

**Cas3 nuclease activity is regulated by an  
iHDA1 region located at the interface  
between HD and RecA1 domain**

Liu He



**University of  
Nottingham**  
UK | CHINA | MALAYSIA

This thesis is submitted for the degree of Doctor of  
Philosophy at the University of Nottingham

February 2022

## **Abstract**

CRISPR Cas3 is a crucial protein effector for type I CRISPR interference. It couples actions of helicase and nuclease domains to eliminate mobile genetic elements (MGEs) in response to the recognition of protospacer sequences on foreign DNA by the Cascade complex that recruits Cas3. The DNA cleavage efficacy of Cas3 determines whether or not the host cell survives infection caused by MGEs. Cascade stimulates Cas3 DNA hydrolysis activity, and Cas3 possesses structure and functions characteristics that are Cascade-independent and may be additionally modulated by other factors. These challenges previous understanding of Cas3 mechanism in CRISPR interference and in potentially other aspects of bacterial physiology.

This work reported in this thesis investigated Cas3 structure-function. A crucial Cas3 region, termed iHDA1, is highlighted for its role in determining whether DNA can access the Cas3 HD nuclease sites. This iHDA1 region accommodates the Trp-406 amino acid that can cause a steric block in the DNA channel within Cas3; therefore, we define this tryptophan amino acid as a 'gate'. Extrinsic factors also impact Cas3 function, we propose by stimulating the iHDA1 region to undergo conformational changes. As reported here, these are temperature, the CRISPR adaptation protein Cas1, and a predicted post-translational modification targeting a predicted redox

switch in the iHDA1 region. In addition, a novel protein interaction between Cas3 and DNA polymerase I was discovered.

## **Acknowledgments**

I would like to express my sincere gratitude to Dr Edward. L Bolt, my supervisor, for offering me the chance to study the interesting CRISPR Cas3 project and for all the support that Dr Bolt has provided since 2016. Without his great efforts, it is hard to image I could make any progress on my project and acquire those necessary skills to become a scientist.

I am grateful to all my colleagues, collaborators and friends. Special thanks to Dr Tom Killelea for helping me with experiments and for his contribution to the publications. Tom has set a good example of being a team player and brilliant scientist for me, thank you very much Tom.

My parents have given me the greatest support they can offer. No words can describe how much I appreciate their efforts. Though I live a life which is completely different from their true will, they still tried their best to help me. I would also thank myself for all the hard work I have done, and did not give up on trying when I came across difficulties. In addition, I would like to mention those adorable dogs bring me a lot of joy, thank you Qiuqiu, Ossie, Crumble and Ruby.

Lastly, I would like to thank the University of Nottingham for offering me the Vice-Chancellor's Scholarship for Research Excellence (International) for providing the funding for my project.

# Table of Contents

<b>Abstract</b>	<b>i-ii</b>
<b>Acknowledgments</b>	<b>iii</b>
<b>Contents</b>	<b>iv-x</b>
<b>List of Figures</b>	<b>xi-xv</b>
<b>List of Tables</b>	<b>xvi</b>
<b>Abbreviations</b>	<b>xvii</b>

## **Chapter 1: Introduction to CRISPR-Cas systems and the Cas3 effector protein**

1.1	Overview of CRISPR-Cas systems.....	1
1.1.1	Discovery and development of the CRISPR-Cas system.....	1
1.1.2	The major processes of CRISPR-Cas systems: adaptation, expression and interference.....	4
1.1.2a	CRISPR Adaptation.....	7
1.1.2b	Expression of crRNA.....	11
1.1.2c	CRISPR Interference reactions.....	12
1.1.3	The interaction of CRISPR-Cas and DNA repair systems.....	14
1.1.4	Applications of CRISPR-Cas.....	17
1.2	The Cas3 interference protein.....	19
1.2.1	Diversity in Cas3 proteins of Type I CRISPR-Cas system.....	21

1.2.2	The nuclease activity of Cas3 compared with other nuclease Cas proteins...	23
1.2.2a	Structural basis of Cas3 for cleaving nucleic acids.....	23
1.2.2b	Metal-ion dependency of HD Cas3 domains.....	24
1.2.2c	Modulation of Cas3 nuclease by Cascade and perhaps Cas1.....	25
1.2.2d	Comparing I-E Cas3 and I-F Cas2-3.....	26
1.2.2e	Comparing Cas3 and Cas9 nucleases.....	27
1.2.3	Cas3 translocase-helicase activity.....	28
1.2.4	Cas3 interacts with Cas proteins - the formation of PAC.....	31
1.2.5	High-temperature protein G (HtpG) and Cas3 in CRISPR interference.....	33
1.3	Potential non-canonical roles of Cas3 in non-CRISPR-Cas systems.....	34
1.3.1	Cas3 interactions with non-CRISPR proteins.....	34
1.3.2	The interplay between Cas3 and RNA.....	34
1.3.3	Cas3 implicated in Biofilm formation.....	37
1.4	Outstanding questions about Cas3.....	37
1.4.1	Cas3 reveals temperature-dependent activities.....	38
1.4.2	MGE pre-spacer DNA fragments: Are they really generated by Cas3?.....	40
1.4.3	Cas1-2 may assist Cas3 in conducting bidirectional DNA cleavage on MGEs from the protospacer site.....	41
1.5	Aims and objectives of this Ph.D research.....	43
 <b>Chapter 2: Materials and Methods</b>		
2.1	Bacterial strains.....	45
2.2	Cloning.....	45

2.2.1 Plasmids for protein overexpression.....	47
2.2.2 Site-direct mutagenesis.....	49
2.2.3 Plasmids for analysis of Cas3 in vivo protein-protein interaction.....	50
2.2.3a Plasmids used in proximity-dependent biotin identification (Bio-ID) system.....	50
2.2.3b Plasmids used in the Bimolecular fluorescence complementation system.....	50
2.3 Protein expression and purification.....	52
2.3.1 Preparing His <sub>x6</sub> -MBP-Cas3 proteins.....	53
2.3.2 Expression and purification of MBP-Cas3 proteins.....	55
2.3.3 Expression and purification of His <sub>x6</sub> -Cas3.....	56
2.3.4 Expression and purification of Cas2 <sup>E9A</sup> .....	58
2.3.5 Expression and purification of PaeCas2-3.....	58
2.3.6 Expression and purification of PaeCas1.....	59
2.4 Design and preparation of DNA substrates.....	60
2.5 Cas3 catalytic assays.....	63
2.5.1 Cas3 Nuclease assay.....	63
2.5.2 In vitro interference assay.....	64
2.5.3 Cas3 ATPase assay.....	65
2.6 Electrophoretic mobility shift assay (EMSA).....	66
2.7 Polymerase DNA primer extension (PE) assay.....	66
2.8 Spacer integration (SPIN) assays.....	67

2.9 Coupled primer extension (PE)-SPIN assay.....	68
2.10. Polyacrylamide gel electrophoresis (PAGE) for DNA and Protein.....	68
2.10.1 Non-denaturing PAGE.....	68
2.10.2 Denaturing gel electrophoresis.....	69
2.10.3 Blue native-PAGE.....	70
2.10.4 Sodium dodecyl sulphate (SDS) PAGE.....	70
2.11 Western blotting to detect proteins.....	71
2.12 Protein pull-down assays.....	73
2.13 Liquid chromatography-mass spectrometry (LC-MS).....	73
 <b>Chapter 3: Molecular mechanism of the interface of Cas3 HD and RecA1-like domains (iHDA1) in temperature-dependent nuclease activity</b>	
3.1 Summary.....	74
3.2 The importance of the interface between the Cas3 HD and RecA1-like domains (iHDA1), and its conserved tryptophan residues.....	76
3.3 Protein purification of wild type EcoCas3 and mutants.....	80
3.4 Selection of DNA substrate used for diagnosing Cas3 nuclease activity.....	81
3.5 E. coli Cas3 tryptophan-406 is a determinant for differential nuclease activity at 30 °C and 37 °C.....	86
3.6 Distinct DNA degradation products produced by EcoCas3 and the mutants.....	91
3.7 DNA binding affinities of Cas3 Trp-mutants.....	92
3.8 EcoCas3 ATPase activity is independent of temperature change.....	94

3.9 Cas3 nicking assay and a Cas3-Cascade in vitro CRISPR interference assay.....	96
3.9.1 Introducing mutants in the iHDA1 region promotes the nicking activity of EcoCas3.....	97
3.9.2 EcoCas3 tryptophan mutants liberate EcoCas3 from functional repression by Cascade, for uncontrolled DNA degradation.....	99
3.10 Discussion.....	100

**Chapter 4: Comparative analysis of two divergent type I Cas3 proteins identified a hypothetical cysteine redox switch, and novel protein-protein interaction**

Part I: Comparative analysis of PaeCas2-3 and EcoCas3 identified a hypothetical cysteine redox switch in EcoCas3 iHDA1 region.....	109
4.1 Purified PaeCas2-3 is a nuclease on the model DNA fork.....	109
4.2 PaeCas2-3 is nuclease active at both 30 °C and 37 °C.....	112
4.3 PaeCas2-3 Trp-526 does not interact with the Ic motif, therefore it may not be able to block ssDNA from accessing its DNA catalytic site in the HD domain....	113
4.4 The reducing agent DTT inhibits the PaeCas2-3 function, but stimulates EcoCas3.....	118
4.5 EMSA shows PaeCas2-3 binds to DNA substrates of different types.....	122
4.6 Discussion for Part I.....	123
Part II: Analysis of EcoCas3 with proteins that potentially target the EcoCas3 iHDA1 region.....	124

4.7 EcoCas1 represses EcoCas3 nuclease activity but not hyperactive nuclease EcoCas3W406A.....	125
4.8 HtpG promotes the nuclease activity of EcoCas3 and EcoCas3W406A at 37 °C.....	128
4.9 Discussion.....	130

**Chapter 5: Biochemical analysis of the Cas3 oligomeric state and novel protein interactions**

5.1 Summary for Part I.....	139
Part I: Identifying Cas3 in oligomer and monomer states.....	140
5.2 Cas3 purified in two forms, distinguished in size exclusion chromatography .....	140
5.3 The Cas3 oligomer (Cas3_E1) revealed features biochemically distinct from monomeric Cas3 (Cas3_E2).....	143
5.4 Different affinity tags fused to Cas3 have no impact on the formation of Cas3 oligomer.....	148
5.5 Bioinformatics analyses of potential intrinsically disordered protein regions within E. coli Cas3.....	149
5.6 Discussion for Part I.....	154
Part II: Cas3 in oligomer state reveals novel protein interactions.....	156
5.7 Validation of DNA polymerase III alpha and DNA polymerase I activities in CRISPR-Cas primed adaptation.....	156
5.8 A primer extension assay to analyse DNA synthesis by DnaE.....	159
5.9 DNA polymerase I co-purified with Cas3.....	160

5.10 Removal of Pol I during Cas3 purification.....	168
5.11 Less DNA elongation is produced by DnaE when it is incubated with an excessive amount of Cas3.....	172
5.12 DnaE influences spacer integration by Cas1-2.....	175
5.13 Cas3 physically interacts with Pol I, but it has no impact on DNA synthesis by Pol I.....	181
5.14 Cas1-2 was not capable of using DNA fragments generated by Pol I as pre-spacers.....	184
5.15 Discussion for Part II.....	185
 <b>Chapter 6: Conclusion and future work</b>	
6.1 The iHDA1 region accommodates a Trp-406 'gate' amino acid and a cysteine redox switch.....	189
6.2 Cas3 phase separation-formation and dissociation.....	191
6.3 Detecting the interaction between Cas3 and DNA polymerases using Bimolecular fluorescence complementation (BiFC) assay.....	193
6.4. BioID could be adopted to screen broader participators of Cas3.....	195
6.5 Cas3 and beyond.....	197
<b>References</b> .....	200
<b>Appendices</b> .....	217

## List of Figures

Figure 1.1 Two classes of CRISPR-Cas systems.....	5
Figure 1.2 The main events in an I-E subtype CRISPR-Cas system.....	6
Figure 1.3 Crystal structure of the Cas1-Cas2-DNA complex.....	8
Figure 1.4 Schematics of protospacer integration by Cas1-2 assisted by IHF.....	9
Figure 1.5 Cascade and Cas9 form R-loop on MGEs.....	14
Figure 1.6 Cas3 structure and its individual functional domains.....	20
Figure 1.7 Diverse Cas3 in Type I CRISPR-Cas systems.....	22
Figure 1.8 Schematics of two Cas3 subtypes nicking non-targeting DNA at different sites.....	27
Figure 1.9 The anti-Pcas RNA encoding gene overlaps the non-coding strand of the Cas3-encoding gene.....	36
Figure 1.10 Temperature-dependent Cas3/CRISPR functions in <i>E. coli</i> and the <i>E. coli</i> $\Delta$ hns strain.....	39
Figure 1.11 Illustration showing the mechanisms of Cas3 unidirectional and potential bidirectional DNA degradation.....	43
Figure 2.1 The construction of CRISPR 1s.....	63
Figure 3.1 Structural positioning of two invariant tryptophan residues at the interface of Cas3 HD and RecA1 domains—a key role in Cas3 nuclease function for Trp-406.....	79
Figure 3.2 Purification of EcoCas3 proteins used in this study.....	81
Figure 3.3 Cas3 nuclease activity on M13ssDNA and DNA fork substrates.....	85

Figure 3.4 EcoCas3 tryptophan mutants revealed distinct DNA degradation ability, the EcoCas3 <sup>W406A</sup> stands out as a hyperactive DNA nuclease.....	89
Figure 3.5. EcoCas3 <sup>R662G,R674G</sup> degrades DNA fork at 30 °C.....	91
Figure 3.6 Distinct DNA degradation products are produced by EcoCas3 and tryptophan mutants.....	92
Figure 3.7 EMSAs showing wild-type EcoCas3 and EcoCas3 <sup>W406A</sup> forming DNA-protein complexes.....	94
Figure 3.8 The EcoCas3 ATPase activity is independent of temperature change.....	96
Figure 3.9 EcoCas3 nicks M13dsDNA in the presence and absence of Cascade.....	98
Figure 3.10 In vitro Cascade-Cas3 interference against M13rfDNA at 30 and 37 °C.....	100
Figure 3.11 Repositioning of a highly conserved pair of tryptophan residues Trp-230/406 in E. coli Cas3 protein.....	102
Figure 3.12 Concluding summary of the Trp-406 nuclease 'gate' model.....	107
Figure 4.1 The structural and amino acid sequence alignment of PaeCas2-3 and EcoCas3.....	109
Figure 4.2 The nuclease activity of PaeCas2-3 was analysed against DNA fork substrate.....	111
Figure 4.3 PaeCas2-3 is an active nuclease at both 30 °C and 37 °C.....	113
Figure 4.4 PaeCas2-3 Trp-526 and EcoCas3 interact with adjoining amino acids.....	116

Figure 4.5 Potential amino acids in PaeCas2-3 that may cause steric block to DNA accessing HD catalytic site.....	117
Figure 4.6 DTT inhibits nuclease activity of PaeCas2-3 but assists EcoCas3 nuclease activity.....	120
Figure 4.7 Two adjacent cysteine residues Cys-403 and Cys-404 tandem the Trp-406 'gate' residue.....	121
Figure 4.8 EMSA assay showing wild-type PaeCas2-3 forming a stable protein-DNA complex with different substrates.....	123
Figure 4.9 EcoCas3 nuclease activity in the presence of EcoCas2 <sup>E9A</sup> and EcoCas1.....	128
Figure 4.10 EcoCas3 nuclease activity in the presence of chaperon protein HtpG.....	130
Figure 5.1 Size exclusion chromatography (SEC) analysis showing Cas3 polymer (Cas3_E1) and monomeric Cas3 (Cas3_E2).....	142
Figure 5.2 Cas3_E1 and Cas3_E2 revealed distinct DNA hydrolysis capacity, which was stimulated by MgCl <sub>2</sub> rather than NiCl <sub>2</sub> .....	146
Figure 5.3 The Cas3 <sup>D75G</sup> oligomer and Cas3 <sup>D452A</sup> oligomer exhibited reduced nuclease activity on the DNA fork substrate.....	147
Figure 5.4 Blue Native PAGE showing dissociation of His <sub>x6</sub> -Cas3 oligomer by adding ATP.....	149
Figure 5.5 The possible IDPRs in E. coli Cas3 predicted by PONDR VL-XT.....	153
Figure 5.6 A model of CRISPR interference in which Cas3 recruits DnaE to synthesise DNA fragments, which are then processed by Cas1-2 into a pre-spacer.....	158

Figure 5.7 DnaE exhibited DNA synthetic activity on ELB40P/41 substrate, as revealed by primer extension assay.....	160
Figure 5.8 Primer extension occurred in samples containing Cas3.....	162
Figure 5.9 Reverse transcriptase inhibitor AZT did not block DNA synthetic activity generated by the DNA polymerase co-purified with Cas3.....	164
Figure 5.10 DNA polymerase was co-purified with EcoCas3 and PaeCas2-3.....	167
Figure 5.11 Primer extension analysis could detect a minimum of 0.6 nM Pol I added to the reaction.....	169
Figure 5.12 Separating Pol I from Cas3.....	171
Figure 5.13 DNA synthesis by DnaE was moderately reduced by Cas3.....	174
Figure 5.14 EMSA showing DnaE did not form a stable protein-DNA complex.....	175
Figure 5.15 The PE reaction using Pol I alongside Cas1-2, for the development of a combined PE-SPIN assay that coupled DNA synthesis to spacer integration.....	177
Figure 5.16 PE-SPIN assay using Pol I and DnaE to produce synthesised DNA substrates as the pre-spacer source for Cas1-2 to capture.....	180
Figure 5.17 Pull-down assay using MBP-Cas3 and Pol I indicated Cas3 and Pol I formed a physical interaction in vitro.....	182
Figure 5.18 Cas3 did not affect primer elongation by Pol I.....	183
Figure 5.19 No spacer integration by Cas1-2 occurred when using Pol I to produce DNA fragments as a pre-spacer source.....	185

Figure 6.1. BiFC assay for analysing Cas3-DNA polymerase interaction during CRISPR interference.....	195
Figure 6.2. BioID assay for detecting broader participators of Cas3.....	197

## List of Tables

Table 1. List of cell strains used in this work.....	45
Table 2. Primers used for cloning.....	49
Table 3. Primers used for mutagenesis.....	50
Table 4. Primers used for developing in vivo analysis on Cas3.....	52
Table 5. Buffers used when purifying His <sub>x6</sub> -MBP-Cas3 proteins.....	55
Table 6. Buffers used when purifying MBP-Cas3 proteins.....	56
Table 7. Buffers used for purifying PaeCas2-3 and PaeCas1.....	59
Table 8. DNA oligos used for biochemical analysis.....	62
Table 9. Disordered regions predicted by PONDR VL-XT.....	154

## Abbreviations

ATP	adenosine triphosphate
BSA	bovine serum albumin
Cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defence
CRISPR	cluster regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
dsDNA	double-stranded DNA
DTT	dithiothreitol
Eco	<i>Escherichia coli</i>
EDTA	ethylenediamine tetra-acetic acid
EMSA	electrophoresis mobility shift assay
gRNA	guide-RNA
HD	histidine-aspartate
Hepes	N-2-hydroxyethylpiperazine- <i>N'</i> -2-ethanesulfonic acid
iHDA1	interface between Cas3 HD and RecA1-like domains
LC-MS	liquid chromatography-mass spectrometry
Pae	<i>Pseudomonas aeruginosa</i>
PAGE	polyacrylamide gel electrophoresis
PE	primer extension
PMSF	phenylmethylsulfonyl fluoride
PVDF	polyvinylidene difluoride
SDS	sodium dodecyl sulfate
SF2	superfamily-2
SPIN	spacer integration
ssDNA	single-stranded DNA
TBE	tris-borate-EDTA
TEMED	N,N,N',N'-tetramethyl ethylenediamine
Tfu	<i>Thermobifida fusca</i>
Tris	tris(hydroxymethyl)aminomethane
Tter	<i>Thermobaculum terrenum</i>

## **Chapter 1**

# **Introduction to CRISPR-Cas systems and the Cas3 effector protein**

### **1.1 Overview of CRISPR-Cas systems**

The CRISPR-Cas system is an adaptive immunity system in prokaryotes. Using a whole-genome sequencing database, bioinformatic analysis has identified CRISPR-Cas systems in 90 % of archaeal and 40 % of bacterial clades(1,2). However, functional analyses of each CRISPR-Cas system is still limited, which is a problem to understand their diversity of functions as immune systems. The current understanding of CRISPR-Cas is that it has two major components for maintaining its 'immunity' against foreign DNA and preventing the host cell from lethal infection: CRISPR-associated (Cas) proteins and a DNA locus recording 'immunity memory', termed Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)(1,3). Section 1.1.1 provides more details regarding the interplay of the functions of the CRISPR-Cas system.

