

Analysis of DNA binding and translocation by the *E. coli* CRISPR immunity protein Cas3

Liu He (4273472)

The School of Life Science

The University of Nottingham

Content

1. Introduction.....	4
1.1 Overview of background and discovery of CRISPR-Cas systems.....	4
1.1.1 General description of CRISPR-Cas systems.....	4
1.1.2 Divergent types of CRISPR-Cas systems.....	5
1.1.3 Major molecular events in CRISPR-Cas immunity.....	6
1.2 Cas3 and Cas3' proteins.....	11
1.3 Aims of study.....	13
2. Materials and Methods.....	14
2.1 Competent cells.....	14
2.2 Cloning.....	15
2.2.1 Polymerase Chain Reaction (PCR)	15
2.2.2 DNA purification.....	16
2.2.3 Digestion.....	17
2.2.4 Gel extraction.....	17
2.2.5 Ligation.....	17
2.2.6 Transformation.....	17
2.3 Preparation of plasmid stock.....	17
2.4 Site directed mutagenesis.....	18
2.5 Protein over-expression.....	18
2.6 Protein purification.....	19
2.7 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)	19
2.8 DNA substrates.....	20
2.9 Denaturing gel electrophoresis.....	22
2.10 Beer-Lambert law.....	23
2.10.1 A ₂₈₀ and protein concentration.....	23
2.10.2 A ₂₆₀ and DNA substrate concentration.....	23
2.11 Nuclease assays.....	24
3. Results.....	25
3.1 Cloning of <i>E. coli</i> Cas3 and Cas3' proteins	25
3.2 Mutagenesis.....	26
3.3 Over-expression and purification of <i>E. coli</i> Cas3 and Cas3' proteins	27
3.4 Nuclease activities of purified Cas3 and Cas3' on M13 ssDNA and forked DNA substrates.....	28
3.5 The influence of chemically modified DNA on Cas3' nuclease activity at forks.....	33

3.6	The influence of chemically modified DNA on Cas3 nuclease activity at forks.....	36
4.	Discussion.....	39
4.1	Cas3, Cas3' and Cas3'-H74GD75G purifications.....	39
4.2	Nuclease activities of Cas3 and Cas3' on unmodified and modified DNA FORK-2B.....	40
5.	Bibliography.....	44
6.	Media Recipes.....	48
6.1	Luria-Bertani Broth (LB) and LB Agar.....	48
6.2	10X Tris Borate EDTA (TBE) Buffer.....	48
6.3	10X SDS PAGE Buffer.....	48

Abstract

Cas3 is essential for CRISPR-Cas immunity in *Escherichia coli* and is conserved across Class 1, type I CRISPR-Cas systems. Cas3 is a metal-dependent nuclease and ATP-dependent helicase that is required for CRISPR interference reactions, the stage of immunity in which a Cascade complex delivers crRNA to a complementary sequence in invader DNA to form an R-loop, trigger nuclease degradation of the invader, and stimulate adaptation ("priming"). Cas3 nuclease activity resides in an HD-domain (called Cas3' here) and Cas3 translocation activity is catalysed by superfamily-2 helicase. The nuclease activity of Cas3 is crucial for interference and a preliminary analysis of its mechanism is presented here. We generated purified *E. coli* Cas3, Cas3' and nuclease inactive Cas3' proteins to study their activities on ssDNA and dsDNA substrates in comparison with similar substrates containing precisely positioned chemical modifications to the DNA backbone or bases. We observed that methyl-phosphonate and abasic DNA modifications had significant inhibitory effects on Cas3' and Cas3 nuclease activities, and that there were some interesting differences in activities of these two enzymes when confronted with phosphorothioate DNA. This gives clues about the mode of action for Cas3' and Cas3 when interacting with DNA for catalytic activity. The new work described will form the basis of more detailed kinetic analyses of Cas3' and Cas3 nuclease activities in future work.

1 Introduction

1.1 Overview of background and discovery of CRISPR-Cas systems

CRISPR-Cas is a prokaryotic adaptive immunity system that is widespread throughout the Bacterial and Archaeal domains of life. It is estimated that CRISPR-Cas is present in about 85% of archaeal species and 50% of bacterial (Grissa et al. 2007; Richter et al. 2012), data based on whole genome sequencing analyses rather than functional analysis of each CRISPR-Cas system. The two major components of a functional CRISPR-Cas system are a DNA locus called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas proteins (CRISPR associated proteins). The molecular function and interplay of these two major components of CRISPR-Cas are described more later.

1.1.1 General Description of CRISPR-Cas systems

Figure 1 A shows a cartoon of the genetic arrangement of a generalised CRISPR-Cas system, and Figure 1 B shows the CRISPR-Cas locus of *Escherichia coli* (*E. coli*) K-12 strain MG1655, the focus of this MRes study. CRISPR DNA comprises repeats that may number from very few (26 in *E. coli* subtype, e.g. MG1655) to many (728 in *Thermotoga neapolitana* DSM4359) (Haft et al. 2005) and may vary in the length of each repeat from about 23 to 47 base pairs (Philippe Horvath and Barrangou 2010). In *E. coli* CRISPR-1 the repeats are 27 base pairs in length. Repeat DNA sequences are highly conserved, reflecting their molecular processing by Cas enzymes, described later. The DNA repeats are separated by “spacer” DNA sequences of 21-47 base pairs depending on different CRISPR subtypes, for example, spacers in *E. coli* are about 33 base pairs while those in *Aeropyrum pernix* are approximate 42 base pairs (Haft et al. 2005). As was identified in 2005 (Pourcel et al. 2005), the spacers act as a repository for DNA sequences that were acquired from mobile genetic elements (MGEs). As we will see, the transcription of CRISPR DNA to RNA

is the fundamental principle that underlies CRISPR-Cas immunity through the sequence matches between CRISPR spacers and the invader.

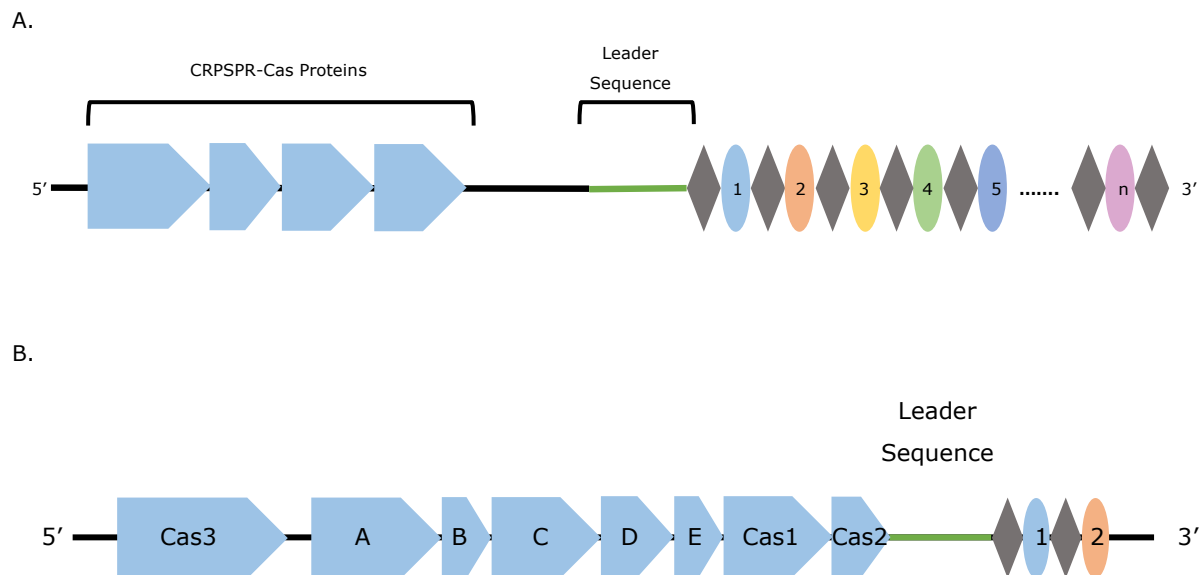


Figure 1. (A) Cartoon of a generalised CRISPR-Cas structure. Genes encoding Cas proteins vary according to class of CRISPR-Cas system and the sub-types. Cas proteins were originally identified as being "CRISPR-associated proteins" in (Jansen et al. 2002). A "Leader sequence" is shown, which is A-T rich and includes a promoter for transcription of repeats and spacers into CRISPR RNA (crRNA)(Jansen et al. 2002; F. J. Mojica et al. 2009). The CRISPR locus comprises repetitive DNA of varying length depending on species (repeats, shown as grey diamond) and variable DNA regions (spacers, coloured ovals and labelled with number 1 to 5) that differ in number widely, depending on species (P. Horvath et al. 2008; Pourcel et al. 2005). (B) The CRISPR-Cas locus of *E. coli* K-12 MG1655. Labelled, as blue block arrows, are genes encoding Cas3 nuclease-helicase, Cascade complex (comprising proteins CasA, CasB, CasC, CasD and CasE), and Cas1 and Cas2 adaptation proteins.

1.1.2 Divergent types of CRISPR-Cas systems

Generally, CRISPR systems can be divided into two major clades called Class 1 and Class 2 (Makarova et al. 2015). This division is based on their differing mechanisms and proteins for effecting crRNA based immunity during interference reactions, described more below: Class 1 systems utilize multi-protein crRNA complexes (e.g. Cascade), but Class 2 systems utilize a single protein (e.g. Cas9) (Burstein et al. 2017; Koonin et al.

2017). CRISPR-Cas subtypes within these major classes show great diversity in their protein components arising and enriched by the diverse threat and environments encountered by each strain and species (Bondy-Denomy et al. 2015; Makarova et al. 2011). CRISPR Class 1 comprises Types I, III and IV. Type I is most common type among prokaryotes and is characterised by the presence of Cas3 protein (Koonin et al. 2017). Class I Type I is further sub-divided into various subtype, including I-A, I-B, I-C, I-U, I-D, I-E, I-F and I-F variant (Koonin et al. 2017). Unlike CRISPR Class 1, CRISPR Class 2 are characterised by a single multidomain protein effector and contains three subtypes, type II, V and VI.

1.1.3 Major molecular events in CRISPR-Cas immunity

The major molecular events that enable and catalyse CRISPR immunity are usually categorised into three stages, called adaptation, transcription/crRNA maturation, and interference (Amitai and Sorek 2016), summarised in Figure 2.

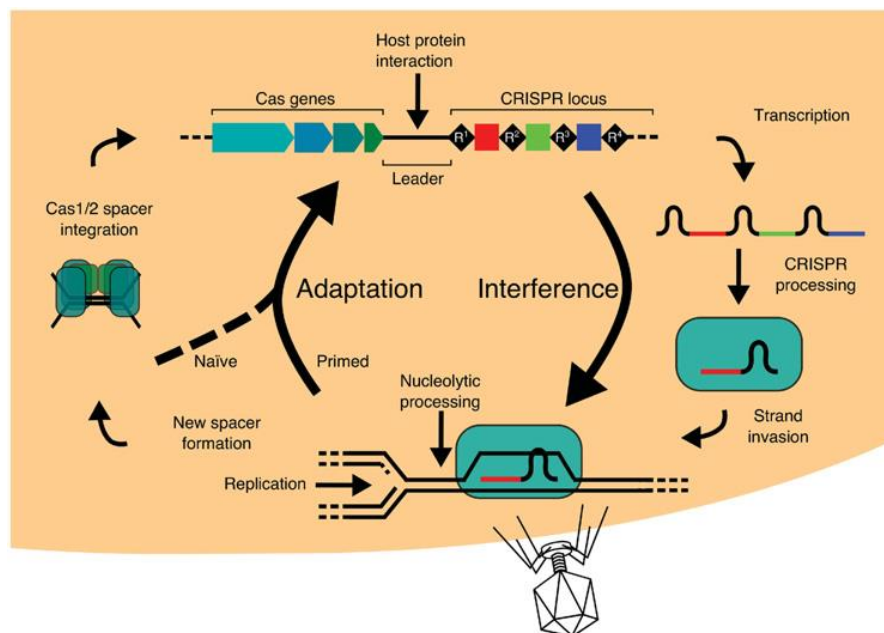


Figure 2. Illustration of the stages of CRISPR-Cas immunity adapted from Killelea et al (Killelea and Bolt 2017). The major general processes of CRISPR-Cas are shown, with particular emphasis on adaptation linked to interference, as is described in the main text.

(a) Adaptation

The CRISPR spacer DNA sequences are derived from MGE DNA during the process of adaptation. In adaptation, summarised in Figure 2 and 4, Cas1 and Cas2 are essential factors (Yosef et al. 2012), Cas1 and Cas2 can form a stable complex and participate in adaptation while interacting with other proteins (e.g. Cas3) (van Erp et al. 2017). During adaptation, the Cas1-Cas2 complex (Figure 3) is able to capture short foreign DNA sequences (spacers) and then catalyse integration. This requires recognition and location of palindromic motif in repeats, so that captured DNA can be inserted into a specific position in CRISPR locus (Beloglazova et al. 2008; Moch et al. 2017; Nunez et al. 2014; Rollie et al. 2015). Following spacer integration, new spacers are flanked by gaps, which are then filled by DNA polymerase I (PolA) (Ivancic-Bace et al. 2015).

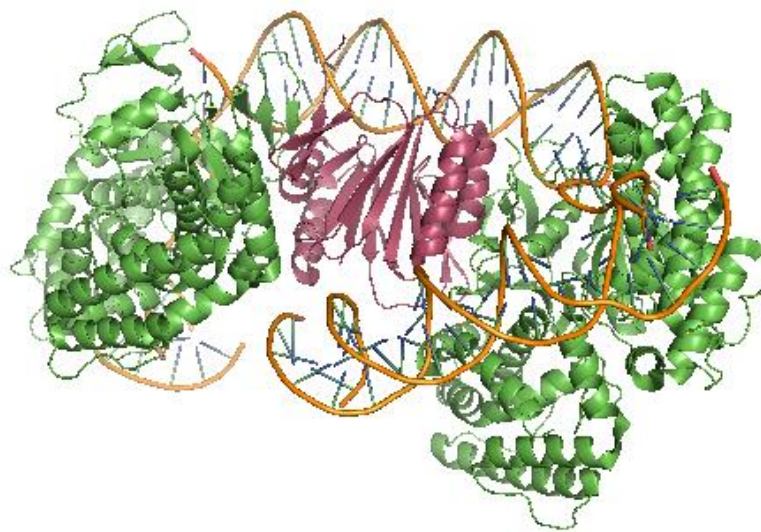


Figure 3. Crystal structure of Cas1-Cas2-DNA complex (PDB code: 5VVJ, produced by PyMol). This image shows a double stranded DNA bound to *E. coli* Cas1-Cas2 complex, as displayed above, Cas1 (305 amino acid residues) is marked in green and Cas2 (95 amino acid residues) is in red, while duplex DNA backbone is labelled in orange.

