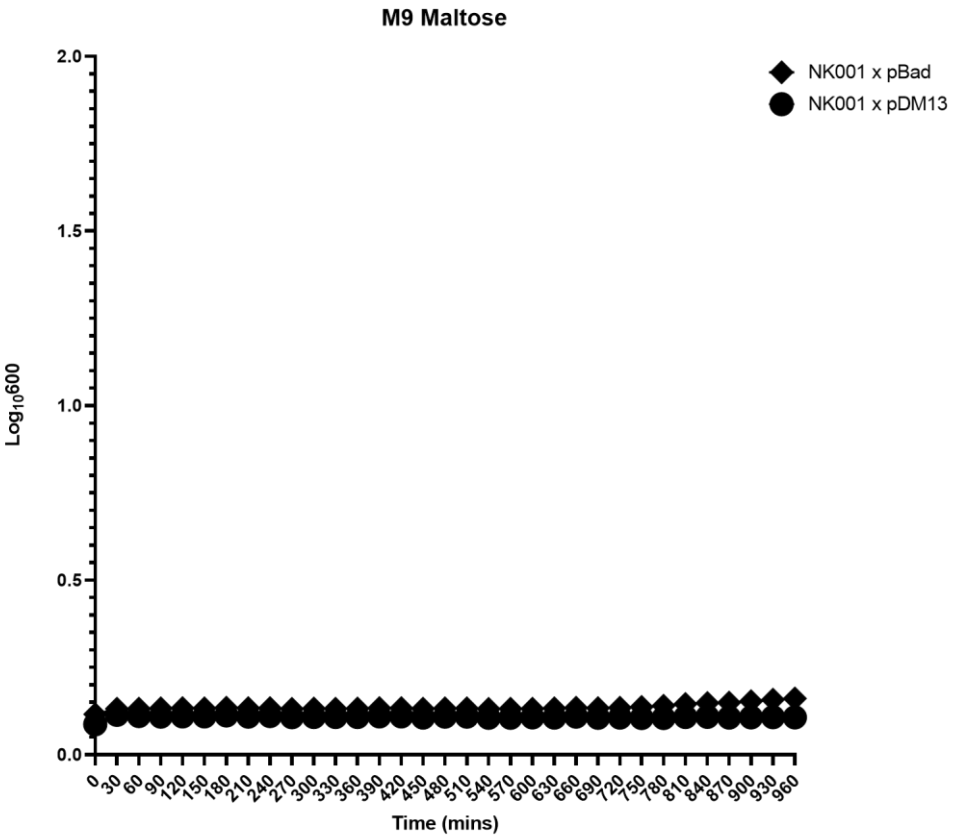
C


M9 Glucose + Casamino acids
MG1655 WT
NK001
Log10600
2.0
1.5
1.0
0.5
0.0
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960
Time (mins)

D


LB medium
NK001 x pBad
NK001 x pDM13
Log10600
2.0
1.5
1.0
0.5
0.0
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960
Time (mins)

E


M9 Maltose
NK001 x pBad
NK001 x pDM13
Log10600
2.0
1.5
1.0
0.5
0.0
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960
Time (mins)

F

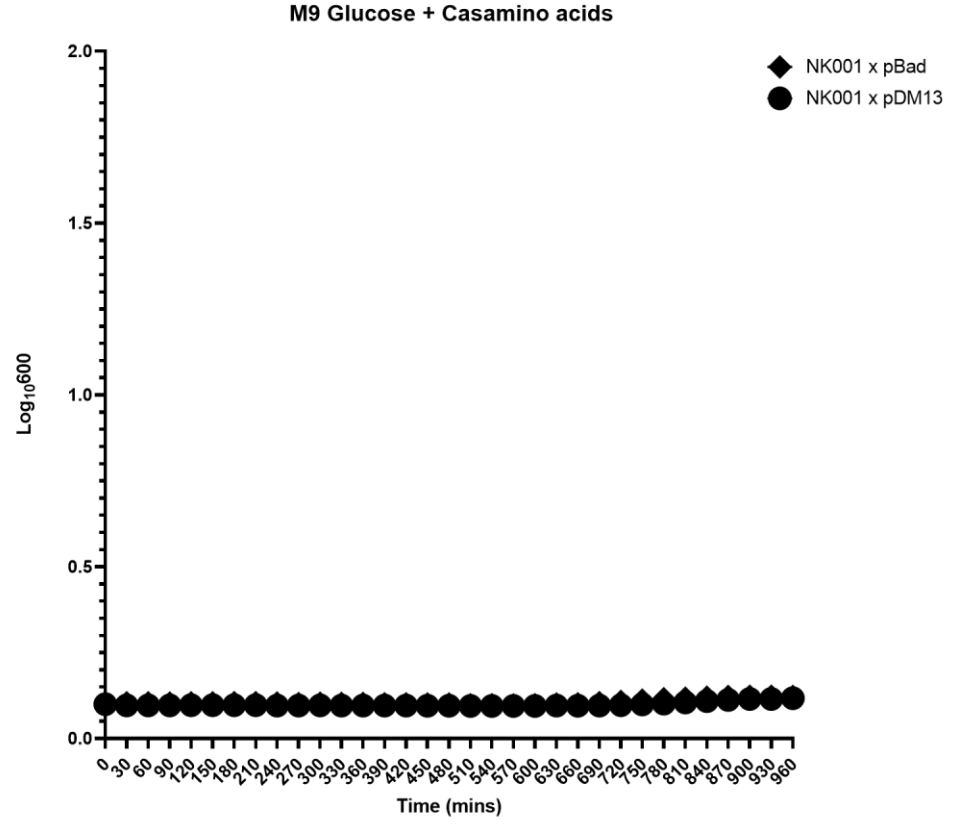

M9 Glucose + Casamino acids
NK001 x pBad
NK001 x pDM13
Log10600
2.0
1.5
1.0
0.5
0.0
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960
Time (mins)

Figure 3.2 **Rate of MG1655 and NK001 cell growth by media**

Growth curves were generated by measuring OD_{600} absorbance values at regular time intervals during overnight incubation. (A-C) Growth of WT MG1655 and NK001 cell strains in varying nutrient medias after 16 hours overnight incubation at 37°C with shaking. Mean values were calculated from assays performed in duplicate. The standard deviation from the mean is indicated by error bars, calculated using Prism 9 (GraphPad) software. (D-F) Comparative growth rates of NK001 WT and with the antiP*cas* gene restored via plasmid. All growth media contained a final concentration of 0.2% L-arabinose to induce the antiP*cas* gene and ampicillin to retain plasmids.

Tables shown below provide an insightful overview of the comparative growth phenotypes of MG1655 WT and NK001 cells in media supplemented with maltose.

Table 20-21. Comparisons of MG1655 and NK001 cell growth in M9 Maltose media

Bacterial strain	OD_{600} range	Means and SD	Backbone vector	0.2% L-arabinose final
MG1655	0.7045	0.3942 , 0.2558	N/A	-
NK001 (1)	0.3215	0.2132 , 0.1074	N/A	-
NK001 (2)	0.0451	0.1363 , 0.0088	pBadHisA	+
NK001 (3)	0.0273	0.1077 , 0.0042	pDM13	+

NK001 (1) cell growth in maltose	
Time (mins)	Mean OD_{600} readings
0	0.128
30	0.086
60	0.087
90	0.089

120	0.091
150	0.093
180	0.099
210	0.105
240	0.111
270	0.122
300	0.13
330	0.139
360	0.147
390	0.159
420	0.167
450	0.172
480	0.184
510	0.193
540	0.204
570	0.221
600	0.236
630	0.253

660	0.272
690	0.287
720	0.308
750	0.319
780	0.336
810	0.35
840	0.363
870	0.375
900	0.386
930	0.403
960	0.408

3.4.2 Phage λ*vir* sensitivity assay

Evaluation of the *in vivo* functionality of the RNA-encoding gene and potential impact on CRISPR-Cas catalysed phage immunity was assayed by examining λ*vir* plaque formations. Results show the NK001 strain displayed a significantly higher resistance to phage incursion at both 30°C and 37°C than MG1655 WT. Viral plaques were visible on NK001 cell lawns at dilutions from 10^{-1} to 10^{-3}, however cell lysis was inhibited at dilutions of 10^{-4} and 10^{-5}. In

comparison, clear plaques were visible at dilutions of 10^{-1} through to 10^{-5} on MG1655 cell lawns, showing an increased sensitivity to phage invasion.

The appearance of the viral plaques on NK001 cell lawns indicated signs of cell lysis and regrowth over the plaques, as can be seen from a distinctive pale overlaying bacterial film. This may suggest an initial phage sensitivity in the cells before undergoing a resistance reaction, a form of cellular "on" switch.

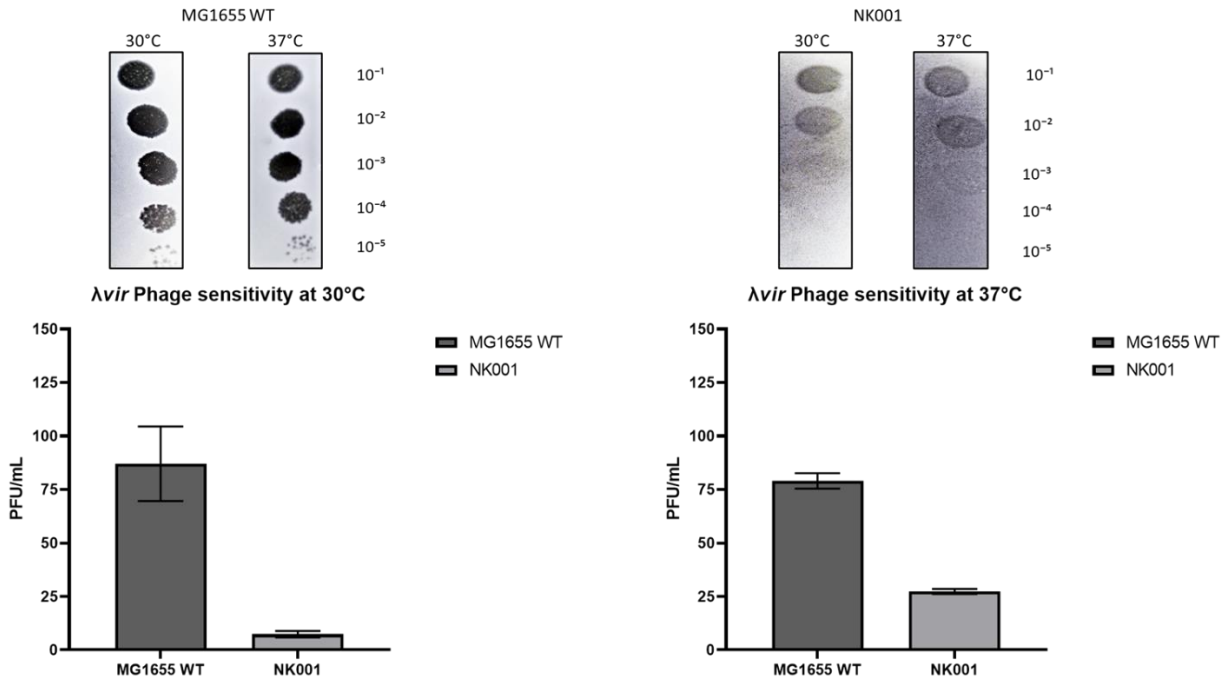


Figure 3.3 **NK001 cells exhibit a distinctive phage resistant phenotype compared to MG1655 WT**

E. coli cell lawns of strains MG1655 and NK001Δ*anti-cas* were infected with phage dilutions of 10^{-1} to 10^{-5} and incubated o/n at 30°C and 37°C as indicated. Bar charts represent means and SD of the number of plaque forming units (PFU/mL) from three independent experiments. Corresponding photographs above show phage plaques in serial dilutions on bacterial lawns. Results showed NK001 cells displayed a significantly higher resistance and pronounced phage resistant phenotype when challenged, at both 30 and 37°C in comparison to MG1655 cells. Plaques on NK001 lawn cells exhibited a faint "milky" overlay, which may be evidence of cell lysis before fresh cellular overlay.

A more robust phage response may also be physiological, based off the lytic behaviour of the phage-challenged cells. An alteration in cellular morphology triggered by the antiP*cas* gene deletion may cause the cell to exhibit a resistant phenotype, by way of a physiological defence against phage, rather than a stronger CRISPR immune response as MG1655 does not have an active CRISPR immune response due to an intact H-NS.

3.4.3 NK001 cell viability and SOS response

NK001 cells challenged by λ*vir* showed an initial lysis stage, and regrowth over viral plaques, the morphology of the cells were subsequently compared to NK001 cells unchallenged by phage. NK001 cell cultures were recovered from phage plaques at 10^{-1} dilutions from LB agar plates incubated at 30°C and 37°C overnight, subsequently cultures were aseptically streaked out onto fresh LB media and incubated overnight at 37°C. Colonies of NK001 plaque cells exhibited a distinct phenotypic morphology in comparison to NK001 cells unchallenged by λ*vir*, phage challenged cells formed very small pale colonies in clusters, in comparison to large colonies of NK001 unchallenged cells.

The distinct cellular morphology of phage-challenged cells enabled informed conjecture that this was triggered by an SOS response; if the absence of the RNA-encoding gene creates a mutant phenotype, it may also result in DNA repair activity that manifests physiologically as an SOS response. Further experimental analysis was conducted to determine if challenged cells were exhibiting a physiological change due to an SOS response triggered by phage.

Optical density readings of LB broth-cultured NK001 strains were taken every 30 minutes for a duration of 150 minutes, viability spot tests were performed to determine if NK001 cells were physiologically stressed, and exhibited a low viability. This methodology is described in

detail in Chapter 2. Cells undergoing physiological stress are filamented and elongated compared to stable healthy cells in optimal conditions, resulting in high OD readings but low viability.

OD_{600} readings from challenged and unchallenged cultures followed an identical growth trend, with minor fluctuations in individual cell density readings as represented by error bars (*Figure 3.4* A) .Colony counts from phage challenged cells were not significantly different from unchallenged cells, showing that phage challenged cells were viable (*Figure 3.4* B). This data confirms that NK001 cells were not exhibiting an SOS response as a by-product of phage invasion.

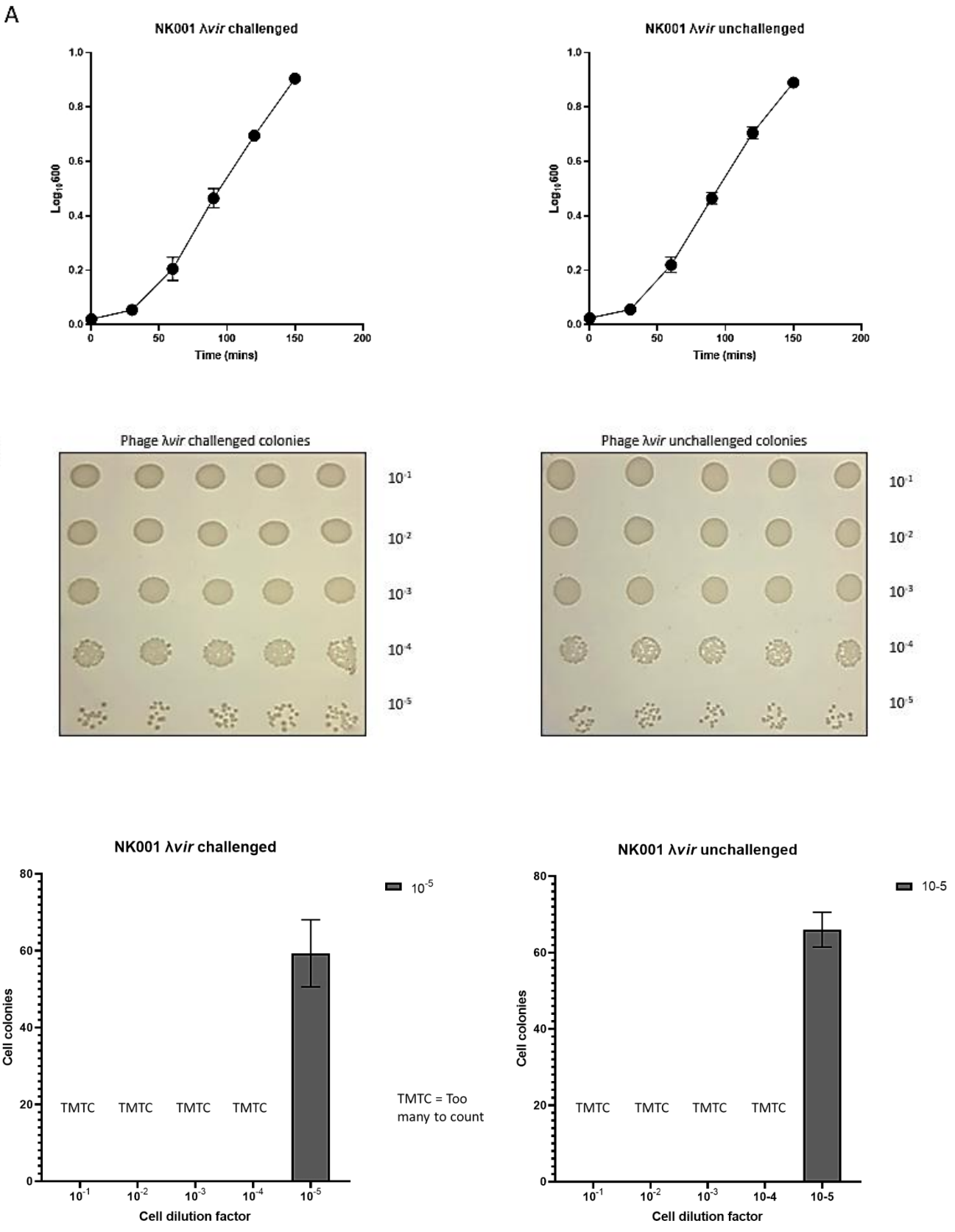

A
NK001 λvir challenged
Log10600
Time (mins)
NK001 λvir unchallenged
Log10600
Time (mins)
B
Phage λvir challenged colonies
10-1
10-2
10-3
10-4
10-5
Phage λvir unchallenged colonies
10-1
10-2
10-3
10-4
10-5
C
NK001 λvir challenged
Cell colonies
TMTC
TMTC
TMTC
TMTC
10-5
Cell dilution factor
TMTC = Too many to count
NK001 λvir unchallenged
Cell colonies
TMTC
TMTC
TMTC
TMTC
10-5
Cell dilution factor

Fig 3.4 *cont.*

Figure 3.4. **Analysis of SOS response and viability in NK001 λ*vir* challenged and unchallenged cells**

Cell growth and viability were analysed in the NK001 RNA deletion strain to detect a physiological SOS response, due to differences in cellular morphology between cells challenged with phage. (A) Optical density readings were taken every 30 minutes for NK001 cultures until OD_{600} reached 0.9. (B) Cell viability of NK001 phage challenged cells was not impacted, both cell cultures exhibited normal cell viability. Assays repeated in triplicate and incubated at 37°C o/n (C) Corresponding number of viable cell colonies counted from viability spot plates.