#### **1.1.1 Discovery and development of the CRISPR-Cas system**

Both people who are either working on CRISPR-related research and those who are not are under the impression that the first discovery of the CRISPR was made by Mojica *et al.*(4), in archaea,

## L. He Chapter 1

but its existence was first noticed by Ishino *et al.* in *E. coli* in 1987 (5). Yoshizumi Ishino found five almost identical DNA sequences each of 29 nt that were spaced out by 32 nt DNA sequences. However, why those repeated DNA sequences existed in the bacteria genome and what role they played was unknown. In 1993, Francisco Mojica found serial DNA sequences of similar features enriched in an archaeon termed *Haloferax mediterranei* (4). Among the benefits from the development of the DNA sequencing technique was the discovery of more similar clusters of DNA repeats in more prokaryotic species, which was then named CRISPR by a Dutch group in early 2002 (6). Together with the denomination, this group suggested the existence of CRISPR-associated (Cas) genes close to the CRISPR locus, to which the Cas genes and CRISPR locus may be functionally related. In the same year, the same Dutch group published a paper and used the SPacers Interspersed Direct Repeats (SPIDR) to define the DNA loci they had previously named CRISPR (7). The former name has been used in preference to the latter as the common name of the DNA repeats locus since then.

The first breakthrough in understanding CRISPR was achieved in 2005-2006 when scientists eventually matched the 'spacer' sequence in the CRISPR locus with the bacteriophage genomic DNA(3,7,8). This finding led to a hypothesis that these CRISPR and Cas genes may be the components of an adaptive immunity system in prokaryotes. Another theory to explain the role of CRISPR was

that it is a novel DNA repair system in bacteria (9). Based on current understanding of this system, that idea is not entirely wrong, for two reasons: First, the new spacer integration during CRISPR adaptation will cause DNA damage on the host genome, so this system is likely to couple with DNA repair pathways (10,11); second, CRISPR-Cas systems co-exist with certain DNA repair systems, indicating a close connection between those two systems (12-14). More descriptions regarding CRISPR and DNA repair appear in Section 1.1.3. In the same period, a non-canonical Cas gene was identified that encodes a large protein and is now known as Cas9, the most famous CRISPR-Cas protein (15).

In the following years, more evidence was proposed and confirmed the CRISPR-Cas is an adaptive immunity system (16-20). In addition, more features of the CRISPR-Cas system were defined, including the role of CRISPR RNA (crRNA) in the recognition of mobile genomic elements (MGEs), the existence of trans-activating CRISPR RNA (tracrRNA) in the CRISPR-Cas9 system, and biochemical features of Cas proteins such as the CRISPR Cas3 protein in this study (17,21-27).

Since 2013, CRISPR-related research has split in two directions: the fundamental study of the CRISPR-Cas system and the application of CRISPR-Cas as a novel gene-editing tool. Three independent groups separately announced their application of

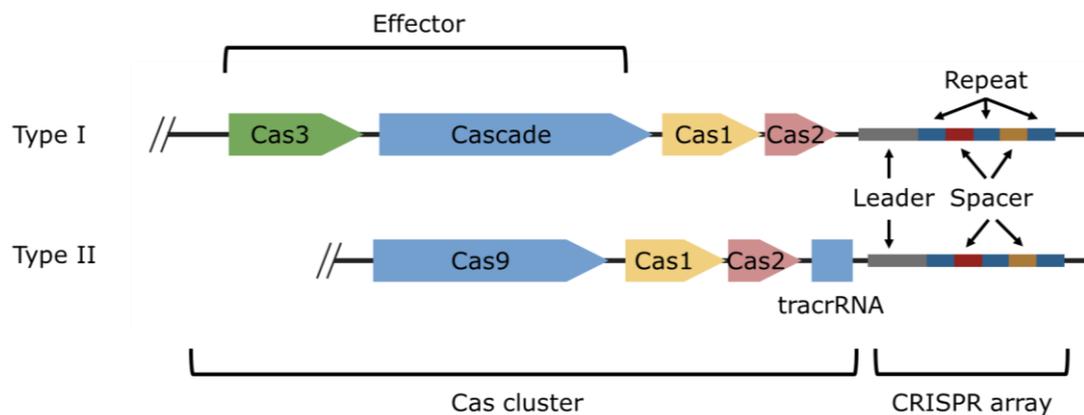
CRISPR-Cas9 as a genome-editing tool in mammalian cell lines (human and mouse) and in yeast (28-30). Only three years passed from the first published application of the CRISPR-Cas system in eukaryotes to the commercialisation of the first CRISPR-Cas edited fungus (mushroom) (31). The applications of CRISPR-Cas are described in greater detail in Section 1.1.4.

Though using CRISPR-Cas as a gene-editing tool draws the most attention from scientists, the fundamental studies on revealing the insight of the CRISPR-Cas system and its coordination with other intracellular biological networks are equally important. The comprehensive studies of the biochemical features of Cas proteins can promote the understanding of how and why prokaryotes adopt CRISPR-Cas as an adaptive immunity system and help with the exploitation of CRISPR-Cas for different purposes.

### **1.1.2 The major processes of CRISPR-Cas systems: adaptation, expression and interference**

CRISPR-Cas systems contain three parts: the Cas protein effectors that are against MGEs, the spacer integration proteins Cas1 and Cas2, and the spacer-repeat array that preserves the 'immunity memory' regarding MGEs previously encountered by the host. The complexity of the Cas proteins is used for classification analysis distinguishing between CRISPR-Cas systems (1,19,32). The

mainstream classification distinguishes CRISPR-Cas into two classes by the number of crRNA-dependent interference effectors in each system (1). Class 1 CRISPR-Cas contains multiple effectors to execute MGEs recognition and degradation, whereas Class 2 systems rely on one large protein module to eliminate MGEs. Though CRISPR interference effectors exist throughout all CRISPR-Cas systems, they vary in their functions (21,26,33-35). Depending on enzymatic activities of the CRISPR effector modules, the Class 1 systems can be further divided into three subtypes: Type I (Cascade-Cas3 complex) (33), Type III (Csm/Cmr complex) and Type IV (Cas5-Cas7 complex, or Csf3-Csf2) (36). The Class 2 systems also have three subtypes: Type II (Cas9), Type V (Cas12) and Type VI (Cas13) (37,38).

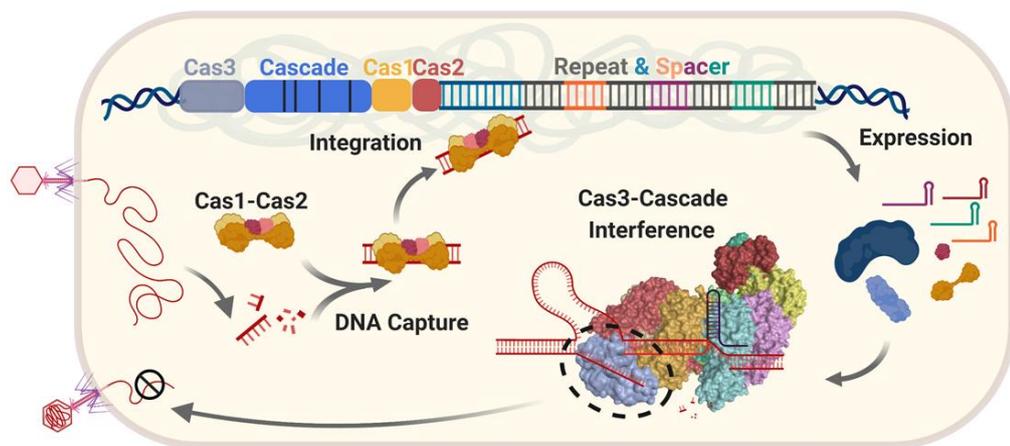


**Figure 1.1. Two classes of CRISPR-Cas systems**

*These images show the structure of two well-studied CRISPR-Cas systems, Subtypes I-E and II-C1. They both contain Cas proteins and the CRISPR locus. Leader sequence, repeats and different spacers are denoted.*

Though CRISPR-Cas systems vary in their components, they share one common feature: the spacer integration proteins Cas1 and Cas2 (39), which exist in all identified CRISPR-Cas systems (1,2). This phenomenon suggests their presence is likely to be a conserved feature for the adaptive immune system in prokaryotes. More details regarding the involvement of Cas1 and Cas2, together with other interactors for those two proteins, appear in Section 1.1.2a.

CRISPR-Cas immunity against MGEs is divided into three stages: Cas1-2 participated adaptation, expression of crRNA and Cas proteins, and CRISPR interference (Figure 1.2 (40)). Section 1.1.2a presents the details of each step, mainly using a canonical CRISPR-Cas Type I-E system as an example.



**Figure 1.2. The main events in an I-E subtype CRISPR-Cas system**

*DNA from an MGE (bacteriophage shown here) is captured and integrated into a CRISPR locus by "Adaptation", catalysed by Cas1-Cas2 proteins helped by various non-Cas host proteins. Transcription of CRISPR generates RNA that after further processing is called CRISPR RNA (crRNA). This is loaded into a multi-protein effector complex, forming a "Cascade" that targets crRNA to the invader DNA. Cascade recruits Cas3 to targeted*

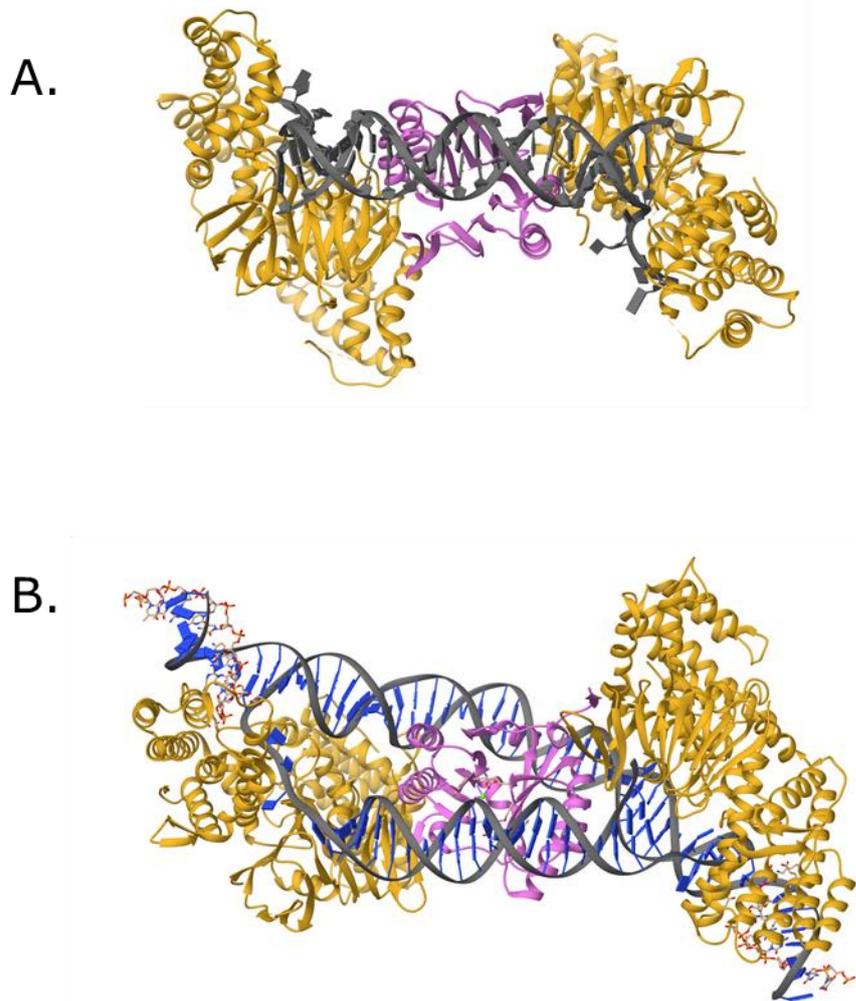
## L. He Chapter 1

*DNA, forming the "Interference" complex that degrades DNA and in so doing can provide DNA for capture by Cas1-Cas2. The Cas3 in the interference complex is marked.*

### **1.1.2a CRISPR Adaptation**

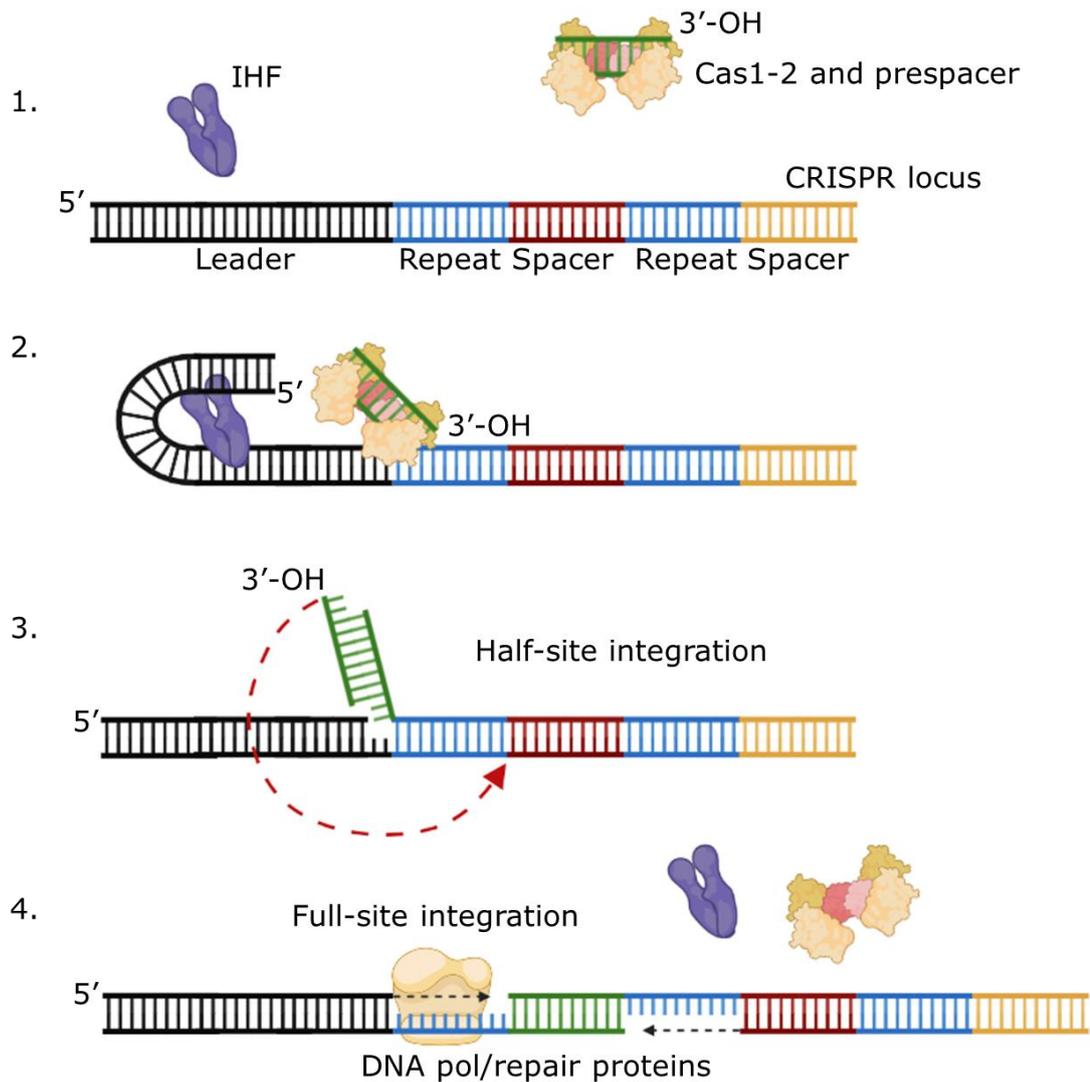
CRISPR adaptation are processes that culminate in integration of protospacers, a portion of MGEs, into a CRISPR locus by Cas1-2, assisted by integration host factor (IHF) (39,41). This process also requires other host non-Cas proteins, including those of DNA repair pathways; for example, for gap-filling polymerases and nucleases, DNA ligases (10) (Figure 1.3). Depending on the stage preceding the occurrence of adaptation, it is divided into naïve adaptation and primed adaptation: The former is used to denote spacer acquisition by Cas1-2, which is independent of CRISPR interference, whereas the latter is facilitated by CRISPR interference.