Adaptation arises through two pathways, primed adaptation (or priming) and naïve adaptation. During primed adaptation, Cas1-Cas2 complex interacts with interference complexes (e.g. Cascade and Cas3) that

causes elevated rates of adaptation (Nunez et al. 2014; Redding et al. 2015). Foreign DNA replication is blocked by a Cascade R-loop complex and then degraded into small single-stranded DNA fragments by Cas3. These fragments can then form a DNA-Protein complex with Cas1-Cas2, ready for spacer integration (Moch et al. 2017; Musharova et al. 2017). Some evidence suggests that Cas1, Cas2 and Cas3 can bind and form a complex, though Cas3 function is not necessary for spacer integration (Fagerlund et al. 2017). Apart from Cas3 and Cascade, RecG and PriA have also been shown to participate in primed adaptation (Ivancic-Bace et al. 2015).

In *E. coli*, Cas1 and Cas2 alone are sufficient to carry out adaptation that leads to integration of new spacers into CRISPR locus (Yosef et al. 2012). This is called naïve adaptation, because it does not require recognition of an MGE DNA by an interference protein complex. However, naïve adaptation, like primed adaptation, does require host cell protein factors that are not Cas enzymes. In *E. coli* these include integration host factor (IHF) (Nunez et al. 2015) and possibly the enzymatic functions of RecBCD and DNA Polymerase I (PolA) (Ivancic-Bace et al. 2015; Levy et al. 2015). The role of IHF (Nunez et al. 2016) is to interact with DNA sequences in the CRISPR leader and bend the DNA to make it more accessible for integration by Cas1 and Cas2. Specifically, the bent DNA (generated by IHF and A-T rich sequence in CRISPR leader) allows Cas1-Cas2-protospacer complex to recognise the first CRISPR repeat and insert spacer into the proper location. PolA is thought to function in gap filling as mentioned above. The role of RecBCD is far from clear, it may under some circumstances provide DNA fragments from an MGE for capture by Cas1-Cas2. A few studies have highlighted the importance of non-Cas proteins for fuelling adaptation. In one study (Levy et al), it was noted that acquisition of new spacers depends on RecBCD's priority of processing foreign DNA, this preference is based on RecBCD identifying self and foreign DNA by recognizing a certain DNA sequence called Chi

site. For example, host genome has high density of Chi sequence, which limits RecBCD's function, in contrast to foreign DNA (Levy et al. 2015).

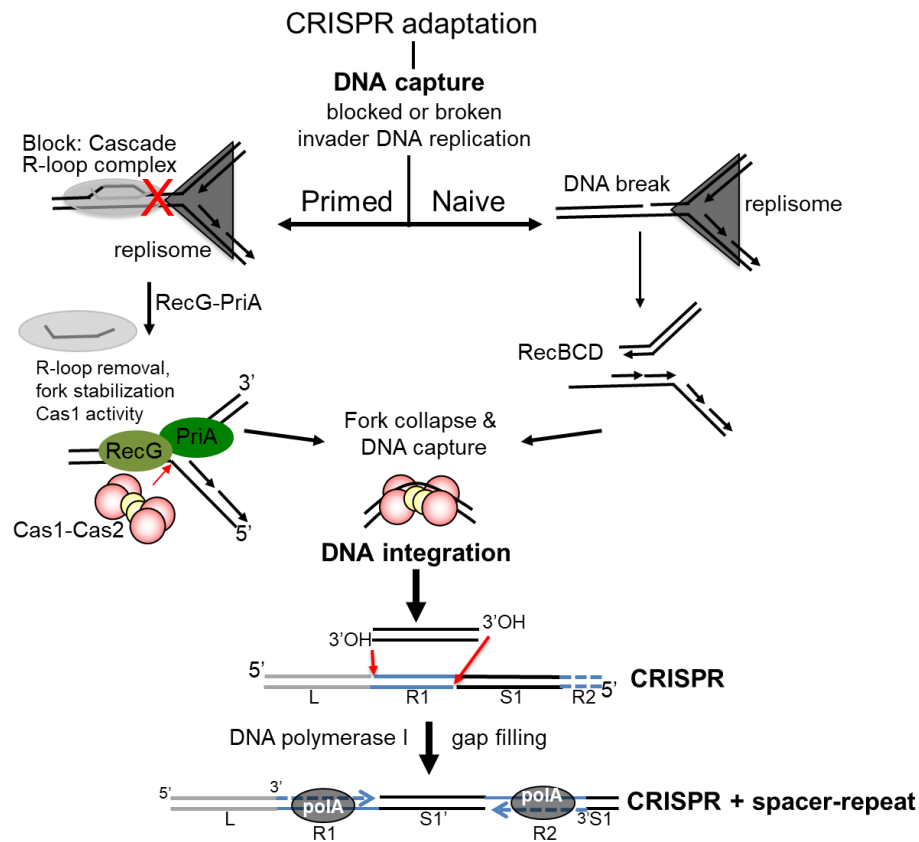


Figure 4. Summary model of primed and naive adaptation in *E. coli*. This figure is taken from Ivancic-Bace et al 2015 and shown how CRISPR proteins capture DNA and inserts it into the CRISPR locus.

(b) Processing of CRISPR RNA.

Adaptation described above generates immunity by expanding CRISPR loci with captured fragments of protospacer DNA, converting them into spacers. Transcription of a CRISPR locus into RNA is required for interference complexes (described below) to target invader MGEs. Transcription of CRISPRs was first detected in *Haloferax mediterranei* (F. Mojica et al. 1993) but at the time the functional significance of this was not known. Later, this process was identified as an early stage of crRNA guided CRISPR targeted immunity (Ryan N Jackson et al. 2014b; van der Oost et al. 2014; Ye and Zhang 2016). In type I and III CRISPR system,

precursor crRNA (pre-crRNA) is first transcribed and then processed by endoribonuclease (e.g CseE and Cas6) to generate mature crRNA (contains invader targeting sequence) (Hale et al. 2009; Wang et al. 2011; Ye and Zhang 2016). The crRNA comprises parts of the repeat DNA sequence and a complete spacer sequence, therefore giving it a complementary sequence to MGE invaders. The crRNA generated during the expression process is loaded into interference complexes for pairing with MGE DNA during the interference process.

In addition to crRNA, type II CRISPR-Cas immunity requires a trans-activating CRISPR RNA (tracrRNA) that base pairs with crRNA to allow assembly of the Class 2 interference protein, Cas9, which is famous for biotechnological applications. tracrRNA is necessary in crRNA maturation, it has a 25 nucleotide stretch (which is complimentary to the CRISPR repeats) and can form duplex stranded RNA (dsRNA). This process then triggers endoribonuclease RNase III cleavage and generates mature crRNA in the presence of CsnI (Charpentier et al. 2015; Deltcheva et al. 2011).

(c) Interference.

In type I systems (e.g *E. coli* CRISPR-Cas), the interference mechanism relies on the multi-protein and crRNA ribonucleoprotein complex "Cascade". The Cascade-crRNA complex scans for complementary sequences in MGEs and recognizes protospacer-associated motifs (PAM). crRNA and DNA base pair via an R-loop leading to a series of dynamic changes to the structure of Cascade, which expose a subunit protein called CasA, allowing the recruitment of Cas3 (van Erp et al. 2017). Hence, MGE DNA can be unwound by Cas3 in the presence of ATP and fully degraded into small DNA fragments by Cas3 metal-dependent cleavage (Brouns et al. 2008; Sinkunas et al. 2011)

Unlike type I effector complexes, type II systems are dependent on a tracrRNA to stabilize effector-crRNA in a system exemplified by Cas9

(Barrangou et al. 2007; Chylinski et al. 2013; Gasiunas et al. 2012; Karvelis et al. 2013; Koonin et al. 2017). In contrast to this the type V CRISPR system utilises an effector named Cpf1, which differs from type II as it does not require tracrRNA when cleaving foreign DNA (Zetsche et al. 2015). The effectors of the type II (Cas9) and type V (Cpf1) systems both have a RuvC domain, which displays nuclease activity (Koonin et al. 2017). Cas9-crRNA-tracrRNA complex form R-loops, as with Cas9, Cpf1 can also form R-loop, but it does not require tracrRNA (Fonfara et al. 2016; Stella et al. 2017). DNA degradation relies on Cas9, it has nuclease sites (HNH and RuvC-like site) and can create double stranded break (Chylinski et al. 2013; Deltcheva et al. 2011; Rath et al. 2015)

1.2 Cas3 and Cas3' Proteins

As discussed above, Cas3 (Figure 5) and its subtypes, are widespread throughout type I CRISPR-Cas systems and are essential for CRISPR immunity in *E. coli* due as a mechanism for degrading MGE DNA (Haft et al. 2005; Makarova et al. 2006; Makarova et al. 2015; Sinkunas et al. 2011). Cas3 has a N-terminal histidine-aspartate (HD) nuclease and superfamily 2 (SF2) helicase domains (Haft et al. 2005; Makarova et al. 2006). Cas3 HD domain has a typical signature H-HD-D motif, which refers to a metal-dependent nuclease and has been observed to function as an exonuclease (Beloglazova et al. 2011). Alongside the nuclease domain, Cas3 possesses a type A SF2 helicase which functions as an ATP-dependent DNA unwind capacity (Beloglazova et al. 2011; Sinkunas et al. 2011).

The nuclease activity of Cas3 was first demonstrated on ssDNA (Sinkunas et al. 2011) and since then, Cas3 has been observed cooperating with different protein/complex during CRISPR interference stage, for example, Cascade and Cas1-Cas2. As discussed above, during interference Cascade first scans foreign DNA and recognizes PAM site. PAM recognition by Cascade initiates foreign DNA unwinding and crRNA R-loop formation. Conformational changes in Cascade (Hayes et al. 2016; van Erp et al.

2017) expose the CasA subunit, facilitating recruitment of Cas3 allowing it to create a nick on the displaced DNA strand. Single strand DNA is wrapped inside Cas3 helicase and its 3' end is directly carried to the Cas3 nuclease core catalysis site (Huo et al. 2014). Cas3 unwinds DNA in presence of ATP and translocate along single strand DNA from 3' to 5' end (Sinkunas et al. 2011). During Cas3 translocation, DNA is degraded into small DNA fragments that may be relevant to adaptation, and those fragments are processed by Cas1-Cas2 complex and inserted into the CRISPR locus (Musharova et al. 2017; Nunez et al. 2014).

As mentioned above *E. coli* Cas3 is formed from multiple domains that are covalently fused, however, in other species these domains can also be observed to be separately encoded (Haft et al. 2005). The Cas3 HD nuclease, also known as Cas3', is recognised as a Cas3 subtype and can be found in bacteria and archaea such as *Methanocaldococcus jannaschii* (*M.jannaschii*) where it is encoded as an individual protein. *M.jannaschii* Cas3' shows 3' to 5' exonuclease activity on ssDNA (Beloglazova et al. 2011). However, the diversity of Cas3 structure across species and whether this diversity causes any mechanistic changes is seldom discussed. Further analysis is required to determine how Cas3' and Cas3 catalyse DNA degradation and if different Cas3 structures (fused/non-fused) have an impact on CRISPR immunity among various species.

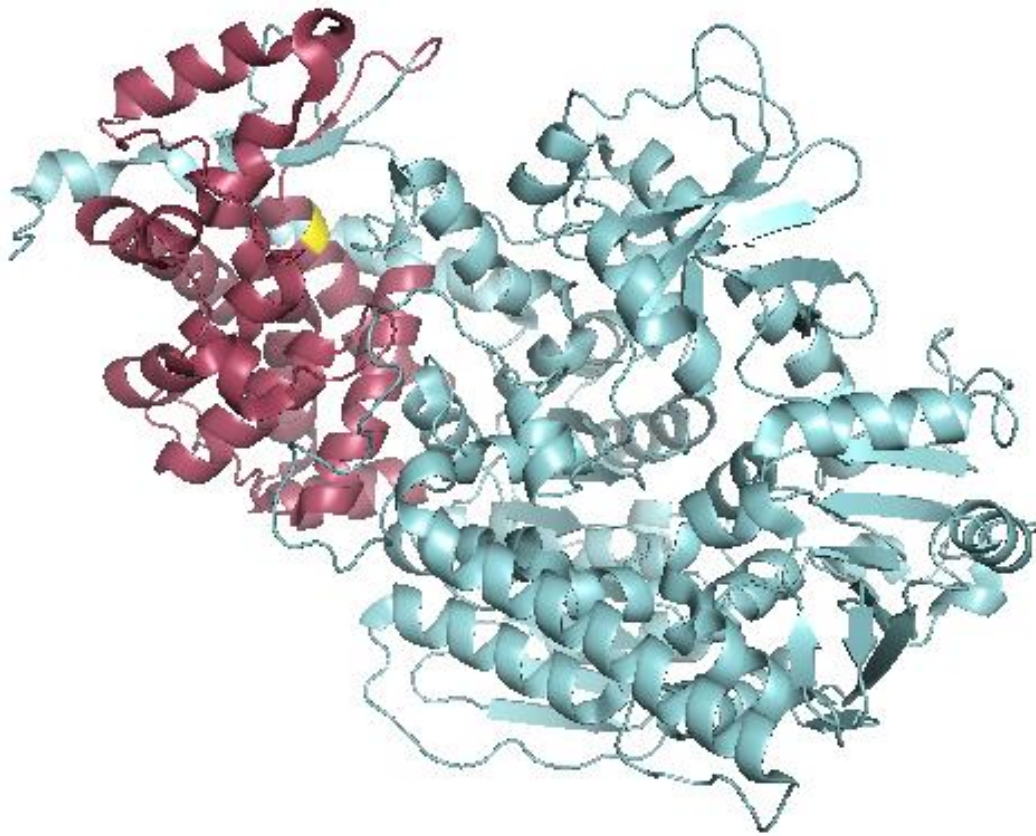


Figure 5. Structure of *Thermobaculum terrenum* Cas3 (PDB: 4Q2C, produced by PyMol). *Thermobaculum terrenum* Cas3 is a protein with multiple functions as discussed above, it has 944 amino acid residues and multiple domains. HD domain, which preforms nuclease function is marked in red, while helicase domains SF2 is in cyan. Cas3 is a metal-dependent nuclease, in this case it relies on Ni^{2+} (e.g. *E. coli* Cas3 requires Mg^{2+}) and the Ni^{2+} binding site has been labelled in yellow (top left).

1.3 Aims of Study

The overall aim of this study is to analyse Cas3 helicase (SF2) tracking along DNA strands, otherwise known as translocating, and the unwinding of double stranded DNA. This study was divided into two aspects. The first aspect is to generate soluble wild-type Cas3 from *E. coli* using a His-tag purification system. The protein is then able to carry out stable nuclease activity on single-stranded DNA, for example a single-stranded circular DNA substrate called M13mp18. The second aspect focuses on the specific translocating and unwinding activity of Cas3 on DNA substrates. This

activity of Cas3 can be analysed by using chemically modified DNA substrates, one for example includes the introduction of abasic sites, this will be addressed in further detail in a following section. By introducing chemical modification sites, it might also reveal interaction of Cas3 and damaged DNA. As mentioned above, Cas3 is a ATP-depend helicase. In this study, all reactions were conducted without any ATP, so the Cas3 translocating with ATP will need to be explored in future work. Methods are based on previous work completed by members from the labs of Bolt, Sinkunas, Beloglazova, Howard and Raney (Beloglazova et al. 2011; Howard et al. 2011; Raney 2014; Sinkunas et al. 2011).