3.4.4 Primed adaptation in the CRISPR-1 array

Experimental attempts to simulate primed adaptation in vivo were not successful, despite a series of trials infecting IIB969 cells with the pNK01 λ*R* gene construct. An explanation for why this series of assays may have been unsuccessful, and possible future experimentation in this area is given in Chapter 5.

3.6 Summary

Knockout of the antiP*cas* gene was successfully achieved using the Lambda RED recombineering system, a novel strain named NK001 was generated for further downstream experimentation to examine the effect of absent anti-*cas* RNA on live *E. coli* cells.

In overnight growth assays with WT MG1655 cells in different nutrient medias, it was found that the NK001 strain exhibited a characteristically strong growth phenotype in media supplemented with maltose. A putative explanation for this is given in Chapter 5.

The NK001 strain was also found to exhibit a more resistant phenotype when challenged by phage invasion. Further analysis of the phage resistant phenotype of the NK001 strain was performed by determining comparative viabilities of λ*vir* challenged and unchallenged cells, to test a notion of the manifestation of a stress response in NK001. NK001 cells exhibited normal growth and high viability after spot testing, exhibiting no signs of an SOS response after inducement.

Overexpression of the RNA-encoding gene was performed in an H-NS deletion strain to simulate primed adaptation *in vivo* and determine any associated effect on CRISPR immunity from the antiP*cas* gene. Both λ*vir* and the pNK01 construct were used to infect cells and stimulate primed adaptation. Colony PCR results from lysed culture showed no expansion of the paternal CRISPR array in infected cells, this was due to issues with adapting the protocol to the pNK01 construct, and will be briefly addressed in Chapter 5.

Chapter 4: *In vitro* and bioinformatics analyses of the anti-*cas* RNA and activity with *Eco*Cas3

4.1 Introduction

AntiP*cas* is an anti-sense promoter encoding a gene that overlaps with the 3` terminal end of *cas3* in the *E. coli* CRISPR locus. It generates a short transcript of ncRNA 150-200nt in length, of which the functional significance is unknown. Previous research has indicated it forms an elaborate secondary folded structure, which is reminiscent of those seen in regulatory RNA classes, such as riboswitches and ribozymes. Cas3 nuclease-helicase has been shown to interact with nonCRISPR RNA molecules[25], presenting an intriguing pathway for potential interaction with uncharacterised ncRNA.

The interaction of the anti-*cas* RNA transcript with the *Eco*Cas3 nuclease-helicase has never been previously tested, therefore presenting a novel opportunity for investigating the interplay between Cas3 and ncRNA using a range of *in vitro* biochemical analyses to determine the potential modulating effect of short ncRNA on Cas3-mediated dsDNA cleavage.

To improve the characterisation of the anti-*cas* RNA sequence, a range of bioinformatics analyses were performed to generate statistically accurate models of the RNA secondary structure and determine sequence conservation and orthology to known species of bacterial small regulatory RNA's.

4.2 RNA synthesis

As previously described in the methodology chapter, anti-*cas* RNA was synthesised for use as a substrate using the pDM13 plasmid as template DNA. Synthesis was achieved using primers containing the T7 promoter to amplify a region spanning across the antiP*cas* gene and 3'

protein coding region of *cas3*. The RNA was transcribed from the top strand of the DNA, and had a resulting molecular weight of 51.1138 kDa.

Purified RNA substrates were visualised through electrophoresis in a 10% denaturing urea PAGE gel, and migrated true to size, exhibiting no large secondary structure due to urea (*Figure 4.1*). In addition, the Cyanine 5 fluorescent dye was conjugated to the template DNA during *in vitro* transcription to generate a labelled RNA substrate for use in protein binding and RNA synthesis assays. Cyanine 5 emits a far-red visible light, with an excitation and emission peak of 651nm and 670nm respectively, enabling the efficient detection of the RNA substrate.

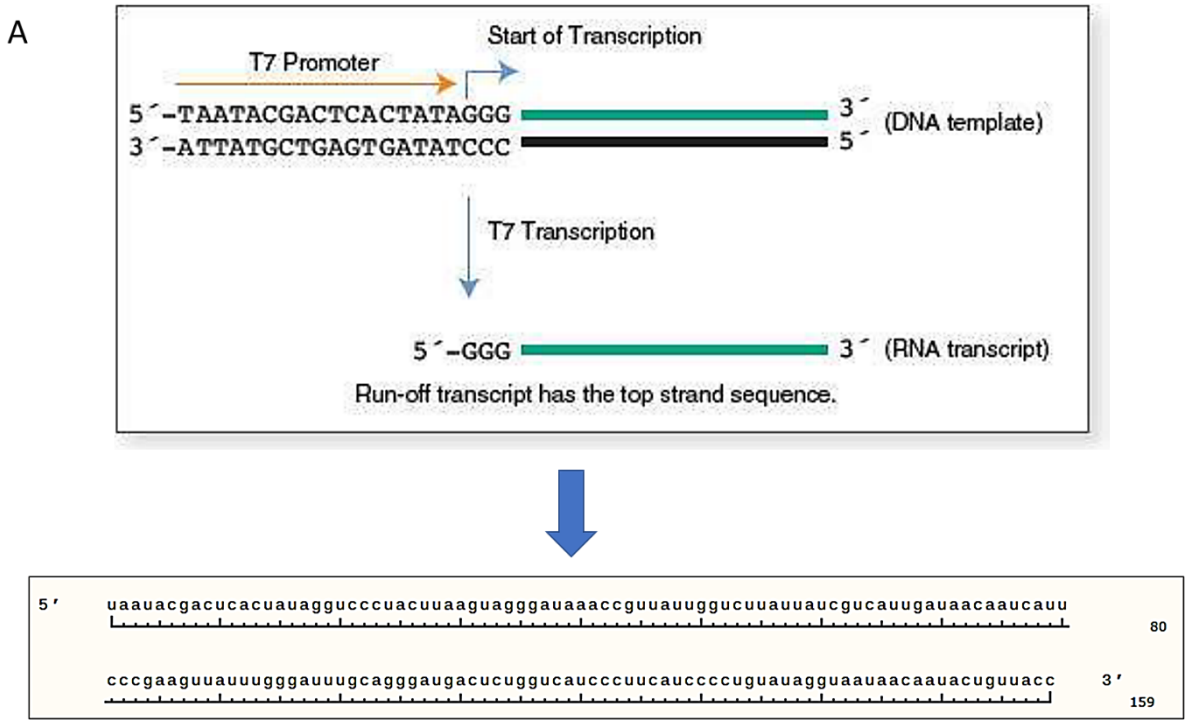
4.3 EcoCas3 purification

Purified monomeric and polymeric *Eco*Cas3-MBP stocks were prepared according to protocol in [65], and were provided to me by Dr. Liu He. Prior to use in subsequent biochemical reactions, the *Eco*Cas3-MBP protein was visualised on an 8% SDS-page gel to confirm the correct molecular size. The Cas3 protein was previously expressed from the pBailey plasmid backbone, and tagged to a 45 kDa MBP protein at the N terminal [65]. After electrophoresis, *Eco*Cas3 was visible at true size of 143.5 kDa (*Figure 4.2*), which can be clearly inferred from the adjacent CPS protein standard ladder.

4.4 Biochemical analyses

To investigate the relationship and establish a basis of potential modulation of Cas3 nuclease activity by anti-*cas* RNA, it was first examined if the RNA could bind to Cas3. Once this was determined, more complex biochemical assays were performed to measure the effect of the anti-*cas* RNA substrate on dsDNA cleavage and unwinding by Cas3 WT and mutant proteins, and whether anti-*cas* could inhibit the transcription of *cas3*. The nascent synthesised

substrates were visualised after electrophoresis on a 1xTBE 5% denaturing PAGE gel post-stained with ethidium bromide. Both substrates were fully denatured and migrated true to size, corresponding to the predicted RNA transcript length described in [24].



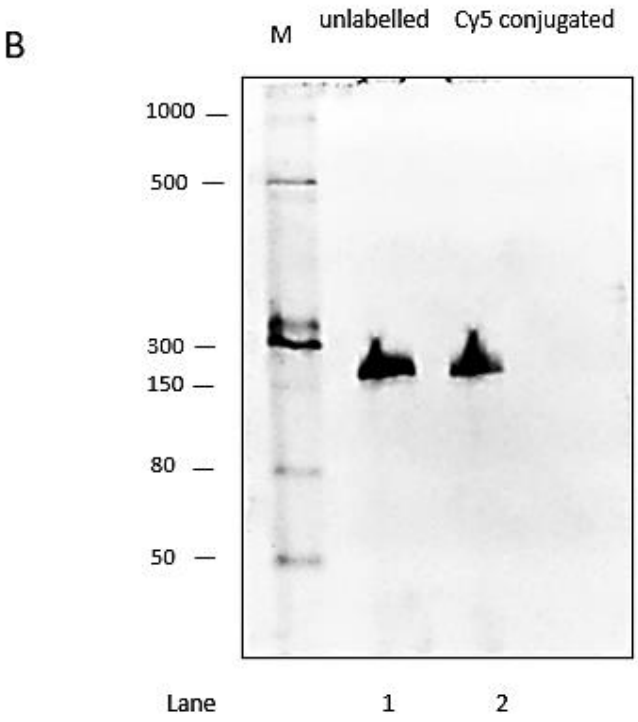


***Figure 4.1 In vitro* synthesis of the anti-*cas* RNA**

(A) RNA transcript synthesised from a PCR amplification of the pDM13 template using T7 RNA polymerase.(B) Ethidium bromide (EthBr) stained 5% 1xTBE acrylamide urea PAGE gel with Low Range ssRNA ladder (NEB) and nascent anti-*cas* RNA substrates. Samples denatured at 95°C prior to gel loading. RNA transcripts are 159 nt in size.

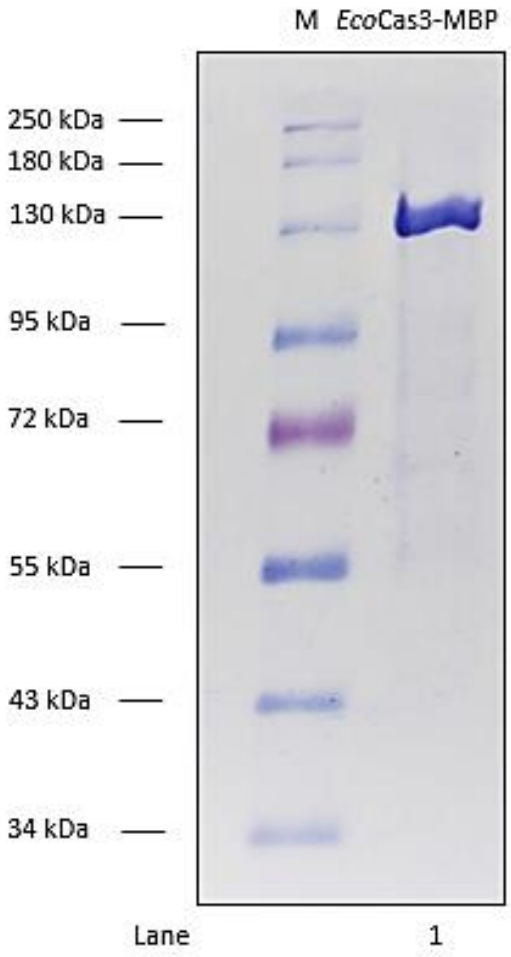


***Figure 4.2* EcoCas3-MBP visualised on SDS-PAGE gel**

Coomassie blue-stained 8% SDS-PAGE gel showing migration position and molecular size of Cas3-MBP protein used in this research. Approximate protein size is 140.8 kDa, with EcoCas3 being 100.5kDa and MBP 40.3kDa. The molecular size can be inferred from the adjacent CPS protein standard ladder (NEB).

4.4.1 Protein-Nucleic Acid assays

4.4.1.1 EMSA

A series of Electrophoretic Mobility Shift Assays (EMSA) were performed to determine if Cas3 could bind to anti-*cas* RNA substrate, by titrating Cas3 in increasing concentrations whilst maintaining an RNA constant molar concentration. *Eco*Cas3 WT was used initially and repeated with a nuclease-inactive mutant Cas3 protein as a control.

Results were imaged using the Amersham Typhoon™ Biomolecular Imager (GE healthcare). Anti-*cas* RNA and *Eco*Cas3 formed a weakly associated complex, as very faint band shifts were observed above the baseline migration position seen in the NP control lane. Intriguingly, faint degradation of the RNA transcript band was observed in reactions with higher Cas3 nanomolar concentrations (*Figure 4.3* A).The assay was repeated with identical experimental conditions and the catalytically inactive mutant protein $Cas3^{D75G}$, this was done to determine if RNA band degradation was caused by Cas3 and not residual tagged proteins remaining after the initial Cas3 WT purification process (*Figure 4.3* B). The results showed no band degradation, confirming that Cas3 was degrading anti-*cas* RNA by a form of binding mechanism.

Binding interactions between Cas3 and anti-*cas* RNA were repeated using agarose gel electrophoresis with a low buffer concentration and high voltage during the run duration. This methodology was optimised for visualising short, labelled RNA sequences, enabling a greater resolution of minor band shifts and reduction of nucleic acid diffusion into the gel matrix. Assay conditions were maintained as in the previous experiment run on a native PAGE gel. Samples were mixed with 8µL Orange G loading dye 1 prior to gel loading. Electrophoresis was performed at 12.5 V/cm for a duration of 20 minutes in 0.5XTB buffer. Results showed

more resolved band shifts above the baseline RNA position, this may be indicative of a formed bound complex between Cas3 and the anti-*cas* RNA substrate (*Figure 4.4* A).

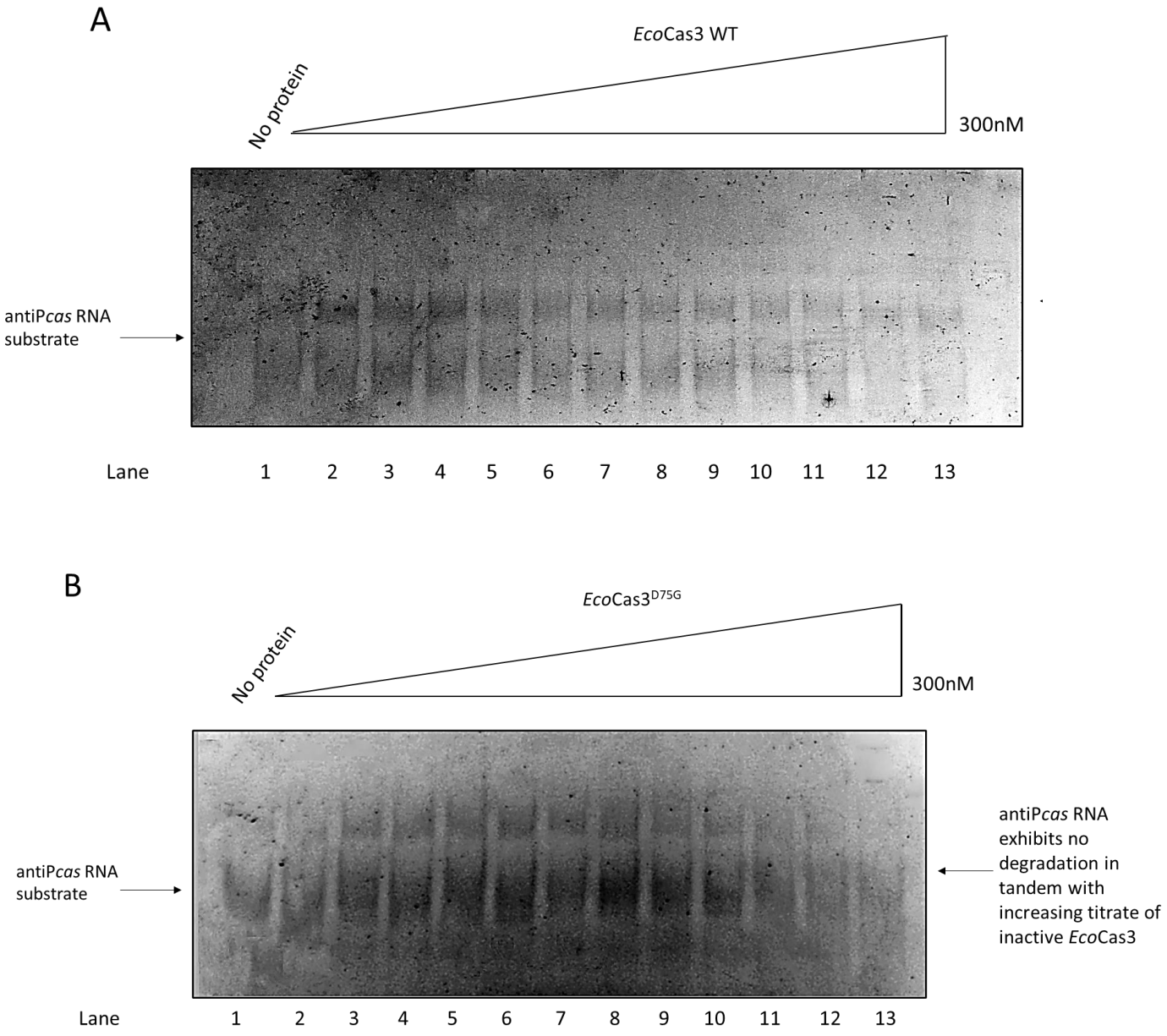


Figure 4.3 **EMSA with *Eco*Cas3 and anti-*cas* RNA**

(A)5% acrylamide 1xTBE Native PAGE gel analysis showing interaction of *Eco*Cas3 WT with constant 25nM Cy5 labelled anti-*cas* RNA substrate. Reactions were incubated for 60 minutes at 30°C in buffer containing 50mM Tris-HCl pH 7.5, 100mM NaCl, 0.1 mg/mL BSA, and 20mM DTT. Minor anti-*cas* RNA substrate degradation was apparent that increased in correlation with *Eco*Cas3 molar concentrations. (B) The assay was repeated with the catalytically inactive Cas3^{D75G} mutant, to control for non-specific RNA degradation from residual tagged proteins in the Cas3 elutant after purification.