In *E. coli*, a functional Cas1-2 complex for new spacer integration is composed of four molecules of Cas1 and two molecules of Cas2 (Figure 1.3). This complex is necessary for the recognition and capture of DNA fragments from MGEs. The protospacer bound by Cas1-2 is processed by the nuclease activity of Cas1 into an integration-favourable format with two 3'hydroxyl groups at both ends, for following new spacer insertion that will take place at the junction of an AT-rich leader sequence and the first repeat sequence (Figure 1.4).



**Figure 1.3. Crystal structure of the Cas1-Cas2-DNA complex**

*(A, B) These images show Cas1-2 complexes bind to double-strand DNA and this protein-DNA complex in a fully integrated stage. Cas1 and Cas2 are denoted in yellow and orchid, double-strand DNA (protospacer) is coloured in grey, and the target sequence (part of CRISPR locus) is denoted in grey. (A). Figure is adapted from PDB: 5DS6 (42). (B). Figure is adapted from PDB: 5XVP (43).*



**Figure 1.4 Schematics of protospacer integration by Cas1-2 assisted by IHF**

*This shows that Cas1-2 integrate protospacer into the CRISPR locus, with the assistance of IHF. IHF interacts with and bends the leader sequence, generating an integration-favourable substrate conformation. Protospacer integration is then achieved at the leader-spacer junction. This process may recruit DNA repair proteins (DNA pols and ligase) to the CRISPR locus for gap-filling and joining DNA backbones. Figure adapted from Rollie et al. and Nussenzweig et al (41,44).*

The Cas1-2 facilitated new spacer integration also requires IHF binding to the leader sequence (39,41,45). This DNA-protein

## L. He Chapter 1

interaction bends the DNA strand consequently induce Cas1-2 to conduct spacer integration. Half-integration is achieved as one strand of protospacer joins the forward DNA strand of the CRISPR locus at the leader-spacer junction. Then, the free strand of protospacer joins the reverse strand of the CRISPR locus at the repeat-spacer junction. Spacer integration may consequently recruit DNA repair proteins, for instance, the DNA polymerase I and ligase, to the CRISPR locus for gap-filling and joining DNA backbones (10). Notably, though this model is well-accepted, how Cas1-2 select DNA substrates as new 'immunity memory' for the host cell during naïve adaptation and primed adaptation is still unknown

The mechanism of inserting new spacers between the leader sequence and the first CRISPR repeat determines one critical dynamic feature of the CRISPR locus. The first few spacers flanking the leader sequence (leader end) are correlated to the MGEs the host cell has encountered most recently, whilst the spacers located in the trailer end are considered to be the oldest spacers, perhaps acquired years previously (46). Therefore, the spacers at the leader end are more diverse between strains of the same species, and they are often used to diagnose MGEs that have recently infected host cells and generate evolutionary analysis between CRISPR-Cas systems among different bacteria and archaea species (47).

### **1.1.2b Expression of crRNA**

Integrated 'immunity memories', or spacers, need to be converted into small RNAs for CRISPR interference effectors that recognise MGEs. They are termed CRISPR RNA (crRNA) in Type I CRISPR-Cas and guide RNA (gRNA) in Type II CRISPR-Cas (22,48,49). The gRNA is a fusion generated by base-pairing of crRNA and trans-activating CRISPR RNA (tracrRNA) (22,50,51).

During the CRISPR-Cas immunity expression stage, CRISPR DNA is first transcribed into precursor crRNA (pre-crRNA). This RNA fragment is then processed into mature crRNA by endoribonucleases. During this process, RNase III participates in cleavage of pre-crRNA and tracrRNA complexes in type II systems (22); while in type I and III systems, the Cas6 endonuclease contributes to crRNA maturation (52-55). A recent study also highlighted Cpf1, Cas5d, Csd1 and Csd2 and their roles in pre-crRNA processing (55-60). In addition, deletion of the *cas3* gene disrupts crRNA maturation, suggesting Cas3 may contribute to this process, though the mechanism is still unknown (61).

The mature crRNA consists of two parts, the sequence that is complementary to protospacer on MGE and the RNA sequence generated from the CRISPR repeat sequence (62). The crRNA and alone do not help with identifying invaders carrying protospacers. To fulfil their function, those crRNA fragments are assembled with

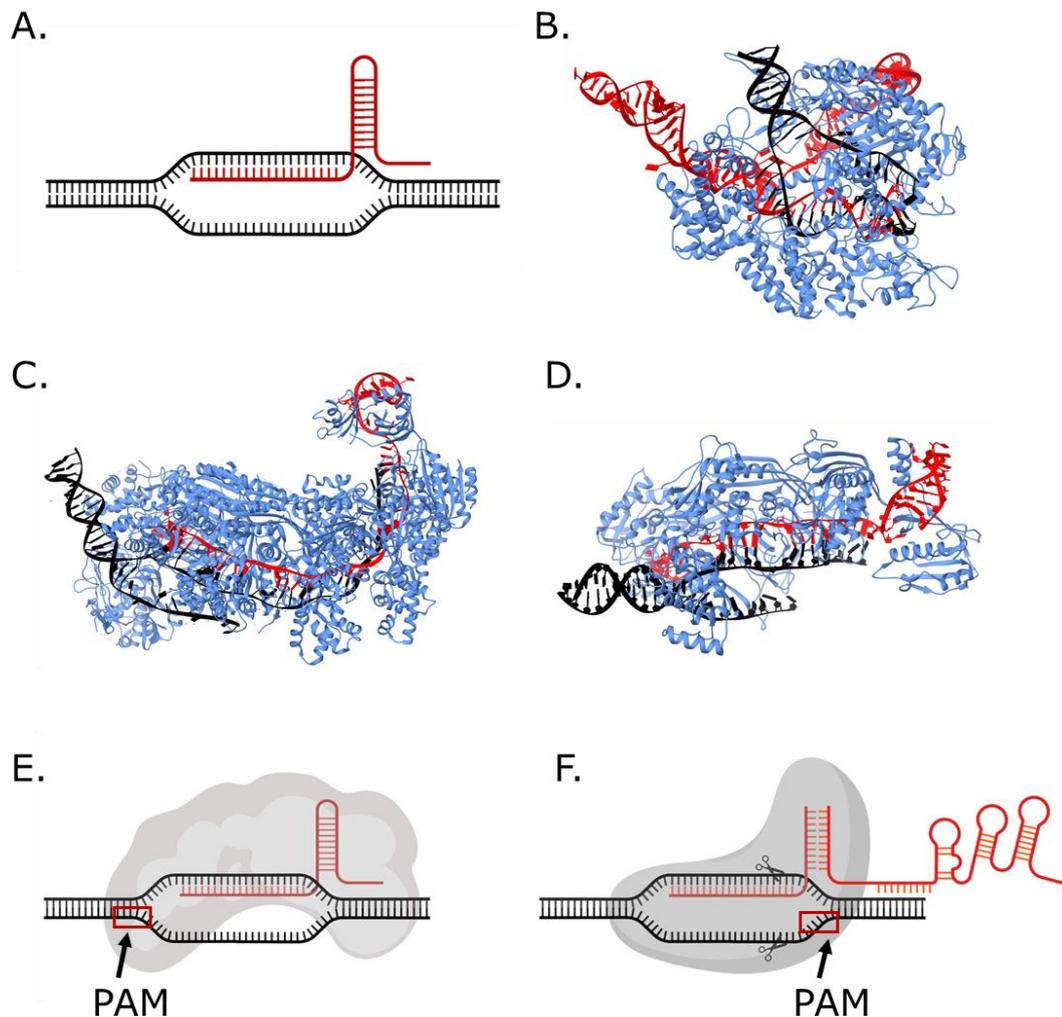
CRISPR surveillance effectors, for instance, the Cascade in Type I CRISPR and the Cas9 in Type II CRISPR. The combination of interference protein and crRNA is crucial for CRISPR-Cas to carry out efficient screening of DNAs and diagnose invaders (33,51,62-64).

### **1.1.2c CRISPR Interference reactions**

In the I-E subtype CRISPR-Cas system, what is critical to the effectiveness of Cas3 as a nuclease during CRISPR interference is that Cascade first validates a bona fide target DNA (23,33,65). Cascade identifies a specific DNA sequence, the protospacer adjacent motif (PAM) (66-69), at the target site, followed by stabilised DNA-binding from RNA-DNA base-pairing, forming a Cascade R-loop that is 'locked' onto the target DNA (70) (Figure 1.5). This formation of an R-loop on MGE will facilitate conformational change of Cascade (65). In this state, the conformation of the Cse1 component of Cascade exposes its interacting regions for recruiting Cas3. Then, Cas3 loads onto the ssDNA, which is the non-targeting strand that is generated in R-loops, but interaction with the Cse1 subunit of Cascade channels its nuclease activity for CRISPR immunity(65,71). Analysis using Hydrogen-Deuterium Exchange (HDX) coupled to mass spectrometry identified alpha helix H1 of Cse1 as a critical docking and/or activation site for interaction with Cas3 (72), which may be the molecular trigger for degradation of

targeted DNA by the Cas3 HD-nuclease activity (73,74). By interacting with Cascade, Cas3 effectively eliminates the non-targeting strand of MGEs and generates DNA deletion upstream away from the targeted protospacer (75). Notably, recent studies on primed adaptation and spacer distribution indicate that Cas3 may interact with the DNA sequence downstream of the protospacer, though the mechanism is still unknown (76). A more detailed description regarding the DNA cleavage directionality of Cas3 appears in Section 1.4.3.

In the Type II CRISPR-Cas system, the interference effector Cas9 undergoes a similar process to recognise protospacer via R-loop formation and then conduct DNA nicking at both strands of MGEs (21,62) (Figure 1.5). This will generate double-strand breaks on MGEs, in contrast to the Cas3-Cascade interference that causes large-range DNA deletion (75). More details appear in Section 1.1.3.



**Figure 1.5 Cascade and Cas9 form R-loop on MGEs.**

DNA strands are coloured in black. RNA, including crRNA (in C, D and E) and gRNA (in B and F) are red. Proteins in B, C and D are blue; in E and F are grey. (A). This shows a basic structure of an R-loop, including a DNA-RNA hybrid and a non-template strand. (B, C and D). These show R-loop formation by *S. pyo* Cas9 (B, PDB: 5F9R), *E. coli* I-E Cascade (C, PDB: 5H9F) and *S. put* I-F Cascade (D, PDB: 5O6U). (E and F). PAM sites in type I (in E)

### 1.1.3 The interaction of CRISPR-Cas and DNA repair systems

The interplay of CRISPR-Cas and DNA repair systems is crucial during CRISPR immunity and adaptation (10-13). Though the

## L. He Chapter 1

underlying mechanism is still unknown, this phenomenon has been observed during genetic and biochemical analysis and indicated by a bioinformatic study of the distribution of the CRISPR-Cas system and DNA repair pathways in prokaryotes (12).

CRISPR adaptation of the Class I system has shown a correlation with DNA repair systems. In *E. coli*, the homologous recombination protein complex RecBCD facilitates Cas1-2 naïve adaptation (11,12). Though the underlying mechanism is unknown, what can be conformed now is this naïve adaptation occurrence is dependent on RecBCD helicase rather than nuclease function (11). Additionally, one of the DNA repair proteins, RecG, primosomal protein A (PriA) and DNA polymerase I are also involved in CRISPR adaptation (10,77-79).

A similar interaction between CRISPR and DNA repair systems is identified in CRISPR I-F subtypes. The *Pseudomonas aeruginosa* (*Pae*) RecD protein is necessary for primed adaptation, whilst the RecG contributes to adaptation in the I-F system but is dispensable (80). In the I-A CRISPR subtype, a putative DNA-binding protein termed Csa3a revealed a regulation function on DNA repair pathways (81,82). The deletion of the gene (*csa3a*) encoding this CRISPR-associated protein factor consequently reduces transcription of DNA damage repair (DDR) genes, thus creating a DNA damage sensitive cell phenotype. Moreover, the Csa3a revealed a specific

## L. He Chapter 1

binding affinity to the promoters of DDR genes (82). These results provide direct evidence linking CRISPR-Cas and DNA repair systems.

Type II CRISPR-Cas may couple with DNA repair proteins. For instance, the CRISPR adaptation-related proteins Csn2 in II-A and Cas4 in II-B (also present in Type I) systems share similar features with a few DNA repair proteins. Csn2 adopts a similar DNA-binding and sliding mechanism to the non-homologous end-joining protein Ku (83-86). The Cas4 carries a RecB-like portion; this similar protein structure can be found in a double-strand break repair protein AddB and RecBCD (87,88).

The association between CRISPR-Cas and the distribution of DNA repair pathways may explain why CRISPR-Cas systems exist in some prokaryotes but are discarded in others (12). Bioinformatic analysis has shown the co-existence of I-E CRISPR and RecBCD systems in Proteobacteria, but this subtype is less frequent to present in Firmicutes (absence of RecBCD) (12). This result is consistent with observations during genetic analysis, whereas the II-A subtype is often absent in Firmicutes carrying non-homologous end-joining systems (13). These two opposite associations between CRISPR subtypes and specific DNA repair pathways indicate that the two systems may be more related than we expected, though the underlining mechanism of their collaboration is still unknown.

#### **1.1.4 Applications of CRISPR-Cas**

The most well-known application of Cas proteins from CRISPR-Cas systems is as genome editing tools, utilized for base-editing, creating genetic insertions or deletions ('indels'), diagnostic kits for RNA or DNA detection and for regulation of gene transcription (CRISPRi and CRISPRa).

Cas9 is most commonly adapted and utilised for genome editing. Since the first application of Cas9 in eukaryotic cells was conducted, more Cas9 variations have been generated, adapted for specific genetic editing applications (89,90). In general, Cas9 variations can be divided into four different types: 1. those with enhanced specificity, 2. recognising different PAM sequences, 3. truncated Cas9, fully functioning but containing fewer amino acids than wild-type Cas9, and 4. Cas9 fusion in which Cas9 is fused with other proteins with specific functions (91,92), for instance the Cas9-POLD3 is Cas9 fused with human DNA polymerase delta subunit 3 (91), this is for boosting the efficacy of genomic editing by Cas9, which is achieved by utilising POLD3 fusion to initiate DNA repair.

In addition to Cas9, the interference enzyme Cas12a from the Class II Type V system (93-96), and Cascade-Cas3 in Class 1 Type I (75,97,98), have revealed potential as gene-editing tools. The mechanism of Cas9 and Cas12a editing systems is different from the Cascade-Cas3 system. The former creates DNA double-stranded

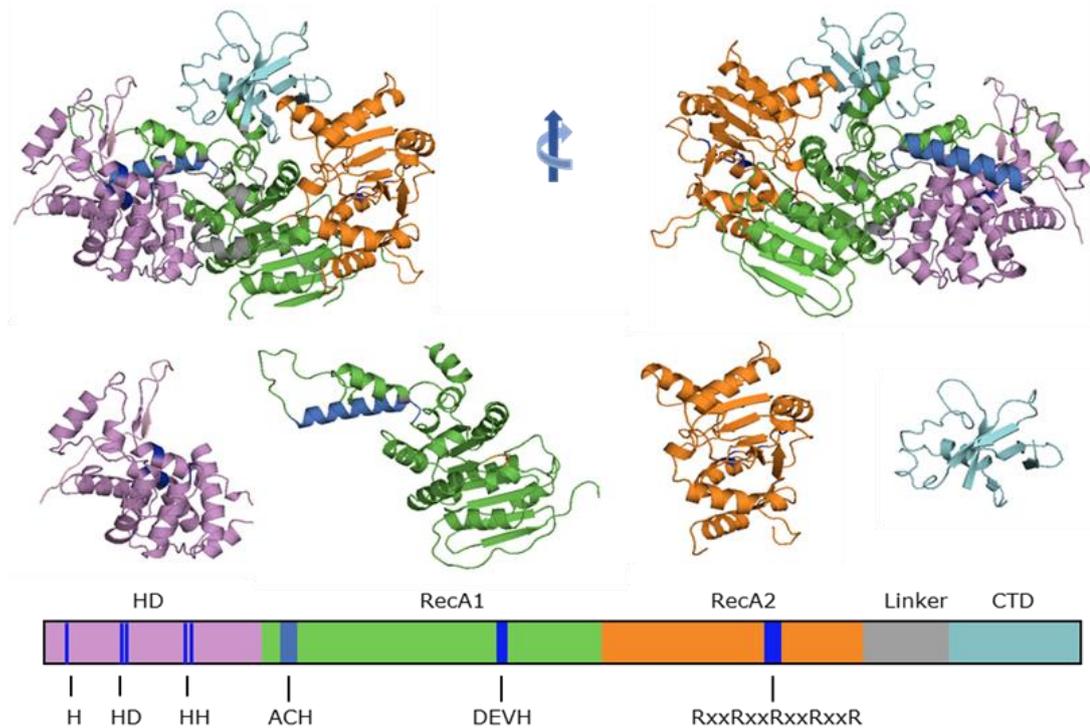
## L. He Chapter 1

breaks on targeting DNA strands and relies on a non-homologous end-joining DNA repair mechanism to generate gene knockouts, or homologous recombination to insert donor DNA (34,93), whereas the Cascade-Cas3 system can produce unidirectional DNA cleavage on the DNA strand over several kilobase-pairs. Thus, the Cascade-Cas3 system generates long-range DNA deletions (75).

Genomic editing using CRISPR requires effector interference proteins capable of forming an R-loop for targeting site recognition and cleaving DNA (nuclease active), whereas nuclease defective CRISPR effectors that have intact R-loop formation function can be applied in many ways. For example, the Cas protein R-loop formation-based SHERLOCK (99), DETECTR (100) and FELUDA (101) are emerging tools for developing rapid diagnostic tools. The most recent application in this aspect is the SHERLOCK-CRISPR-SAR-CoV-2 kit, which has been approved by the Food and Drug Administration (FDA, USA) for emergency usage for detection of SARS-CoV2 infection (102). In addition, some genetic studies utilise CRISPRi and dCas9 to interact with assigned promoters via R-loop formation, thus preventing the transcription of certain genes. The dCas9 is used as a roadblock (103). It has been adapted to study the function of a designed base deaminase (BD)-T7 RNA polymerase fusion protein (104).