In order to study DNA degradation with and without translocase and Cas3 active sites, Cas3' (as a comparison) and its mutant Cas3'-H74GD75G (negative control) were designed and generated. As discussed above, Cas3' is modified so that it is only able to catalyse DNA degradation but is unable to translocate along DNA and therefore differences in DNA cleavage between Cas3' and Cas3 are expected.

Cas3 is an essential factor for CRISPR-Cas immunity in bacteria and archaea (Reeks et al. 2013; Sorek et al. 2013; Westra et al. 2012). Clarifying the chemical mechanism of Cas3 will help expand our understanding of the role of Cas3 in bacterial immune activity and how it interacts with other proteins (for example, Cascade and Cas1-Cas2). Analysis of change of Cas3 function in different reaction conditions can facilitate a better understanding of the bacterial response to both the initial phage infection and phage escape from host immunity within the system.

2 Materials and Methods

2.1 Competent cells

DH5 α and BL21AI competent cells were prepared for the relevant experiments by the following method. Cells were grown at 37 °C with shaking until the optical density (OD_{600nm}) reached 0.6 and then spun

down at 5 000 rpm for 5 min. Cells were resuspended in ice cold 0.1 M CaCl₂ and cells were incubated on ice for 1 h before being pelleted and resuspended in 0.1 M CaCl₂. Glycerol was added to the cells to a final concentration to 30% w/v for storage (-80 °C freezer).

2.2 Cloning

The DNA sequence of Cas3' (also known as Cas3 HD) was amplified from *E. coli* MG1655 genomic DNA by PCR and subsequently cloned into pBAD/HisA vector via XhoI and EcoRI site. The subsequent plasmid (pLH1) was used to express N-terminal poly-histidine (His_{6x}) tagged Cas3' (referred to as Cas3').

2.2.1 Polymerase Chain Reaction (PCR)

The PCR primers used for amplifying *E. coli* Cas3' (listed in table 1) were purchased from Sigma-Aldrich. Primer design was based on sequence comparison of Cas3' and Cas3 (Figure 6), and induced XhoI and EcoRI restriction site (followed by a stop code) separately into produce terminals. Vent® DNA polymerase and 10X ThermoPol Reaction Buffer was purchased from England Biolabs® Inc (NEB), dNTPs were from Sigma-Aldrich. The reagents were mixed as follows: 0.25 µM forward and reverse primers, 0.25 µM dNTPs, 2 units of Vent polymerase, 1X ThermoPol Buffer and 15 nM MG1655 genomic DNA, and diluted to a final reaction volume of 40 µL by sterile distilled water (SDW).

Cas3' gene was amplified by using Veriti™ 96-Well Thermal Cycler (ThermoFisher). Fully denature template at 95 °C for 1 minute (at first time, then 30 sec in cycles) initially and then cool down reaction temperature to 60 °C and continued for 10 sec to anneal forward and reverse primers into template, extension process was took at 72 °C for 1 min, repeated denaturation-annealing-extension process for 30 cycles. PCR products were mixed with purple loading dye (NEB) and analysed using a 1% agarose, 1X TBE gel.

Purpose	Name	Sequence	Restriction Site
Cloning E. coli Cas3' into pBAD/HisA	Cas3'F	5'GCCGCTCGAGGAACCTTTTAAATATATATGCCATT	XhoI
	Cas3'R	5'ATTTGAATTCATACCAACAGTCAGCAAGCGA	EcoRI

Table 1. Primers for cloning pLH1. This pair of primers can generate PCR product that has full length of Cas3' sequence and a XhoI site in 5'-end and no start code, and a EcoRI site and a stop code in 3'-end.

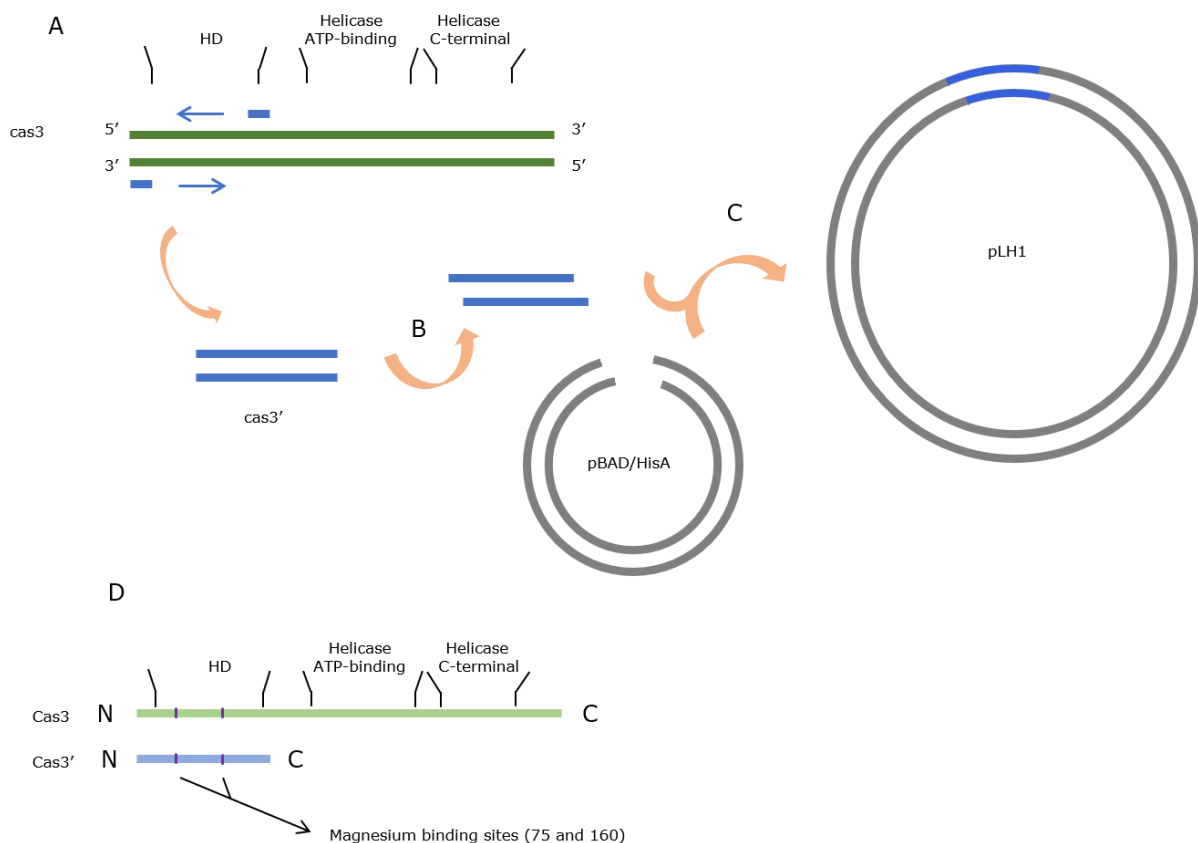


Figure 6. pLH1 cloning and protein alignment of Cas3 and Cas3'. (A)(B) and (C) were three steps of generating pLH1, (A) *cas3'* DNA sequence was amplified from *E. coli cas3* (also known as *ycgB*), primers were described in Table 1. (B) PCR product and pBAD/HisA vector were digested by XhoI and EcoRI-HF restriction enzymes, and (C) pLH1 was generated by insert-vector ligation. (D) alignment of linear Cas3 and Cas3' protein. Cas3' in this study covers full length of *E. coli Cas3* HD domain, with 2 magnesium binding sites were shown in purple.

2.2.2 DNA purification

PCR products were purified by QIAquick PCR Purification Kit (from Qiagen Inc) as described in the manufacture's protocol.

2.2.3 Digestion

10X CutSmart® Buffer, restriction enzymes XhoI and EcoRI-HF and Alkaline Phosphatase, Calf Intestinal (CIP) (all of them were purchased from NEB). pBAD/HisA vector was purchased from Invitrogen.

Target DNA was digested in reaction contains 10 µL purified PCR product, 1X CutSmart Buffer, 2 units of restriction enzymes XhoI and EcoRI-HF and made up to 50 µL reaction volume with SDW. Digested vector was generated by follows, 1 ng pBAD-HisA, 1X CutSmart Buffer, 2 units XhoI and EcoRI-HF, 3 units CIP and made up to 50 µL reaction volume by SDW. Reactions were incubated at 37 °C for 1 h.

2.2.4 Gel extraction

Digestion products were loaded into wells of 1% agarose, 1X TBE gel and collected by using QIAquick Gel Extraction Kit (QIAGEN).

2.2.5 Ligation

T4 DNA Ligase and T4 DNA Ligase Reaction Buffer (from NEB) were used to carry out ligations, insert and vector were mixed as follows: 50 ng digested vector and insert DNA, 1X T4 DNA Ligase Reaction Buffer, 400 units of T4 DNA ligase and SDW made up final 20 µL reaction volume. The ligation was incubated at 16 °C for 1 h.

2.2.6 Transformation

1 µL of plasmid was mixed with 100 µL of competent cells (DH5α or BL21AI) and incubated mixture on ice for 30 min, cells were heat shocked at 42 °C for 90 sec then followed by 5 min on ice. Cells were grown up by adding 900 µL fresh LB broth and incubating at 37 °C for 1 h with shaking. Cells were then spun down for 1min at 13 000 rpm and resuspend in 200 µL LB, before being plated onto an LB agar plate which contains the relevant antibiotic. Cells grew at 37 °C overnight.

2.3 Preparation of plasmid stock

Plasmids were collected by QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. After DNA was eluted from DNA binding

column, a NanoDrop™ 2000/2000c Spectrophotometers (Thermo Scientific) was used to determine the plasmid concentration. Plasmids were analysed by double restriction enzyme digestion (Figure 9) and sequencing (service provided by Source Bioscience). Sequencing result then analysed by alignment.

2.4 Site directed mutagenesis

pLH1 was used as the template to create site specific mutations in the HD domain to generate plasmid pLH4 which contained the Cas3' with the mutations H74G and D75G (mutant protein named Cas3'-H74GD75G).

Primers (Table 2) were designed using NEBaseChanger_{v1.2.6} and purchased from Sigma-Aldrich. pLH4 was generated followed the manufactures protocol of Q5® Site-Directed Mutagenesis Kit (NEB) and transformed in competent cells as describe above.

Purpose	Name	Sequence
Creating <i>E. coli</i> Cas3'	HDmutaF	5'CATTGCTCTTGGTGGTATTGGAAAGTTTGATATAC
HD mutant Cas3'-H74GD75G	HDmutaR	5' AAAAATAACAGCCAGGCC

Table 2. Primers used for generating site specific mutations in pLH1

2.5 Protein over-expression

Proteins were expressed in *E. coli* BL21AI cells (carrying the relevant vectors). Cells were grown up in LB broth with 100 µg/ml ampicillin and 0.1% w/v glucose at 37 °C till OD_{600nm} reached 0.6 to 0.8, then cells were induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) till final concentration was 1 mM and 0.2% w/v L-Arabinose. The protein was overexpressed for 20 h at 18 °C and then spun down at 5000 rpm for 10 min in an Avanti® J-26 XP centrifuge by using JLA10.500 rotor (Beckman Coulter Life Sciences). Pelleted cells were resuspended in 25 mL filtered lysis buffer (20 mM Tris pH8.0, 10 % glycerol, 500 mM NaCl, 10 mM Imidazole, 2M ammonium sulfate 0.1 % Tween 20 and one EDTA free Protease Inhibitor Cocktail tablet (Roche and then flash frozen in dry ice before storage at -80 °C.

2.6 Protein purification

Biomass was defrosted on ice, each 5 mL of biomass was lysed using a Vibra-Cell™ Ultrasonic Liquid Processor (Sonics & Materials, Inc). During the sonication process the biomass was kept on ice to retain a low temperature and sonicated for 10 sec and then paused for 10 sec for 3 min for each 5 mL suspension. Then samples were clarified at 50 000 rpm for 30 min (Avanti® J-26 XP and JA25.50 rotor, Beckman Coulter Life Sciences). The supernatant was collected and loaded onto a 5 mL HisTrap HP column connected to ÄKTA start purification system (both from GE Healthcare). The column was charged with Charge Buffer (filtered 0.1 M NiCl_2 solution) and equilibrated with Buffer A (20 mM Tris pH8.0, 10% glycerol, 500 mM NaCl and 10 mM Imidazole) before loading sample onto it. The protein was eluted with buffer B (20 mM Tris pH8.0, 10% glycerol, 500 mM NaCl and 500 mM Imidazole) via linear gradient from 0 % to 100 % Buffer B (increasing Imidazole concentration) and collected in 1 mL fractions each for analysis by SDS Page (detailed in 2.7). The protein was dialysed in 2 L Dialyse Buffer (20 mM Tris pH8.0, 35% glycerol and 50 mM NaCl) at 4 °C for at least 10 h before being aliquoted and then flash frozen and stored in -80 °C. Calculation of protein concentration was based on the Beer-Lambert Law (2.9.1).

2.7 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Purified proteins were analysed by standard denaturing SDS-PAGE. 8% and 12.5% polyacrylamide gels were used to identify Cas3 and Cas3' or H74G&D75G protein. 20 µL of each protein samples containing 10 mM dithiothreitol (DTT) and 1X SDS Loading Dye (50 mM Tris-HCl, pH 6.8, 2 % SDS, 10 % glycerol, 1 % β -mercaptoethanol, 12.5 mM EDTA and 0.02% bromophenol blue) was heated at 95 °C for 10 min before 12 µL of sample was then loaded onto gels and ran for 60 min at 150 V. Proteins were coloured by Coomassie Stain (0.1% Coomassie, 10% acetic acid,

40% methanol) and then destained in destain buffer (20% methanol, 10% acetic acid) to wash out extra colour.

2.8 DNA substrates

5' end labelled DNA oligo MW12 and forked DNA substrates was used to test Cas3' and Cas3 cleavage capacity. Bioinformatics analyses of Cas3 possesses that it has a polar nuclease cleavage activity from 3'- end and moves toward to 5'-end on single stranded DNA (Beloglazova et al. 2011; Sinkunas et al. 2011). According to this chemical characteristic, DNA Fork2B and other chemical modified forks based on basic structure of Fork2B has been designed and used in protein function analysis.