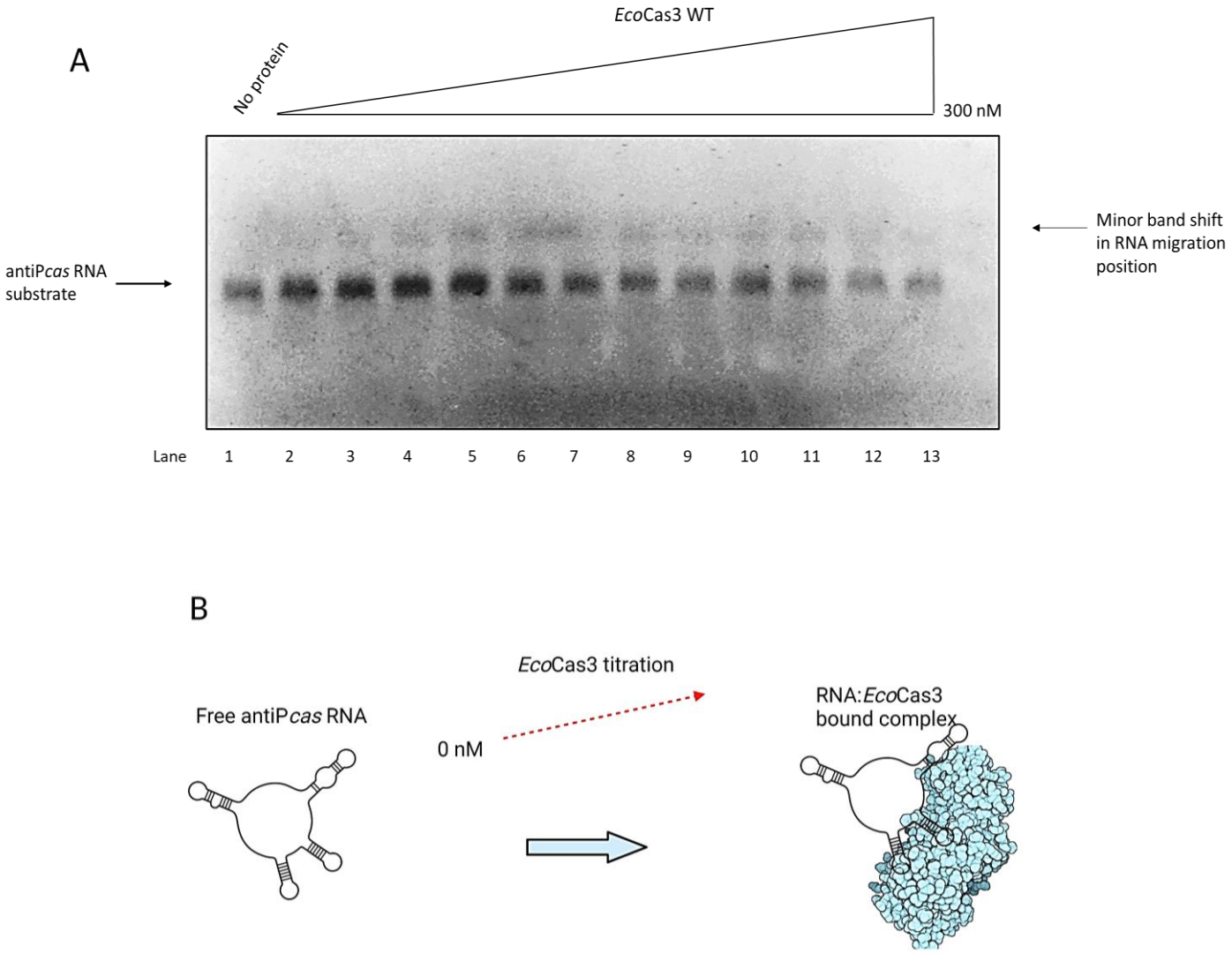


Figure 4.4 **Agarose gel electrophoresis of anti-*cas* RNA and *Eco*Cas3 EMSA**

(A)2.5% 0.5XTB agarose gel stained with EthBr showing anti-*cas* RNA substrate migration after incubation with *Eco*Cas3. Assay conditions were maintained and repeated in an identical manner to reactions loaded onto Native PAGE gels. Anti-*cas* RNA was incubated at a constant concentration of 30nM with a Cas3 molar gradient. Putative sign of a bound complex formed during incubation was observed, as minor band sifts are visible (B) Illustration shows the formation of a bound complex of the anti-*cas* RNA with *Eco*Cas3.

4.3.1.2 Nuclease assays

It was hypothesised that the presence of the anti-*cas* RNA substrate may modulate Cas3 dsDNA degradation and nuclease activity *in vitro*, by competitive inhibition with DNA for access to the binding channel. If both RNA and DNA substrates are captured and translocated along the binding channel, Cas3 cleavage on dsDNA substrates may be significantly reduced.

Preliminary assays were performed to determine if Cas3 nuclease activity on dsDNA was affected in different experimental conditions, and therefore gauge a range of optimal conditions needed to control for both Cas3 functional stability and anti-*cas* RNA structural formation (*Figure 4.5* panels A-B). Cas3 was also incubated with labelled RNA substrates to determine if Cas3 could cleave dsRNA and display any nucleotide specificity (*Figure 4.5* panels C-D).

Sequential assays showed cleavage of the labelled RNA substrate by Cas3 WT, optimal cleavage occurred at a protein concentration of 500nM at 30°C, with reduced activity at 37°C, *(Figure.4.5* panels E-F) Cas3 WT displayed a low affinity for saline conditions, which was controlled for by the use of KoAc and to measure if nuclease activity was repressed at high salinity. The use of KoAc was also to determine any difference in how the degree of RNA secondary structure formation could impact Cas3 nuclease activity, thus resulting in a greater binding of the substrate in the ssDNA channel. The assay was repeated with the catalytically inactive $Cas3^{D75G}$ mutant, results showed significantly less band degradation at 30 and 37°C, replicating the data from[62]. The hypernuclease $Cas3^{w406A}$ mutant displayed maximal cleavage product at increasing concentrations, at both 30 and 37°C assay conditions.

Anti-*cas* RNA and *Eco*Cas3 WT protein were preincubated in nuclease-free water at 30 or 37°C for 10 minutes before the reaction was initiated with $MgCl_2$. Incubation continued for 4 hours before reactions were stopped with the addition of Proteinase K and RNase T1. Results showed no visible decrease in oligonucleotide duplex substrate cleavage with the addition of anti-*cas* RNA (*Figure 4.5* panel I, lanes 6-7, 13-14). This is contrast to what was predicted based on the preceding assays showing Cas3 dual cleavage of DNA and RNA substrates. A potential explanation for this result is provided in Chapter 5.

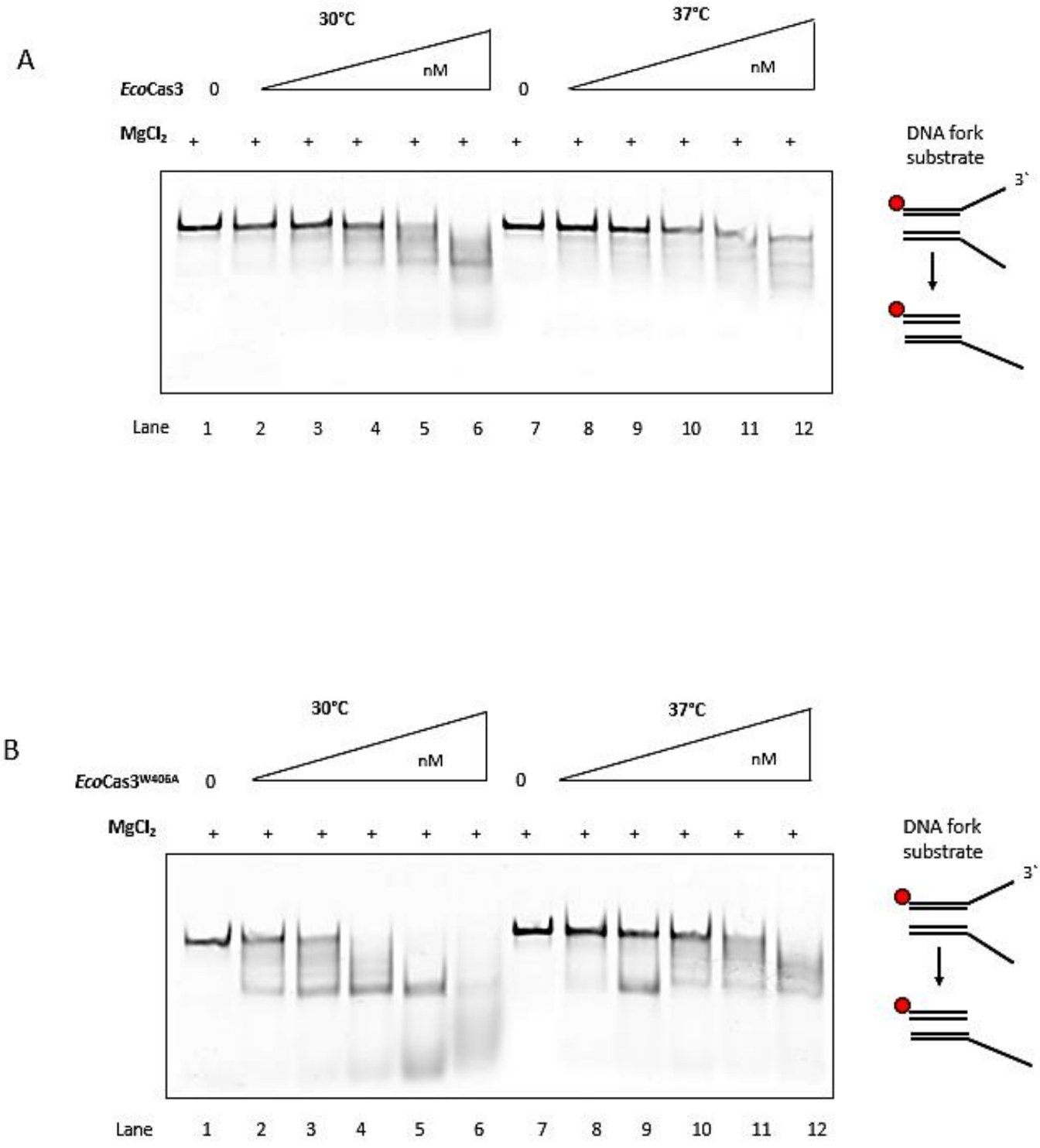
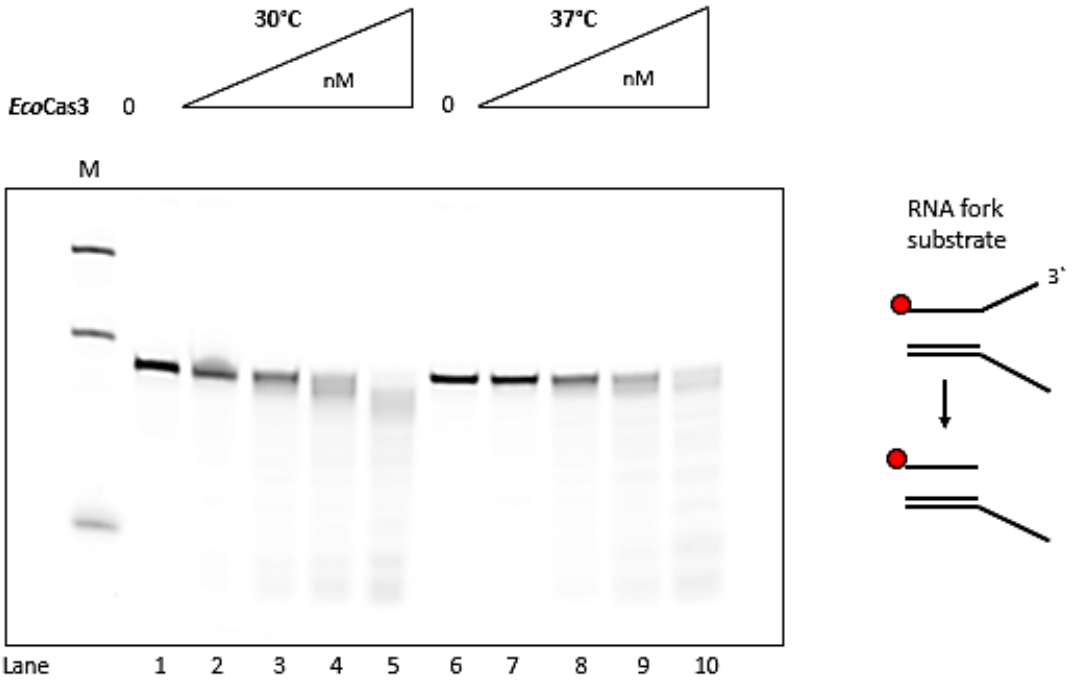

A
30°C
37°C
EcoCas3
0
nM
0
nM
MgCl2
+ + + + + + + + + + + +
DNA fork substrate
3`
Lane 1 2 3 4 5 6 7 8 9 10 11 12
B
30°C
37°C
EcoCas3W406A
0
nM
0
nM
MgCl2
+ + + + + + + + + + + +
DNA fork substrate
3`
Lane 1 2 3 4 5 6 7 8 9 10 11 12

Fig 4.5 *cont.*

C



D

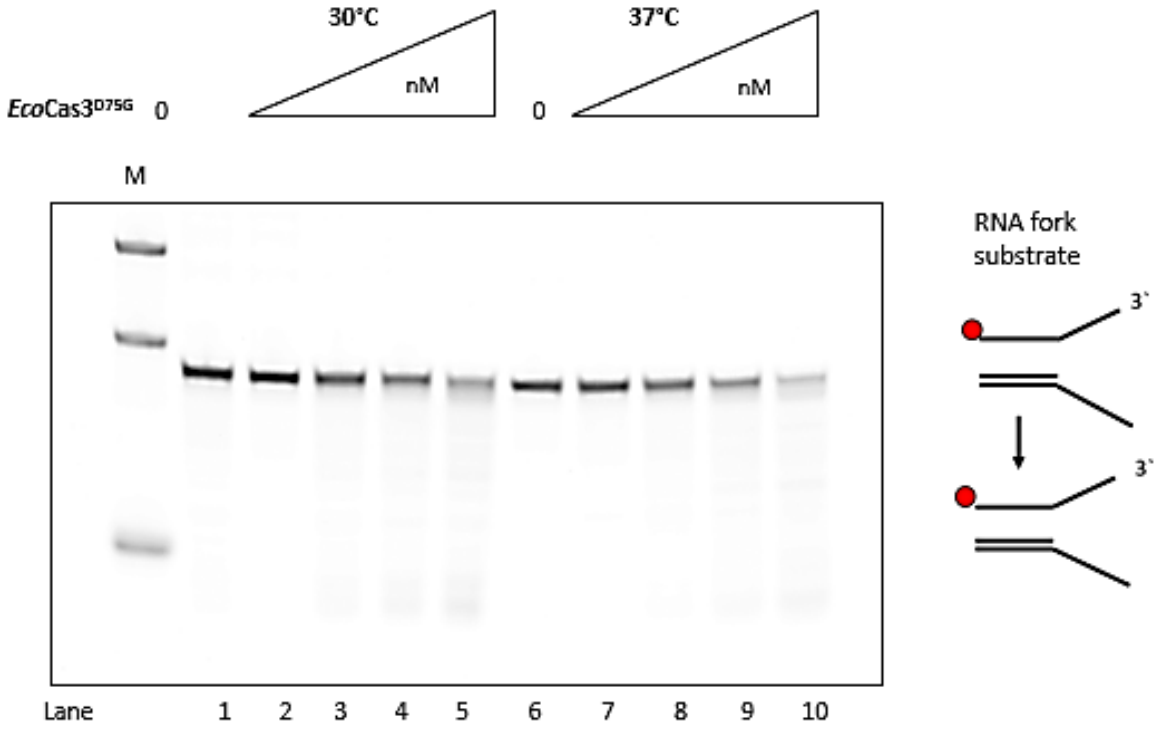


Fig.4.5 *cont*.