## **1.2 The Cas3 interference protein**

Cas3 is an essential component of Type I CRISPR-Cas systems that repel MGEs (74,105-108). Cas3 is an ATP-dependent single-strand DNA (ssDNA) translocase/helicase that in many CRISPR systems is fused to an HD-nuclease domain (Figure 1.6). Its helicase portion contains RecA1, RecA2 and a C-terminal domain (CTD) with no known function. A linker polypeptide connects RecA2 and CTD; all three components along with RecA1 pack together and form a hydrophobic DNA tunnel for transferring ssDNA from helicase domains to the HD-nuclease domain (71,109). The helicase domains of Cas3 reveal classic features for characterising superfamily-2 (SF2) helicases (110,111), including SF2 specific conserved motif III (amino acids S<sup>483</sup>A<sup>484</sup>T<sup>485</sup>) in the Cas3 RecA1 domain and IV (amino acids Q<sup>661</sup>R<sup>662</sup>xG<sup>664</sup>R<sup>665</sup>xxR<sup>668</sup>) in the Cas3 RecA2 domain (112). The structure and function of Cas3 are described in detail in the next paragraph.



**Figure 1.6 Cas3 structure and its individual functional domains.**

*PHYRE2-predicted Escherichia coli Cas3 structure, modeled from Cas3 structures solved from Thermobaculum terrenum (PDB 4Q2C) and Thermobifida fusca (PDB 4QQW, 4QQX and 4QQY), represented in two orientations and with a corresponding cartoon of primary sequence presented below. We highlight the HD domain (purple), two RecA-like domains (green and orange) and the accessory C-terminal domain (CTD, pale blue). Active sites comprising the Asp-His HD domain and the amino acid DEVH motif of one RecA-like domain for ATP-hydrolysis are highlighted in red spheres within the structures, and marked on the cartoon primary structure. Also marked are the prominent solvent-exposed alpha helix (ACH) and the arginine rich IV motif.*

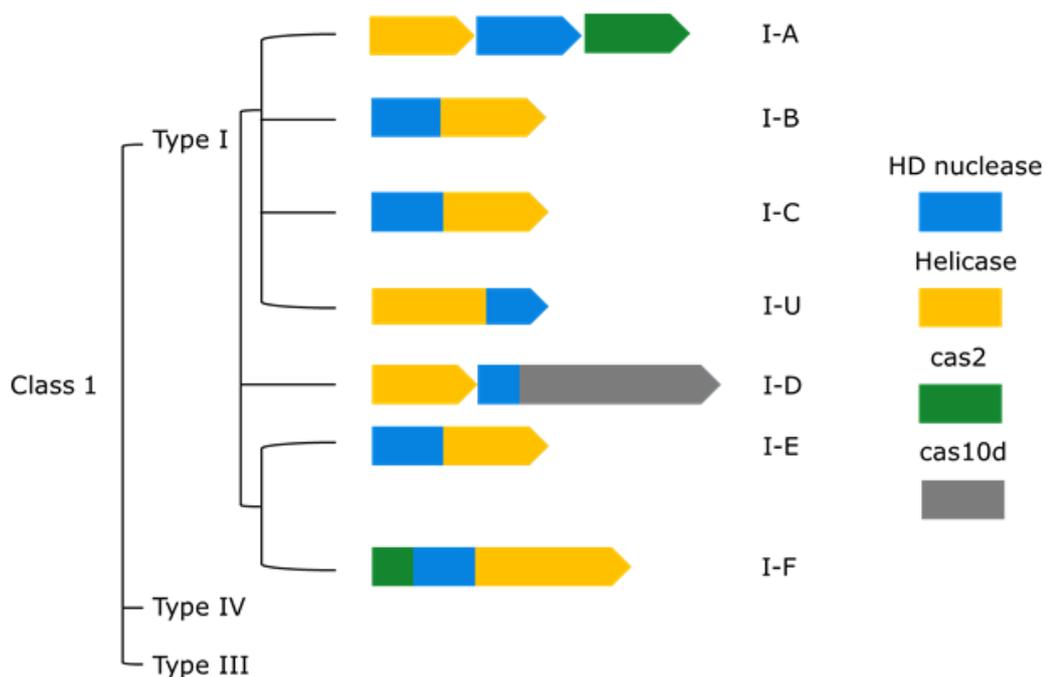
Cas3 was first highlighted in 2002 during in silico analyses of prokaryotic genomes that identified its superfamily-2 helicase motifs (as 'COG1203') and its association with other proposed nucleic acid processing enzymes located alongside repeat DNA sequences (6,113). After the created term 'CRISPR' had begun to be used more

broadly to denote the DNA repeats-spacers cluster, the protein COG1203 became a CRISPR-associated protein, Cas3 (6). The discovery and establishment of CRISPR biology and biotechnology proposed the hypothesis that Cas3 is essential for the CRISPR-Cas system, which is an RNA-interference-based prokaryotic immunity system (23,24,33,59,114). In 2008, it was proved a necessary effector for CRISPR interference, in cooperation with the 'Cascade' complex comprising proteins CasA, B, C, D and E and crRNA (115). Research subsequently to define biochemical activities of Cas3 are covered in Section 1.2.2. However, outstanding questions remain in Cas3; these will be described in Section 1.4. In Chapters 3-5, I report my attempts to reveal some of the unclarified features of Cas3.

### **1.2.1 Diversity in Cas3 proteins of Type I CRISPR-Cas system**

The 'signature' of Type I CRISPR-Cas systems is the Cas3 protein and the ribonucleoprotein complex Cascade (1). Cas3 reveals the most differences in protein organisation between subtypes (Figure 1.7). In subtypes I-B, I-C and I-E, Cas3 proteins exist in a protein fusion form, that is an N-terminus Cas3 HD-nuclease domain fused to a C-terminal helicase domains (33,116,117). Whereas the Cas3 fusion in the I-G subtype is an exception, its nuclease and helicase regions are fused in reverse (1). Some fused Cas3 HD-nuclease-helicase proteins are further fused at the N-terminus with the Cas2

adaptation protein, apparent in some *Yersinia* and *Pseudomonas* bacteria of the I-F subtype (118-120). Conversely, in some archaea, Cas3 functions are split into distinct HD-nuclease (Cas3'') and translocase/helicase (Cas3') proteins encoded by different open reading frames (1,2,23,48). Such fission Cas3 proteins are observed in I-A and I-D subtypes. And finally, some fissile Cas3'' HD nucleases are themselves fused with Cas10d protein at the C-terminus, in I-D subtype CRISPR identified in some cyanobacteria (1). The HD-nuclease-helicase fusion Cas3 is the most common form and is the focus of this thesis.



**Figure 1.7. Diverse Cas3 in Type I CRISPR-Cas systems**

This shows a summary of Cas3 proteins in a different format in known Type I CRISPR-Cas systems. Figure adapted from Makarova, et al. (2020) (1).

## **1.2.2 The nuclease activity of Cas3 compared with other nuclease Cas proteins**

### **1.2.2a Structural basis of Cas3 for cleaving nucleic acids**

Cas3 nuclease activity destroys MGEs recognised by the Cascade-crRNA surveillance complex, as part of processes called 'CRISPR interference' (Section 1.1.2). In *E. coli* Cas3, the HD-nuclease domain is named after the crucial functional amino acids histidine (H) and aspartate (D). The HD nuclease domain of Cas3 hydrolyses DNA or RNA (23,33), utilizing five key active site amino acids (H<sup>30</sup>, H<sup>74</sup>, D<sup>75</sup>, H<sup>159</sup> and H<sup>160</sup>, *E. coli* numbering) (Figure 1.6). The specific roles each amino acid exerts are not fully understood yet, but the Asp-75 and His-159/160, together with the conserved Asp-229, may be necessary for metal ions binding and interaction with water molecules. This is apparent from aligning the structure of *E. coli* Cas3 with its homologous *Pae*Cas2-3 (PDB: 5B7I, 5GQH) in the I-F CRISPR-Cas subtype.

In fusion Cas3 proteins (e.g., *E. coli* Cas3), the HD-nuclease domain is fused to RecA-like domains characteristic of superfamily-2 helicases (Figure 1.6) (110,111). These co-operate to deliver ATP-dependent ssDNA translocation and degradation (33,64). Several conserved features of Cas3 enzymes are likely to be important for the coordination of ssDNA from RecA domains into the HD-active site, including a prominent solvent-exposed alpha helix (ACH), and

## L. He Chapter 1

tyrosine/tryptophan residues located at the interface of HD-RecA1 domains (iHDA1) and elsewhere (40,71,109,116). However, this coordination between Cas3 helicase and nuclease domains may not be necessary for maintaining the efficacy of CRISPR interference. Splitting I-C Cas3 fusion into helicase and nuclease portions and co-overexpression of them can still conduct effective MGEs elimination during CRISPR interference (116). This indicates the interaction between two Cas3 functional domains is less likely to be necessary for interference, but this domain-domain coordination may be essential for other unknown mechanisms.

### **1.2.2b Metal-ion dependency of HD Cas3 domains**

Cas3 HD-nuclease function can be supported by a variety of metal ions (23,121-123). Crystal structures of Cas3 display HD-active sites bound to iron, manganese and calcium (PDB 3M5F) (71,109). Biochemistry of Cas3 *in vitro* indicates nuclease activity stimulated by nickel, manganese and cobalt and inhibited by iron, and cobalt was necessary to observe stable interaction between Cas3 and Cascade in one system at least (69,74,123).

Cas3 subtypes seem to have preferences for different metal ions in order to exert their nuclease activities. *E. coli* Cas3 (I-E subtype) has shown stalled DNA hydrolysis when it is incubated with magnesium ( $Mg^{2+}$ ), but it becomes an active nuclease when it is

## L. He Chapter 1

interacted with nickel ( $\text{Ni}^{2+}$ ) and cobalt ( $\text{Co}^{2+}$ ) (74,122). The other Cas3 proteins, including *T. fusca* Cas3 (I-E), *B. halod* Cas3 (I-C) and *P. ae* Cas2-3 (I-F) are all able to cleave DNA when their HD active sites are bound to  $\text{Mg}^{2+}$  (109,116,118,120). Why the *E. coli* Cas3 is an exemption when utilising  $\text{Mg}^{2+}$  is still unknown.

### **1.2.2c Modulation of Cas3 nuclease by Cascade and perhaps Cas1**

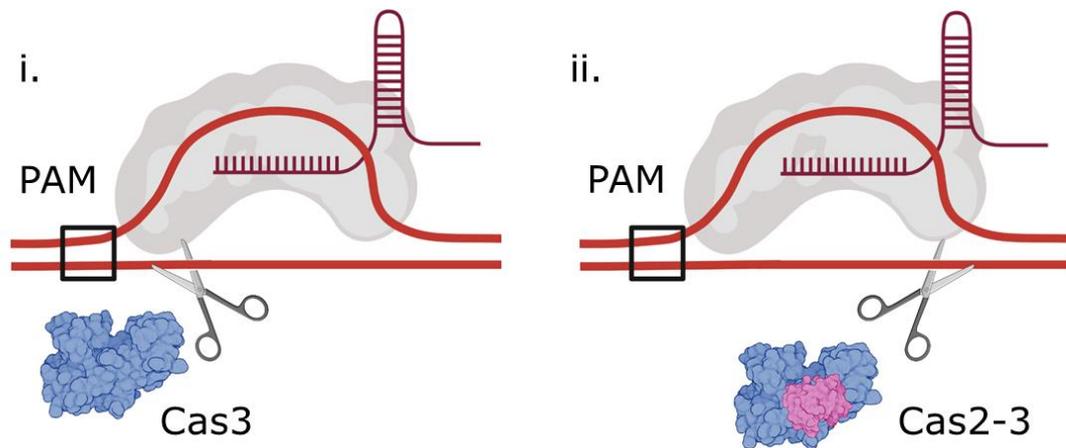
In cells in natural CRISPR systems, Cas3 nuclease is targeted to MGEs by Cascade identifying MGE. Cascade identifies a DNA PAM sequence at a target site (66), followed by 'seed' stabilised DNA-binding from RNA-DNA base-pairing, eventually forming an R-loop that is 'locked' onto the target DNA (26,33,69,70,74,124). In this state, the conformation of the Cse1 component of Cascade reveals Cas3 interacting regions. Cas3 will load onto the ssDNA that is generated in R-loops, but interaction with the Cse1 subunit of Cascade channels its nuclease activity for CRISPR immunity (64,65,124,125). Analysis using Hydrogen-Deuterium Exchange (HDX) coupled to mass spectrometry identified alpha helix H1 of Cse1 as a critical docking and/or activation site for interaction with Cas3 that may be the molecular trigger for degradation of targeted DNA by the Cas3 HD-nuclease activity (72-74). This interaction with Cascade can increase Cas3 nuclease efficacy, as revealed in I-E, I-C and I-F CRISPR subtypes (116,118,122).

## L. He Chapter 1

In addition, Cas1 may play a regulatory role over Cas3 nuclease activity and reveal an opposite effect compared to what Cascade does to Cas3. When *P.ae* Cas2-3 conducts DNA degradation on MGEs targeted by Csy (the Cascade in subtype I-F), its nuclease activity is reduced in the presence of *P.ae* Cas1 (118). Taking into account that the *P. ae* Cas1 and Cas2-3 form a stable Cas1-Cas2-3 complex, reduced Cas2-3 nuclease function is likely caused by protein-protein interaction with Cas1. Since no further study regarding Cas1 inhibits Cas2-3 function has been conducted, mechanism and verification for this is still unclear.

### **1.2.2d Comparing I-E Cas3 and I-F Cas2-3**

Cas3 proteins in different CRISPR-Cas subtypes have the same functional domains, but there is substantial plasticity in how the Cas3 proteins are functioning. Studies of I-E Cas3 proteins in *E. coli* and *Streptococcus thermophilus* (*S. ther*) suggest that Cas3 preferentially nicks DNA close to the PAM identified by Cascade (64,122). However, the *P. ae* Cas2-3 fusion protein, in the I-F subtype, favours the distal end of the PAM site for DNA nicking. In addition, these sites selected by *P. ae* Cas2-3 are CG-rich sequences (126), rather than the NTT (where N is any nucleotide) *E. coli* sequence Cas3 prefers (64,107) (Figure 1.8).



**Figure 1.8 Schematics of two Cas3 subtypes nicking non-targeting DNA at different sites.**

(i, ii) MGE and crRNA are denoted in red, Cascade is grey, Cas3 is in blue, and Cas2 fusion is in orchid. (i). This shows I-E Cas3 prefers nicking PAM proximal sequence. (ii). This shows I-F Cas2-3 nicks at distal end of the PAM site.

### 1.2.2e Comparing Cas3 and Cas9 nucleases

The Type I and II CRISPR-Cas systems adopt distinct MGEs surveillance-interference mechanisms. Type I CRISPR utilises Cascade to identify non-self-nucleotides that recruits Cas3 to destroy the invader DNA. In contrast, Cas9, in Type II CRISPRs, is a single polypeptide that couples MGE recognition and DNA cutting (1,2).

The DNA hydrolysis mechanism is different ways in Cas3 and Cas9. Cas3 uses the HD-nuclease domain and targets the non-targeting strand on MGEs (26,64), Cas9 has two nuclease domains, a RuvC-like active site that cleaves the targeting DNA and an HNH-like active site that uses the non-targeting DNA as substrates (21,62). The

consequence of Cas9 nuclease is generation of DNA double-strand breaks, but Cas3 produces long range DNA degradation, *T. fusca* Cascade-Cas3 (I-E subtype) conducts unidirectional DNA deletion up to 77.7 kb away from the Cascade-recognition (PAM and R-loop) site (127), whilst *P. ae* I-C subtype Cascade-Cas3 generates bidirectional DNA deletion up to 424 kb on genomic DNA (98).

### **1.2.3 Cas3 translocase-helicase activity**

Cas3 loads onto ssDNA and translocates it powered by ATP-hydrolysis (128). This can separate DNA duplex strands as a helicase and may be able to displace other DNA-binding proteins during translocation (128-131). After the DNA nicking by Cas3, Cas3 translocates DNA with 3' to 5' directionality (33), generating ssDNA gaps from a Cascade-R-loop formed on duplex DNA, which is detectable in single-molecule experiments using GFP-RPA (73). Single-molecule studies of Cas3 from *E. coli* and *T. fusca* show translocation over thousands of nucleotides at mean velocities of 89-316 bp/s (70,73). *E. coli* Cas3 was found to translocate away from Cascade rapidly, whilst Cascade remained bound to the target site (73,122). In this system, the interaction of Cas3 with Cascade, and its subsequent translocation away from Cascade, nucleolytically destroys the substrate (ssDNA) onto which Cas3 loads at the target site, so no re-association of Cas3 and that same Cascade was

## L. He Chapter 1

observed (122). However, when another I-E *T. fusca* Cas3-Cascade was used, half of the translocating Cas3 molecules remained associated with Cascade (70). In this system, a sustained association of Cas3-Cascade may remain essentially fixed in place, as DNA moves relative to it by 'reeling' or 'looping' powered by ATP-dependent ssDNA translocation of Cas3 (70,74). The dissociation of *T. fusca* Cas3 from Cascade then resulted in Cascade re-binding to DNA target sites.

The apparently distinct Cas3 translocation activities, observed in two I-E subtypes either moving away from or binding to Cascade, may be artifacts attributed to the different protein tags and experimental conditions used in those two studies; Cas3 translocating away from Cascade-R-loop complex (73) was fused to a chitin-binding domain at its C-terminus as an affinity tag for protein purification. This additional polypeptide may disturb the function of the Cas3 CTD domain, which has been shown to facilitate Cas3 interaction with Cascade, reducing Cas3-Cascade interaction affinities (109,116). In contrast, the Cas3 with an MBP-tag fused to its N-terminus, 'reels' DNA and has an undisturbed CTD (70,74). The opposite conclusions proposed in those two publications highlight the role of the Cas3 CTD domain in maintaining stable Cas3-Cascade interaction, which is crucial for conducting CRISPR interference (116).

## L. He Chapter 1

Reeling of DNA through Cas3 resembles mechanisms of DNA translocation by the DNA replication-recombination-repair helicases PcrA and RecBCD, although with significant differences in DNA translocation velocity and step-size. DNA targeting by Cascade is pivotal for Type I CRISPR systems because it triggers destructive catalysis of Cas3. Initial collision of Cascade complexes with DNA then requires transfer to target sites defined by a trinucleotide PAM sequence—5'-AAG in *E. coli*—that will direct Cascade to 'lock-on' to DNA in the presence of a complementary crRNA-DNA sequence for R-loop formation (66).