Fork2B is a branched DNA substrate, it contains Cy5 labelled (5' terminal) double strand part (24 bp) and two single stranded flaps with 3'-end and 5' ends (Table 3). Methyl-phosphonate Fork2B (Table 3) was introduced into a modified deoxyadenosine triphosphate (methyl-phosphonate dATP, figure 7 A) in duplex part (labelled stand), which generates a neutral backbone linkage from phosphodiester linkage to methyl-phosphonate backbone linkage and leads to a resistant of nuclease cleavage activity (Sarin et al. 1988). Phosphorothioates ACT and GCT were induced in this study, they replaced original oligodeoxynucleotides and then change phosphodiester backbone linkage to phosphorothioate type (Kurreck 2003). This modification in labelled stand endues a nuclease resistance character in certain location, for Phosphorothioate-1 Fork2B (ACT), the modified site is located in fork flap close to 3'-end, and modified GCT is posited in duplex part of phosphorothioate-2 Fork2B (Figure 7 B, Table 3). Abasic-1 Fork2B and abasic-3 Fork2B were generated by using two 5'-end Cy5 labelled DNA oligos which were inserted an abasic furan (referred to as abasic) in 3' flap and duplex separately (Figure 7 C, Table 3). Abasic is a simulation of deoxyribonucleotide but without nucleotide base, this special structure can influence 3' to 5' exonuclease degradation.

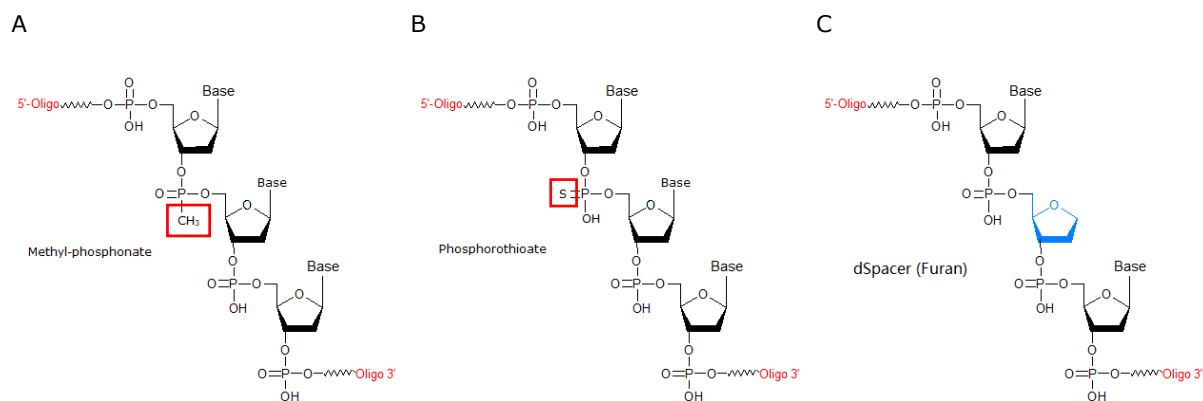


Figure 7. Chemical modification structure. (A) DNA backbone of Methyl-phosphonate amino acid residue. (B) DNA backbone of Phosphorothioate amino acid residue. (C) amino acid residue structure of dSpacer (Furan). Note: Images were from <http://www.genelink.com> and have been re-edited.

The oligonucleotide used to make the DNA forks in Table 3 were annealed at a 1:1.2 ration of labelled to unlabelled oligonucleotide. The annealing reaction contained 5 μ M labelled oligo, 6 μ M unlabelled oligo(s), 1X Annealing Buffer (10 mM Tris pH7.5, 50 mM NaCl and 1 mM EDTA) and was made up to 25 μ L with SDW. Incubation took place at 95 $^{\circ}$ C for 5 min, before reducing the temperature to 20 $^{\circ}$ C and allowing to cool slowly (about 8h to 12h). 5 μ L Orange G loading dye (80% glycerol and Orange G) was added to the annealed DNA substrate before annealed DNA substrate was separated from unannealed oligonucleotides using a 10 % native polyacrylamide gel. DNA substrates were diffused from native gel in 100 μ L Elution Buffer (4 mM Tris pH8.0 and 10 mM NaCl) at 4 $^{\circ}$ C for at least 48 h and then concentrated using a Savant™ SPD131DDA SpeedVac™ Concentrator (Thermofisher) to about 25 μ L. The DNA substrate absorbance at A_{260} was measured using a NanoDrop™ 2000/2000c Spectrophotometer and the concentration of the DNA substrate was calculated using the Beer–Lambert law (described below).

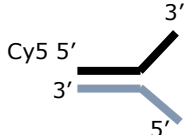
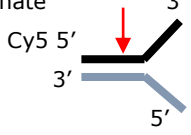
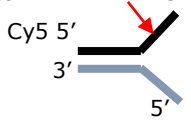
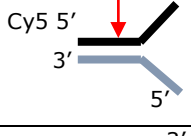
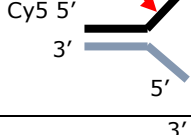
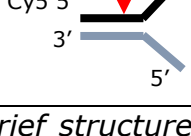
Name	Structure	Oligo sequence
Fork2B		Oligo 1: Cy5 5'-TCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Methyl-phosphonate Fork2B		Oligo 1: Cy5 5'-GTCGGATCCTCTAGACAGCA ^A TCCATGATCACTGGCACTGGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGTATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Phosphorothioate-1 Fork2B		Oligo 1: Cy5 5'-GTCGGATCCTCTAGACAGCTCCATGATCACTGGCA ^{ACT} GGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Phosphorothioate-2 Fork2B		Oligo 1: Cy5 5'-GTCGGATCCTCTAGACA ^{GCT} TCCATGATCACTGGCACTGGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Abasic-1 Fork2B		Oligo 1: Cy5 5'-GTCGGATCCTCTAGACAGCTCCATGATCACTG [^] GCACTGGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Abasic-3 Fork2B		Oligo 1: Cy5 5'-GTCGGATCCTCTAGACAGCTC [^] CATGATCACTGGCACTGGTAGAATTCGGC Oligo 2: 5'-CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA

Table 3. Brief structure and chemical modified sites of DNA Fork2 and original oligonucleotides. Note: ↓ indicate chemical modification site; ^A is Methyl-phosphonate modified oligonucleotides; ^{ACT} and ^{GCT} are the Phosphorothioate Oligonucleotides; [^] represent the Abasic (dspace).

2.9 Denaturing gel electrophoresis

10 % polyacrylamide denaturing gel was used to analyse tests based on Cy5 labelled oligo. It contains following reagents: 10 mL Acrylamide (40 % w/v, AccuGel 19:1 was purchased from National Diagnostics), 4 mL of 10X TBE, 16.8 g urea, 2 mL formamide and 9.5 mL H₂O. Mixture was incubated at 37 °C for about 1 to 5 min with shaking till urea dissolved, while the mix still cold, which about 20 to 30 °C, and then add 10 % APS 100 µL and 50 µL TEMED. Washing out all the Acrylamide trapped in wells and warming up gel by pre-running at 5 w per gel for 1 h till gel temperature reach 40-60 °C. Samples were incubated with Stop Buffer (mentioned above) for 5 min and then mixed with 10 µL Formamide

Loading Dye (70 % glycerol, 20 % formamide and Orange G) before loaded into wells. Gel was running for 3 h at 5 w for each gel, till the orange dye reached the bottom of the gel. DNA ladder for running denaturing gel contains 500 nM four 5' Cy5 labelled oligos with different length (85 nt, 49 nt, 30 nt and 20 nt). Loading 4 µL for each time.

2.10 Beer-Lambert law

Formula:

$$A = \epsilon * l * c$$

A: Absorbance

ϵ : Extinction coefficients ($M^{-1} * cm^{-1}$)

l: Pathlength (given as 1 cm)

c: molar concentration ($mol * L^{-1}$)

2.10.1 A₂₈₀ and Protein Concentration

Protein A₂₈₀ was measured by Nanodrop and blanked using Lysis Buffer. Molar absorption coefficient was calculated by uploading protein sequence to an online analysis tool ProtParam (<http://web.expasy.org/protparam/>). Protein data was showed in Table 4.

	Cas3'	H74G&D75G	Cas3
Extinction coefficients ($M^{-1} * cm^{-1}$)	77920	77920	183230
Molecular weight (Da)	30681.46	30543.34	104815.57

Table 4. Protein extinction coefficients and molecular weight.

2.10.2 A₂₆₀ and DNA Substrate Concentration

DNA substrate A₂₆₀ was measured by Nanodrop and blanked by Elution Buffer. Extinction coefficients were calculated by uploading protein sequence to an online analysis tool OligoAnalyzer 3.1 (<https://www.idtdna.com/calc/analyzer>). Data was showed in Table 5.

Name	Extinction coefficients (M ⁻¹ *cm ⁻¹)
Fork2B	967 300
Methyl-phosphonate Fork2B	1 001 500
Phosphorothioate-1 Fork2B	978 600
Phosphorothioate-2 Fork2B	978 600
Abasic-1 Fork2B	978 600
Abasic-3 Fork2B	978 600

Table 5. DNA substrate extinction coefficients.

2.11 Nuclease assays

To analyse whether the purified proteins possessed nuclease or limited nuclease activity, several DNA substrates were supplied to proteins and incubate in different reaction conditions. Reactions were carried out in 15 μ L volumes of 1X buffer O (50 nM Tris-HCl pH7.5, 10 mM MgCl₂, 100 mM NaCl, 0.1 mg/mL BSA) with either 4 nM M13mp18 single stranded circular DNA (M13ssDNA, purchased from NEB) or 25 nM Cy5 labelled DNA substrate (TK12 DNA oligo and DNA forks), and Cas3', Cas3 and Cas3'-H74GD75G separately. Protein was added from 0 nM to 1000 nM and the reactions were incubated at 37 °C for 1 min to 1 h. Reactions were stopped by adding 3 μ L stop buffer (0.5 % SDS, 40 mM EDTA and 2 mg/mL Proteinase K, and then incubated at 37 °C for 5 min. Cleavage products of M13ssDNA were analysed by running a 1% agarose gel (140 V for 50 min), and DNA oligo/forks degradation products were investigated by running a 10% denaturing gel (5 W for 3 h) after mixed samples with 5 μ L Orange G loading dye and 10 μ L Formamide loading dye separately. Agarose gels were imaged by U: Genius3 (Syngene), denaturing gels were scanned by Fuji FLA3000 machine (purchased from FujiFilm). To analyse denaturing gels, GelAnalyser2010a was used to count degradation products amount, and time course graphs were generated by using GraphPad Prism 7.

3 Results

3.1 Cloning of *E. coli* Cas3 and Cas3' proteins

The aim of the project was to begin to assess the mechanism of *E. coli* Cas3 as a translocase and nuclease, by using a forked DNA substrate that is chemically modified in its backbone or bases at precise locations. Therefore, it was necessary to generate purified Cas3 (translocase/nuclease) and Cas3' (nuclease) proteins. Cas3 protein had already been generated in a number of different forms in The Bolt lab, including as a hexa-histidine tagged protein made from arabinose inducible expression from a pBad plasmid (named pAH4, unpublished), as described in the methods. Purification of Cas3 is described below.

Cas3' was generated anew by PCR and cloning of the *cas3'* fragment into pBad/HisA empty vector, generating new recombinant plasmids (named pLH1) for expression of the protein in *E. coli*. PCR reaction products and DNA used for cloning are showed in figure 8. The methods described alignment of Cas3 amino acid sequences that were used to determine primer design for genetic dissection of Cas3 to give isolated Cas3'.

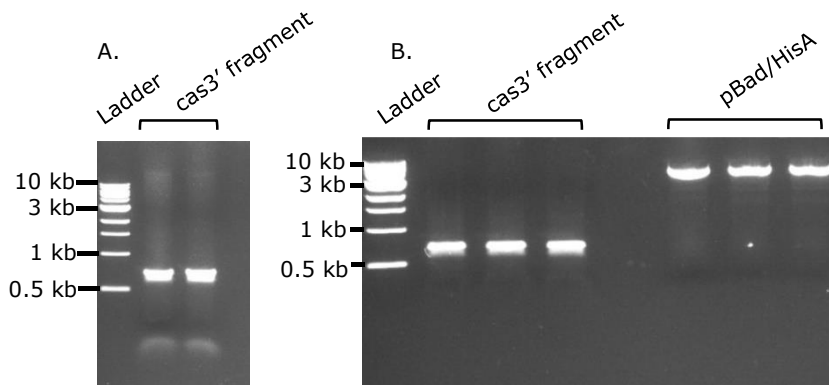


Figure 8. Ethidium bromide stained agarose gels showing DNA used for cloning of *E. coli cas3'*, the nuclease encoding part of Cas3. (A) PCR product from amplification of *cas3'*, consistent with the 713 bp that was expected. (B) *XhoI* and *EcoRI*-HF digested *cas3'* PCR product alongside identically digested pBAD/HisA, for cloning of *cas3'* under the control of the arabinose inducible promoter of the plasmid vector, and for expression of Cas3' protein with a hexahistidine tag. Successful cloning in this way created new plasmid pLH1.

Following successful cloning of *cas3'* to form pLH1, the resultant plasmids for expression of Cas3' and Cas3 were analysed together on a single DNA gel, after appropriate digestion with restriction enzymes (Figure 9).

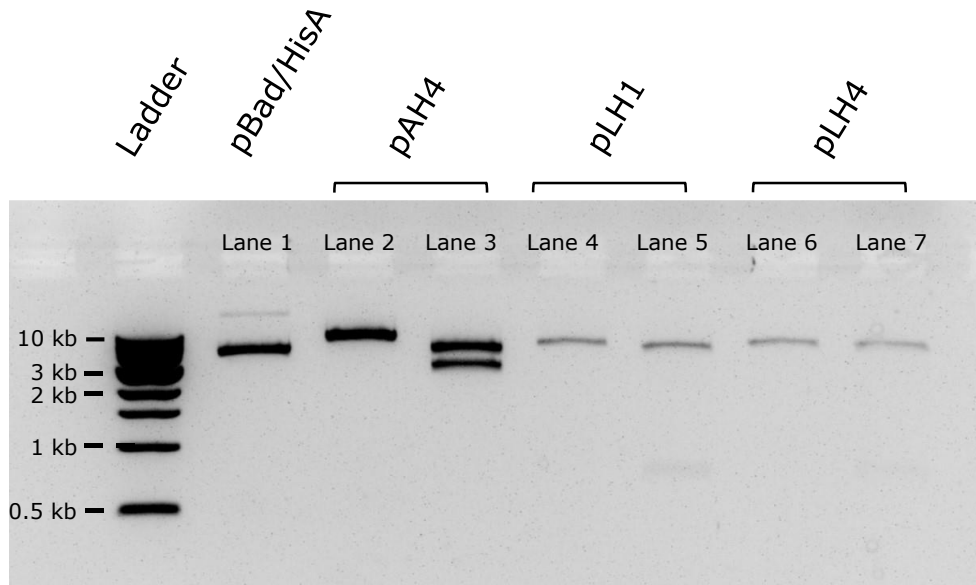


Figure 9. Summary of plasmids used for protein expression in this work. Each plasmid is shown twice as indicated, either as undigested DNA (lanes 2, 4 and 6) or after digestion with appropriate restriction endonucleases, *XhoI* and *EcoRI*-HF, (lanes 3, 5 and 7) to release cloned *cas3* (pAH4), *cas3'* (pLH1) or *cas3'*-H74GD75G (pLH4).