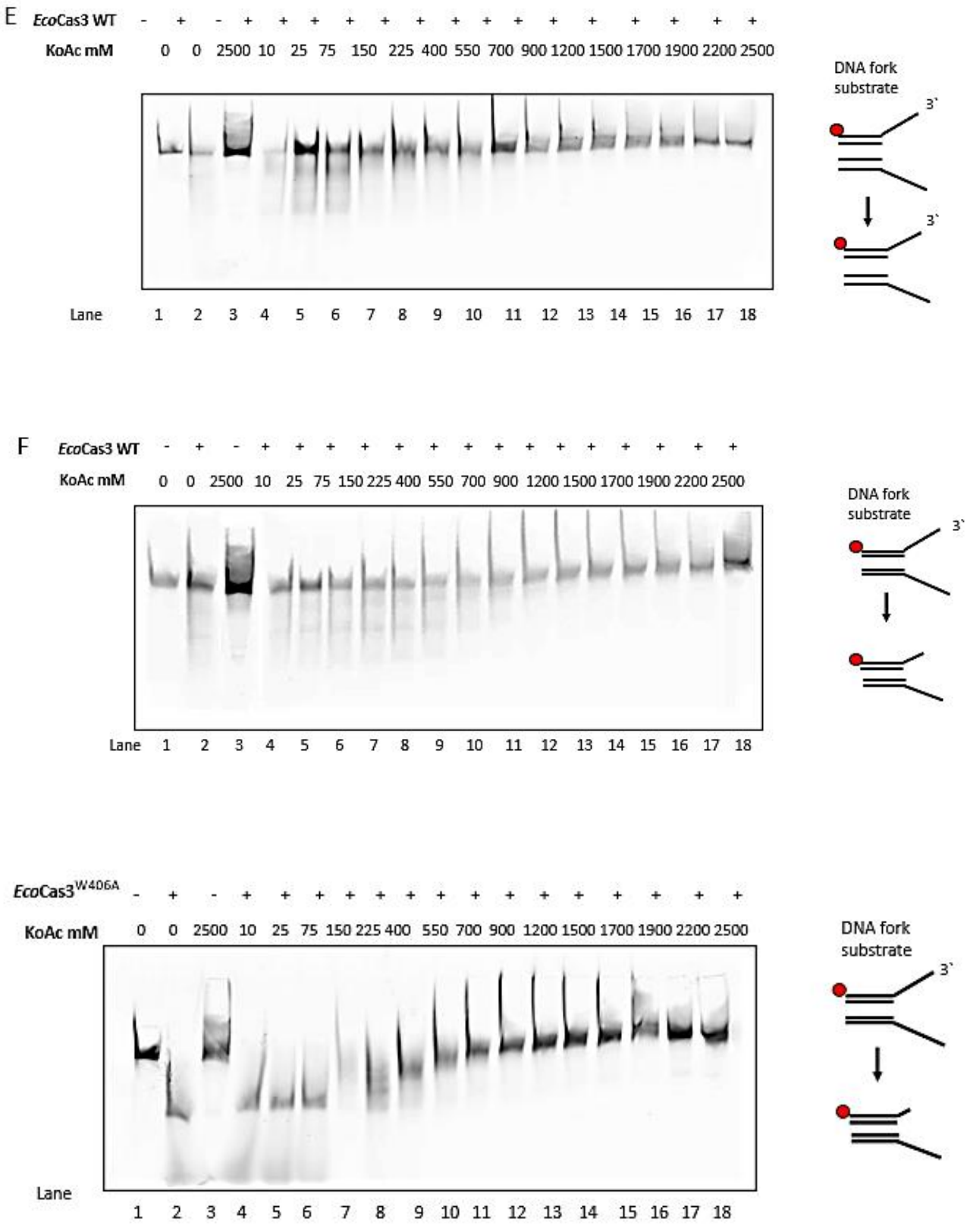
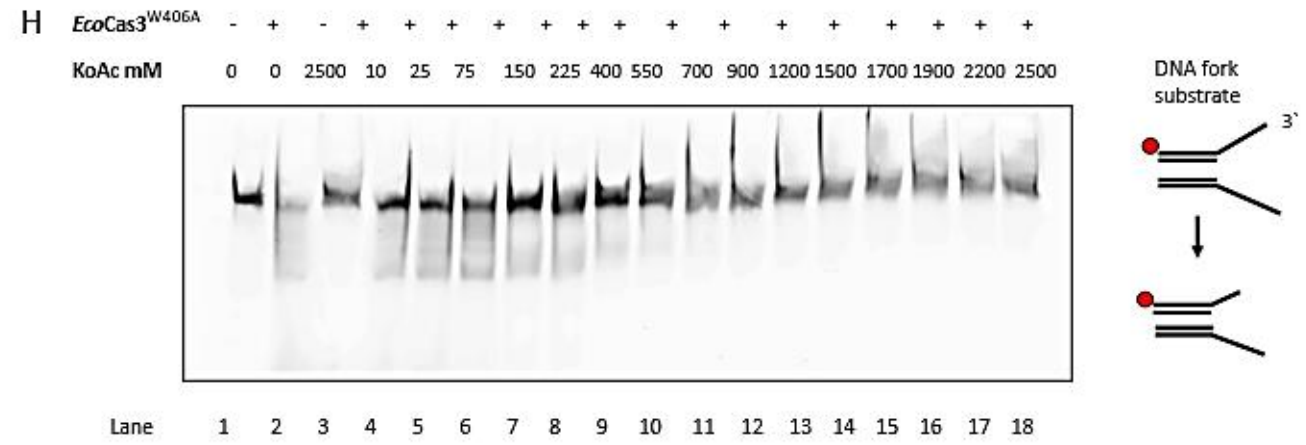

E
EcoCas3 WT - + - + + + + + + + + + + + + + + +
KoAc mM 0 0 2500 10 25 75 150 225 400 550 700 900 1200 1500 1700 1900 2200 2500
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
DNA fork substrate
3`
3`
F
EcoCas3 WT - + - + + + + + + + + + + + + + + +
KoAc mM 0 0 2500 10 25 75 150 225 400 550 700 900 1200 1500 1700 1900 2200 2500
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
DNA fork substrate
3`
G
EcoCas3W406A - + - + + + + + + + + + + + + + + +
KoAc mM 0 0 2500 10 25 75 150 225 400 550 700 900 1200 1500 1700 1900 2200 2500
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
DNA fork substrate
3`

Fig. 4.5 *cont.*


H
EcoCas3W406A - + - + + + + + + + + + + + + + + +
KoAc mM 0 0 2500 10 25 75 150 225 400 550 700 900 1200 1500 1700 1900 2200 2500
DNA fork substrate
3`
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

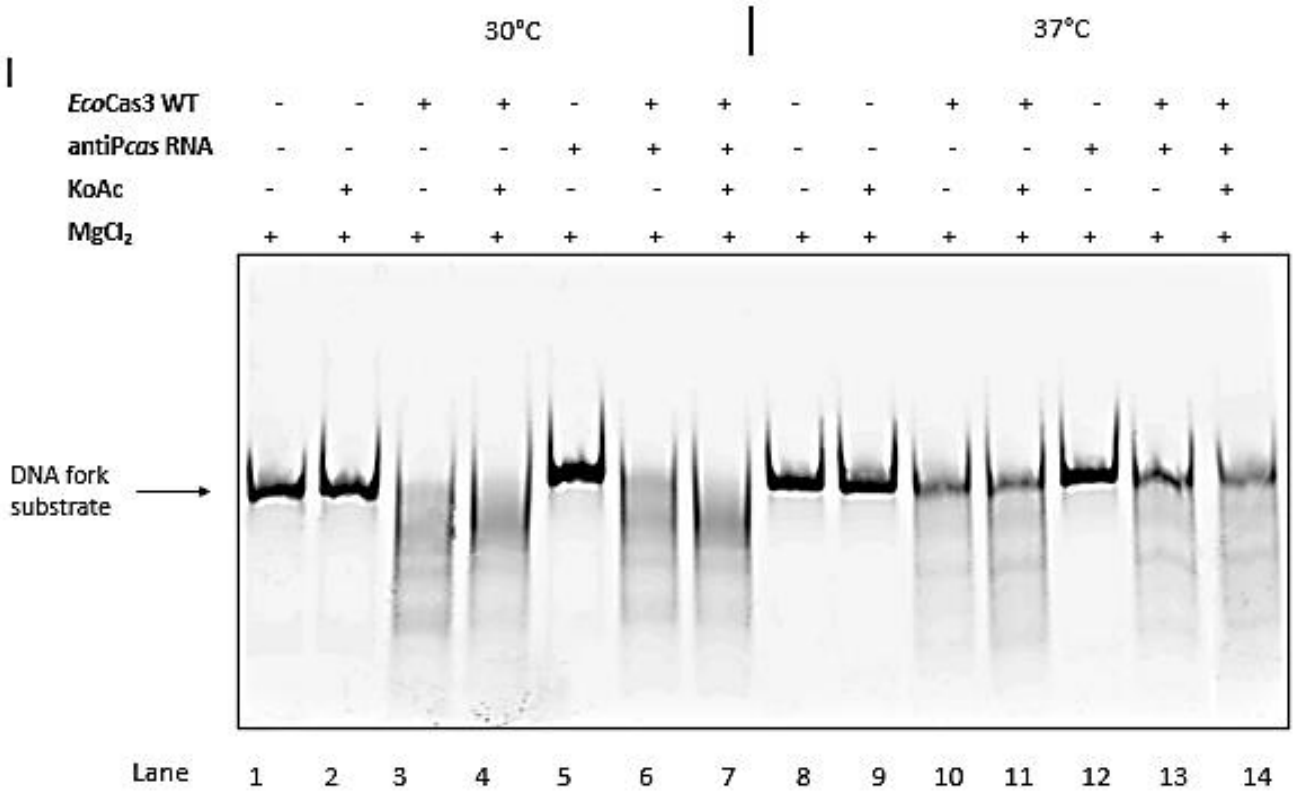

I
30°C
37°C
EcoCas3 WT - - + + - + + - - + + - + +
antiPcas RNA - - - - + + + - - - - + + +
KoAc - + - + - - + - + - + - - +
MgCl2 + + + + + + + + + + + + + +
DNA fork substrate
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14

Fig 4.5 *cont.*

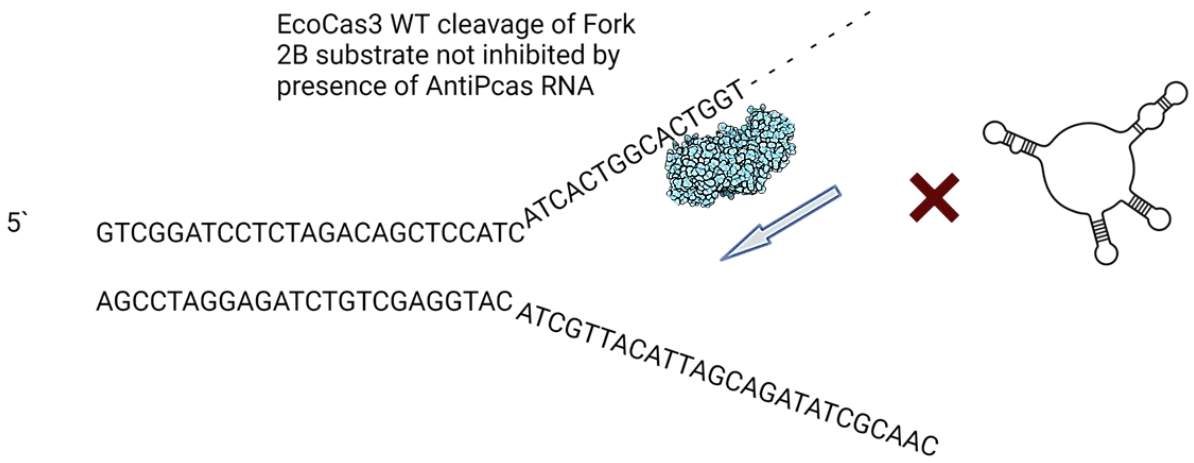


Figure 4.5 **Nuclease assays on dsRNA and dsDNA Fork substrates**

(A-H) Preliminary assays with *Eco*Cas3 WT, Cas3^{W406A}, and Cas3^{D75G} proteins on Cy5 Fork 2B DNA and RNA substrates (I) Fork 2B DNA substrate cleavage by Cas3 WT protein in reactions with and without the addition of 25nM anti-*cas* RNA. All reactions were initiated upon the addition of Buffer O (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 100mM NaCl, 0.1 mg/mL BSA, 20mM DTT), and stopped by adding 50mM Tris pH 8.0, 100mM EDTA, 5 mg/mL Proteinase K and 1% (*w/v*) SDS, followed by incubation for 15 minutes at 30 and 37°C. RNase T1 was added at a volume of 15µL, followed by further incubation at 37°C for 30 minutes. Anti-*cas* RNA caused no visible inhibition of Cas3 nuclease activity. The corresponding illustration depicts this reaction.

4.3.1.3 Helicase unwinding assay

*Eco*Cas3 possesses helicase activity that originates from the HD structural domain, a conserved feature inherited from super-family 2 helicases. This domain enables double-stranded oligonucleotide unwinding, as has been shown in previous research on RNApol molecules and R-loop dissociation [24-25]. It was proposed that the addition of anti-*cas* RNA may have an additional modulating effect on Cas3 helicase activity of duplex substrates, therefore a preliminary assay was performed to the determine the baseline strand unwinding activity of Cas3. Results showed no strand unwinding of the target Fork 2B DNA substrate by Cas3,

irrespective of the addition of complementary cold-trap DNA. Monomeric *Eco*Cas3 exhibits minimal helicase activity in functional *in vitro* assays, this has been reported in previous research[65]. Due to this, any modulating effect of the anti-*cas* RNA on Cas3 helicase activity would not be visible, therefore no subsequent assay was performed with the addition of anti-*cas* RNA. Alternative fork DNA substrates were not utilised to measure *Eco*Cas3 comparative strand unwinding, due to previous data showing Cas3 exhibiting similar minimal helicase activity [65].

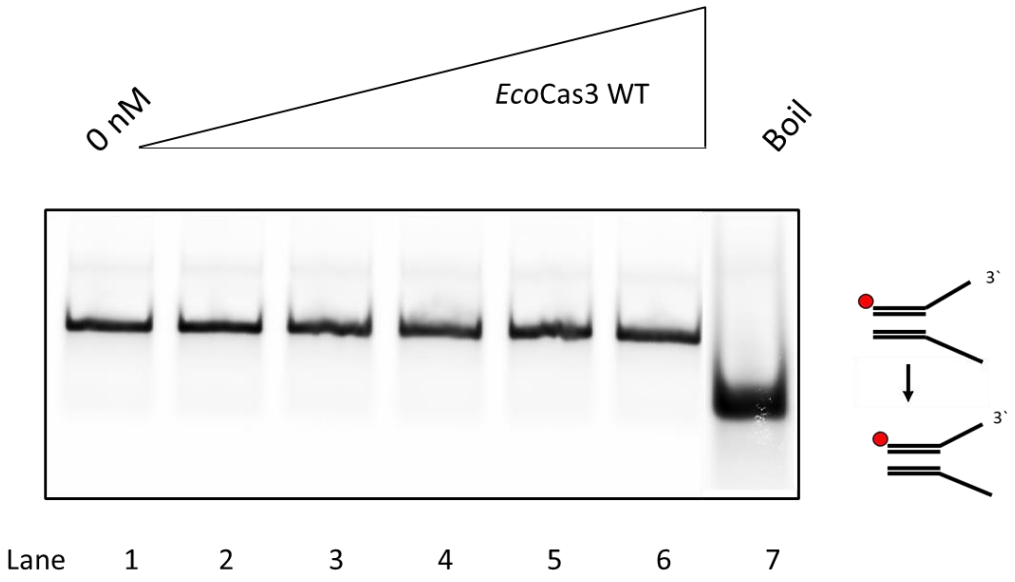


***Figure 4.6 Eco* Cas3 helicase unwinding assay on dsDNA Fork 2B**

10% Native PAGE gel showing Cas3 helicase activity on Fork2B . *Eco*Cas3 WT was preincubated with 25nM of Fork2B substrate at an increasing gradient (0nM, 10nM, 25nM, 50nM, 100nM, 200nM) at 37°C. A reaction master-mix containing 25nM Fork2B DNA, 5mM $MgCl_2$, 5mM ATP, 5mM DTT and 1X "helicase buffer" was used to initiate Cas3 activity, with 2.5µM of cold-trap MW12 DNA immediately added before reaction initiation. After 30 minutes incubation, reactions were quenched by 1X stop solution containing Proteinase K. *Eco*Cas3 WT exhibited no visible strand unwinding activity on target Fork 2B substrate, (Lanes 2 - 6) compared to the DNA-only control.

4.3.2 Anti-*cas* effect on *cas3* mRNA transcription

Computational analysis yielded a complex secondary structure of the anti-*cas* RNA, which exhibited similarity to that of bacterial riboswitches, thus formulating a research question of

whether the RNA was capable of inhibiting the nascent transcription of *cas3* mRNA if bound to cellular metabolites such as Mg^{+}. Template DNA from the pDM13 anticas plasmid was amplified using PCR, and served as a defined span of 407bp in length containing the *cas3* gene present on the plasmid, and also included the complete antiP*cas* gene sequence with the upstream promoter. Primers were used to ensure the length of the *cas3* segment in pDM13 plasmid was enough to be viably inhibited by the addition of anti-*cas* RNA, the *cas3* mRNA transcript length is 204 nt, whereas the anti-*cas* RNA is 159 nt in length.