DNA target site identification by interference complexes follows biophysical principles established for other site-specific DNA targeting enzymes, which are conceptualised or described in several possible ways (132,133). For the Cascades, current areas of disagreement arise from analyses of target site recognition. In TIRF and FRET studies, *E. coli* Cascade favoured three-dimensional (3-D) diffusion to sample physical space for DNA target sites, analogous to 'jumping' at DNA (73,134). On sensing a PAM, this Cascade dwells to establish if R-loop formation is possible by RNA-DNA base-pairing. In this model the PAM allows Cascade to discern kinetically whether the DNA is a bona fide target, or not (134). A similar study (70) using *T. fusca* Cascade favoured facilitated diffusion of Cascade over one-dimensional (1-D), in which Cascade essentially 'slides' or 'hops' along DNA whilst determining target sites. This highlighted the

importance to the process of five conserved lysine/arginine amino acid residues within the Cascade structure, in contrast to the *E. coli* Cascade in which three of these five residues are present, a factor that may explain the discrepancy in behaviours of these two Cascades. Another factor could be that analysis of the 1-D target site search by *T. fusca* used DNA that lacked an identifiable RNA-DNA sequence complementarity that would allow R-loop formation, despite PAMs being present. In any event, successful engagement of Cascade with DNA target recruits the Cas3 translocase-nuclease.

#### **1.2.4 Cas3 interacts with Cas proteins - the formation of PAC**

DNA reeling by PcrA displaces RecA protein from DNA during DNA replication-repair (135), giving precedence to a possible role for Cas3 reeling in removing bound proteins during CRISPR interference reactions. This aspect of Cas3 is particularly interesting; biophysical data indicates that *T. fusca* Cas3 and Cas3-Cascade are ineffective at removing ectopic protein roadblocks from DNA (70), including RNA polymerase barriers that are most relevant to bacterial chromosome dynamics (136). However, the addition of the CRISPR adaptation complex Cas1-Cas2 to the Cas3 Cascade, forming a new machine called the primed adaptation complex (PAC), triggered the displacement of RNA polymerase 63 % of the time (70). In this

system, it is not known if the association of Cas1-Cas2 activates the motive power of Cas3 as a DNA translocase or if Cas1-Cas2 itself modifies some aspect of the PAC or the roadblock to facilitate its displacement. In *Pseudomonas*, the nuclease activity of the fused Cas2-Cas3 protein is inhibited strongly by association with Cas1, but this effect is counteracted when the Cas1-Cas2/3 complex encounters Cascade (called 'Csy' in this system) that is bound to a target DNA (118).

Physical and functional interaction of the *T. fusca* PAC and Cas1-Cas2/Cas3/Cascade in other systems more generally provides the important biological role linking interference with DNA fragment capture and integration reactions of 'adaptation', recently reviewed (137-139). Physical and functional coupling of interference with adaptation in a PAC can re-cycle DNA from invader into CRISPR loci, updating immunity in 'primed' or 'targeted' adaptation (107,140-147). FRET measurements indicate that Cas3 DNA reeling is stimulated by Cascade but that Cas3 nuclease activity is weak when compared to nuclease activity recorded from bulk/ensemble biochemical assays (74). However, in this study, no measurements were made of the PAC by the addition of Cas1-Cas2. It would be useful to determine if nuclease activity within the PAC is stimulated relative to Cas3-Cascade and, if so, whether this is attributable to Cas3, Cas1-Cas2, or both; data suggest that Cas1-Cas2 and Cas2 alone have nuclease activity (11,148,149). This would help to

understand further the molecular events underpinning the various stages of primed adaptation.

### **1.2.5 High-temperature protein G (HtpG) and Cas3 in CRISPR interference**

Cas3 function as an essential part of CRISPR immunity requires chaperone protein HtpG (150) in *E. coli* and probably in other bacteria, although HtpG is not widely distributed across archaea. Functional interaction between Cas3 and HtpG is observed *in vivo*; CRISPR immunity against lysogenization by  $\lambda$  prophage at 32°C was lost when HtpG was lacking ( $\Delta$ htpG), an effect suppressed by Cas3 over-expression. The efficacy of CRISPR immunity against  $\lambda$  phage infection is temperature-dependent; cells lacking HNS protein ( $\Delta$ hns) that are engineered to target  $\lambda$  phage are effective at resisting infection at 30°C but not at 37 °C (108) (Figure 1.9 B, section 1.4.1), but cells lacking HtpG and HNS are phage-sensitive. Over-expression of Cas3 or HtpG rescued this phage sensitivity, but only at 30°C: The same cells remained sensitive at 37 °C (108). The reason for Cas3 instability in  $\Delta$ hns cells at 37 °C is not known but may be useful in revealing other roles for Cas3 in cells, and a better understanding of the molecular basis for interaction between Cas3 and HtpG will be helpful.

## **1.3 Potential non-canonical roles of Cas3 in non-CRISPR-Cas systems**

### **1.3.1 Cas3 interactions with non-CRISPR proteins**

A genome-wide protein interaction study (151) using >4000 immobilised *E. coli* proteins identified seven proteins that physically interacted with hexahistidine tagged Cas3 (at that time called YgcB): GroL, DnaE, GyrA, AceE, IbpA, CrfC, and MdoD each co-eluted with Cas3 immobilised on Ni-NTA beads. More generally across bacteria and archaea, functional interactions of CRISPR systems with DNA repair systems are important for the efficient functioning of CRISPR immunity, especially for adaptation reactions that build CRISPR loci from captured invader or host DNA fragments (10,12,79,148,152,153).

### **1.3.2 The interplay between Cas3 and RNA**

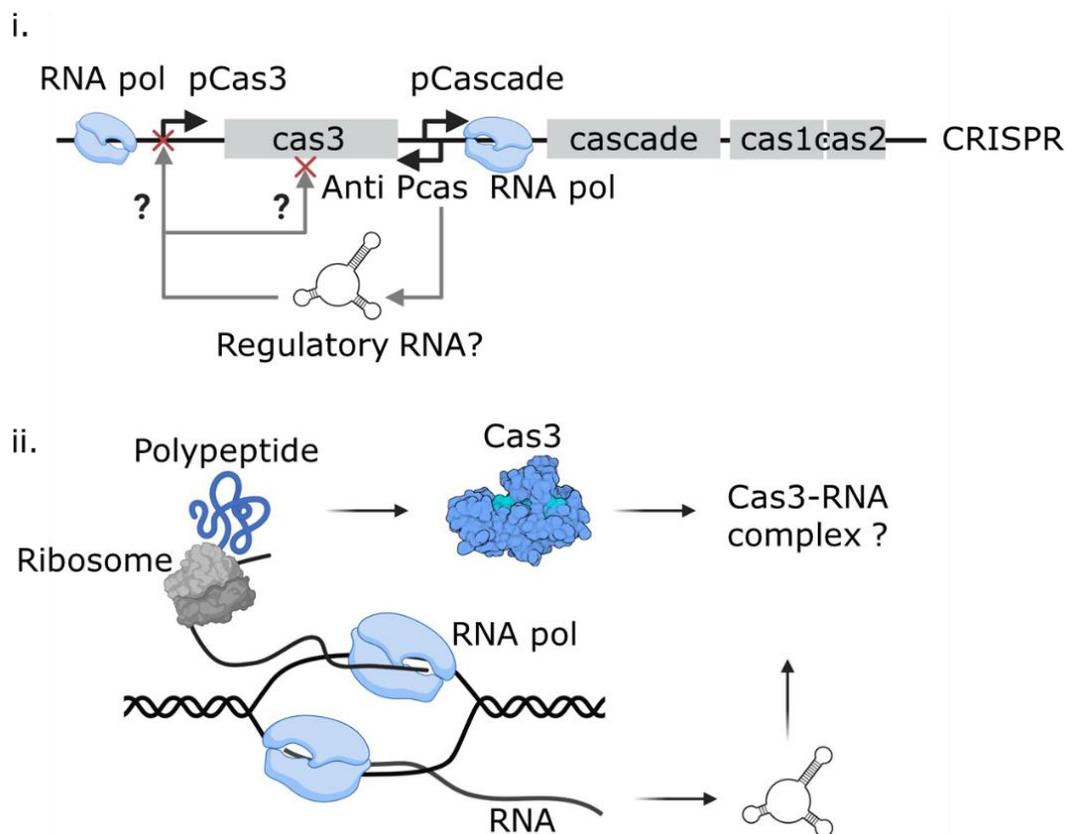
An interesting set of observations implicates Cas3 in several other cellular events that centre on or are at least associated with the regulatory processing of RNA molecules. Ectopic over-expression of Cas3 in *E. coli* cells stimulated uncontrolled DNA replication of plasmids from their ColE1 replicons but had no such effect on non-ColE1 plasmids (154). These plasmids use an R-loop at their ori sites to initiate replication formed by plasmid-encoded RNA molecules,

## L. He Chapter 1

one that primes replication (RNAII) and the other that prevents priming (RNAI) (154). Cas3 stimulates priming of replication, dependent on its ATPase/translocase activity independently of Cascade. It is proposed that Cas3 can specifically target RNAI-RNAII molecules for dissociation, liberating RNAII to pair with DNA at ori and priming replication. This unusual property of Cas3 to stimulate uncontrolled plasmid replication is temperature-dependent, increased at 37 °C and decreased at 30 °C (Figure 1.10 A). Reduced or inhibited ATPase/helicase activity of Cas3 at 30 °C was found responsible for this effect, and  $\Delta$ hns mutation was found indirectly to decrease plasmid copy number at 30 °C (155).

Purified Cas3 from *E. coli* and the archaeal species *Methanothermobacter thermautotrophicus* can manipulate RNA molecules *in vitro*. In common with helicase/translocase enzymes of the RecQ family (156-158), Cas3 can anneal nucleic acid strands as well as unwind them. Purified Cas3 is especially robust at annealing RNA strands into plasmid DNA (i.e., R-loop formation) (24). This activity required the Cas3 HD domain but is independent of ATPase activity. HD domain-catalysed nicking of DNA is thought to facilitate RNA-DNA annealing by relaxing plasmid supercoiling. It is not known what, if any, specific role this may have in cells, but it does support ideas that Cas3 may transmit RNA signals into cellular networks more widely than CRISPR systems. Analysis of the organisation of the *E. coli* MG1655 CRISPR-Cas system identified a promoter, 'anti-

Pcas', located just downstream of the Cas3 gene coding sequence, which generates detectable RNA transcript of 150–200 nucleotides that would overlap with the Cas3 transcript (25) (Figure 1.9). The authors of that study predict that the anti-Pcas transcript may form an elaborate folded structure reminiscent of riboswitches and other regulatory RNAs. The significance of this, and any role for Cas3 in its putative function, is not known.



**Figure 1.9. The anti-Pcas RNA encoding gene overlaps the non-coding strand of the Cas3-encoding gene.**

Panel i denotes a few crucial promoters for transcription of Cas genes and the anti-Pcas promoter for generating a potential regulatory RNA, which may play a role in regulating Cas3 gene transcription. Panel ii shows a hypothesis that Cas3 and anti-Pcas RNA may form a protein-RNA complex.

### **1.3.3 Cas3 implicated in Biofilm formation**

Regulatory roles for CRISPR systems are identified in several contexts (159-161). The DNA targeting capabilities of Cascade, and observations that CRISPR systems frequently capture 'self' DNA from both host chromosomes and 'non-self' invader DNA, indicate that CRISPR systems may regulate gene function as transcription factors (162,163). Cas3 influences group behaviour of *Pseudomonas* cells—formation of biofilms is inhibited in response to lysogeny by phage DMS3, a response that requires Cas3 to be fully functioning for HD-nuclease and helicase activities (164). In these cells, chromosomal sequences partially match sequences of DMS3 that might allow binding of Cascade sufficient to trigger the DNA nicking step of Cas3 and some ssDNA formation at the Cascade target site. In this model, the ssDNA generated may be enough to trigger the loading of the DNA recombinase RecA, triggering a DNA repair "SOS" response that is known to be able to block biofilm formation (165).

### **1.4 Outstanding questions about Cas3**

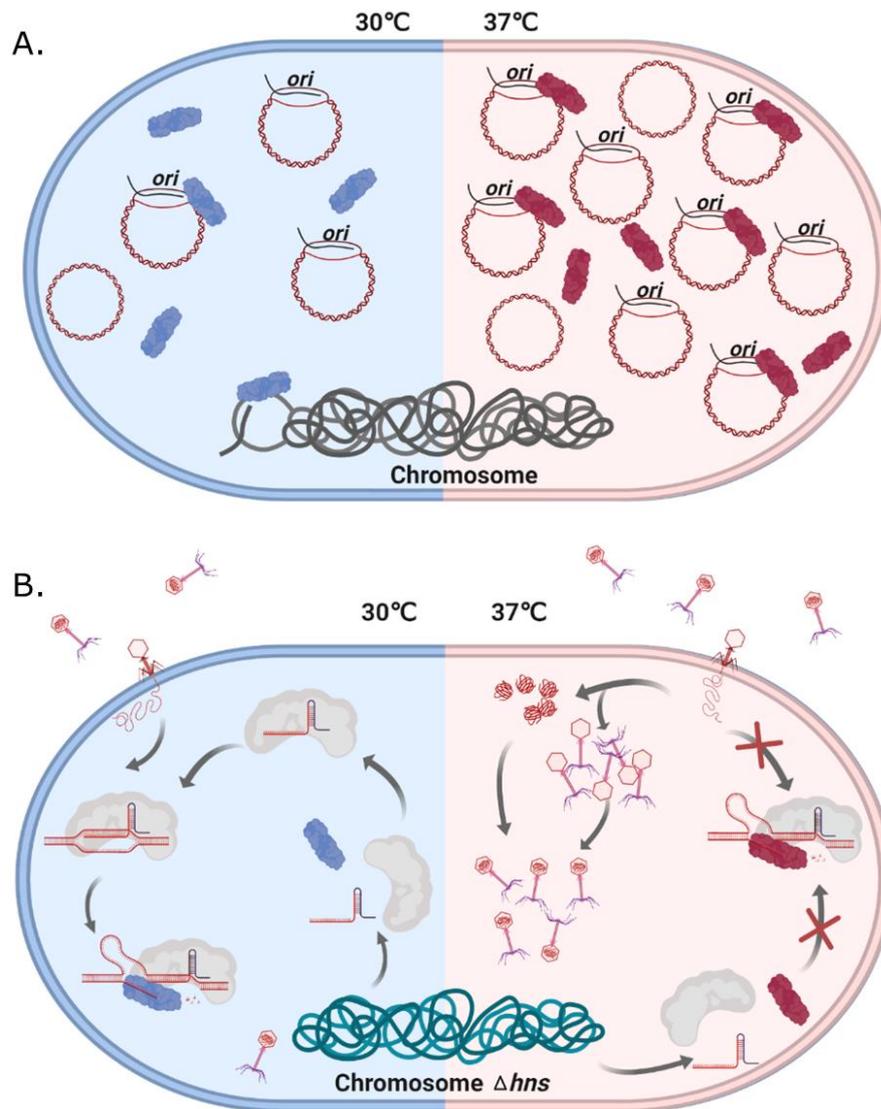
Cas3 is firmly established as an essential component required for CRISPR immunity in model systems (e.g., *E. coli* Cas3) (26,33). It's biochemical mechanisms as a helicase and translocase, both in isolation and as part of complexes with Cascade, has been extensively studied and utilised for genome editing in eukaryotic

cells (127). However, interactions and contributions of Cas3 to prokaryotic physiology more broadly, and how these impact on its biochemical activities are much less clear (40). A few outstanding questions regarding this are described in Section 1.4.1.

#### **1.4.1 Cas3 reveals temperature-dependent activities**

In Section 1.3.2, a Cas3 study was briefly introduced. This study revealed Cas3 facilitates runaway replication effects on plasmids, which is ColE1-dependent (154). Notably, this function of Cas3 occurs at 37 °C but not at 30 °C (155). This suggests Cas3 may undergo a 'switch-like' functional change that is temperature-dependent. The same phenomenon was then observed during the study of CRISPR interference against  $\lambda$  phage (108). In  $\Delta hns$  *E. coli* cell strain, which overcomes the downregulation of protein binding protein H-NS on the endogenous CRISPR-Cas system, the interference efficacy provided by Cascade-Cas3 shows temperature dependency.  $\Delta hns$  strains infected with  $\lambda$  phage gave more plaques at 37 °C than the same analysis at 30 °C. These results indicate that CRISPR interference, at least when focussed on Cas3, is sensitive to temperature. Further investigation suggested that this repression of CRISPR interference at 37°C can be restored if over-expressing Cas3 or a chaperon HtpG that can maintain a functional level of Cas3 *in vivo* (108). All these clues indicate the enzymatic functions of Cas3

may be temperature-dependent (Figure 1.10), and the effect may involve protein structure rearrangement within Cas3 at distinct temperatures.



**Figure 1.10. Temperature-dependent Cas3/CRISPR functions in *E. coli* and the *E. coli*  $\Delta hns$  strain**

(A). Cas3 stimulates uncontrolled ColE1 plasmid replication in a temperature-dependent manner. ColE1 plasmid yield was stimulated by Cas3 in cells at 37°C but not at 30°C. This requires a functional helicase domain. We speculate that this phenomenon may indicate a possible change in Cas3 conformational 'state', illustrated by colouring Cas3 in blue 'demotivated' or red 'motivated'. At 37°C, 'motivated' Cas3 may interact

*with R-loop formation in ori by dissociating RNA II from the complementary strand, and lead to increased plasmid replication in vitro. (B). Temperature impacts CRISPR function in E. coli cells lacking H-NS. This  $\Delta hns$  strain at 30°C can defend against invader DNA during phage infection, even though Cas3 is in an 'unmotivated' state. However,  $\Delta hns$  strain cannot survive under infection pressure at 37°C.*

#### **1.4.2 MGE pre-spacer DNA fragments: Are they really generated by Cas3?**

The close connection between CRISPR interference and primed adaptation generates an ambiguous consent that the Cas3 nuclease activity on MGEs generates short DNA fragments for Cas1-2 to capture and process into new spacers. The study by Kunne, T. *et al.* (107) is cited most often as providing evidence for this. They presented data showing that Cas3 nuclease generates DNA products with NTT sequence at the 3' end, including the CTT PAM sequence, and that incubating Cas1-2 with Cascade, Cas3 and target DNA can produce DNA substrates similar to the 'half-site integration' product. However, the *in vitro* interference-adaptation assay did not include the IHF, so it was missing one of the important cofactors for completing prime adaptation in a natural state. In addition, they did not address an important factor: Cas3 generates ssDNA fragments whilst Cas1-2 mediated integration uses dsDNA. So, what causes this transition from ssDNA to dsDNA?

The theory of Cas1-2 utilising DNA degradation products generated by Cas1-2 seems to contradict the fact that decay of

foreign DNA loss is concurrent with primed adaptation. A comprehensive study carried out by Musharova, O. *et al* (67) on all possible PAM variants revealed a phenomenon indicating that primed adaptation is correlated with delayed foreign DNA loss that represents a low efficacy of Cas3 nuclease, whilst rapid CRISPR interference/MGEs degradation leads to less or no occurrence of primed adaptation. If Cas1-2 relies on Cas3 to acquire DNA fragments for primed adaptation then repression of Cas3 nuclease would cause a disadvantage for spacer integration by Cas1-2 and consequently lead to the occurrence of less adaptation, which contradicts what was observed in Musharova, O. *et al* (67). Therefore, how exactly does Cas1-2 acquire new spacers for primed adaptation? A hypothesis is proposed in Chapter 5 to explain this.