3.2 Mutagenesis

In addition to wild type Cas3 and Cas3' proteins, we also wished to study the effects of targeted mutations on DNA nuclease activities. The mutations made to inactivate nuclease activity were based on previously published ones (Sinkunas et al. 2011; Westra et al. 2012): Residues histidine-74 and aspartate-75 were both altered to glycine. Cas3'-H75GD75G were introduced as described in the methods and resultant proteins were over-expressed (used pLH4 plasmid, figure 9) and purified as for Cas3' wild type, described below.

3.3 Over-expression and purification of *E. coli* Cas3 and Cas3' proteins.

Cloned *E. coli* proteins Cas3 and Cas3' and Cas3'-H74GD75G were expressed for purification with hexahistidine tags, and each protein could be generated in sufficient quantities to allow purification through 5 mL HisTrap HP column. A summary gel showing purified Cas3, Cas3' and Cas3'-H74GD75G proteins is in figure 10. Protein sizes observed in SDS-PAGE were slightly smaller than predicted: Cas3 ran at about 90 kDa rather than 104.8 kDa as expected, while Cas3' and Cas3'-H74GD75G was about 23 kDa instead of about 30 kDa. It was thought that the hexahistidine tag can influence protein gel mobility and cause this error, it was acceptable (Brady et al. 1990; Shirai et al. 2008)

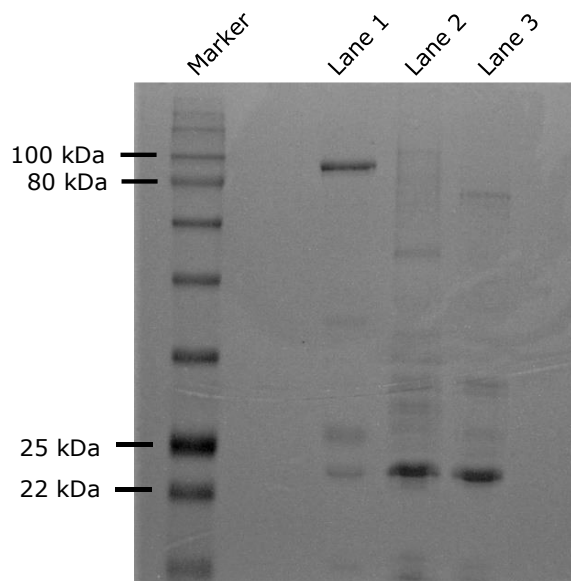


Figure 10. Cas3, Cas3' and Cas3'-H74GD75G proteins. 1 ug of each purified protein was loaded onto a 12.5% acrylamide SDS gel as shown. Contaminating proteins observed in lane 1 (Cas3 are proteolytic breakdown products of Cas3, noted previously. Lanes 2 and 3 shows, respectively, purified Cas3' and Cas3' mutant proteins. These also contained some contaminating proteins, which are discussed later.

3.4 Nuclease activities of purified Cas3 and Cas3' on M13 ssDNA and forked DNA substrates

Cas3 and Cas3' proteins possess a nuclease HD-domain centred on histidine and aspartate (HD) residues, which we also targeted for mutagenesis to generate Cas3'-H74GD75G. Inclusion of the mutant protein in nuclease assays was to act as a negative control for catalytic activity, ensuring that nuclease activity observed from Cas3' is attributable to this protein and not from potentially any contaminating protein. We first assessed the expected nuclease activity of each of these three proteins on two different substrates: M13 closed covalent circular (CCC) single stranded DNA (ssDNA), and a flayed duplex forked DNA molecule that we call FORK-2B (more details were described in methods). M13 ssDNA has been used previously to show activity of various nucleases, including Cas3 and Cas3' (Beloglazova et al. 2011; Sinkunas et al. 2011). FORK-2B is described in detail in the methods, and was chosen because (a) it has both ssDNA and duplex DNA regions for potential nuclease activity, (b) it has 3' ssDNA end, for 3' to 5' directional translocation by Cas3, and (c) because it's small size (two oligonucleotides each of 50 bases) made it amenable and cost-effective to synthesise derivative fork 2 molecules containing precise chemical modifications to the sequences.

Figure 11 showed the outcome of a typical nuclease assay on M13 DNA, showing concentration-dependent nuclease digestion of DNA from Cas3' and Cas3. This result was expected, although we note that Cas3' seems to be a more productive nuclease than Cas3 (compared lane 5 with lane 10 in figure 11). There was much less or no detectable nuclease activity from Cas3'-H74GD75G, which was also expected because this protein has lost two key catalytic residues from the HD domain active site.

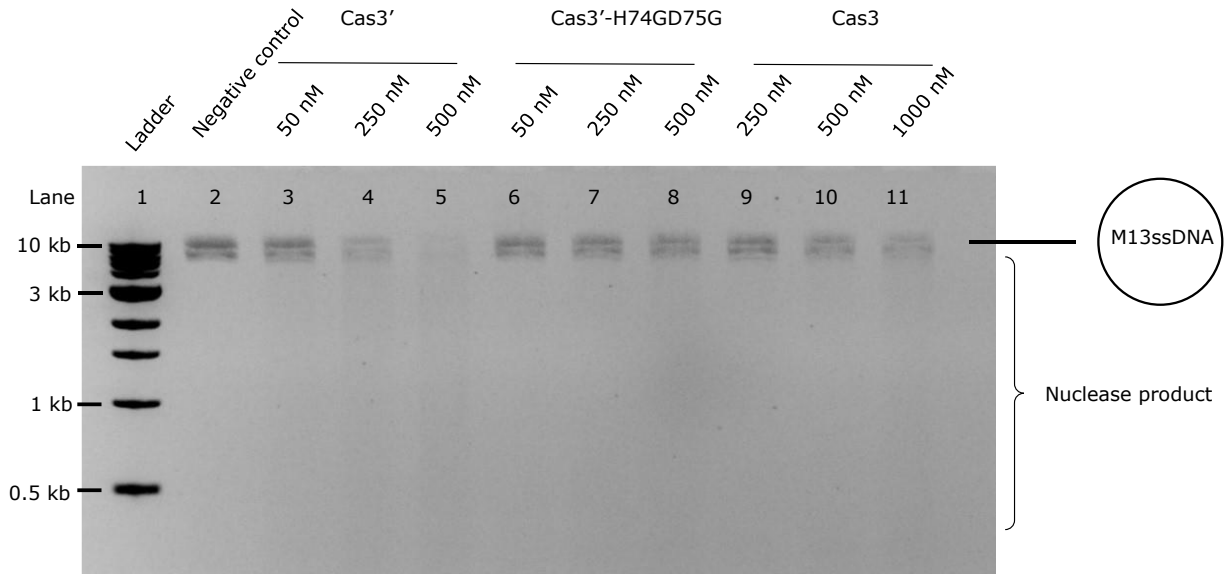


Figure 11. Nuclease activity of Cas3', Cas3'-H74GD75G and Cas3 when assayed with M13ssDNA at 37°C for 1 h in buffered magnesium chloride. Degradation products were analysed by running 1% agarose gel containing ethidium bromide and imaged by UV light.

In these assays on M13 ssDNA, 10 mM Mg²⁺ was used in standard buffer, as has been described before to analyse Cas3 nuclease activities. Because this gave substantial nuclease activity the same conditions were used to next assay for digestion of a DNA oligonucleotide (MW12, 50 nt length) or of DNA in FORK-2B, as preliminary analysis for using chemically modified forks. Note that these nuclease assays did not contain ATP that would be necessary for translocation of Cas3 along DNA. Therefore, as much as possible, we were analysing nuclease activity independently of translocation activity in this work. Cas3 or Cas3' were incubated in increasing concentration with the 50 nucleotides long MW12, which has a Cy5 label located at the 5' end, as indicated in figure 12 A. Both Cas3' and Cas3 gave robust nuclease activity but with significantly different pattern of products (Figure 12A, compare lanes 2-5 to lanes 7-10). Products of Cas3' were overwhelmingly longer than 30 nucleotides, with a trace of shorter, 20 nucleotide products. This is contrasted by products of

Cas3, which were predominantly of the shorter (20 nt) type. More analysis of this is given in the discussion section.

Nuclease activity of Cas3' and Cas3 proteins were next investigated in assays using FORK-2B as a substrate, made by annealing 5'Cy5 MW12 to an unlabelled oligonucleotide (MW14) to generate a flayed duplex that has 25 nt regions of ssDNA and a 25-base pair dsDNA region, shown in figure 12 B and detailed in Methods. Both Cas3' and Cas3 were active as nucleases on this fork, and again we observed differing pattern of nuclease products (Figure 12 B, compare lane 6 with lane 14). Interestingly, products of Cas3'-H74GD75G comprised the removal of only a few nucleotides from the 3'-end of Fork-2B to 5'-end and generated products which about 30 nt to 49 nt (Figure 12 C, lane 7 and 8). It also shows limited nuclease activity compared with Cas3' (Figure 12 C, lane 2 and lane 6). We also analysed nuclease degradation of FORK-2B by Cas3, Cas3' and Cas3'-H74GD75G at fixed concentrations (200 nM) as a function of time (Figure 13 A), with representative gels corresponding to these assays shown in figure 13 B and 13 C.

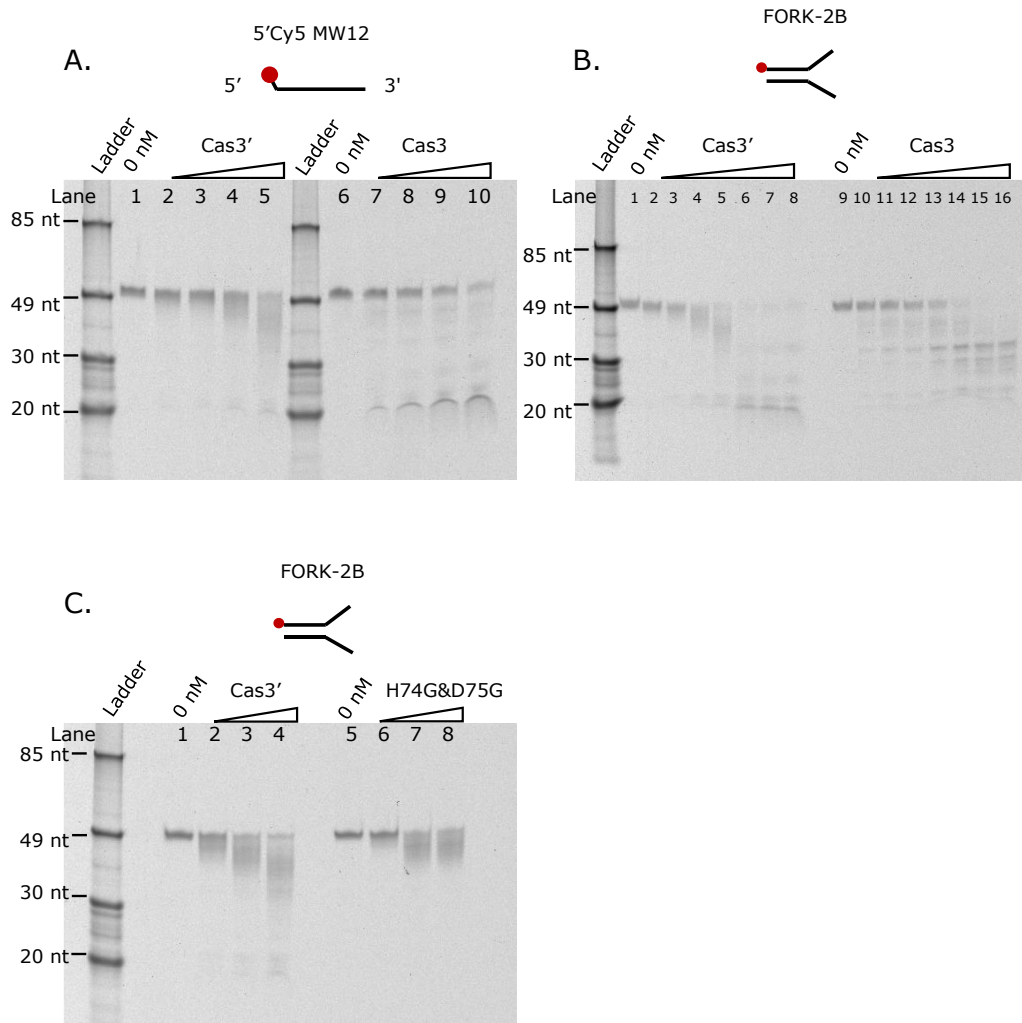
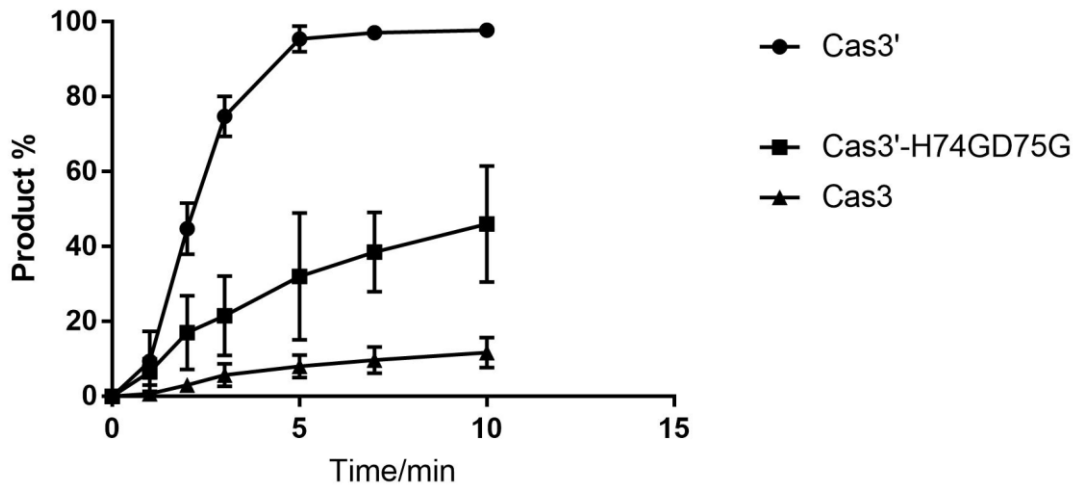
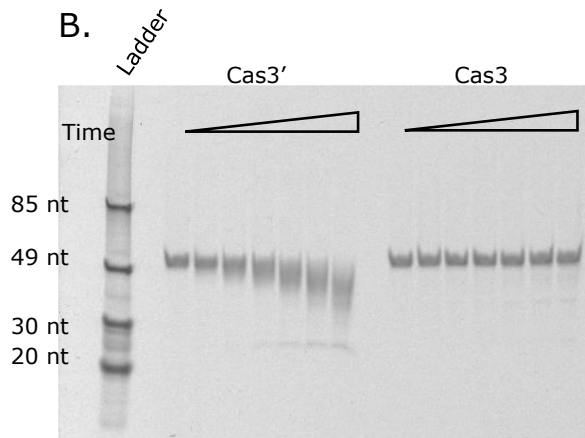


Figure 12. Nuclease cleavage of Cy5 labelled ssDNA or FORK-2B DNA by Cas3', Cas3 and Cas3'-H74GD75G. All reactions were in 10 mM MgCl₂ in reaction buffer detailed in methods. All gels are 10% acrylamide urea denaturing gels, and a marker of known DNA fragments lengths is indicated. (A) and (B) show nuclease activities on, respectively, ssDNA and FORK-2B (each at 25 nM). In (A) Cas3' was included at 0, 25, 50, 100 or 250 nM, Cas3 was included at 0, 100, 250, 500 or 1000 nM. Reactions were taken at 37 °C for 10 min (analyse Cas3' nuclease activity) or 1 h (analyse Cas3 nuclease activity). (B) Cas3' was included at 0 nM, 10 nM, 25 nM, 50 nM, 100 nM, 150 nM, 200 nM or 250 nM, reactions were incubated at 37 °C for 10 min; Cas3 was included at 0 nM, 50 nM, 100 nM, 150 nM, 250 nM, 500 nM, 750 nM, or 1000 nM, reactions were incubated at 37 °C for 1 h. (C) show result of Cas3' and Cas3'-H74GD75G cleave FORK-2B. Both Cas3' and cas3'-H74GD75G were included at 0 nM, 25 nM, 100 nM and 200 nM, 25 nM FORK-2B was exited in each reaction and samples were incubated at 37 °C for 10 min.