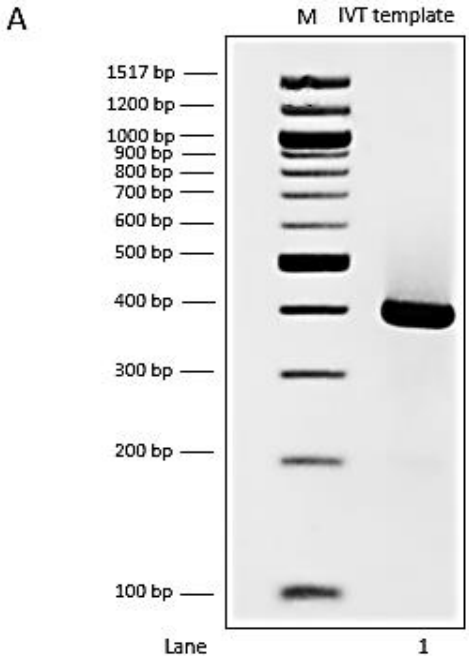


***Fig 4.7** cont.*

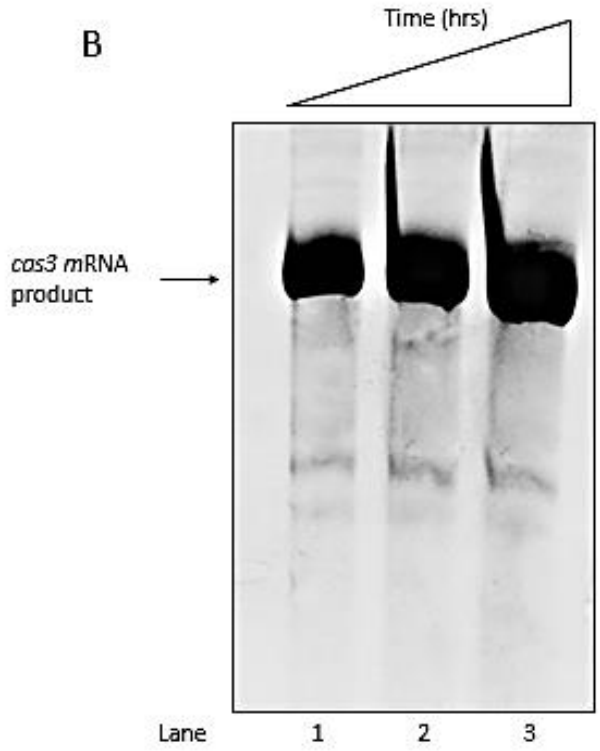
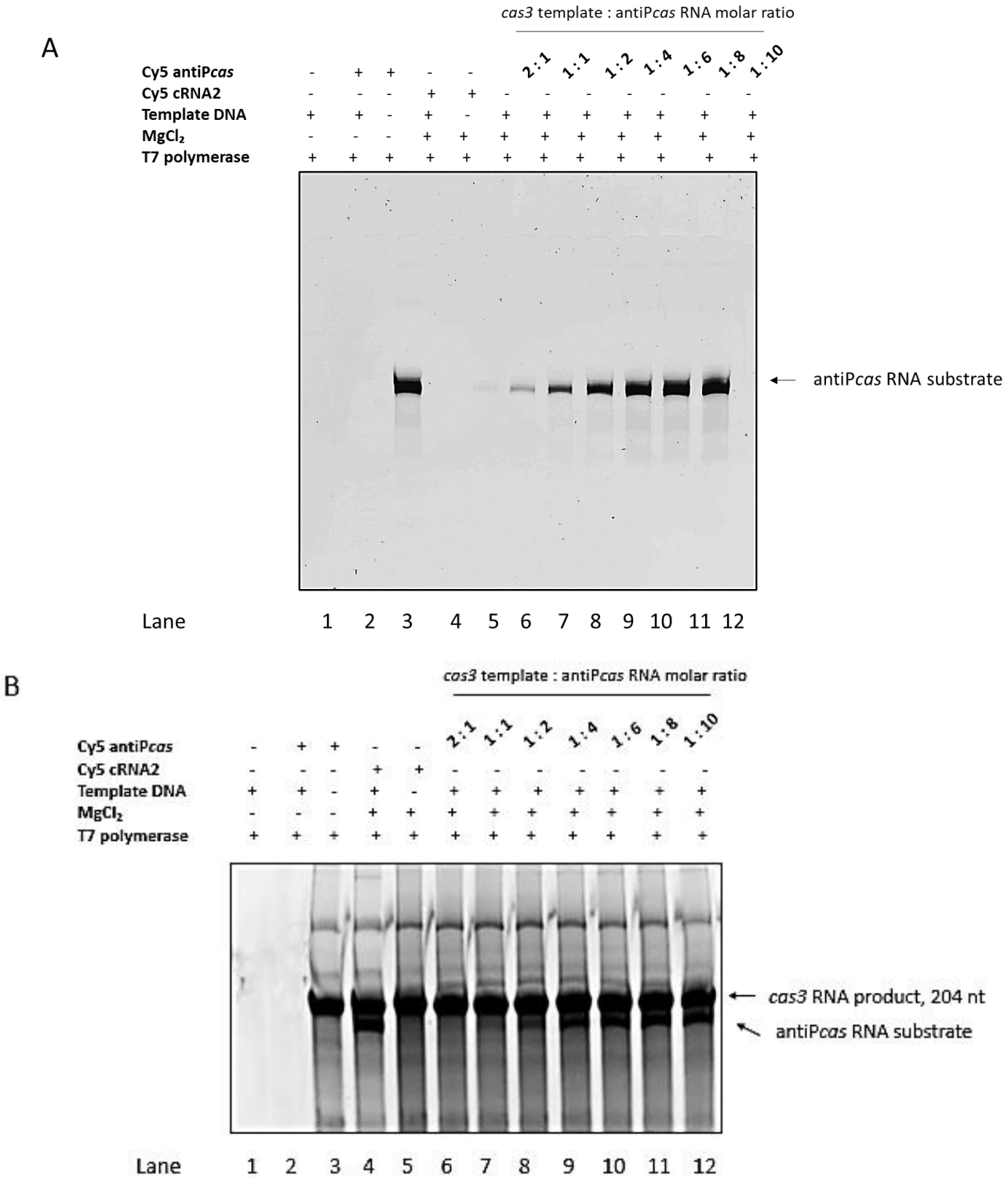


***Figure 4.7* Amplified PCR product of DNA template sequence**

(A)*cas3* transcription template successfully amplified by PCR. Reactions mixed with 1X Purple Gel loading dye (NEB #B7025S) prior to loading in a 2% TBE agarose gel for electrophoresis. Approximate size of the amplified template is 407 bp (B)Reactions contained template DNA only, incubated at 37°C for different durations to determine the optimal incubation period, 1hr to 2hrs 30 minutes. This was done to ensure transcription products were visible after an extended time period, as the T7 polymerase may displace the RNA in reactions with prolonged incubation [45]

Reactions were prepared on ice prior to incubation at 37°C for a duration of 45 minutes. Control reactions contained template DNA only, anti-*cas* RNA, and a Cy5 RNA control termed cRNA2 preincubated with template DNA without the addition of $MgCl_2$ or T7 polymerase respectively. After incubation, the assay was quenched upon the addition of RNase T1 and DNase 1 and incubated at 37°C for a further 15 minutes. Reactions were then immediately loaded for electrophoresis through a 5% denaturing PAGE gel. Cy5 labelled anti-*cas* RNA was present in a clear gradient of band intensity, corresponding to the molar gradient when titrated in (*Figure* 4.8 A). The *cas3* nascent mRNA migratory position was visualised by post-

staining with SYBR Gold nucleic acid stain and UV imaging. *cas3* mRNA transcription was visibly not inhibited by the presence of anti-*cas* RNA in an increasing molar gradient (*Figure* 4.8 B).



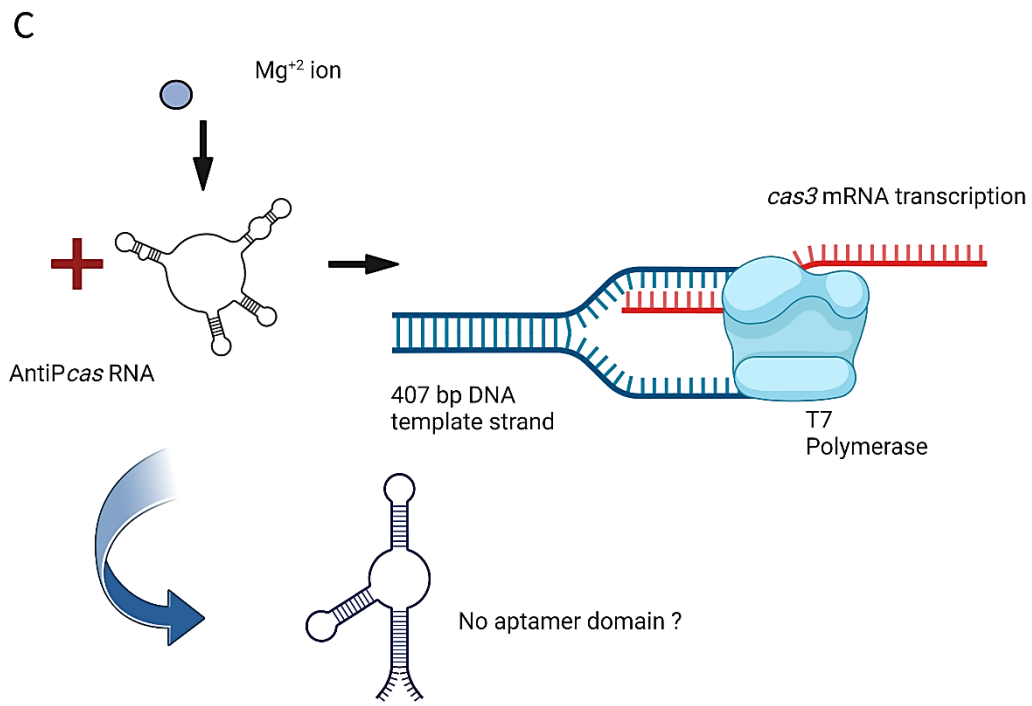


***Figure 4.8 cas3* mRNA transcription was not inhibited by anti-*cas* RNA**

(A) 5% denaturing PAGE gel showing concentration gradient of Cy5 labelled anti-*cas* RNA substrate (B) SYBR Gold stained gel imaged under UV light showing nascent transcribed unlabelled *cas3* mRNA transcripts after gel migration. *cas3* mRNA is approximately 204 nt in length. Lanes 4-5 show Cy5 labelled control cRNA2 substrate at a high position on the gel after migration. ssRNA ladder not visible after SYBR stain due to poor migration (C) Illustration showing process of *cas3* RNA transcription, not inhibited with anti-*cas* RNA present and with $MgCl_2$ ions.

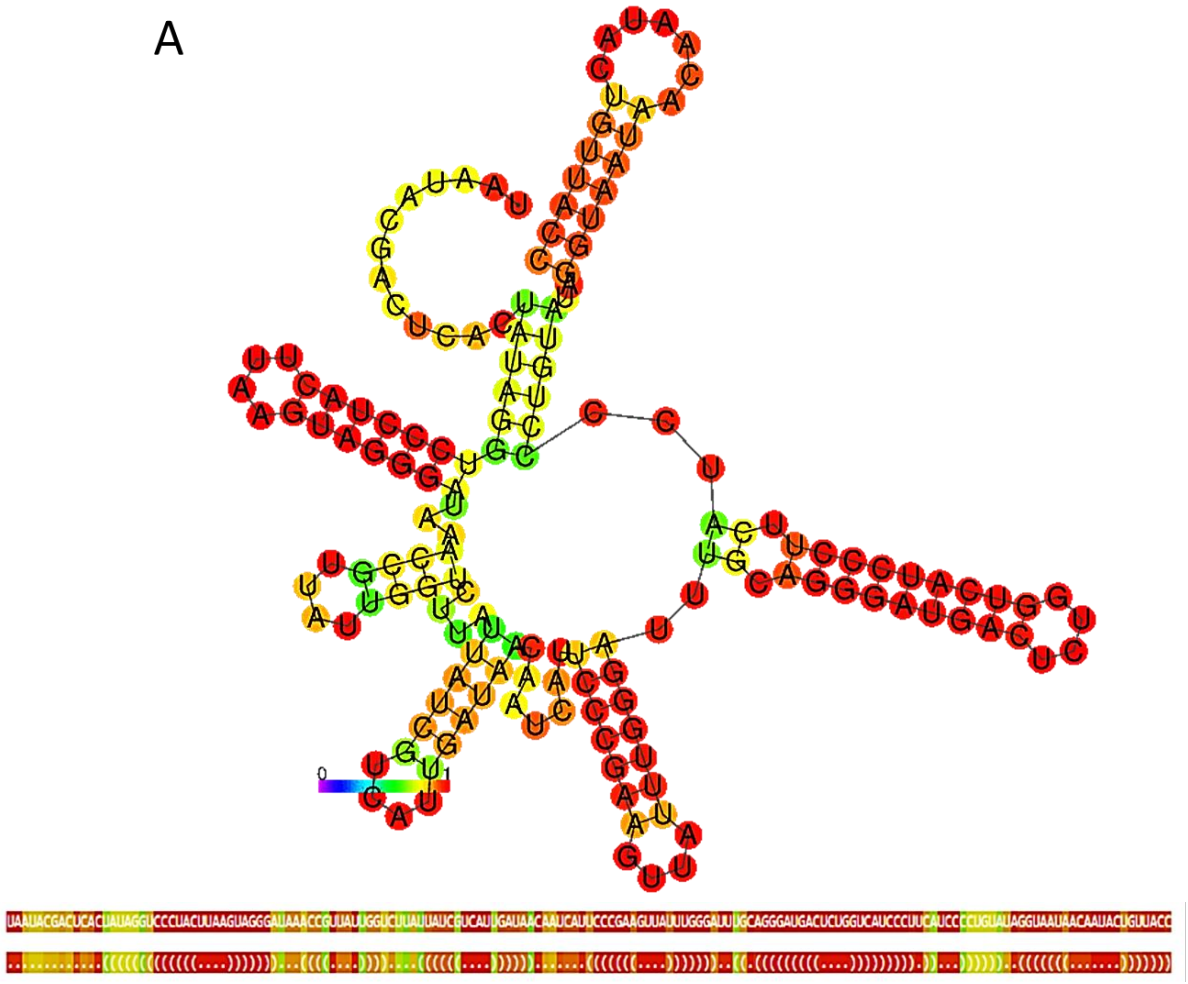
This result offers some insight into the structural capability of the anti-*cas* RNA, indicating that it is unable to inhibit mRNA transcription in the presence of Mg^+ ions, a behaviour not characteristic of riboswitches. Unimpeded *cas3* mRNA synthesis may additionally have occurred due to a weak binding affinity of the anti-*cas* RNA to the DNA template, highlighting the need for a prolonged incubation period in order for optimal binding to occur.

4.4 Structural modelling and MSA

Characterisation of the anti-*cas* RNA and comparative analysis with known regulatory RNA species was performed using bioinformatics. A full sequence of the anti-*cas* RNA was used to generate a minimum free energy (MFE) model of the folded secondary structure, using the Vienna RNA FoldHacker software programme as previously described in Chapter 2. Models were also made of high quality well-annotated sequences from classes of bacterial ribozymes, retrieved from the Rfam database. Interspecies comparison of ribozyme secondary structures with that of the anti-*cas* RNA showed commonly shared hairpin loops, suggesting a degree of sequence orthology and conservation of looped domains.

Multiple sequence alignment (MSA) analyses were performed using this pool of ribozyme sequences with the anti-*cas* RNA to further determine sequence alignment and identity, using the Clustal Omega software. Results showed the anti-*cas* RNA displayed sequence homology to top hits of RNase P class B, hatchet, twister P5 and twister-sister ribozymes. The sequence conservation and identity percentage scores for each class are summarised in table 23. Additionally, motif site detection in the anti-*cas* RNA sequence identified a conserved motif present in one of the hairpin loop domains, defined as an Sm binding site. This motif enables the association of the RNA with Sm proteins, necessary for pre-mRNA splicing, as is found in catalytically active self-cleaving ribozymes. The motif is primarily found in eukaryotic ribozymes, but conserved sites are present in bacterial ribozymes.