#### **1.4.3 Cas1-2 may assist Cas3 in conducting bidirectional DNA cleavage on MGEs from the protospacer site**

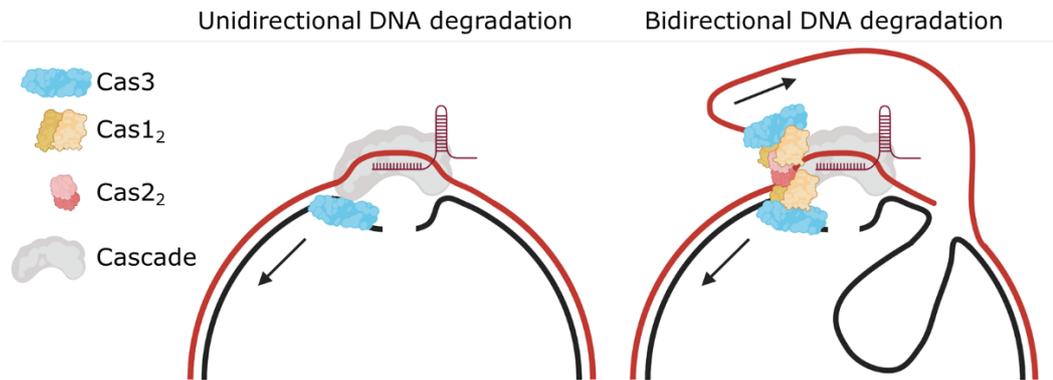
The mechanistic question about Cas3 and Cas1-2 as the source of pre-spacer fragments does not negate the clearly defined link between Cascade-Cas3 interference and Cas1-2 integrating new spacers, 'primed adaptation'. By an unknown mechanism, Cas1-2 captures DNA fragments located upstream and downstream of the protospacer when the MGE is targeted and processed by wild-type Cascade and Cas3 (47,76,147). In contrast, if CRISPR interference is carried out using a Cas3<sup>K320N</sup> mutant, the new spacers inserted

into the CRISPR locus are only complementary to the DNA sequence upstream of the protospacer. This Cas3 mutant was discovered by Nikita Vaulin in Peter the Great St. Petersburg Polytechnic University. A hypothesis proposed by Nikita Vaulin *et al.* (from conference poster, unpublished) suggested Cas3 may adopt a change of moving direction, thus facilitating DNA cleavage on strands flanking the protospacer, but this hypothesis is contradicted by the Cas3's polarity when it exerts catalytic functions on DNA (5' to 3') (23).

In addition, Cas3 bidirectional DNA cleavage is only occurred when CRISPR interference is alongside Cas1-2 (147); when Cas1-2 is absent in the system, Cas3 only conducts unidirectional DNA degradation (127). For example, when adapting Cascade-Cas3 as a genomic editing tool in eukaryotic cells, Cas3 generates large-range unidirectional DNA cleavage, as described in Morisaka, H. *et al* (127). This study, together with the discovery made by Nikita Vaulin, suggests that Cas1-2 is likely to contribute to the Cas3 bidirectional DNA cleavage.

Figure 1.11 shows a predicted model for Cas3-catalysed bidirectional DNA degradation done with the assistance of Cas1-2. One molecule of Cas1-2 and two molecules of Cas3 may form a complex, in which two Cas3 proteins generate DNA degradation on both the targeting strand and the non-targeting strand on MGE. This combination of Cas1-2-Cas3 is inspired by the structural analysis of the I-F CRISPR-Cas system. Two individual studies described by

Rollins, M. F. *et al.* (118) and Fagerlund, R. D. *et al* (120) show that in the I-F system, the Cas1 and Cas2-3 forms a complex containing Cas1<sub>4</sub>-Cas2-3<sub>2</sub>. Meanwhile, as revealed by *in vitro* analysis, this protein complex is capable of catalysing spacer integration.



**Figure 1.11. Illustration showing the mechanisms of Cas3 unidirectional and potential bidirectional DNA degradation**

*Cas3* has been shown to eliminate the upstream DNA sequence of the targeting site recognised by Cascade-crRNA (left). The right side shows a hypnotised primed adaptation complex (PAC) assembled by Cascade, Cas3 and Cas1-2, which may carry out a bidirectional DNA cleavage mechanism on MGEs. Cas3 translocation directions are denoted using arrows. The targeting strand is coloured red, and the non-targeting strand is black.

## 1.5 Aims and objectives of this Ph.D research

The studies presented here sought to gain more insight about the outstanding questions summarised in Section 1.4. Using the *E. coli* Cas3 from the I-E subtype CRISPR system, the main research focused on understanding the temperature-dependent effect related to Cas3 summarised in Section 1.4.1. This effect has been observed via genetic analysis of Cas3, but biochemical analysis was lacking. To this end, we purified *E. coli* Cas3 protein to interrogate its

## L. He Chapter 1

enzymatic function with focus on allosteric or intrinsic functional regulation of Cas3. To assist with this, Cas3 mutants were made to compare with wild-type Cas3 protein of their biochemical features.

Completion of the above tests identified and characterised a crucial region in Cas3 that is capable of achieving conformational rearrangement, thus facilitating Cas3 'switch-like' functional change regarding environmental temperatures. A hypothesis was then proposed suggesting that this conformational change of the Cas3 region may correspond to that stimulated by protein-protein interaction. To validate this theory, Cas2-3 fusion protein in the I-F CRISPR system was introduced in this study, for comparison of biological characteristics with its close homologue *E. coli* Cas3.

Another research topic included in this thesis is related to the question of how Cas1-2 acquires a new spacer during primed adaptation, which was proposed in Section 1.4.2. Due to limited time, this part of the work has not yet been completed. However, preliminary data from investigating this gave interesting results that are reported here, with the hope to be able to develop these with further research studies.

## Chapter 2

### Materials and Methods

#### 2.1 Bacterial strains

**Table 1. List of cell strains used in this work**

Name	Genotype	Company
DH5 $\alpha$	<i>F- <math>\phi</math>80lacZ<math>\Delta</math> M15 <math>\Delta</math> (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK- mK+) phoA supE44 <math>\lambda</math>- thi-1 gyrA96 relA1</i>	Invitrogen
BL21AI	<i>fhuA2 [lon] ompT gal [dcm] <math>\Delta</math>hsdS</i>	NEB

#### 2.2 Cloning

DNA amplifications by the polymerase chain reaction (PCR) utilised Vent DNA polymerase (New England Biolabs, NEB, M0254S), unless mentioned otherwise. For a 50  $\mu$ L PCR reaction, it contains 50 ng template DNA (15 ng DNA for doing mutagenesis), 200 nM of forward primer, 200 nM reverse primer, 200  $\mu$ M of each dATP, dTTP, dCTP and dGTP, 1X ThermoPol Reaction Buffer (NEB, B9004S) and 1 unit of Vent DNA polymerase. For doing mutagenesis using Q5 high-fidelity DNA polymerase (NEB, M0491S), a 50  $\mu$ L PCR reaction contains 15 ng template DNA, 500 nM of forward primer, 500 nM reverse primer, 200  $\mu$ M of each dATP, dTTP, dCTP and dGTP, 1X Q5 Reaction Buffer (NEB, B9027S) and 1 unit of Q5 DNA polymerase.

## L. He Chapter 2

To amplify target DNA, the thermocycling conditions are adapted from recommended DNA polymerase usage protocols supplied by NEB. The one for doing PCR using Vent DNA polymerase could be found in the link below:

<https://international.neb.com/protocols/2012/09/06/protocol-for-a-routine-vent-pcr-reaction>

The thermocycling conditions for PCR using Q5 DNA polymerase could be found in the link below:

<https://international.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491>

PCR products are further analysed on 1X TBE 1-2 % agarose gel. NEB 100 bp or 1kb DNA marker was used as a comparison.

For cloning via restriction digestion and DNA ligase, target DNA (insert) was processed by assigned restriction enzymes in following condition: 1X Cutsmart buffer (NEB, B6004S), PCR products (target DNA included), 4 units of restriction enzyme 1 and 4 units of restriction enzyme 2. Vector DNA was processed using the same method but with extra 1 unit of Alkaline Phosphatase (NEB, M0290) added in. Reaction was taken at 37 °C for 1 h, before separating target DNA from non-target DNA products by running 1-2 % agarose gel. Target DNA was then extracted from gel slice by using Wizard(R) SV Gel and PCR Clean-Up Kit (Promega) and eluted in ddH<sub>2</sub>O. Ligation of insert and vector was carried out in following condition: Molar ratio of insert : vector = 3:1, 1X T4 DNA Ligase Buffer and 80

## L. He Chapter 2

units of T4 DNA ligase (NEB, M0202S). DNA ligation was taken at 16 °C for 1 h or 20 °C for 20 min accordingly. Sample was then transformed into DH5 alpha chemically competent cells, and cells containing target plasmids were selected using Luria-Bertani (LB) agar plate containing relevant antibiotics.

For doing site-directed mutagenesis, target DNA was processed by DpnI in a reaction containing 1X Cutsmart buffer, PCR products (target DNA included) and 1-unit DpnI. This step is necessary for removal of template DNA (methylated DNA). Reaction was taken at 37 °C for 1 h followed by target DNA extraction using the same method described in above paragraph. DNAs were then incubated with 2 units of T4 Polynucleotide Kinase in 1X T4 DNA ligase buffer at 37 °C for 30 min before adding 80 units of T4 ligase and further incubating sample at 16 °C for 1 h or 20 °C for 20 min. Sample was then transformed into DH5 alpha chemically competent cells.

The recipes of LB liquid broth and LB agar medium used in cloning and other bacteria culture in this study could be found in the links below

[http://cshprotocols.cshlp.org/content/2006/1/pdb.rec8141.full?text\\_only=true](http://cshprotocols.cshlp.org/content/2006/1/pdb.rec8141.full?text_only=true)

<http://cshprotocols.cshlp.org/content/2009/3/pdb.rec11683.full>

### **2.2.1 Plasmids for protein overexpression**

The gene encoding Cas3 (*ygcB*) was amplified from *E. coli* MG1655 genomic DNA and cloned into pMalc2 (NEB) and pBAD-HisA

## L. He Chapter 2

(Thermo Fisher Scientific, UK) to construct plasmid pAH1 and pAH4 for overexpression and purification of Cas3 with an N-terminal maltose-binding protein (MBP) tag (pAH1) and poly-histidine tag (pAH4). The above two plasmids were made by Dr Jamieson Howard (University of Nottingham) in 2013. The plasmid for purification of His<sub>6</sub>-MBP-Cas3 protein was described in Mulepati *et al* (122), denoted in the following content as pBailey. Truncated *ygcB* was amplified from pAH1 and cloned into pMalc2 to construct plasmid pCas3<sup>ΔCTD</sup> for expressing Cas3 without its C-terminal domain (ΔCTD). Plasmid for the expression of *E. coli* Cas2<sup>E9A</sup> pTK40 was provided by Dr Tom Killelea (University of Nottingham). *E. coli ygcB*, together with genes encoding poly-histidine tag and MBP tag separately were sub-cloned into a modified FBP-Dasher-SV1\_origin\_TnGmPrpIJ vector to construct the pHis-Cas3-Pstrain and pMBP-Cas3-Pstrain for expressing *E. coli* Cas3 in *Pseudomonas putida* (NCIMB 9494) and *Pseudomonas fluorescens* (NCIMB 9046) strains. The primers used for amplifying the target genes are shown in Table 2. The vector used in this part of the work was reconstructed, with the addition of NotI, BamHI and SacI restriction sites, by mutagenesis (Section 2.2.2). Primers are listed in Table 3. These expression materials, including the vector and cell strains, were kindly donated by Dr Stuart Wood (Nanna Therapeutics, Cambridge, UK).

The genes encoding *Pseudomonas aeruginosa* (*Pae*) Cas1 and Cas2-3 were amplified from *Pae* UCBPP-PA14 genomic DNA and



**Table 3. Primers used for mutagenesis.**

Name	5' to 3'
<i>EcoCas3W149A_F</i>	TTATGAGTCCgcgTTTCCATGGGTAGAGGC
<i>EcoCas3W149A_R</i>	GGATGAGGAGCGGCATCA
<i>EcoCas3W152A_F</i>	CTGGTTTCCAgctGTAGAGGCCG
<i>EcoCas3W152A_R</i>	GACTCATAAGGATGAGGAG
<i>EcoCas3W230A_F</i>	GCTTGCTGACgctTTAGGCTCCTGG
<i>EcoCas3W230A_R</i>	GAGCAAAAACCTGCTAAC
<i>EcoCas3W406A_F</i>	TGTTGTCAGgctTTGTCACAAAGCAATAAGAAAG
<i>EcoCas3W406A_R</i>	TGAACCCACGCTTCTTCT
<i>EcoCas3R662GR671G_F</i>	catcgccatcatggcAAATATCGTCCCGCTGGT
<i>EcoCas3R662GR671G_R</i>	taaacggcccaatccTTGGAAAAGCAAATCTGCAG
<i>EcoCas3D75G_F</i>	TTTAATTGTTggtGAAGTTCATGC
<i>EcoCas3D75G_R</i>	ACACTTCGACCAATTCCC
<i>EcoCas3D452A_F</i>	TGTTTTAATTGTTgctGAAGTTCATGCT
<i>EcoCas3D452A_R</i>	CTTCGACCAATTCCTCAAACCA
<i>PaeCas2-3D124G_F</i>	GCTGTTCCACggtATCGGCAAGG
<i>PaeCas2-3D124G_R</i>	GCTGCCATCACCGTCAGC
InsertNotI_F	ccgcGCTGCAGCCACCTGCGCC
InsertNotI_R	ccgcAAGCTTGGCCTTGACGGCC

### 2.2.3 Plasmids for analysis of Cas3 in vivo protein-protein interaction.

#### 2.2.3a Plasmids used in proximity-dependent biotin identification (Bio-ID) system

The Bio-ID vector pTK47 was constructed by Dr Tom Killelea (University of Nottingham) using the gene (*birA*) encoding *Aquifex aeolicus* biotin-protein ligase<sup>R118G</sup> (Bio-ID) and pBAD-HisA vector. The Cas3 gene (*ygcB*) was cloned into pTK47 to construct pBioID-Cas3 for producing the *E. coli* Cas3 protein fused with a Bio-ID

protein at its N-terminus. The primers used in this part of the study are shown in Table 4.

### **2.2.3b Plasmids used in the Bimolecular fluorescence complementation system**

In this part of the study, a yellow fluorescent protein, mVenus, was split into mVenusN and mVenusC at amino acid 154/155 and fused into two target proteins separately to diagnose protein-protein interaction *in vivo*. The separation of the mVenus protein was carried out based on the description in Dillard *et al* (70). The genes encoding mVenusN and mVenusC were amplified from mVenus N1 (Addgene 27793) and cloned into pBAD-HisA and pRSFDuet1 (Novagen) to construct two new vectors: pBAD-mVenusN and pRSFDuet-mVenusC. The mVenus N1 plasmid was from Prof. Uwe Vinkemeier (University of Nottingham).

*E. coli ygcB* was sub-cloned into pBAD-mVenusN (pBAD-mVenusN-Cas3) for expressing Cas3 fused with the mVenusN at its N-terminus. The genes encoding *E. coli* DNA polymerase I (*polA*) and *E. coli* polymerase III catalytic subunit alpha (*dnaE*) were amplified from *E. coli* MG1655 genomic DNA and cloned into pRSFDuet-mVenusC to construct pRSFDuet-mVenusC-PolI and pRSFDuet-mVenusC-DnaE. Those two plasmids were used for the overexpression of Pol I and DnaE fused with an N-terminal mVenusC. All the plasmids were verified by Sanger sequencing.

**Table 4. Primers used for developing in vivo analysis on****Cas3**

Name	5' to 3'
BioIDN-Cas3_F	ATCGGAGCTCGAACCTTTTAAATATATATGCC
BioIDN-Cas3_R	ATATGCGGCCGCTTATTTGGGATTTGCAG
mVenusN-NheI_F	ATATGCTAGCATGGTGAGCAAGGGCGAG
mVenusN-NheI_R	ATATGCTAGCGGCGGTGATATAGACGTTGTG
mVenusC-NcoI_F	ATATCCATGGATGGACAAGCAGAAGAACGGC
mVenusC-NcoI_R	ATATCCATGGCCTTGTACAGCTCGTCCATG
RSFDuet_EcoRI_polA_F	ATATGAATTCAGTTCAGATCCCCAAAATCCAC
RSFDuet_NotI_polA_R	ATATGCGGCCGCTTAGTGCGCCTGATCCCAG
RSFDuet_EcoRI_DnaE_F	ATATGAATTCATCTGAACCACGTTTCGTAC
RSFDuet_SalI_DnaE_R	ATATGTCGACTTAGTCAAACCTCCAGTTCCA

**2.3 Protein expression and purification.**

The methods adopted to purify the *EcoCas3*, *PaeCas1* and *PaeCas2-3* proteins used in this thesis are mentioned in the chapters focusing on the analysis of relative proteins. The *EcoCascade*-crRNA complex in Chapter 3 Section 3.9, the *EcoCas1* fused to a Strep-tag used in Chapter 4 Section 4.7, the *EcoCas1-2* fused with a poly-histidine tag used in Chapter 5 Section 5.12 and 5.14, and the integration host factor (IHF) were purified by Dr Tom Killelea (University of Nottingham). The plasmid required for the expression of the high-temperature protein G chaperone HtpG was provided by Dr Ivana Ivančić-Baće (University of Zagreb). The HtpG used in Chapter 4 Section 4.8 was purified by Dr Edward Bolt (University of Nottingham) using HiTrap Heparin HP affinity column, Cibacron Blue 3G-A Sepharose and Hydroxyapatite column. This protein was then

## L. He Chapter 2

further purified using a Superdex S200 gel filtration column (GE healthcare) in a buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol. The eluted proteins were dialysed in a buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 30 % (v/v) glycerol, stored at -80 °C. The DNase I used in Chapter 3 was purchased from NEB. The DnaE used in Chapter 6 was a gift from Dr Katie Jameson (University of York). The *E. coli* DNA polymerase I used in Chapter 6 was purchased from NEB. This product does not fuse to any protein tag, based on the feedback from Dr Alvin Lee in the NEB Technical Support department. Product specification can be found via the link provided by Dr Alvin Lee.