A.



B.



C.

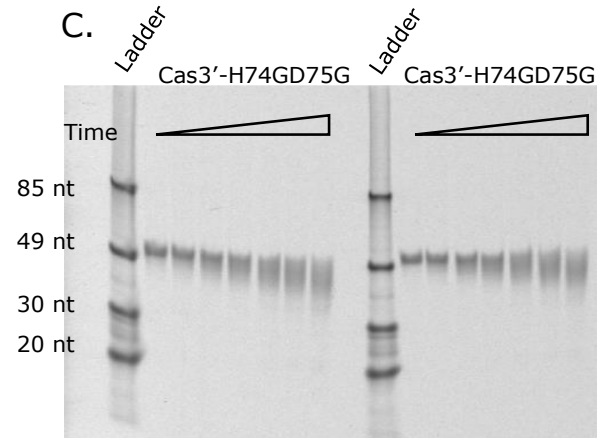


Figure 13. Time course analysis of nuclease cleavage on FORK-2B by Cas3', Cas3 and Cas3'-H74GD75G. All reactions were in 10 mM MgCl₂ in reaction buffer detailed in methods. Samples were taken at 0 min, 1 min, 2 min, 3 min, 5 min, 7 min and 10 min. All gels are 10% acrylamide urea denaturing gels, and a marker of known DNA fragments lengths is indicated. (A) Plot graph shows time course of Cas3', Cas3 and Cas3'-H74GD75G cleaving FORK-2B. 200 nM protein and 25 nM FORK-2B were used in reactions. The average of three independent experiments were showed as values and the error bars represented standard deviation. (B) and (C) are representative gels that were used to generate the data in the graphs shown (B) 200 nM Cas3' and Cas3 were incubated (separately) with 25 nM FORK-2B at 37 °C. (C) 200 nM Cas3'-H74GD75G were incubated with 25 nM FORK-2B at 37 °C.

3.5 The influence of chemically modified DNA on Cas3' nuclease activity at forks

The use of chemically modified DNA substrates can give insights into the molecular mechanism of enzyme function (Raney 2014). We used defined modifications to FORK-2B to first investigate for any effect on nuclease activity of Cas3' (Figure 14). These chemical modifications are named and their positions indicated above the gel panel in figure 14, and are showed in detail in the methods (Table 3). Each assay used the same buffer and magnesium conditions as in figure 12, and the concentration of Cas3' that was used in this assay (200 nM) corresponds to lane 7 in figure 12 B. As expected, Cas3' showed multiple nuclease products from FORK-2B (Figure 14, lane 2). A very similar pattern of nuclease products was observed from Cas3' activity on fork abasic-3 derivative (Figure 14 A, lane 12), but there were some subtle and quite obvious differences in nuclease products of the other chemically modified forks. In particular, degradation products of FORK-2B derivatives phosphorothioate-1 and abasic-1 reflected a strong inhibition of Cas3' exonuclease activity, with accumulation of a single nuclease product of about 30 nt in length. This product would correspond to the position of the respective chemical modifications within the ssDNA region of these forks, indicating that these modifications are substantially inhibiting the nuclease processing of ssDNA by Cas3', as is discussed more later. The effect of the modifications in FORK-2B derivatives methylphosphonate-2 and phosphorothioate-2 were less pronounced (Figure 14, lane 4 and 8), however, there was a very light band in lane 4 (between 20 nt and 30 nt) and in lane 8 (about 20 nt). Although the chemical modifications affected the pattern of cleavage products, there was little perceivable difference in *overall* nuclease activity of Cas3'. This can be explained by that chemical modification was believed can block helicase/translocase activity rather than nuclease activity, so that Cas3' still can identify DNA substrate while tolerate the

chemical changes and then cleave single stranded DNA from 3'-end (Figure 15 A) and is discussed more later.

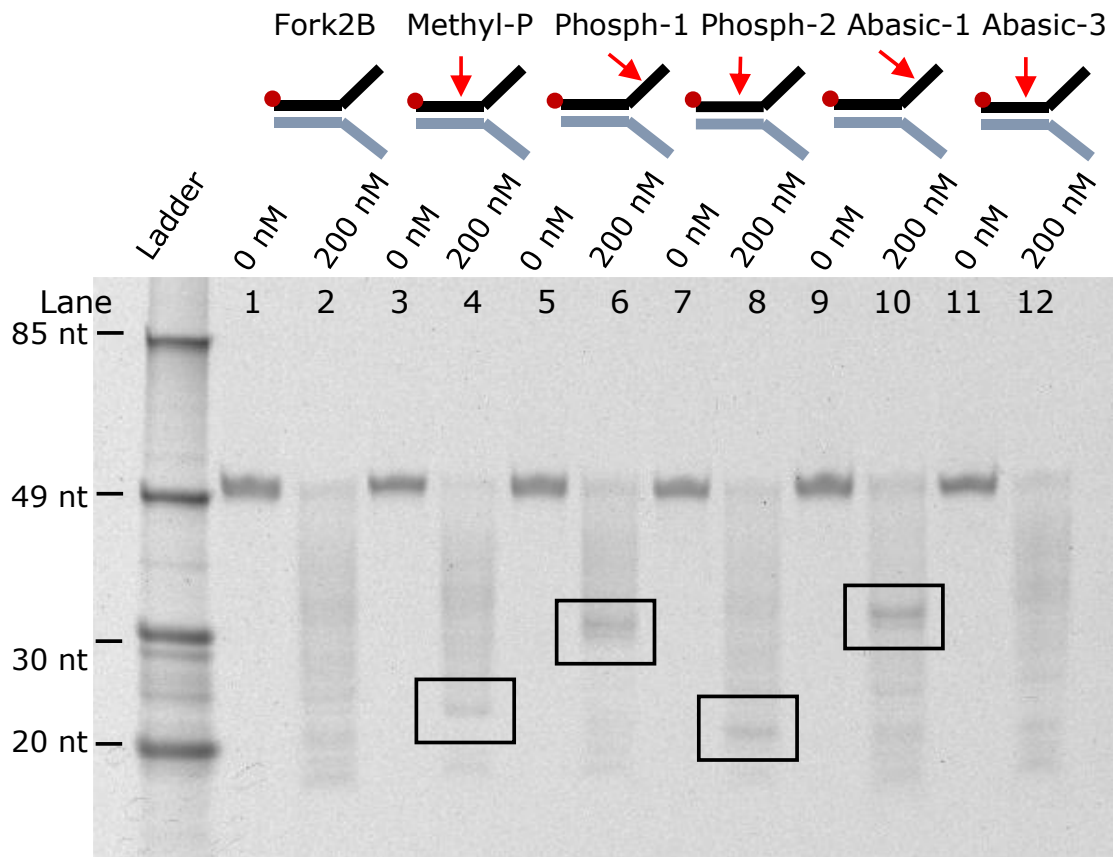


Figure 14. Cas3' nuclease activity assessed on chemically modified forks. All reactions were in 10 mM MgCl₂ in reaction buffer detailed in methods. 200 nM Cas3' was incubated with 25 nM FORK-2B or its variants separately. Reactions were incubated at 37°C for 10 min and results were visualized by running 10% acrylamide urea denaturing gel. Note: Methyl-phosphonate FORK-2B, phosphorothioate-1 FORK-2B, phosphorothioate-2 FORK-2B, abasic-1 FORK-2B and abasic-3 FORK-2B was abbreviated to Methyl-P, Phosph-1, Phosph-2, Abasic-1 and Abasic-3 and marked above the image with general structures (Cy5 label showed as red dot, which arrow pointed out the modified location). More DNA fork details were described in methods.

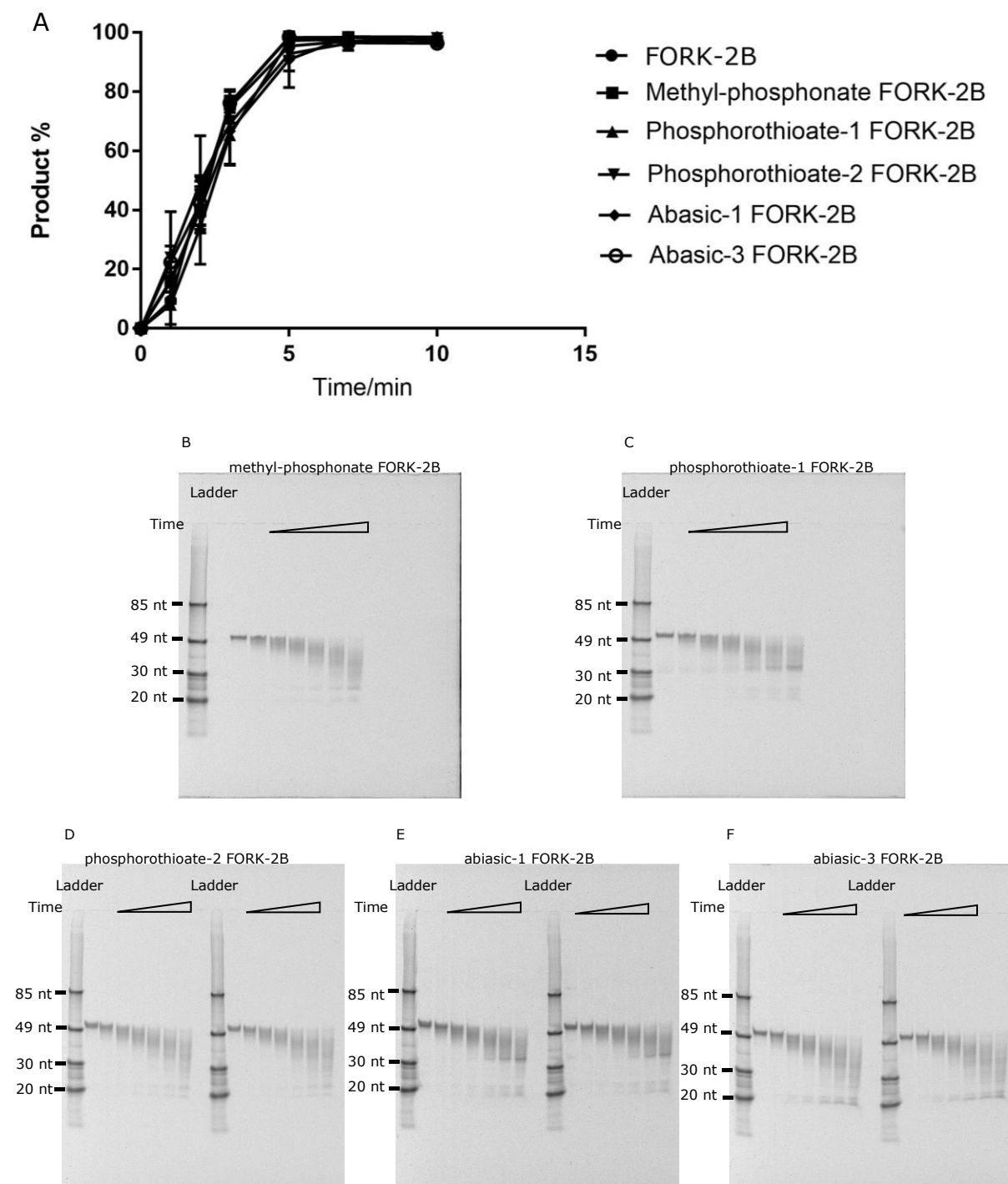


Figure 15. Time course of Cas3' cleave chemically modified forks. All reactions were in 10 mM MgCl₂ in reaction buffer detailed in methods. 200 nM Cas3' was incubated with 25 nM FORK-2B or its variants separately. Reactions were incubated at 37 °C and samples were collected at 0 min, 1 min, 2 min, 3 min, 5 min, 7 min and 10 min. Results were visualized by running 10% acrylamide urea denaturing gel. (A) Plot graph shows time course of Cas3' cleaving FORK-2B and its derivatives. The average of three independent experiments were showed as

values and the error bars represented standard deviation. (B) (C) (D) (E) and (F) were part of the time course data, which generated plot graph above. As marked above each image, (B) shows Cas3' cleave methyl-phosphonate FORK-2B; (C) shows Cas3' cleave phosphorothioate-1 FORK-2B; (D) shows Cas3' cleave phosphorothioate-2 FORK-2B; (E) shows Cas3' cleave abasic-1 FORK-2B; (F) shows Cas3' cleave abasic-3 FORK-2B.

3.6 The influence of chemically modified DNA on Cas3 nuclease activity at forks

As described in figure 12 B that Cas3 was able to cleave FORK-2B, then it was incubated with chemically modified forks separately in same reaction condition (10 nM Mg^{2+} and no ATP) matched the assays in lane 13 (250 nM Cas3) and lane 16 (1000 nM Cas3) (Figure 12 B). This experiment was to analyse the present of helicase/ translocase interacts with nuclease domain (showed as Cas3', figure 14) and various type of chemical modifications.

Cas3 cleaved different forks and generated multiple length of products as indicated by figure 16. Compared lane 6, lane 12 and lane 18 with lane 3 (non-modified FORK-2B), the modifications in FORK-2B duplex part did not have obvious influence on products generation. However, chemical modifications on single-stranded flaps showed strong blocking capacity on Cas3 translocation from 3'-end to 5', for example, phosphorothioate restriction reflected by a clear band, which about 30 nt and marked in lane 9, and similarly, abasic sites blocking showed as a diffusion zone between 30 nt and 49 nt (more analysis in discussion section). Though chemical modifications on ssDNA had impact on Cas3 translocation, it seems that they did not have obvious influence on total nuclease activity of Cas3 (Figure 17 A).