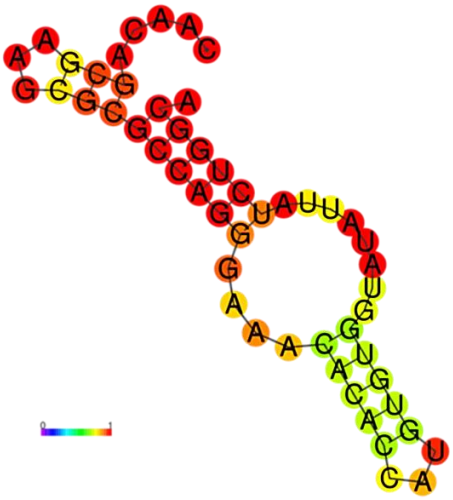
A



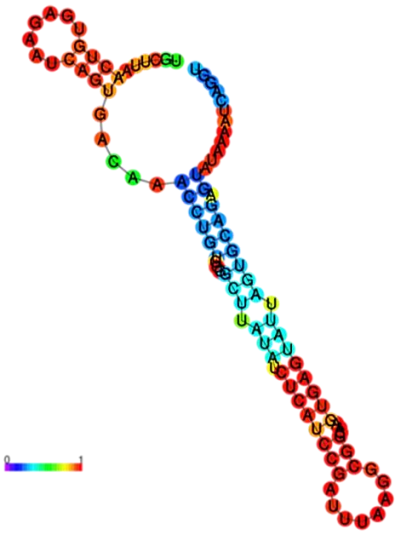
Hairpin

Hatchet

B I

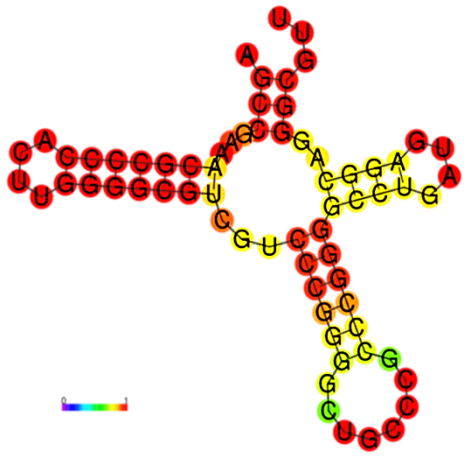


II



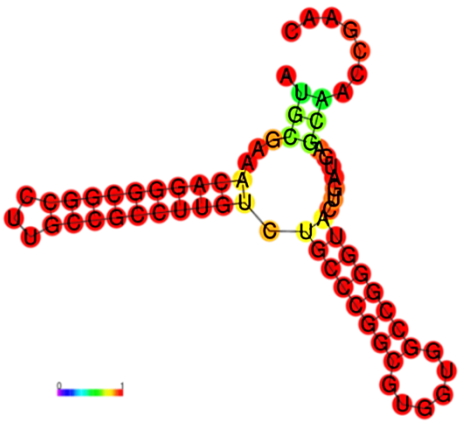
Hammerhead type II.1

III



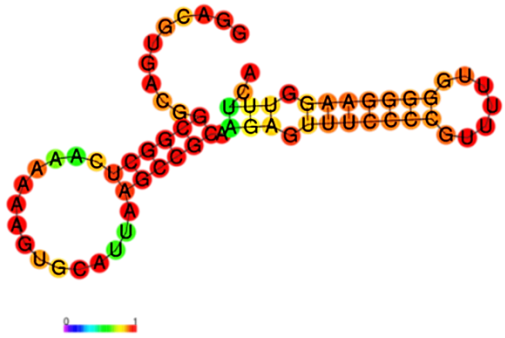
Hammerhead type II.2

IV



Pistol

V



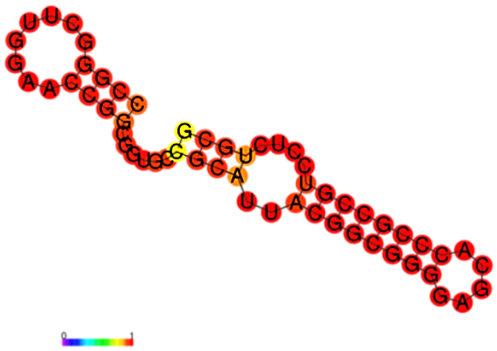
Ragath13

VI

Fig 4.9 *cont.*

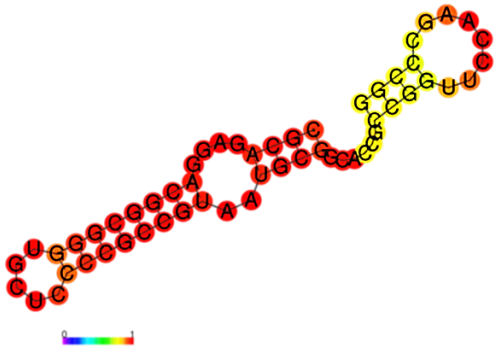
Twister P5

VII



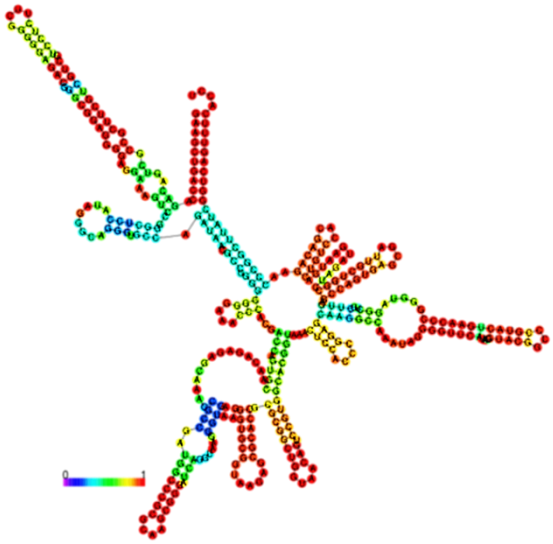
Twister sister

VIII



Rnase P class A

IX



Rnase P class B

X

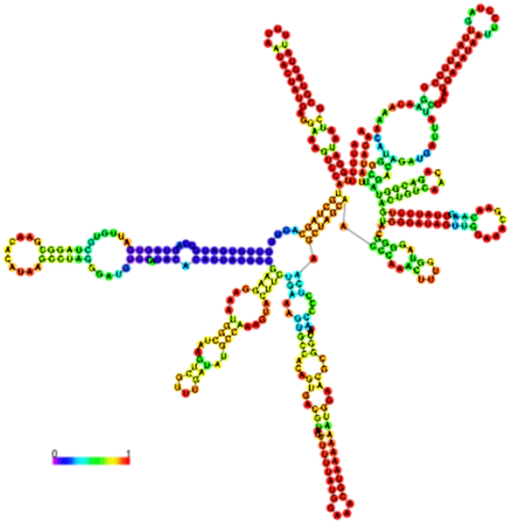
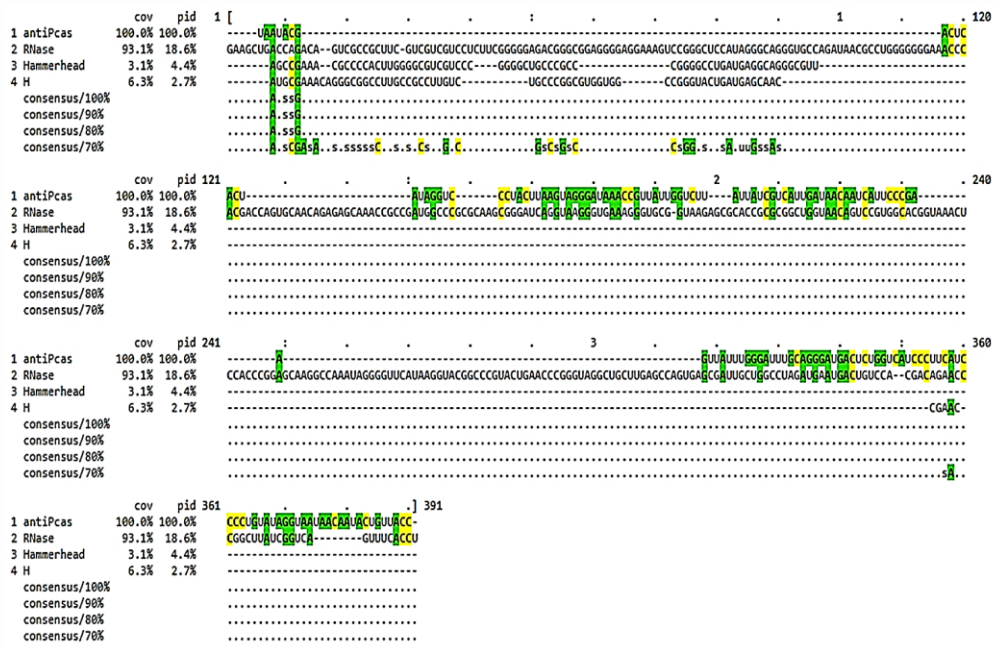


Fig 4.9 *cont.*

C

Key:
2. RNase P
3. Hammerhead type II
4. Hammerhead type II .1



D

Key:
8. RNase P class B

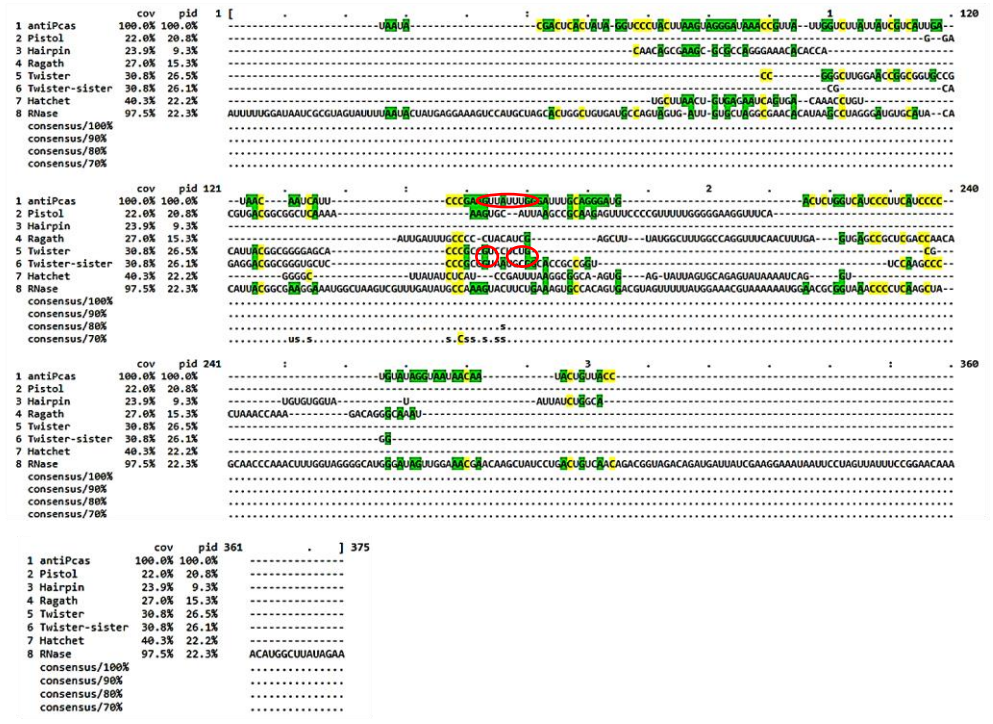
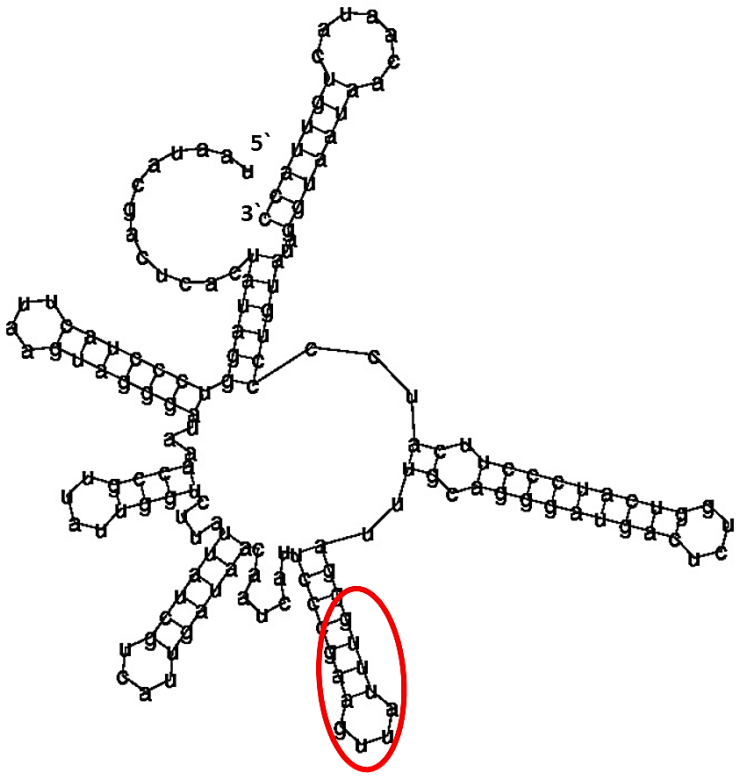


Fig 4.9 *cont.*

E



5` uaauacgacucacuauaggucccuacuuaaguagggauaaaccguuauug 50
gucuuauuaucgucauugauaacaaucauucccgaaguuauuuggggauuu 100
ugcagggaugacucuggucaucccuucaucccccuguauagguaauaacaau 150
acuguuacc 3`

Figure 4.9 **Anti-*cas* secondary structure and homology to regulatory catalytic RNA molecules**

(A) MFE (Minimum Free Energy) model of the anti-*cas* RNA secondary structure, indicating a high statistical probability of hairpin loops (B i-x)Statistical modelling of well-annotated bacterial ribozyme classes showed high probabilities of similarly well-defined hairpin loops (C-D) Clustal Omega MSA analysis with bacterial ribozymes showed the anti-*cas* RNA exhibited a high level of sequence homology with the RNase P class B ribozyme, in addition to Twister and Hatchet species. Twister and Twister-sister sequences exhibited partial homology with conserved site in anti-*cas* sequence (E) RNA Analyzer software used to identify conserved motifs in the anti-*cas* RNA sequence. Results showed a basic model of the folding structure of the RNA sequence, and identified an uracil-rich conserved SMSITE motif. This motif is typically found in catalytic self-cleaving RNA families, and is key for enabling RNA splicing.

Table 23. Anti-*cas* sequence id% and conservation with top hit ribozymes

Ribozyme	Conservation %	ID %
RNase P class B	93.1	18.6
Hatchet	40.3	22.2
Twister P5	30.8	26.5
Twister sister	30.8	26.1

4.5 Phylogeny

The homology and sequence similarity of the Anti-*cas* RNA molecule to well-annotated classes of bacterial ribozymes was also analysed using MEGA11 phylogeny software, using an input pool of previously used ribozyme sequences in FASTA format.

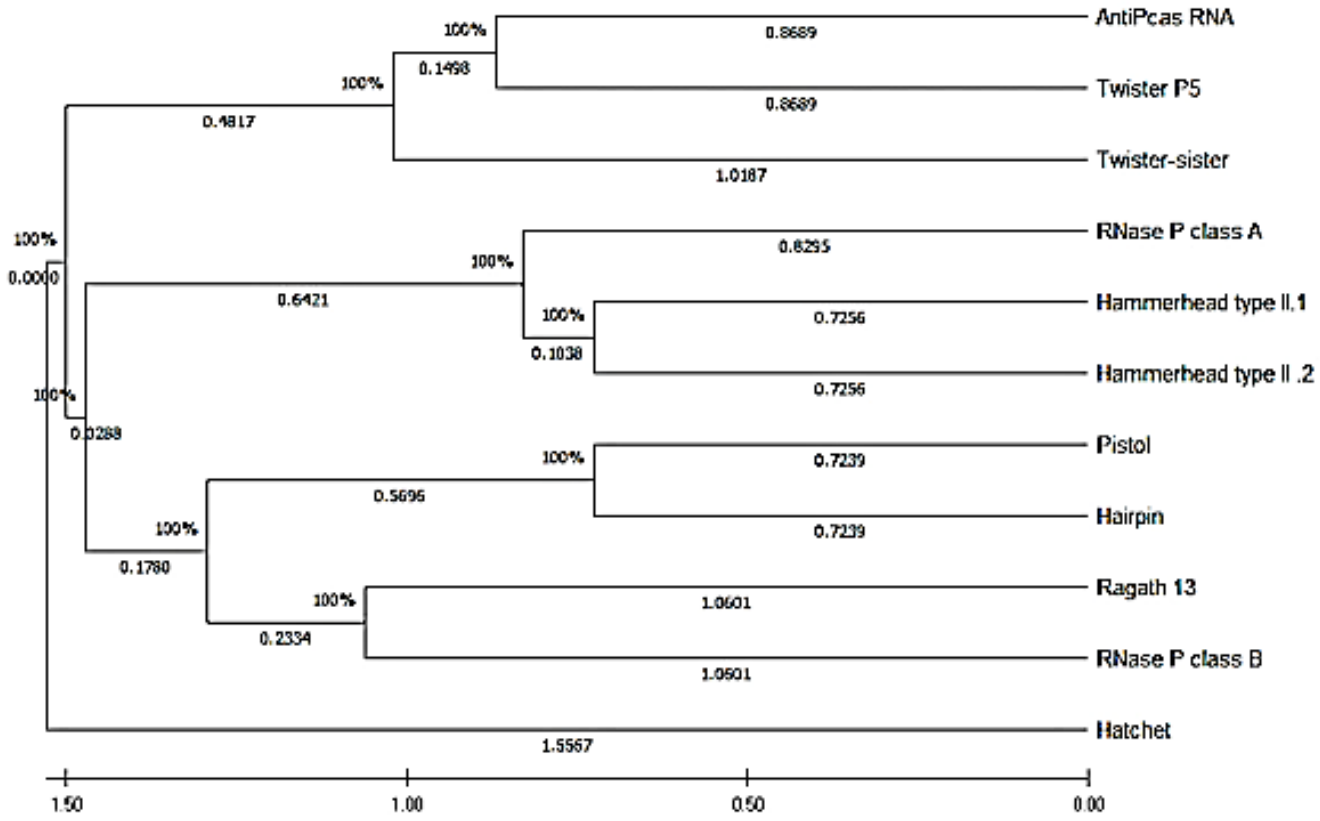


Figure 4.10 **Phylogeny of the anti-*cas* RNA to bacterial ribozymes**

Phylogenetic tree of maximum likelihood showing anti-*cas* homology to bacterial ribozymes, generated using the MEGA11 software package. The evolutionary history was inferred using the UPGMA method. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown above the branches. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method, and are in the units of the number of base substitutions per site. This analysis involved 11 nucleotide sequences, all positions containing gaps and missing data were eliminated. There were a total of 52 positions in the final dataset, the mean value of evolutionary divergence was 2.79

4.6 Summary

Biochemical analyses of the interplay between the anti-*cas* RNA and *Eco*Cas3 endonuclease show evidence of a stable bound complex formation after conducting electromobility shift assays. Minor band shifts in the RNA baseline position are visible in tandem with increasing *Eco*Cas3 concentration, suggesting at an affinity between the RNA and protein.

We conclude no evidence for any modulation on Cas3 nuclease activity exerted by the anti-*cas* RNA, this is in contrast to data from preliminary analyses showing dual cleavage of DNA and dsRNA substrates by *Eco*Cas3. It was hypothesized that competitive inhibition of Cas3 nuclease activity on the Fork 2b dsDNA substrate would occur due to Cas3 cleaving the anti-*cas* RNA in addition, thus occupying the ssDNA binding channel at the iHDA1 interface. Reasons for why this activity was not seen will be explored in Chapter 5.