[https://international.neb.com//media/catalog/specifications/m/0/m0209s\\_l\\_v1.pdf?rev=faff84bfa0f142b3a6de48f368cd580a&hash=398803F0D978EF4D49758DC51C31320C](https://international.neb.com//media/catalog/specifications/m/0/m0209s_l_v1.pdf?rev=faff84bfa0f142b3a6de48f368cd580a&hash=398803F0D978EF4D49758DC51C31320C).

The following sections contain information on the protein overexpression and cell lysate preparation used before the protein purification. The buffers used during the protein purification are also listed.

### **2.3.1 Preparing His<sub>x6</sub>-MBP-Cas3 proteins**

The method used to purify His<sub>x6</sub>-MBP tagged *EcoCas3* and mutants Cas3<sup>W149A</sup>, Cas3<sup>W152A</sup>, Cas3<sup>W230A</sup>, Cas3<sup>W406A</sup> and *EcoCas3*<sup>R662G,R674G</sup> is discussed in Chapter 3 Section 3.3.

The plasmid pBailey, and those containing mutations, were transformed into BL21AI and selected on agar containing 100 µg/mL

## L. He Chapter 2

ampicillin. The cells were grown in LB broth with 100 µg/mL ampicillin at 37 °C until the absorbance at 600 nm (OD<sub>600</sub>) reached 0.3; L-arabinose 0.2 % (w/v) and IPTG 0.5 mM were added at this point to induce protein overexpression. Growth continued for another 3 h at 37 °C before the cells were resuspended in a lysis buffer containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 % (v/v) glycerol and 0.5 mM phenylmethylsulphonyl fluoride (PMSF) and stored at –80 °C. The 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; it was sonicated for 10 sec and then paused for 10 sec for 3 min for each 5 mL suspension. Then samples were clarified at 48,000 g for 1 hour (Avanti® J-26 XP and JA25.50 rotor, Beckman Coulter Life Sciences). The sample was centrifuged at a relative centrifugal force of 48,000 g for 1 h at 4 °C. His<sub>x6</sub>-MBP-Cas3 was collected from the cell lysate using the ÄKTA Start Chromatography System (GE Healthcare, Sweden) attached to Ni-NTA resin, an MBPTrap HP 1 mL column and a 1 mL HiTrap Heparin HP affinity column separately. The resin and columns were purchased from GE Healthcare (UK). The buffers used in the purification are listed in Table 5.

Dialysing the purified protein for at least 12 h is necessary; otherwise, the acquired protein sample has no nuclease activity.

**Table 5. Buffers used when purifying His $\times$ 6-MBP-Cas3 proteins**

Buffer	Reagents	Resin/Column
Buffer A	20 mM Tris-HCl pH 8.0, 10 mM imidazole, 200 mM NaCl, 10 % (v/v) glycerol	Ni-NTA resin
Buffer Wash	20 mM Tris-HCl pH 8.0, 10 mM imidazole, 1 M NaCl, 10 % (v/v) glycerol	Ni-NTA resin
Buffer B	20 mM Tris-HCl pH 8.0, 250 mM imidazole, 200 mM NaCl, 10 % (v/v) glycerol	Ni-NTA resin
Buffer MA	20 mM Tris-HCl pH 8.0, 100 mM NaCl	MBPTrap HP column
Buffer MB	20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM Maltose	MBPTrap HP column
Buffer HA	20 mM Tris-HCl pH 8.0, 100 mM NaCl	Heparin HP affinity column
Buffer HB	20 mM Tris-HCl pH 8.0, 1 M NaCl, 10 % (v/v) glycerol	Heparin HP affinity column
Dialyse Buffer	20 mM Tris-HCl pH 8.0, 200 mM NaCl, 30 % (v/v) glycerol	N/A

### 2.3.2 Expression and purification of MBP-Cas3 proteins

Details of purifying MBP tagged *EcoCas3* and its mutants Cas3<sup>D75G</sup> and Cas3<sup>D452A</sup> is discussed in Chapter 5. When preparing the biomass, plasmid pAH1 and those carrying mutations were transformed into DH5 $\alpha$  (Invitrogen) and selected on agar containing 100  $\mu$ g/mL ampicillin. The cells were grown in LB broth with 100  $\mu$ g/mL ampicillin at 20 °C until absorbance at 600 nm (OD600) reached 0.1.

## L. He Chapter 2

IPTG was added to a final concentration of 0.5 mM to induce protein overexpression. Meanwhile, glucose 0.2 % (w/v) was added to repress the expression of alpha-amylase. Growth continued for another 24 h at 20 °C. At 12 hours of induction, fresh glucose was added to the culture for the second time to achieve a final concentration of 0.2 % (w/v). At the end of the protein expression, the cells were collected using the same method as mentioned previously in Section 2.2.1. The cells were resuspended in a lysis buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 % (v/v) glycerol and 0.5 mM PMSF and stored at –80 °C. Before the protein purification, the biomass was processed as previously described in Section 2.2.1. The buffers used during purification via amylose resin (NEB) and the HiLoad 16/600 Superdex 200 pg preparative SEC column (GE Healthcare) are listed in Table 6.

**Table 6. Buffers used when purifying MBP-Cas3 proteins**

Buffer	Reagents	Resin/Column
Buffer MA	20 mM Tris-HCl pH 7.5, 200/500 mM NaCl, 1 mM DTT	Amylose Resin
Buffer MB	20 mM Tris-HCl pH 7.5, 200/500 mM NaCl, 10 mM Maltose, 1 mM DTT	Amylose Resin
Buffer E	20 mM Tris-HCl pH 7.5, 200/500 mM NaCl, 1 mM DTT, 5 % Glycerol (v/v)	Superdex 200
Dialyse buffer	20 mM Tris-HCl pH 7.5, 200 mM NaCl, 30 % Glycerol (v/v)	N/A

*Note: 200 mM NaCl in Buffer MA, MB and Buffer E were used for the purification described in Chapter 5. Those containing 500 mM NaCl were adopted during the purification as described in Chapter 6.*

### **2.3.3 Expression and purification of His<sub>6</sub>-Cas3**

Plasmid pAH4 was transformed into BL21AI and selected on agar containing 100 µg/mL ampicillin. The cells were grown in LB broth with 100 µg/mL ampicillin at 37 °C until the absorbance at 600 nm (OD<sub>600</sub>) reached 0.6; L-arabinose 0.2 % (w/v) was added at this point to induce protein overexpression. Growth continued for another 12 h at 20 °C. The cells were collected and resuspended in a lysis buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 % (v/v) glycerol and 0.5 mM PMSF, and were stored at –80 °C. Before the protein purification, the biomass was processed as previously described in Section 2.2.1. The supernatant was collected and loaded onto a 5 mL HiTrap HP column connected to an ÄKTA start purification system. The column was charged with NiCl<sub>2</sub> and equilibrated with Buffer A (20 mM Tris-HCl pH 8.0, 10 mM imidazole, 500 mM NaCl and 10 % (v/v) glycerol) before the sample was loaded onto it. The protein was eluted with buffer B (20 mM Tris-HCl pH 8.0, 500 mM imidazole, 500 mM NaCl and 10 % (v/v) glycerol) via a linear gradient from 0 % to 100 % Buffer B (increasing imidazole concentration). The eluted protein was dialysed in 2 L dialyse Buffer (20 mM Tris pH 8.0, 10 % glycerol (v/v) and 100 mM NaCl) at 4 °C for at least 10 h. The dialysed protein was then loaded onto a HiTrap Heparin HP affinity column which has pre-equilibrated using a buffer containing 20 mM Tris pH 8.0, 5 % glycerol (v/v) and 50 mM NaCl. The Cas3 protein could not bind to the heparin, so it directly flowed

through the column. This step was adopted to remove other DNA-binding proteins, for example, DNA nucleases. Cas3 was then dialysed into a buffer containing 20 mM Tris pH 8.0, 30 % glycerol (v/v) and 100 mM NaCl, and stored in -80 °C.

### **2.3.4 Expression and purification of Cas2<sup>E9A</sup>**

*E. coli* Cas2<sup>E9A</sup> mutant was overexpressed and purified using the same method described in Ivančić-Baće *et al* (10). In this study, HiLoad 16/600 Superdex 200 pg column was used instead.

### **2.3.5 Expression and purification of PaeCas2-3**

BL21AI cells containing plasmid for the expression of *PaeCas2-3* were grown in LB broth with 100 µg/mL ampicillin at 37 °C until the absorbance at 600 nm (OD<sub>600</sub>) reached 0.5. The cell culture was chilled immediately on ice, with shaking, until its temperature reached 16 °C. L-arabinose 0.2 % (w/v) and IPTG (1 mM) were added, at this point, to induce protein overexpression. Growth continued for another 12 h at 16 °C. The cells were collected on a lysis buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl NaCl, 5 % (v/v) glycerol and 0.5 mM PMSF, and stored at -80 °C. Before protein purification, the biomass was processed as previously described in Section 2.2.1. Poly-histidine tagged *PaeCas2-3* was purified from a cell lysate using the Ni-NTA and ÄKTA start purification system. The primary extracted proteins were dialysed to

reduce the NaCl concentration in the sample solution and then loaded into a HiTrap Heparin HP affinity column. The *PaeCas2-3* bound to the heparin column was eluted directly and concentrated to 0.5 mL volume using a Vivaspin 6 centrifugal concentrator MWCO 5 kDa (Cytiva). The protein was further purified at 4 °C using a Superdex 200 Increase 10/300 GL column (GE Healthcare) connected to an ÄKTA Purifier 10 FPLC System. The eluted proteins were directly flash frozen and stored at -80 °C. The buffers used during the purification are listed in Table 7.

**Table 7. Buffers used for purifying *PaeCas2-3* and *PaeCas1*.**

Buffer	Reagents	Resin/Column
Buffer A	20 mM Tris-HCl pH 7.5, 10 mM imidazole, 500 mM NaCl, 10 % (v/v) glycerol	Ni-NTA resin
Buffer B	20 mM Tris-HCl pH 7.5, 250 mM imidazole, 500 mM NaCl, 10 % (v/v) glycerol	Ni-NTA resin
Dialyse buffer	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT, 10 % (v/v) glycerol	N/A
Buffer HA	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT, 5 % (v/v) glycerol	Heparin HP affinity column
Buffer HB	20 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM DTT, 5 % (v/v) glycerol	Heparin HP affinity column
Buffer E	20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM DTT, 5 % (v/v) glycerol	superdex S200

### **2.3.6 Expression and purification of PaeCas1**

BL21AI cells containing the plasmid for the expression of *PaeCas1* were grown in LB broth with 50 µg/mL ampicillin at 37 °C until the absorbance at 600 nm (OD<sub>600</sub>) reached 0.6. L-arabinose 0.2 % (w/v); IPTG (0.5 mM) was added at this point to induce protein overexpression. Growth continued for another 3 h at 37 °C. The cells were collected and resuspended using the same method as described in Section 2.2.1. The *PaeCas1* fused to a poly-histidine tag was purified using the same method as applied to product *PaeCas2-3* (Table 7). Instead, the *PaeCas1* eluted from the heparin column was directly dialysed in a buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 30 % (v/v) glycerol and stored.

### **2.4 Design and preparation of DNA substrates.**

The M13ssDNA and M13mp18 RF I DNA (M13dsDNA) used in Chapter 3 were both purchased from NEB. Other substrates mentioned in this thesis were prepared using customised DNA oligos from Merck; the oligo sequences are shown in Table 8.

The labelled DNA substrates, including the DNA forks (Chapter 3, 4 and 5) and the ELB40Primer/41 substrate for doing primer extension (Chapter 5), were constructed by annealing 5 µM of each assigned oligonucleotide in Annealing Buffer comprising 10 mM Tris pH 7.5, 50 mM NaCl and 1 mM EDTA. The DNA was incubated at

## L. He Chapter 2

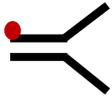
95 °C for 10 min and allowed to cool to 20 °C overnight. The annealed DNA fork was separated from the free oligonucleotides by migrating the sample onto a 10 %(v/v) acrylamide 1x TBE gel. The band containing the forked substrate was excised, with the annealed substrate being eluted in elution buffer containing 20 mM Tris pH 8.0 and 50 mM NaCl, and stored at -20 °C.

The pre-spacer substrates TK24/25 with 5' overhangs and 70 bp ELB40/41 were prepared in the same way as described above but used in spacer integration (SPIN) assays (Chapter 5) directly after quenching.

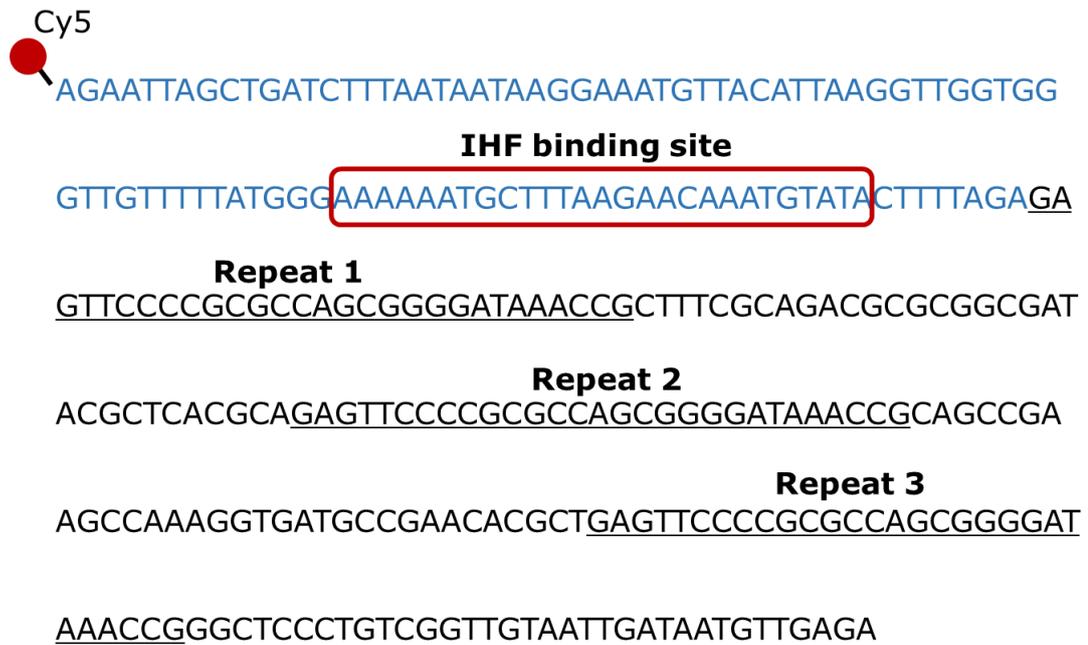
The CRISPR 1s substrates used In Chapter 5 Section 5.12 and 5.14 were prepared differently from the method described above. In this method, a 284 bp DNA (CRISPR 1s) was amplified using a pair of primers from a plasmid pJRW2 containing the DNA sequence of CRISPR 1s. The DNA products were further purified using a QIAquick PCR Purification Kit and stored in distilled water. CRISPR 1s contains an IHF binding site and three repeat sequences (Figure 2.1). The primers for the DNA amplification were fused with either Cy5 or Cy3 label, as shown in Table 8.

The oligos, including 70 nt ELB41-5'Cy5, ELB40Primer-5'Cy5 and MW12-5'Cy5, were mixed and used as DNA markers in the primer extension assays, as described in Chapter 5.

**Table 8. DNA oligos used for biochemical analysis**

Name	Sequence 5' to 3'	Structure
MW12-5'Cy5	<b>CY5-</b> GTCGGATCCTCTAGACAGCTCCATGAT CACTGGCACTGGTAGAATTCGGC	
MW14	CAACGTCATAGACGATTACATTGCTACA TGGAGCTGTCTAGAGGATCCGA	
MW12	GTCGGATCCTCTAGACAGCTCCATGATC ACTGGCACTGGTAGAATTCGGC	
MW14-5'Cy5	<b>Cy5-</b> CAACGTCATAGACGATTACATTGCTACA TGGAGCTGTCTAGAGGATCCGA	
MW12-5'Cy5	(Sequence is the same as shown above)	
CL4	GCCGAATTCTACCAAGTCCAGTGATCAT GGAGCTGTCTAGAGGATCCGA	
ELB40Primer-5'Cy5	GGAGCTCCCTAGGCAGGATC	
ELB41	GCAGGATCCGTATCCGTAAGTGGAGCT CTTCGAAGGCCATCGTCGCGAACGATC CTGCCTAGGGAGCTCC	
ELB40	GGAGCTCCCTAGGCAGGATCGTTCGCG ACGATGGCCTTCGAAGAGCTCCAGTTAC GGATACGGATCCTGC	
ELB41	(Sequence is the same as shown above)	
ELB41-5'Cy5	<b>Cy5-</b> GCAGGATCCGTATCCGTAAGTGGAGCT CTTCGAAGGCCATCGTCGCGAACGATC CTGCCTAGGGAGCTCC	N/A
TK24	GCAGTCCCCTCGCCTCAGCTACGCTCGT	
TK25	CGTAGCTGAGGCGAGGGGACTGCTGG GC	
Crispr1 F-5'Cy5	<b>Cy5-</b> AGAATTAGCTGATCTTTAATAATAAGG	N/A
Crispr1 R-5'Cy3	<b>Cy3-</b> TCTCAACATTATCAATTACAACCG	N/A

Note: Cy5 labelling is denoted as red dot in substrate structure.



**Figure 2.1. The construction of CRISPR 1s.**

Showing CRISPR 1s, including the leader sequence in blue, and three repeats. A Cy3 labelling is fused to the 5' of the complementary strand so it is not denoted in the figure above.

## 2.5 Cas3 catalytic assays.

### 2.5.1 Cas3 Nuclease assay

*EcoCas3* proteins were incubated with 20 nM DNA forks or M13ssDNA in buffer O containing 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, and 0.1 mg/mL BSA, with the addition of 20 mM DTT. The reactions were incubated at 30 or 37 °C then stopped by adding 50 mM Tris pH 8.0, 100 mM EDTA, 5 mg/mL proteinase K and 1 % (w/v) SDS, followed by incubation at 37 °C for 15 min. To analyse the impact of *EcoCas1*, *EcoCas2* and *htpG* over the Cas3 nuclease activity at 30 or 37 °C, these proteins were added to the

## L. He Chapter 2

Cas3 nuclease assays accordingly, when preparing the sample mixes. The assays using Cas1, Cas2, HtpG and Cas3 were mentioned in Chapter 4. The same analysis of *EcoCas3* was repeated in buffer E containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mg/mL BSA and 150  $\mu$ M NiCl<sub>2</sub>. This exemption was mentioned in Chapter 5. The nuclease assay using *PaeCas2-3* was carried out in the same way only without DTT, as described in Chapter 4.