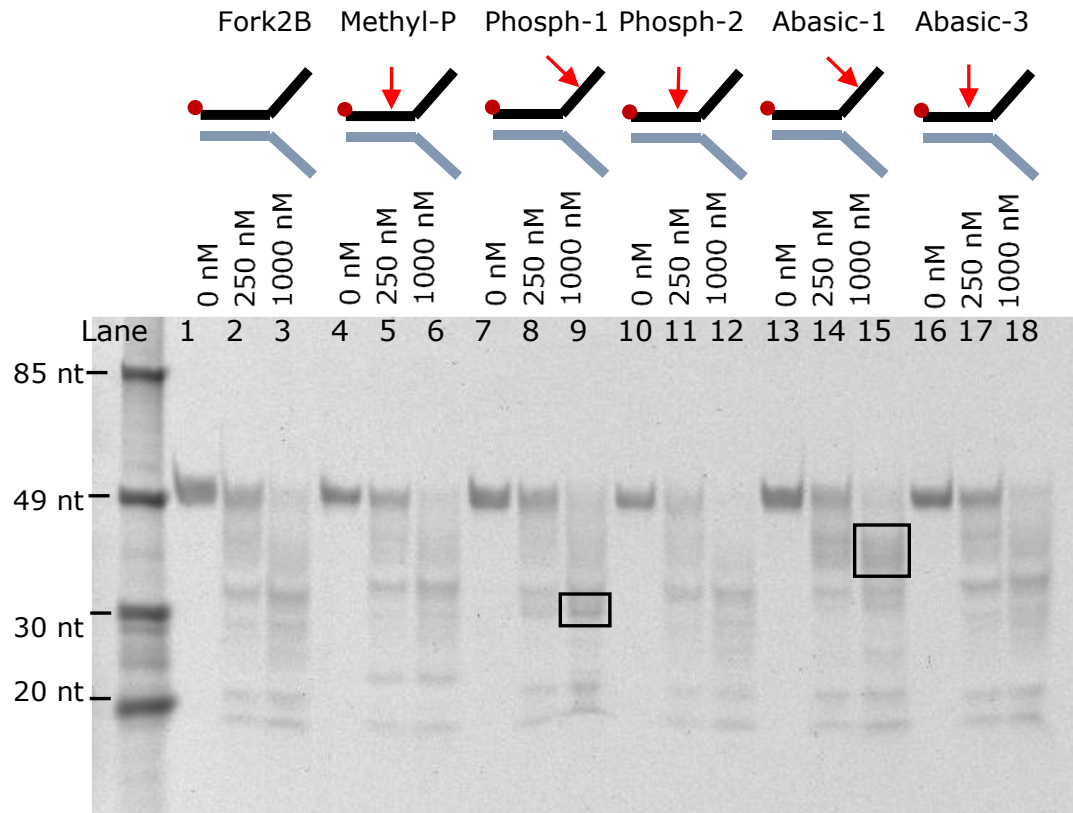


Figure 16. Cas3 cleave chemically modified forks. All reactions were in 10 mM $MgCl_2$ in reaction buffer detailed in methods. 250 or 1000 nM Cas3 was incubated with 25 nM FORK-2B or its variants separately. Reactions were incubated at 37 °C for 1 h and results were visualized by running 10% acrylamide urea denaturing gel. Note: Methyl-phosphonate Fork2B, Phosphorothioate-1 Fork2B, Phosphorothioate-2 Fork2B, Abasic-1 Fork2B and Abasic-3 Fork2B was abbreviated to Methyl-P, Phosph-1, Phosph-2, Abasic-1 and Abasic-3 and marked above the image. For the brief fork structure (details showed in Table 3), Cy5 labelled oligo marked as black, blue part is unlabelled oligo.

A

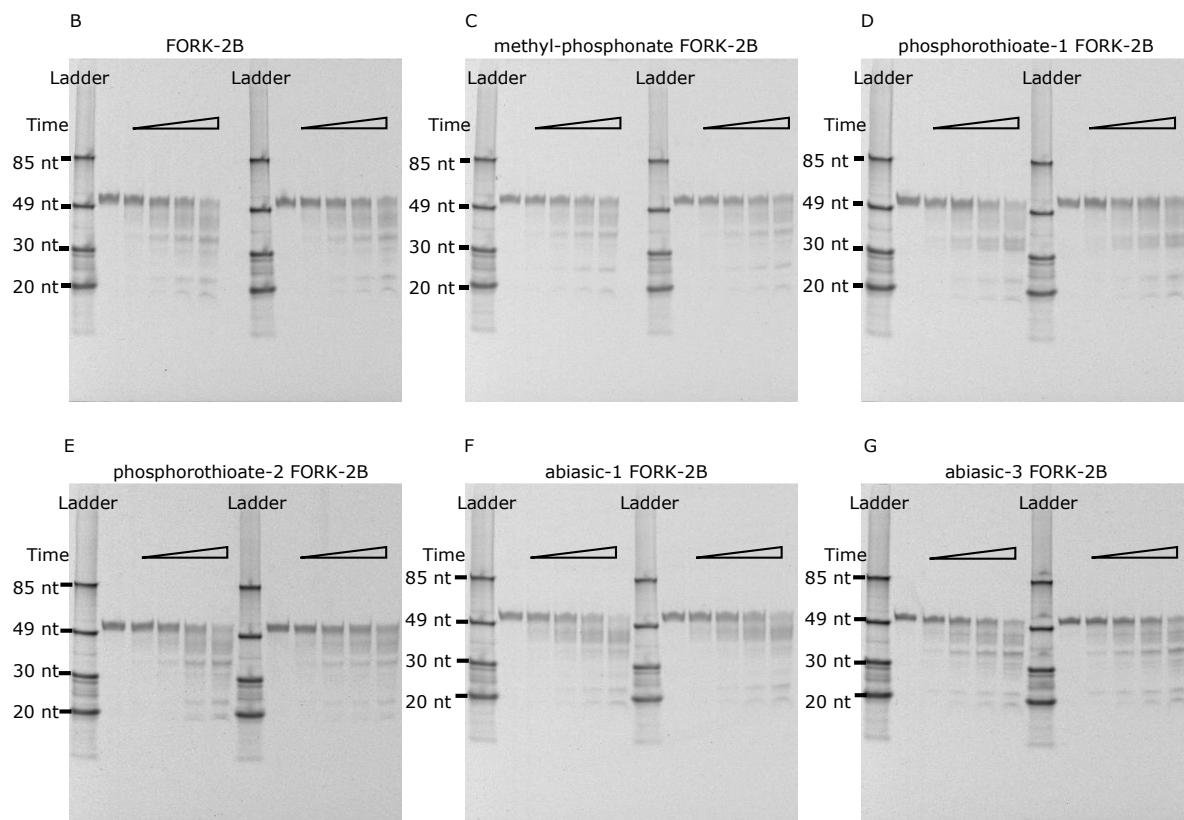
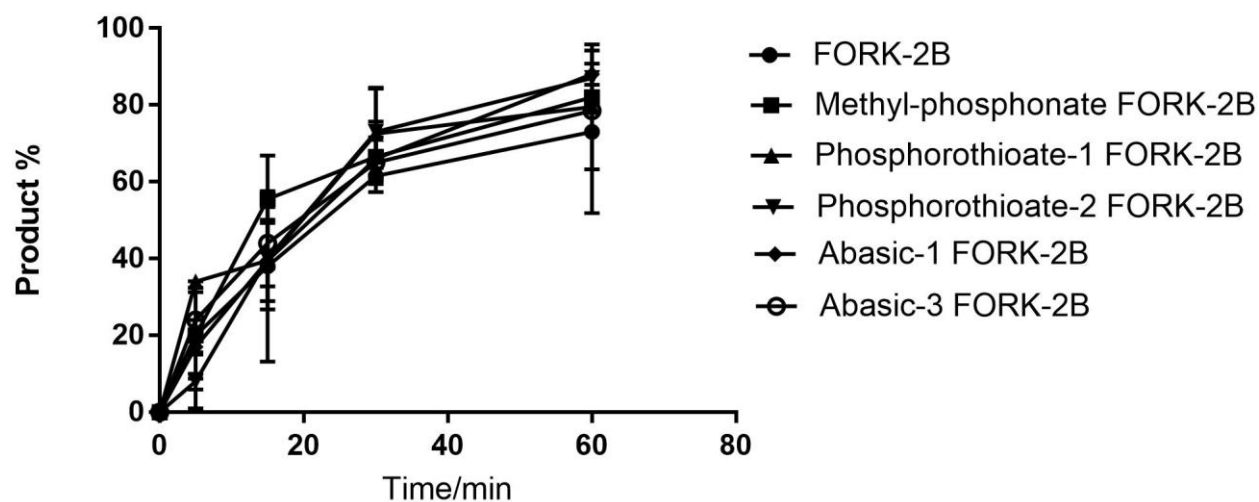


Figure 17. Time course of Cas3 cleave chemically modified forks. All reactions were in 10 mM MgCl₂ in reaction buffer detailed in methods. 1000 nM Cas3 was incubated with 25 nM FORK-2B or its variants separately. Reactions were incubated at 37 °C and samples were collected at 0 min, 5 min, 15 min, 30 min and 60 min. Results were visualized by running 10% acrylamide urea denaturing gel. (A) Plot graph shows time course of Cas3 cleaving FORK-2B and its

derivatives. The average of three independent experiments were showed as values and the error bars represented standard deviation. (B) (C) (D) (E) (F) and (G) were part of data, which generated plot graph above. (B) shows Cas3 cleave Fork2B. (C) shows Cas3 cleave Methyl-phosphonate Fork2B. (D) shows Cas3 cleave Phosphorothioate-1 Fork2B. (E) shows Cas3 cleave Phosphorothioate-2 Fork2B. (F) shows Cas3 cleave Abasic-1 Fork2B. (G) shows Cas3 cleave Abasic-3 Fork2B.

4 Discussion

This work describes the preliminary analysis of the nuclease mechanism of Cas3, as analysed using DNA substrates containing methyl-phosphonate, phosphorothioate or abasic chemical modifications to DNA at precise locations on DNA strands. The analyses included separation and isolation of the HD-like nuclease domain protein of Cas3 (Cas3') away from the superfamily 2 helicase domains, to be able to assess the catalytic nuclease mechanism of Cas3 directly without the influence of the helicase activity. The data obtained can now form the basis of further investigations into the catalytic mechanism of Cas3 as a nuclease and helicase.

4.1 Cas3, Cas3' and Cas3'-H74GD75G purifications

Cas3 and Cas3' proteins were purified in a single column step using Ni²⁺-NTA affinity resin. We included purification of mutant Cas3' protein, Cas3'-H74GD75G, as a control because that was expected to have no, or very little nuclease activity based on studies of Cas3 in other bacteria (Beloglazova et al. 2011; Sinkunas et al. 2011). Our results (Figure 11) indicated that some nuclease activity is apparent from Cas3'-H74GD75G, although much reduced compared to "wild-type" Cas3'. We cannot rule out that one or more contaminating proteins visible in figure 10 may contribute to or cause nuclease activity observed in assays containing Cas3'-H74GD75G. In future preparations of Cas3' it should be possible to remove more contaminating proteins by judicious use of additional FPLC columns, after the Ni²⁺-NTA. This should help to determine if there really

is some residual nuclease activity present in Cas3-H74GD75G. However, the overall outcomes of experiments with chemically modified substrates are not significantly affected by this, because the pattern of nuclease products formed seems to be independent of the weak nuclease trimming by Cas3'-H74GD75G. Contaminants in the purified Cas3 were somewhat different in nature from those of Cas3', appearing mainly after dialysis and storage of the purified protein, and are thought to be proteolytic products of the full-length Cas3 enzyme. Mass spectrometry of both major Cas3' and Cas3 protein bands confirmed their identities.

4.2 Nuclease activities of Cas3 and Cas3' on unmodified and modified DNA FORK-2B

Nucleases catalyse breakage of phosphodiester bonds in a nucleic acid backbone. DNA degradation by nucleases may have directionality arising from their catalytic mechanisms, either from 5' to 3' or from 3' to 5' according to function. The latter is utilized for proofreading activities during DNA replication that remove mismatched base pairs (Reha-Krantz 2010). Degradation of DNA by nucleases relies on interactions of the nuclease with the negatively charged phosphodiester backbone of DNA and with nitrogenous bases, to generate products with 5'-phosphate and 3'-OH ends (Yang 2011). Cas3 and Cas3' (HD nuclease domain) show 3' to 5' degradation polarity against single stranded DNA within circular molecules (e.g. M13) or ssDNA with free ends (Beloglazova et al. 2011; Mulepati and Bailey 2011; Sinkunas et al. 2011). The ATP-dependent helicase activity of Cas3 is also an important characteristic of overall functionality of Cas3 activities in CRISPR-Cas immunity (Howard et al. 2011; Huo et al. 2014; R. N. Jackson et al. 2014a), but in this study, we have focussed on the nuclease activities of Cas3, and its derivative Cas3'.

Cas3 and Cas3' nuclease activity was effective on M13 ssDNA, oligonucleotides and FORK-2B and this was used as the basis for establishing conditions for studying nuclease activities on chemically modified FORK-2B. Cas3 and Cas3' acted differently when incubating with

5'-end labelled ssDNA oligo (5'Cy5 MW12) compared with DNA fork (FORK-2B). Reactions with MW12 for both Cas3 and Cas3' nucleases were efficient at generating multiple products of variable DNA length when viewed on a urea gel, but DNA products of <20 nt were not observed. The major products types of Cas3' and Cas3 on this ssDNA substrate were markedly different, with Cas3' generating much shorter products than Cas3. Similarly, reaction products of Cas3 and Cas3' generated from FORK-2B differed but in the opposite way to those of MW12; in this case Cas3' was able to generate shorter products than Cas3 (Figure 12 B). There is no ATP present in these assays and so translocation by Cas3 along DNA whilst also being a nuclease cannot be the cause. Instead, the differing products observed might be caused by differences in DNA binding positioning and activities by Cas3 compared to Cas3'. Further work in the future to analyse by anisotropy and EMSAs the relative binding of these two proteins to FORK-2B may help to clarify this.

However, the main objective of the project was to compare Cas3 or Cas3' nuclease activities apparent from FORK-2B with their products from chemically modified derivatives of FORK-2B. Modifications to the DNA backbone that were investigated for effect on nuclease activity were methyl-phosphonate (MeP) and phosphorothioate (Pthio) moieties. Because nucleases may require formatting a highly negatively charged penta-covalent intermediate (Yang 2011) from the phosphodiester backbone of DNA, local loss of charge from a methyl-phosphonate (MeP) can inhibit nuclease degradation (Rangarajan and Shankar 2001; Reynolds et al. 1996). In Pthio substrates the non-bridging oxygen atom of the backbone phosphate is replaced by sulfur in a phosphorothioate bond, which can negatively impact on nuclease cleavage by inhibiting hydrolysis of neighbouring phosphate groups. A DNA backbone phosphorothioate also induces alteration to the local positioning of DNA relative to the protein active site, which inhibits cleavage activity of nucleases (Brautigam and Steitz 1998).