It was found that *Eco*Cas3 exhibits minimal helicase unwinding activity on duplex DNA substrates, as has been previously tested in[64,] and described in detail in the PhD thesis of Liu He [65].

Nascent *cas3* mRNA transcription was not inhibited by the presence of anti-*cas* RNA when incubated with $MgCl_2$. This data yields evidence showing that anti-*cas* RNA does not behave similarly to a riboswitch.

A series of computational analyses performed on the anti-*cas* RNA sequence yielded further insights into its homology with known catalytic regulatory ribozyme sequences present across bacterial genomes. Structural modelling analyses identified distinctive domains in the anti-*cas* secondary structure that exhibited close similarity to those modelled in a set of known ribozymes. Sequential MSA analyses performed using the anti-*cas* RNA sequence with ribozymes prevalent in *Escherichia* species and subsequentially in Gramm-negative bacteria showed sequence similarity and conservation with the Twister P5, Twister-sister, Hatchet, and RNase P class B classes.

To conclude, the anti-*cas* RNA molecule has minimal influence on *Eco*Cas3 nuclease-helicase *in vitro* functioning and was unable to inhibit *cas3* mRNA transcription. The RNA exhibited sequence and structural similarity to bacterial ribozyme classes, suggesting a putative catalytic function and role in pre-mRNA synthesis. This can be verified with further robust experimentation, and will be discussed in Chapter 5.

Chapter 5: Discussion and future research

The aims of this study were to elucidate the function and activity of the anti-*cas* RNA molecule and RNA-encoding gene on the CRISPR-Cas effector Cas3 and in CRISPR immunity, using both biochemistry and genetics. Significant data generated during the course of this research could then improve our current understanding of how Cas endonucleases are regulated, therefore leading to a greater understanding of ncRNA activity and allowing for the potential optimisation of Cas3-mediated gene editing. Experimentation with the NK001 strain enabled robust speculation about the wider role of the RNA-encoding gene in cellular physiology and phage defence, and served as a model for testing additional research questions. In addition to the main findings reported here, future experimentation has been discussed with an aim to uncover more about the post-transcriptional modulation of *cas3* and impact on CRISPR immunity by antiP*cas*.

5.1 Impaired cell growth in the anti-*cas* deletion strain

The generation of a novel MG1655 strain with the RNA gene deleted enabled further investigation into the role of the antiP*cas* gene in normal cell growth and viability. It was found that NK001 cells exhibited an impaired growth rate in nutrient medias containing M9 and sugar carbon sources, however growth was not impaired in LB media.

The significantly slower growth of NK001 cells in M9 medias supplemented with different sugar-based carbon sources may indicate impaired cellular metabolic channels and pathways. The intact antiP*cas* gene in WT cells may therefore play a role for maintaining active cell metabolic pathways. When the antiP*cas* gene was restored to NK001 via a plasmid vector, no complementation with the WT phenotype was observed, as slower growth in M9

supplemented medias was recorded over 16 hours, as seen in *Fig.3.2* panel E. Genes introduced via a plasmid however do not always restore a WT phenotype, as complementation can be impacted in cells due to the increased ampicillin concentration in the peptidoglycan layer and associated repression. To counter this, all medias were supplemented with a final concentration of L-arabinose to induce antiP*cas* expression.

Notably, NK001 cells exhibited a strong growth phenotype in M9 media supplemented with maltose as a primary carbon source. This is in stark contrast to WT MG1655 cells, and when the antiP*cas* gene was restored and induced via plasmid delivery, no complementation was measured. A proposed explanation for this distinctive cell phenotype is given in the following section on the physiology of NK001 cells.

5.2 CRISPR immunity and NK001 cellular physiology

Novel findings were gained from conducting plaque sensitivity assays with WT MG1655 strains and the NK001 anti-*cas* deletion strain generated during the course of this research study. The identification of a potential regulatory RNA molecule in the *E. coli* CRISPR locus could play a key role in modulating CRISPR immunity, and enable further elucidation into the gene regulatory activity of ncRNA molecules. Through the generation of a novel knockout strain with the absence of the RNA-encoding gene, it was shown that antiP*cas* exerts a physiological control over the host cell morphology and metabolism, which manifests primarily as a more robust response to phage predation compared to WT MG1655 cells.

The phage robust phenotype exhibited by NK001 cells is a distinctly physiological response, not a change to CRISPR immunity resulting from a mutation in the CRISPR interference mechanism. This can be corroborated from experimental data from collaborators at the University of Zagreb. The deletion of a gene with regulatory roles in cellular physiology may

lead to spontaneous mutations occurring in cellular structures, such as membrane and cell membrane receptors that would render the cell more vulnerable to invasion by phage. Increased permeability of cell membrane can also be triggered by changes in membrane receptors as presented below in table 22.

Table 22. *E. coli* K12 cell membrane receptors Stochastic mutations in the LamB membrane receptors can alter the cell phenotype and lifecycle. Phage adsorption to the extracellular membrane may be impacted by mutations in the LamB receptor.

Receptor name and type	**Associated phage species**	**Mutation**	**Associated cell phenotype**	**Literature source**
LamB	λ, K10, StxPhil	Single amino acid base changes	Decreased permeability of phage DNA into the cell, increased maltose synthesis	Hazelbauer, 1975 Szmelcman and Hofnung, 1975 Roa, 1979 Watarai *et al.,* 1998

Previous research shows that morphological mutations in the LamB cell membrane receptor can impair the ability of λ*vir* MGE to bind prior to DNA injection into the cell, thus providing increased immunity to the host cell from invasion. Additionally, the study indicated the maltose metabolic synthesis pathway was positively impacted, due to LamB becoming enriched when cells were grown in maltose media[74,75,77].

Although a clear cellular phenotype resulting from the deletion of anti-*cas* was observed, the associated cellular pathway is uncharacterised .Robust conclusions from data on WT MG1655 and NK001 strains including collaborative work is that deletion of antiP*cas* could not affect the morphology of the LamB receptor and have no measurable activity on specific phage binding receptors.

5.3 Cas3 is able to form a weak complex with the stable anti-*cas* RNA substrate

Electrophoretic mobility shift assays performed as described in Chapter 4 show that *Eco*Cas3 is able to form a complex with the anti-*cas* RNA molecule, thereby answering a key question first proposed in the initial study that discovered the promoter [40]. Cas3 can bind to a variety of CRISPR-associated RNA molecules, however it has never been previously shown to associate with ncRNA. As discussed in Chapter 1, complexes of proteins with active ribozyme molecules can give useful insights into any impacts on protein functioning.

Future binding assays could be performed to control for anti-*cas* binding to other CRISPR-associated proteins, and to map the precise binding site interface of the RNA onto *Eco*Cas3. One useful control to employ in future EMSA assays could determine if the anti-*cas* RNA is able to bind to the purified *Eco*Cascade complex. As Cascade is a highly specific effector molecule, in theory the anti-*cas* RNA should be unable to form a bound complex, thus showing a strong binding affinity to Cas3. In addition, *in silico* simulation and modelling can aid with identifying singular binding sites on the RNA secondary structure to the Cas3 surface topology and helical chains in the secondary structure. Depending on which site the anti-*cas* RNA binds to, it may show an affinity for certain domains on Cas3.

5.4 The role of anti-*cas* RNA in Cas3-mediated interference

Preliminary assays to measure *Eco*Cas3 nuclease activity on dsDNA and dsRNA substrates yielded intriguing findings; Cas3 WT and Cas3^{W406A} were found to be able to cleave both nucleic acid substrates, showing no nucleotide specificity. The Cas9 endonuclease has been previously shown to be able to cleave RNA[8,25], but this is the first time *Eco*Cas3 has been

shown to do this. This led to a prediction that the ability of Cas3 to cleave dsDNA would be partially or significantly inhibited in the presence of the anti-*cas* RNA substrate, as both substrates would compete to access the ssDNA binding channel. The competitive inhibition of Cas nuclease activity on target invasive DNA has been observed by the action of oligomers and antiCRISPR proteins on Cas9 in bacterial models, however any inhibition by an additional oligonucleotide sequence has not been tested.

After incubation with the anti-*cas* RNA substrate, no significant inhibition of Cas3 nuclease activity on the target dsDNA substrate was observed. Anti-*cas* RNA may have exerted a very weak effect, but not great enough to be clearly visible. This result contrasts with the efficacy of Cas3 cleaving control RNA oligos. This may be explained by the potency of nuclease activity and strand shearing, acting faster on ssRNA oligos than dsDNA replication forks, thus resulting in no inhibition of DNA cleavage.

A study using *S.aureus* Cas9 endonuclease showed reduced cleavage of dsDNA target strands upon the addition of a model RNA molecule[27], the efficacy of competitive inhibition via ssDNA channel binding may therefore depend on the amino acid composition of the nuclease active site and structure of Cas endonucleases. Further research could be performed with mutant Cas9 and Cas3 proteins with deformed binding channels to investigate impaired substrate cleavage, or the use of a synthetic RNA molecule to investigate binding affinity when pre-incubated with Cas proteins.

5.5 Anti-*cas* does not display similar behaviour to riboswitches and was unable to inhibit *cas3* mRNA transcription

The initial characterisation of the antiP*cas* gene prompted conjecture into its functional capabilities, based on the elaborate secondary structure of the ncRNA sequence. A hypothesis posed by the authors of this study [40] stated that a stable anti-*cas* RNA transcript may inhibit the transcription of nascent *cas3* mRNA, which would indicate that the RNA can behave similarly to a riboswitch. As discussed in Chapter 1, in the presence of cellular metabolites such as Mg^{+2}, a functional riboswitch undergoes a conformational change that inhibits nascent mRNA transcription at the 5` UTR.

In the presence of $MgCl_2$, the anti-*cas* RNA was unable to terminate the transcription of *cas3* mRNA, when titrated in an increasing concentration. The inability of the anti-*cas* RNA molecule to inhibit *cas3* transcription may be explained by computational analysis showing that the RNA molecule exhibits structural similarity and shared sequence conservation to known bacterial ribozyme species, which are unable to inhibit transcription.

Further characterisation of the anti-*cas* RNA could be performed by using sequence detection software to identify a conserved structural aptamer domain, which is characteristic of riboswitches. The inability of the anti-*cas* RNA to inhibit *cas3* transcription would be explained by the absence of an aptamer domain in the RNA molecule, as it would be unable to respond to ligand binding and thus undergo no conformational change leading to termination loop expression and inhibition of gene transcription.

The homology of the anti-*cas* RNA to characterised bacterial ribozymes can also be explored further. A more robust method of determining the *in vivo* activity of ribozymes is to conduct reporter assays, whereby the ribozyme sequence is fused to a “reporter” gene that changes activity once expressed as RNA. A loss of reporter RNA activity is correlated with a functioning ribozyme [62]. Anti-*cas* RNA exhibited conserved sequence homology with the Ribonuclease P

class B ribozyme, which catalyses tRNA cleavage and tRNA maturation in the presence of metal ions. This may indicate a role for the anti-*cas* RNA in interacting with innate cellular RNA molecules, and engagement in RNA strand scission. The anti-*cas* RNA also exhibited close evolutionary phylogeny to the Twister P5 species, that rapidly cleaves pre-mRNA sequences. Partial conservation with the SMSITE motif detected in anti-*cas* indicated a close homology with the bacterial Twister P5 ribozyme, however divergence due to speciation could account for base changes across the aligned sequences and in the motif site. The wider significance of a ribozyme in the CRISPR locus, in relation to their functionality, could entail roles in the splicing of pre-mRNA of *cas* genes, however this would need to be tested further.

In addition to the sequence similarity between anti-*cas* and characterised ribozymes, antisense RNA molecules are also unable to inhibit nascent mRNA transcription, this may serve as an additional reason why anti-*cas* is uninhibitory. Recent advancements in NGS and transcriptomics have identified high frequencies of short antisense RNA molecules interspaced throughout prokaryotic genomes, including *Escherichia* species, with known roles in affecting gene regulation in sense-orientated target genes [91]. Transcriptomic RNA inhibition assays conducted in species including *E. coli* showed no effect on adjacent protein-coding gene mRNA by target ncRNA molecules, in general correlation with the *in vitro* assay data of anti-*cas* RNA[91]. While some antisense RNA molecules can affect mRNA translation, for example in *Bacillus subtilis*, no counterpart has as yet been detected in *E. coli* [91,92].

5.6 Future work: examining the potential of the antiP*cas* gene to exert post-transcriptional modulation over *cas3* and CRISPR immunity

AntiP*cas* activity *in vivo*, interaction with H-NS and its homolog StpA

The DNA-binding protein H-NS silences the expression of the antiP*cas* promoter *in vivo*, along with *cas* gene transcription, resulting in the innate integration of foreign MGE DNA into the CRISPR locus. A study performed on CRISPR-associated proteins and expression of the CRISPR immune response also indicated the role of the H-NS homolog StpA in repressing transcription of *cas* genes[86]. StpA cofunctions to tightly regulate CRISPR-Cas expression via promoter control, and performs as a co-backup for H-NS in targeting related *cas* genes[86]. As H-NS has been found to control the expression of the antiP*cas* promoter, further research performed with the expression of StpA in MG1655 WT and antiP*cas* deletion strains could determine any post-transcriptional effect on antiP*cas* gene expression and subsequent modulation on Cas3 activity against invader DNA. This may result in an increased amount of crRNA, or lower in contrast to that purified from cells with activated P*cas*.

ColE1 *ori* over-replication by Cas3 may be reduced in NK001

Further post-transcriptional modulation of *Eco*Cas3 in CRISPR immunity by the antiP*cas* gene could be visible by Cas3 displaying an impaired ability to form R-loops on dsDNA, necessary in CRISPR immunity. This could be tested using genetics, in the NK001 strain, by recording yields of ColE1 *ori* based plasmids compared to purified yields from the MG1655 WT strain.

Data from plasmid yield assays detailed in [25] show the ability of Cas3 to associate with RNA polymerase molecules to over-replicate ColE1 *ori* DNA. The absence of the antiP*cas* gene may therefore impair the ability of *Eco*Cas3 to initiate R-loops catalysed by dual helicase unwinding and further nuclease activity, thus causing a reduced plasmid yield in NK001 cells transformed with ColE1 *ori* vectors. A mechanism for this may be via decreased Cas3 expression levels in the host cell, thus resulting in no pairing and association with RNAPol II enzymes to liberate

ori DNA. An illustration depicting a putative genetic pathway for this is represented below in *Figure* 5.1.

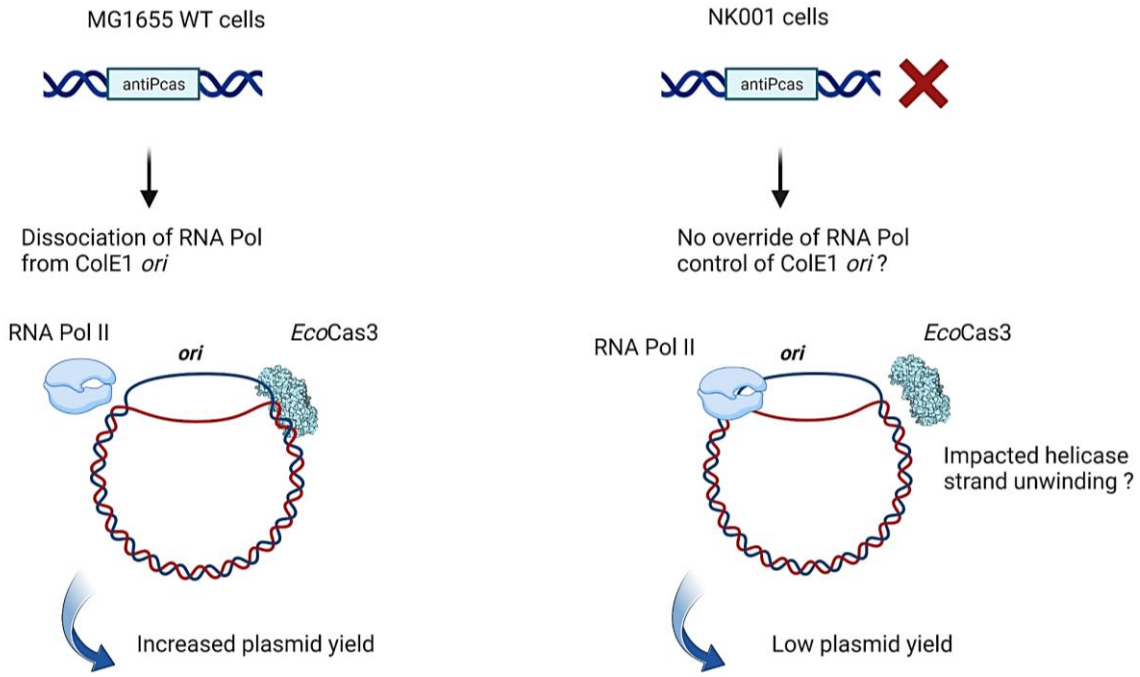


Figure 5.1 **Theoretical pathway by how deletion of the antiP*cas* gene results in post-transcriptional modification of Cas3 helicase unwinding of ColE1 *ori* DNA**

In WT cells, *Eco*Cas3 was able to increase ColE1 plasmid yield by strand unwinding at *ori* and dissociation of RNAPol II from the distal strand. In NK001, *Eco*Cas3 activity may be impacted due to the resulting modulation of *cas3* expression by the absence of antiP*cas*.

<u>AntiP*cas* impact on primed adaptation *in vivo*</u>

Experimental attempts to examine the effect of the overexpressed antiP*cas* gene on primed acquisition in live cells were not successful. This may have been due to impeded cell infectivity by the pNK01 construct, resulting from a low concentration ratio of the construct to the cells. Given sufficient time, the assay could have been repeated to control for this factor and others that may have arisen.