Potentially of interest was that nuclease activities of Cas3' were inhibited by MeP but those of Cas3 did not seem to be impacted (Figure 14 lane 4, figure 16 lane 5 and 6). The position of the major "stall" product of Cas3' nuclease activity on MeP FORK-2B is consistent with the position of the MeP at 25 nt from the Cy5 end of the oligo. Perhaps in Cas3 multiple interactions with the DNA backbone overcome a single MeP but in Cas3' those additional interactions are lost therefore causing a stall. Nevertheless, this result does indicate that the Cas3' nuclease component of Cas3 requires interaction with DNA backbone charge to achieve nuclease activity. Similarly, Cas3' nuclease activity was inhibited by a Pthio in the same position, but Cas3 activity was not. However, we cannot rule out the possibility that the reason we did not observe inhibition of Cas3 by MeP or Pthio at these positions is because Cas3 did not digest DNA as far as that position in FORK-2B in any event.

Most significantly, both Cas3' and Cas3 nuclease activities were strongly inhibited by inclusion of a single abasic site in the ssDNA tail region of FORK-2B (Figure 14 lane 10 and figure 16 lane 14, 15). This modification is located closer to the 3' end of this DNA strand and it seems significant that this ssDNA region may be where Cas3' or Cas3 load into the FORK-2B substrate. Inhibition of Cas3 for this reason would not be surprising given that it is a 3' to 5' ssDNA translocase, but the effect of this abasic site on Cas3' activity is more surprising. Nuclease degradation of DNA requires nucleophilic attack of the scissile phosphate bond (Yang 2011). It is likely that in Cas3' the acidic amino acid side chains interact with water and magnesium ion(s) to generate the nucleophile that is required. The surface charge distribution of Cas3' indicates that it has a negatively charged groove corresponding to active site His- and Asp- residues (Figure 18, coloured in red). It is predicted that these Cas3' active sites catalyse the nucleolytic attack mediating 3'phosphoate and 5' -OH generation. This process might be inhibited by transforming achiral phosphate into chiral type through replacing one non-bridging oxygen

with methyl (methyl-phosphonate) or sulphur (phosphorothioate). In addition, interaction between neutral DNA backbone and charges Cas3' activity sites might also disturb nucleophiles reaching 5' O-P phosphodiester bond. Inhibition of Cas3' by an abasic site may also be related to changes in the DNA backbone interaction with the enzyme, because loss of the base causes local structural perturbation of DNA. However, it could also be that the abasic site helps Cas3' (and Cas3) to form stable bound complex on DNA, and that the loss of the base reduces binding affinity. EMSA and anisotropy methods are proposed to study in more detail the effects of these chemically modified substrates on Cas3' and Cas3 activity.

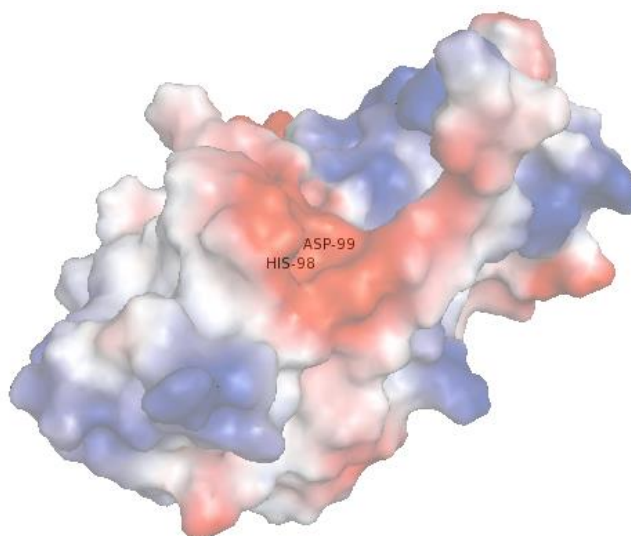


Figure 18. Image illustrating structural features of *Thermobaculum terrenum* Cas3 HD domain taken from PDB 4Q2C. The surface charge distribution is displayed in different colour, negatively charged area labelled in red while positively charged area coloured in blue. Two active site residues in Cas3 HD core catalytic site (red area) are labelled as His-98 and Asp-99.

This study displays various DNA degradation pattern generated by Cas3 and Cas3', which indicate that Cas3 nuclease function might be regulated by other components (e.g. the SF2 helicase domain). And the inhibition of chemical modification suggests Cas3 translocase process and it is different from Cas3'. Nuclease/translocase Cas3 is critical to CRISPR

adaptation and priming, it cooperates with Cascade, Cas1-Cas2 and other factors (Kunne et al. 2016; Moch et al. 2017). Previous works refer that Cas3 is recruited by Cascade conformational dynamics changes, which can trigger nuclease-active Cas3 or nuclease-repressed Cas3 depend on different conformation (van Erp et al. 2017). However, the pathway of Cas3 activity change is still unknown. We observed that Cas3 HD originally has high endonuclease activity, but Cas3 displays endonuclease activity with lower efficiency. This phenomenon might indicate SF2, or part of it, determines Cas3 nuclease type or even efficiency by participating Cas3 conformational dynamics. Except Cascade, Cas1-Cas2 can also recruit Cas3 and regulates Cas3 function (Redding et al. 2015). Therefore, future work could focus on Cas3 mechanism of DNA degradation by adding ATP in reactions or generating Cas3 SF2. And it is also attractive that study Cas3 interaction with other Cas-proteins based on kinetic analysis.

5 Bibliography

- Amitai, G. and Sorek, R. (2016), 'CRISPR-Cas adaptation: insights into the mechanism of action', *Nat Rev Microbiol*, 14 (2), 67-76.
- Barrangou, Rodolphe, et al. (2007), 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315 (5819), 1709-12.
- Beloglazova, N., et al. (2011), 'Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference', *EMBO J*, 30 (22), 4616-27.
- Beloglazova, N., et al. (2008), 'A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats', *J Biol Chem*, 283 (29), 20361-71.
- Bondy-Denomy, J., et al. (2015), 'Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins', *Nature*, 526 (7571), 136-9.
- Brautigam, Chad A and Steitz, Thomas A (1998), 'Structural principles for the inhibition of the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates', *Journal of molecular biology*, 277 (2), 363-77.
- Brouns, Stan JJ, et al. (2008), 'Small CRISPR RNAs guide antiviral defense in prokaryotes', *Science*, 321 (5891), 960-64.
- Burstein, D., et al. (2017), 'New CRISPR-Cas systems from uncultivated microbes', *Nature*, 542 (7640), 237-41.

- Charpentier, E., et al. (2015), 'Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity', *FEMS Microbiol Rev*, 39 (3), 428-41.
- Chylinski, K., Le Rhun, A., and Charpentier, E. (2013), 'The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems', *RNA Biol*, 10 (5), 726-37.
- Deltcheva, E., et al. (2011), 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature*, 471 (7340), 602-7.
- Fagerlund, R. D., et al. (2017), 'Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex', *Proc Natl Acad Sci U S A*, 114 (26), E5122-E28.
- Fonfara, I., et al. (2016), 'The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA', *Nature*, 532 (7600), 517-21.
- Gasiunas, G., et al. (2012), 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proc Natl Acad Sci U S A*, 109 (39), E2579-86.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007), 'The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats', *BMC Bioinformatics*, 8, 172.
- Haft, D. H., et al. (2005), 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes', *PLoS Comput Biol*, 1 (6), e60.
- Hale, C. R., et al. (2009), 'RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex', *Cell*, 139 (5), 945-56.
- Hayes, R. P., et al. (2016), 'Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*', *Nature*, 530 (7591), 499-503.
- Horvath, P., et al. (2008), 'Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*', *J Bacteriol*, 190 (4), 1401-12.
- Horvath, Philippe and Barrangou, Rodolphe (2010), 'CRISPR/Cas, the immune system of bacteria and archaea', *Science*, 327 (5962), 167-70.
- Howard, J. A., et al. (2011), 'Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein', *Biochem J*, 439 (1), 85-95.
- Huo, Y., et al. (2014), 'Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation', *Nat Struct Mol Biol*, 21 (9), 771-7.
- Ivancic-Bace, I., et al. (2015), 'Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity', *Nucleic Acids Res*, 43 (22), 10821-30.
- Jackson, R. N., et al. (2014a), 'Fitting CRISPR-associated Cas3 into the helicase family tree', *Curr Opin Struct Biol*, 24, 106-14.
- Jackson, Ryan N, et al. (2014b), 'Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*', *Science*, 345 (6203), 1473-79.

- Jansen, Ruud, et al. (2002), 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular microbiology*, 43 (6), 1565-75.
- Karvelis, T., et al. (2013), 'crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*', *RNA Biol*, 10 (5), 841-51.
- Killelea, T. and Bolt, E. L. (2017), 'CRISPR-Cas adaptive immunity and the three Rs', *Biosci Rep*, 37 (4).
- Koonin, E. V., Makarova, K. S., and Zhang, F. (2017), 'Diversity, classification and evolution of CRISPR-Cas systems', *Curr Opin Microbiol*, 37, 67-78.
- Kunne, T., et al. (2016), 'Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation', *Mol Cell*, 63 (5), 852-64.
- Kurreck, Jens (2003), 'Antisense technologies. Improvement through novel chemical modifications', *European Journal of Biochemistry*, 270 (8), 1628-44.
- Levy, A., et al. (2015), 'CRISPR adaptation biases explain preference for acquisition of foreign DNA', *Nature*, 520 (7548), 505-10.
- Makarova, K. S., et al. (2006), 'A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action', *Biol Direct*, 1, 7.
- Makarova, K. S., et al. (2011), 'Evolution and classification of the CRISPR-Cas systems', *Nat Rev Microbiol*, 9 (6), 467-77.
- Makarova, K. S., et al. (2015), 'An updated evolutionary classification of CRISPR-Cas systems', *Nat Rev Microbiol*, 13 (11), 722-36.
- Moch, C., et al. (2017), 'DNA binding specificities of *Escherichia coli* Cas1-Cas2 integrase drive its recruitment at the CRISPR locus', *Nucleic Acids Res*, 45 (5), 2714-23.
- Mojica, F. J., et al. (2009), 'Short motif sequences determine the targets of the prokaryotic CRISPR defence system', *Microbiology*, 155 (Pt 3), 733-40.
- Mojica, FJM, Juez, G, and Rodriguez-Valera, F (1993), 'Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites', *Molecular microbiology*, 9 (3), 613-21.
- Mulepati, S. and Bailey, S. (2011), 'Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3)', *J Biol Chem*, 286 (36), 31896-903.
- Musharova, O., et al. (2017), 'Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation', *Nucleic Acids Res*, 45 (6), 3297-307.
- Nunez, J. K., et al. (2015), 'Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity', *Nature*, 519 (7542), 193-8.
- Nunez, J. K., et al. (2016), 'CRISPR Immunological Memory Requires a Host Factor for Specificity', *Mol Cell*, 62 (6), 824-33.

- Nunez, J. K., et al. (2014), 'Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity', *Nat Struct Mol Biol*, 21 (6), 528-34.
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005), 'CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology*, 151 (Pt 3), 653-63.
- Raney, K. D. (2014), 'Chemical modifications of DNA for study of helicase mechanisms', *Bioorg Med Chem*, 22 (16), 4399-406.
- Rangarajan, E Srinivasan and Shankar, Vepatu (2001), 'Sugar non-specific endonucleases', *FEMS microbiology reviews*, 25 (5), 583-613.
- Rath, D., et al. (2015), 'The CRISPR-Cas immune system: biology, mechanisms and applications', *Biochimie*, 117, 119-28.
- Redding, S., et al. (2015), 'Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System', *Cell*, 163 (4), 854-65.
- Reeks, J., Naismith, J. H., and White, M. F. (2013), 'CRISPR interference: a structural perspective', *Biochem J*, 453 (2), 155-66.
- Reha-Krantz, L. J. (2010), 'DNA polymerase proofreading: Multiple roles maintain genome stability', *Biochim Biophys Acta*, 1804 (5), 1049-63.
- Reynolds, Mark A, et al. (1996), 'Synthesis and thermodynamics of oligonucleotides containing chirally pure R P methylphosphonate linkages', *Nucleic acids research*, 24 (22), 4584-91.
- Richter, C., Chang, J. T., and Fineran, P. C. (2012), 'Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems', *Viruses*, 4 (10), 2291-311.
- Rollie, C., et al. (2015), 'Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition', *Elife*, 4.
- Sarin, Prem S, et al. (1988), 'Inhibition of acquired immunodeficiency syndrome virus by oligodeoxynucleoside methylphosphonates', *Proceedings of the National Academy of Sciences*, 85 (20), 7448-51.
- Sinkunas, T., et al. (2011), 'Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system', *EMBO J*, 30 (7), 1335-42.
- Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013), 'CRISPR-mediated adaptive immune systems in bacteria and archaea', *Annu Rev Biochem*, 82, 237-66.
- Stella, S., Alcon, P., and Montoya, G. (2017), 'Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage', *Nature*, 546 (7659), 559-63.
- van der Oost, J., et al. (2014), 'Unravelling the structural and mechanistic basis of CRISPR-Cas systems', *Nat Rev Microbiol*, 12 (7), 479-92.
- van Erp, P. B. G., et al. (2017), 'Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-guide Cascade Complex', *ACS Chem Biol*.

- Wang, R., et al. (2011), 'Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage', *Structure*, 19 (2), 257-64.
- Westra, E. R., et al. (2012), 'CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3', *Mol Cell*, 46 (5), 595-605.
- Yang, W. (2011), 'Nucleases: diversity of structure, function and mechanism', *Q Rev Biophys*, 44 (1), 1-93.
- Ye, Y. and Zhang, Q. (2016), 'Characterization of CRISPR RNA transcription by exploiting stranded metatranscriptomic data', *RNA*, 22 (7), 945-56.
- 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 (12), 5569-76.
- Zetsche, B., et al. (2015), 'Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system', *Cell*, 163 (3), 759-71.

6 Media Recipes

6.1 Luria-Bertani Broth (LB) and LB Agar

For 1 L LB broth, contains 10.0g tryptone, 10.0g NaCl, 5.0g yeast extract and 2ml 1M NaOH solution, made up to final volume of 1 L by adding distilled water.

6.2 10X Tris Borate EDTA (TBE) Buffer

For 1 L TBE buffer, contains 121.1g tris base, 61.8g boric acid and 7.4g EDTA solution, made up to final volume of 1 L by adding distilled water. Dilute to 1 X TBE running buffer when needed.

6.3 10X SDS PAGE Buffer

For 1 L TBE buffer, 30g tris base, 144g glycine and 100ml 10% SDS, made up to final volume of 1 L by adding distilled water. Dilute to 1 X TBE running buffer when needed.