The potential of the antiP*cas* gene to impact CRISPR immunity was proposed due to its overlapping proximity to *cas3* and demonstrated *in vivo* functionality and control by H-NS.

The absence of the antiP*cas* gene may also impact normal rates of spacer acquisition into the host cell, potentially resulting in the inhibition of new spacer integration into the CRISPR array. This could be trialled using the NK001 strain transduced with a WT H-NS deletion strain, and then infected with λ*vir* or the pNK01 construct.

5.7 Conclusions

Through the course of performing this research, new insights into the regulation of ncRNA on the CRISPR immunity effector Cas3 have been uncovered, providing a basis for conducting further research into novel RNA interactions with Cas effector proteins and the optimisation of the *E. coli* Cas3 in biotechnology.

As reported in this work, the anti-*cas* RNA substrate and RNA-encoding gene were found to interact with the CRISPR-Cas3 endonuclease and influence a distinct physiological response to phage incursion and cellular metabolism in the host cell. Bioinformatics analysis showed the RNA displayed sequence characteristics of catalytically active ribozymes, which sheds light onto potential and as yet unmapped regulatory pathways. As previously stated, understanding the role of non-canonical regulatory genomic elements in CRISPR-Cas expression and functionality during CRISPR immunity is essential for identifying and exploiting novel ncRNA-based tools for future use.

Further research is required in order to determine how CRISPR-Cas immunity may be impacted by conserved ncRNA and functional promoters local to the CRISPR locus, I have fleshed out some ideas relating to this herein. As the field of CRISPR-Cas research, and synonymously research into the activity of ncRNA expands, it is promising that RNA-based tools can be developed to aid and enhance Cas-based gene editing, thus expanding the current toolkit.

Acknowledgements

I am extremely grateful to my supervisors Dr. Edward Bolt and Dr. Tom Killelea for their immeasurable support, guidance, and mentorship during the course of this research project. It was an amazing opportunity to be able to undertake original research in a highly acclaimed lab, and I feel very privileged to have done this and share ideas with brilliant scientists. I am also very grateful for the advice and guidance from Dr. Liu He on all things Cas3 related, and in-depth chats about science and lab work in general. I wish her the very best success in her research career.

I would also like to thank Dr. Ivana Ivančić-Baće for sharing her extensive knowledge on antiP*cas* and her willingness to stay in contact with me. I greatly appreciate her suggestions on experimental design with the NK001 antiP*cas* deletion strain, and also for sending over cell strains for the purpose of using them to study overexpression and primed adaptation. I also greatly appreciate the advice and insight from Damjan and his willingness to find time to speak to me during his thesis writing.

The supportive advice and research experience from Andy, Ryan, Beccy, and Ashley have helped me a lot and allowed me to become better at experimental planning, I am very thankful for their time. I wish them the very best of luck and success in their current PhD's and future careers.

I enjoyed sharing ideas with Fiona and Sabs and collectively suggesting advice whenever we hit a roadblock in our projects, teamwork is key.

Lastly, I am very thankful for the support of my family, and I'm humbled by their pride in what I've achieved.

References

1. Makarova, K. *et al.*, 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, **18**, 67-83.

2. Mojica, F. and Rodriguez-Valera, F., 2016. The discovery of CRISPR in archaea and bacteria. *The FEBS Journal*, **283**, 3162-3169.

3. Makarova, K. *et al.,* 2002. A DNA repair system specific for thermophilic Archaea and Bacteria predicted by genomic context analysis. *Nucleic Acids Research,* **30**, 482-496

4. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A., 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, **169**, 5429-5433.

5. Brouns, S. *et al*. 2008. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, **321**, 960-964.

6. Jinek, M. *et al*. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**, 816-821.

7. Barrangou, R. *et al.* 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, **315**, 1709-1712.

8. Godde, J., and Bickerton, A., 2006. The repetitive DNA elements called CRISPRS and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution*, **62**, 718-729.

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

10. Pontarotti, P., 2017. *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*. **3**. Geneva: Springer International Publishing.

11. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V., 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO Journal*, **30**, 1335-1342.

12. Gong, B., Shin, M., Sun, J., Jung, C., Bolt, E., van der Oost, J. and Kim, J., 2014. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences*, **111**, 16359-16364.

13. Morisaka, H. *et al*. 2019. CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nature Communications*, **10**, 11-24.

14. Pawluk, A. *et al*., 2017. Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. *ASM Journals*, **8**, 35-47.

15. Xiao, Y., Luo, M., Dolan, A., Liao, M. and Ke, A., 2018. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science*, **361**, 30-39.

16. Kelley, L., Mezulis, S., Yates, C., Wass, M., and Sternberg, M., 2015. The Phyre2 web portal for protein modelling, prediction and analysis. *Nature protocols*, **10**, 845-858.

17. Mulepati, S., and Bailey, S., 2013. *In Vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *Journal of Biological Chemistry*, **288**, 22184-22192.

18. Zheng, Y. *et al*. 2020. Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering. *Frontiers in Bioengineering and Biotechnology*, **8**, 10-27.

19. Dillard, K. *et al.* 2018. Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell*, **175**, 934-946.

20. Yoshimi, K. *et al.* 2022. Dynamic mechanisms of CRISPR interference by Escherichia coli CRISPR-Cas3. *Nature Communications*, **13**, 451-463.

21. Hochstrasser, M., Taylor, D., Bhat, P., Guegler, C., Sternberg, S., Nogales, E. and Doudna, J., 2014. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences*, **111**, 6618-6623.

22. Dolan, A. *et al.* 2019. Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using Type 1 CRISPR-Cas. *Molecular Cell*, **5**, 936-950.

23. Kyrou, K. *et al*. 2018. A CRISPR-Cas9 gene drive targeting doublesex mutations causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology*, **11**, 1062-1066.

24. Ivančić-Baće, I., Howard, J. L, and Bolt, E.L., 2012. Tuning in to interference: R-loops and Cascade complexes in CRISPR immunity. *Journal of Molecular Biology*, **422**, 607-61.

25. Ivančić-Baće, I., Radovčić, M., Bočkor, L., Howard, J.L., and Bolt, E.L., 2013. Cas3 stimulates runaway replication of a ColE1 plasmid in Escherichia coli and antagonises RNaseHI. *RNA Biology,* **10**, 770-778.

26. Howard, J., Delmas, S., Ivančić-Baće, I., and Bolt, E.L., 2011. Helicase dissociation and annealing of RNA-DNA hybrids by Escherichia coli Cas3 protein. *Biochemical Journal*, **439**, 85-95.

27. Strutt, S., Torrez, R., Kaya, E., Negrete, O., and Doudna, J., 2018. RNA-dependent RNA targeting by CRISPR-Cas9. *eLife,* **75**, 224-241.

28. Zuo, Z., and Liu, J., 2020. Allosteric Regulation of CRISPR-Cas9 for DNA targeting and Cleavage. *Current Opinions in structural Biology*, **62**, 166-174

29. Micura, R. and Höbartner, C., 2020. Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chemical Society Reviews*, **49**, 7331-7353.

30. Bédard, A., Hien, E., and Lafontaine, D., 2020. Riboswitch regulation mechanisms: RNA, metabolites, and regulatory proteins. *BBA Gene Regulatory Mechanisms*, **1863**, 52-61.

31. Stav., S *et al*. 2019. Genome-wide discovery of structured noncoding RNA's in bacteria. *BMC Microbiology*, **19**, 25-43.

32. Mandal, M. and Breaker, R., 2004. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology*, **5**, 451-463.

33. Panchapakesan, S. and Breaker, R., 2021. The case of the missing allosteric ribozymes. *Nature Chemical Biology*, **17**, 375-382.

34. Ferre-D'Amare, A. and Scott, W., 2010. Small Self-cleaving Ribozymes. Cold Spring Harbour Perspectives in Biology, **2**, 354-366.

35. Weinberg, Z., Kim, P., Chen, T., Li, S., Harris, K., Lünse, C. and Breaker, R., 2015. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nature Chemical Biology*, **11**, 606-610.

36. Weinberg, C., Weinberg, Z., and Hammann, C., 2019. Novel ribozymes; discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Research*, **18**, 9480-9494.

37. Roth, A., Weinberg, Z., Chen, A., Kim, P., Ames, T. and Breaker, R., 2013. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nature Chemical Biology*, **10**, 56-60.

38. Huang, R., Han, M., Meng, L., and Xing, C., 2018. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *PNAS*, **17**, 3879-3887.

39. Cañadas, I., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N.P., 2019. RiboCas: A universal CRISPR-based editing tool for *Clostridium*. *ACS Synthetic Biology*, **6**, 1379-1390.

40. Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N. and Wagner, R., 2010. Identification and characterization of E. coli CRISPR-Cas promoters and their silencing by H-NS. *Molecular Microbiology*, **75**, 1495-1512.

41. Waters, L. and Storz, G., 2009. Regulatory RNAs in Bacteria. *Cell*, **136**, 615-628.

42. Bækkedal, C. and Haugen, P., 2015. The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism. *RNA Biology*, **12**, 1071-1077.

43. Yan, Y., Tao, H., He, J., and Huang, S., 2017. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*, **45**, 365-373

44. Yang, D., Wang, Z., Ma, J., Fu, Q., Wu, L., Wang, H., Wang, S., Yan, Y. and Sun, J., 2020. Glycine Cleavage System and cAMP Receptor Protein Co-Regulate CRISPR/cas3 Expression to Resist Bacteriophage. *Viruses*, **12**, 90-108.

45. Patterson, A., Chang, J., Taylor, C. and Fineran, P., 2015. Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Research*, **43**, 6038-6048.

46. Jiang, M., Ma, N., Vassylyev, D. and McAllister, W., 2004. RNA Displacement and Resolution of the Transcription Bubble during Transcription by T7 RNA Polymerase. *Molecular Cell*, **15**, 777-788.

47. Guzman, L., Belin, D., Carson, M. and Beckwith, J., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, **177**, 4121-4130.

48. Tabor, S. and Richardson, C., 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences*, **82**, 1074-1078.

49. Lorenz, R., Bernhart, S., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. and Hofacker, I., 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, **6**, 10-24.

50. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J., Taly, J. and Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, **39**, 13-17.

51. Bailey, T., Johnson, J., Grant, C. and Noble, W., 2015. The MEME Suite. *Nucleic Acids Research*, **43**, 39-49.

52. Rio, D., 2014. Electrophoretic Mobility Shift Assays for RNA–Protein Complexes. *Cold Spring Harbour Protocols*, **4**, 435-440.

53. Ream, J., Lewis, L. and Lewis, K., 2016. Rapid agarose gel electrophoretic mobility shift assay for quantitating protein: RNA interactions. *Analytical Biochemistry*, **511**, 36-41.

54. Migneult, I., Dartiguenave, C., Bertrand, M., and Waldron, K., 2004 Glutaraldehyde: behaviour in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques*, **37**, 21-33.

55. Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S.Y., 2017. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*, **45**, 365-373.

56. Norrander, J., Kempe, T. and Messing, J., 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*, **26**, 101-106.

57. Datsenko, K. and Wanner, B., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, **97**, 6640-6645.

58. Sharan, S., Thomason, L., Kuznetsov, S. and Court, D., 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols*, **4**, 206-223.

59. Sawitzke, J., Thomason, L., Constantino, N., Bubunenko, M., Datta, S. and Court, D., 2007. Recombineering: In Vivo Genetic Engineering in *E. coli*, *S. enterica*, and beyond. *Methods in Enzymology*, **421**, 171-199.

60. Sievers F, Barton GJ, Higgins DG (2020) Multiple Sequence Alignment. *Bioinformatics* **227**, 227-250.

61. Blast.ncbi.nlm.nih.gov. 2022. Nucleotide BLAST.

62. Ivančić-Baće, I., Cass, S., Wearne, S. and Bolt, E., 2015. Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Research*, **43**, 10821-10830.

63. Yosef, I., Goren, M. G. and Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res*. **40**, 5569-5576.

64. He, L., Matošević, Z., Mitić, D., Markulin, D., Killelea, T., Matković, M., Bertoša, B., Ivančić-Baće, I. and Bolt, E., 2021. A Tryptophan 'Gate' in the CRISPR-Cas3 Nuclease Controls ssDNA Entry into the Nuclease Site, That When Removed Results in Nuclease Hyperactivity. *International Journal of Molecular Sciences*, **22**, 2848-2867.

65. He, L., 2022. Cas3 nuclease activity is regulated by an iHDA1 region located at the interface between HD and RecA1 domain. A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy.

66. Saragliadis, A., Trunk, T. and Leo, J., 2018. Producing Gene Deletions in Escherichia coli by P1 Transduction with Excisable Antibiotic Resistance Cassettes. *Journal of Visualized Experiments*, **54**, 20-32.

67. Masek, T., Vopalensky, V., Suchomelova, P. and Pospisek, M., 2005. Denaturing RNA electrophoresis in TAE agarose gels. *Analytical Biochemistry*, **336**, 46-50.

68. He, L., James, M.S.J., Radovic, M., Ivančić-Baće, I., and Bolt, E., 2020. Cas3 Protein – A Review of a Multi-Tasking Machine. *Genes*, **2**, 208.

69. Zlobina, M., Sedo, O., Chou, M., Slepankova, L. and Lukavsky, P., 2016. Efficient large-scale preparation and purification of short single-stranded RNA oligonucleotides. *BioTechniques*, **60**, 75-83.

70. Killelea, T., Hawkins, M., Howard, J., McGlynn, P. and Bolt, E., 2018. DNA replication roadblocks caused by Cascade interference complexes are alleviated by RecG DNA repair helicase. *RNA Biology*, **16**, 543-548.

71. Ceelen, M., 2019. *Phage stock preparation*. [online] Protocols Io. Available at: https://www.protocols.io/view/phage-stock-preparation-4r3l28mp3l1y/v1

72. Majsec, K., Bolt, E. and Ivančić-Baće, I., 2016. Cas3 is a limiting factor for CRISPR-Cas immunity in Escherichia coli cells lacking H-NS. *BMC Microbiology*, **16**, 45-56.

73. Aroonsri, A., Kongsee, J., Gunawan, J., Aubry, D., and Shaw, P., 2021. A cell-based ribozyme reporter system employing a chromosomally-integrated 5` exonuclease gene. *BMC Molecular and Cell Biology*, **22**, 34-46.

74. Hantke, K., 2020. Compilation of Escherichia coli K-12 outer membrane phage receptors – their function and some historical remarks. *FEMS Microbiology Letters*, **367**, 21-30.

75. Hazelbauer, G., 1975. Role of the receptor for bacteriophage lambda in the functioning of the maltose chemoreceptor of Escherichia coli. *Journal of Bacteriology*, **124**, 119-126.

76. Watarai, M., Sato, T., Kobayashi, M., Shimizu, T., Yamasaki, S., Tobe, T., Sasakawa, C. and Takeda, Y., 1998. Identification and Characterization of a Newly Isolated Shiga Toxin 2-Converting Phage from Shiga Toxin-Producing Escherichia coli. *Infection and Immunity*, **66**, 4100-4107.

77. Schwartz, M. and Le Minor, L., 1975. Occurrence of the bacteriophage lambda receptor in some enterobacteriaceae. *Journal of Virology*, **15**, 679-685.

78. Erni, B., Zanolari, B. and Kocher, H., 1987. The mannose permease of Escherichia coli consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage lambda DNA. *Journal of Biological Chemistry*, **262**, 5238-5247.

79. Smith, D., James, C., Sergeant, M., Yaxian, Y., Saunders, J., McCarthy, A. and Allison, H., 2007. Short-Tailed Stx Phages Exploit the Conserved YaeT Protein to Disseminate Shiga Toxin Genes among Enterobacteria. *Journal of Bacteriology*, **189**, 7223-7233.

80. Bradbeer, C., Woodrow, M. and Khalifah, L., 1976. Transport of vitamin B12 in Escherichia coli: common receptor system for vitamin B12 and bacteriophage BF23 on the outer membrane of the cell envelope. *Journal of Bacteriology*, **125**, 1032-1039.

81. Morona, R. and Henning, U., 1986. New locus (ttr) in Escherichia coli K-12 affecting sensitivity to bacteriophage T2 and growth on oleate as the sole carbon source. *Journal of Bacteriology*, **168**, 534-540.

82. Madeira, F., Pearce, M., Tivey, A., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A. and Lopez, R., 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, **50**, 276-279.

83. Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.L. and Weiss, D.S., 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, **497**, 254-257.

84. Vishweshwaraiah, Y., Chen, J., and Dokholyan, N., 2021. Engineering an Allosteric control of Protein function. *American Chemical Society publications*, **7**, 1806-1814.

85. Sternberg, S., Richter, H., Charpentier, E., and Qimron, U., 2016. Adaptation in CRISPR-Cas systems. *Molecular Cell*, **6**, 797-808.

86. Mitić, D. *et al*. 2020. StpA represses CRISPR-Cas immunity in H-NS deficient *Escherichia coli*. *Biochimie*, **174**, 136-143.

87. Bengert, P., and Dandekar, T., 2003. A software toolbox for analysis of regulatory RNA elements. *Nucleic Acids Research*, **31**, 3441-3445.

88. Oliva, G., Sahr, T., and Buchrieser, C., 2015. Small RNA's, 5` UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiology Reviews*, **3**, 331-349.

89. Haltiner-Jones, M., and Guthrie, C., 1990. Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the SM binding site. *EMBO*, **8**, 2555-2561.

90. Tang, W., Hu, J., and Liu, D., 2017. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature Communications*, **8**, 751-759.

91. Lejars, M., Kobayashi, A., and Hajnsdorf, E., 2019. Physiological roles of antisense RNA's in prokaryotes. *Int Journal of Biochem and Mol Bio*, **164**, 3-16.

92. Wade, J. and Grainger, D., 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Reviews Microbiology*, **12**, 647-653.