After the reactions were stopped, all the samples were mixed with either Orange G loading dye 1 containing 80 % (v/v) glycerol and Orange G for electrophoresis in 10 % acrylamide 1x TBE gels (Section 2.10.1), or Orange G loading dye 2 containing 20 % (v/v) glycerol, 78 % (v/v) formamide and Orange G for electrophoresis through 10 % acrylamide denaturing gels (Section 2.10.2).

### **2.5.2 In vitro interference assay**

For the Cas3 nuclease reactions in the presence of Cascade, the Cascade complex was prepared as described in Killelea *et al* (79). The crRNA targets the AGGCCCGCACCGATCGCCCTTCCCAACAGTTG sequence on M13dsDNA. To analyse the Cas3 nicking activity, 100 nM Cascade was first pre-incubated with 50 ng of supercoiled M13dsDNA in buffer O (Section 2.5.1) and added to 20 mM DTT at 30 or 37 °C for 30 min to form an R-loop complex. The Cas3 proteins were added, and the mixture was incubated for a further 120 min

## L. He Chapter 2

before directly mixing it with Orange G loading dye 1 (Section 2.5.1) and loading for electrophoresis through a 1 % (w/v) agarose 1× TAE gel. The DNA migration was carried out using 15 volts for 900 min.

The DNA degradation by interference complex Cascade-Cas3 was stimulated by adding a final concentration of 2 mM ATP during the Cas3 nicking analysis (described above) after the Cas3 was added to the sample. All the reactions were stopped using 50 mM Tris pH 8.0, 100 mM EDTA, 5 mg/mL proteinase K and 1 % (w/v) SDS before Orange G dye was added for electrophoresis in a 1 % (w/v) agarose 1× TAE gel. The DNA migration was carried out using 90 volts for two hours.

### **2.5.3 Cas3 ATPase assay**

The Cas3 ATPase assays followed the method described in Bird *et al* (166), which uses malachite green dye as a reporter for the liberation of the phosphate from the ATP. The reactions took place in buffer O (Section 2.5.1) containing 20 mM DTT supplemented with 20 nM of MW14 oligo and 2.5 mM ATP. The reactions were incubated at either 30 °C or 37 °C for 60 min before adding eight reaction volumes of pre-mixed colour reagent and incubating them at room temperature for 2 min, and then stopping the reaction by adding one volume of 3 % (w/v) sodium citrate. The colour reagent was prepared using six volumes of 0.045 % (w/v) malachite green hydrochloride mixed with one volume of 4.2 % (w/v) ammonium

molybdate in 4 M HCl. The samples were transferred to a 96-well flat-bottom plate (Life Technologies) to measure the phosphate production by dye absorbance at 660 nm. The total amount of phosphate product (nmol) was calculated from a phosphate standard curve plotted using NaH<sub>2</sub>PO<sub>4</sub> at 2–16 nmol in 100 µL reactions that was developed as described above.

## **2.6 Electrophoretic mobility shift assay (EMSA).**

Cas3 proteins were incubated with 20 nM of fork DNA at room temperature for 1 hour. The reactions were carried out in buffer E, which contained 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mg/mL BSA and 20 mM DTT. The samples were mixed with 8 µL Orange G loading dye 1 (80 % v/v glycerol and Orange G) for electrophoresis in 8 % (v/v) acrylamide 1 x TB gel containing 0.13 M Tris-HCl pH 7.6, 45 mM boric acid, 8 % (v/v) acrylamide and 5 mM DTT. A running buffer containing 0.13 M Tris-HCl pH 7.6, 45 mM boric acid, 5 mM DTT, and 0.1 mg/mL BSA was used in the EMSA.

## **2.7 Polymerase DNA primer extension (PE) assay.**

The primer extension (PE) assay, using either DnaE or Pol I and 10 nM DNA substrate ELB40P/41, was carried out in a buffer containing 40 mM Hepes pH 8.0, 10 mM Magnesium Acetate, 0.1

## L. He Chapter 2

mg/mL BSA, 10 mM DTT and 200  $\mu$ M dNTPs. The reactions were incubated at 37 °C for 30 min or 1 hour, accordingly. To stop the reaction, 50 mM Tris pH 8.0, 100 mM EDTA, 5 mg/mL proteinase K and 1 % (w/v) SDS were added. All the samples were mixed with Orange G loading dye 2 containing 20 % (v/v) glycerol, 78 % (v/v) formamide and Orange G for electrophoresis through 10 % acrylamide 1 $\times$  TBE denaturing gels (Section 2.10.2). Cas3 proteins were added to the primer extension assays accordingly, as described in Chapter 5.

### **2.8 Spacer integration (SPIN) assays.**

A standard SPIN assay (10  $\mu$ L reaction), comprised 250 nM Cas1-2 in 40 mM Hepes pH 8.0, 10 mM Magnesium Acetate, 0.1 mg/mL BSA, 10 mM DTT, 1  $\mu$ M IHF, 25 ng CRISPR 1s (acceptor) and a pre-spacer (donor). The reactions were incubated at 37 °C for 1 hour before 50 mM Tris pH 8.0, 100 mM EDTA, 5 mg/mL proteinase K and 1 % (w/v) SDS were added. The samples were mixed with Orange G loading dye 1 (Section 2.5.1) and left for electrophoresis in 5 % (v/v) acrylamide 1 $\times$  TBE gels at 150 volts for 120 min.

### **2.9 Coupled primer extension (PE)-SPIN assay.**

The PE-SPIN assay is a combination of a primer extension followed by a SPIN assay. The Primer extension was carried out as

## L. He Chapter 2

described in Section 2.7, using assigned DNA polymerase, with or without Cas3. At end of the reaction, pre-incubated CRISPR 1s and IHF were added directly to the samples. Cas1-2 was then added to start the SPIN assay. The reactions were stopped and then analysed using the same method as described in Section 2.7. How this combination assay was designed, and the underlying principles are described in Chapter 5.

## **2.10. Polyacrylamide gel electrophoresis (PAGE) for DNA and Protein.**

### **2.10.1 Non-denaturing PAGE**

5 % (v/v) and 10 % (v/v) acrylamide 1× TBE gel was used to analyse the tests based on Cy5 labelled oligos, including nuclease assays, primer extension assays, and SPIN and PE-SPIN assays. For a 5 % (v/v) acrylamide 1× TBE gel of 40 mL, 6.7 mL 30 % Acrylamide/Bis Solution, 37.5:1 (Bio-Rad), 4 mL 10× TBE, 29 mL dH<sub>2</sub>O, 200 µL 10 % (w/v) Ammonium persulphate (APS) and 50 µL Tetramethyl ethylenediamine (TEMED) are used. For a 10 % (v/v) acrylamide 1× TBE gel of 40 mL, 13.3 mL 30 % Acrylamide/Bis Solution, 37.5:1 (Bio-Rad), 4 mL 10× TBE, 22.7 mL dH<sub>2</sub>O, 200 µL 10 % (w/v) APS and 50 µL TEMED are used. The DNA samples, mixed with Orange G loading dye 1, were loaded onto acrylamide 1× TBE gel for electrophoresis in a Protean II tank (Bio-Rad) in a 1×

TBE buffer. The voltage and the migration time were set accordingly. The gels were imaged using the Typhoon™ laser-scanner platform (GE Healthcare) and quantified using a GelAnalyzer (if necessary).

### **2.10.2 Denaturing gel electrophoresis**

10 % acrylamide 1× TBE denaturing gel was used to analyse the nuclease assay and primer extension. In the case of a 40 mL gel mix, it was prepared using 10 mL 40 % (w/v) acrylamide/bis-acrylamide (19:1) (National Diagnostics), 4 mL 10× TBE, 7 M Urea (16.8 g), 9.5 mL dH<sub>2</sub>O, 2 mL formamide, 100 µL 10 % (w/v) APS and 50 µL TEMED. The reagents, apart from the formamide, APS and TEMED, were pre-incubated at 37 °C for about 5 min with shaking until the urea was fully dissolved; the rest of the components were then added.

Before electrophoresis the gel wells were rinsed thoroughly using 200 mL dH<sub>2</sub>O and the gel was warmed up by pre-running it at 5 watts per gel for 1 h in a Protean II tank (Bio-Rad) in 1× TBE buffer until the gel reached a temperature of 40-60 °C. Before loading the samples, the gel wells were rinsed twice using 1× TBE buffer. The samples mixed with Orange G loading dye 2 were loaded onto acrylamide 1× TBE denaturing gel for electrophoresis at 5 watts for 2.5 to 3 h until the loading dye reached the bottom of the gel. Alternatively, after the loading dye migrates into the gel completely, the electrophoresis setting was changed to 50 volts for 780 min.

Gels were imaged using the Typhoon™ laser-scanner platform (GE Healthcare).

### **2.10.3 Blue native-PAGE**

Blue native-PAGE was carried out using 3–12 % Bis-Tris gels, markers and loading buffers (Invitrogen). Cas3 was pre-incubated with cofactors including ATP, MgCl<sub>2</sub> and DNA fork separately on ice for 30 min. The samples were then prepared using a NativePAGE Sample Prep Kit (Invitrogen) based on the method described in the manufacturer's instructions. Gel electrophoresis takes 90 min at 175 volts. The NativeMark Unstained Protein Standard (Invitrogen) was used to denote protein size.

### **2.10.4 Sodium dodecyl sulphate (SDS) PAGE**

Proteins were analysed by denaturing SDS-PAGE. Proteins were mixed with 10 mM dithiothreitol (DTT) and 1X SDS loading dye (50 mM Tris-HCl, pH 6.8, 2 % SDS, 10 % glycerol, 1 % beta-mercaptoethanol, 12.5 mM EDTA and 0.02 % bromophenol blue) were heated at 95 °C for 10 min before being loaded onto gels and run at 150 volts for 60 min (10 % SDS-PAGE gel) or 90 min (12.5 % SDS-PAGE gel).

Proteins were visualised by Coomassie staining (0.1 % Coomassie, 10 % acetic acid, 40 % methanol) destained in 20 % methanol, 10 % acetic acid to wash out the extra colour. For the protein samples prepared for mass spectrometry (MS) analysis (Section 2.13), the

gels containing protein bands were processed using a SYPRO Ruby protein gel stain (Sigma-Aldrich), according to the method described in the manufacturer's instructions.

## **2.11 Western blotting to detect proteins.**

The expressions of BioID-Cas3 and His-mVenusN-Cas3 in BL21AI were tested using western blot. The results are described in Appendix 3 and 4.

An Amersham Hybond P0.45 PVDF blotting membrane (GE healthcare) was soaked in 100% methanol (Sigma-Aldrich), then equilibrated in an anode buffer containing 60 mM Tris (Sigma-Aldrich), 40 mM N-cyclohexyl-3-aminopropanesulfonic acid (CAPS, Sigma-Aldrich) and 15 % (v/v) methanol for at least 30 min. One piece of thick blotting paper (larger than the membrane) was soaked in an anode buffer. Unstained SDS-PAGE gel containing protein samples were equilibrated in a cathode buffer containing 60 mM Tris, 40 mM CAPS and 0.1 % (w/v) SDS for 10 to 30 min. A second piece of thick blotting paper (larger than the membrane) was soaked in a cathode buffer. The protein transfer was carried out using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). First, the dry transfer western blotting system was assembled with the membrane, gel, and the thick blotting papers in the following order: the platinum anode (bottom) platform, the blotting paper soaked in the anode

## L. He Chapter 2

buffer, the equilibrated PVDF membrane, the equilibrated SDS gel, the blotting paper soaked in the cathode buffer. The transfer was carried out at 15 V for 45 min. After the protein transfer, the PVDF membrane was blocked using a blocking buffer containing 4 % (w/v) milk powder in 1× TBST solution at 20 °C for 1 hour or 4 °C for 3 hours. The membrane was then blocked using the fresh blocking buffer containing the primary antibody: For detecting BioID-Cas3, the mouse monoclonal IgG Anti-BioID2 primary antibody (AbCam) was used in a 1:2000 dilution; for detecting His-mVenusN-Cas3, 6x-His Tag Monoclonal Antibody (Biotin) (Ivitrogen) was used in 1 in 1000 dilution. This step was carried out at 4 °C for 12 h with gentle agitation. After incubated with the primary antibody, the membrane was washed three times using ice-cold 1× TBST. The second stage blot was carried out at room temperature for 1 h using a fresh blocking buffer containing the Goat anti-Mouse IgG (H&L) HRP conjugated antibody (Agrisera) at a 1:2500 dilution. The washing step was repeated and the membrane dried before being incubated for 2 min with luminol enhancer solution, which was prepared using an ECL Western Blot Substrate Kit (Promega). After incubation, the PVDF membrane was imaged using a Las-3000 Mini Digital Imaging System (FUJIFILM).

## 2.12 Protein pull-down assays.

The pull-down method is described in Chapter 5. The amylose resin used in the pull-down assay was purchased from NEB. The assay was carried out using the buffers listed below.

**Table 9. Buffers used when purifying MBP-Cas3 proteins**

Buffer	Reagents	Resin/Column
Buffer MA	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT	Amylose Resin
Buffer MB	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM Maltose, 1 mM DTT	Amylose Resin

## 2.13 Liquid chromatography-mass spectrometry (LC-MS).

Purified MBP-Cas3 (Section 2.3.2) samples were sent to the Cambridge Centre for Proteomics (University of Cambridge) for LC-MS analysis. Two samples were prepared for carrying out the LC-MS, including protein in SDS-PAGE gel and protein in solution. A gel band was cut from the SDS-PAGE gel processed by the SYPRO Ruby stain (Section 2.10.4) and sent off for testing immediately. A liquid sample was prepared by dialysing the purified protein sample in a buffer containing 50 mM ammonium bicarbonate at 4 °C for 12 h.

## **Chapter 3**

### **Molecular mechanism of the interface of Cas3 HD and RecA1-like domains (iHDA1) in temperature-dependent nuclease activity.**

#### **3.1 Summary.**

A pre-requisite for CRISPR interference is recognition of mobile genetic elements (MGEs). This is achieved via forming an DNA-RNA hybrid (R-loop) on MGE DNA (or RNA) by an effector protein-RNA complex, which in *E. coli* is the CRISPR-associated complex for antiviral defence (Cascade) bound to a CRISPR RNA (crRNA) (64). The R-loop formation consequently leads to the change of Cascade conformation, recruiting the CRISPR interference effector Cas3 to the specific targeting site on MGE (65). By interacting with Cascade, the Cas3 nuclease-helicase generates wide-scale degradation of MGE DNA thus neutralising the MGE infection. Since Cas3 and Cascade are coordinated during CRISPR interference, it is necessary to investigate the mechanisms of Cascade and Cas3 together and in isolation to establish the factors that influence CRISPR interference, and successful targeting of the MGE.

Using  $\lambda$  phage as the MGE, studies using plaque-reporter assays have revealed an unexpected result: cells expressing Cascade-

## L. He Chapter 3

crRNA complex and Cas3, are sensitive to  $\lambda$  phage lysis at 37 °C, but not at 30 °C (108). This defective CRISPR interference could be restored by ectopic overexpression of either Cas3 or a protein chaperon HtpG, the latter attributed to stabilise Cas3 at a functioning level (150). The differing efficacies of Cascade-Cas3 at various temperatures, and recovery of defective CRISPR interference at disadvantageous temperature by Cas3 stabilisation suggests a temperature-dependent switch-like function for Cas3 at the intracellular level (108). Study on Cas3 secondary structure using circular dichroism (CD) analysis carried out by collaborators from the University of Zagreb provided biophysical evidence to support the above hypothesis (167). In that work, when Cas3 was heated from 20 °C to 55 °C, it was seen to undergo conformational rearrangements and revealed two different secondary structure compositions on either side of the thermal inflexion point at 34 °C. We hypothesised that this conformational change may determine Cas3 nuclease activity, and can eventually affect the efficacy of CRISPR interference. Biochemical analysis was conducted *in vitro* and presented in this chapter using the *E. coli* Cas3 protein (*EcoCas3*). A Cas3 functional regulation model was also presented.

### **3.2 The importance of the interface between the Cas3 HD and RecA1-like domains (iHDA1), and its conserved tryptophan residues.**

Many Cas3 proteins comprise a fusion of an HD nuclease domain with superfamily 2 (SF2) helicase (RecA-like) domains and an additional predicted C-terminal domain (CTD) of unknown function (Figure 3.1 A, B). Two RecA-like domains together with CTD form a conserved DNA tunnel that is capable of continuously transferring single-stranded DNA (ssDNA) towards the HD nuclease domain. However, fitting ssDNA in a DNA catalytic site in the HD domain of Cas3 is not easily achieved. Structural analysis of I-E subtype *T. fusca* Cas3 (109) indicates that the DNA phosphate backbone undergoes extraordinary twisting before reaching the nuclease active site in the HD domain. This DNA kink occurs in the interface between Cas3 HD and RecA1-like domains, that in this thesis refers to as the **iHDA1 region** (Figure 3.1 B). This region includes an ssDNA binding channel in the vicinity of the helicase motifs Ib and Ic, which interacts with the ssDNA in the superfamily two helicases. The Ib and Ic motifs, along with a tryptophan residue, are highly conserved components in the much less conserved iHDA1 region (Figure 3.1 B, C and D), suggesting they are potentially very important for maintaining Cas3 function. Therefore, the iHDA1 region was selected for this study. Besides, it has been predicted to

## L. He Chapter 3

have conformational flexibility and is important for the reorientation of ssDNA before it gains access to the nuclease active site (109). Furthermore, it's positioning is favourable for possibly coordinating cross-talk between Cas3 helicase (RecA1) and nuclease active sites (HD). For these reasons, this study focuses on investigating the biochemical features of this iHDA1 region in Cas3.

Amino acid substitutions were made to Cas3 to reveal the contribution of the iHDA1 region to protein function regulation and identify the underlying factors that may be responsible for the hypothetical temperature-dependent Cas3 nuclease activity. The structure of *E. coli* Cas3 at atomic resolution is not available, so the I-E subtype Cas3-DNA co-structures from *Thermobifida fusca* (PDB 4QQW-Z (109)) was used as a close homologue of the *E. coli* Cas3—30 % amino acid identity to identify areas for mutagenesis. Two tryptophan residues were targeted for their high conservation across bacterial Cas3 proteins (Figure 3.1 D), one located in the iHDA1 (Trp-406 in *E. coli* Cas3 numbering, in RecA1 domain) and the other (Trp-230 in HD domain) located close by (Figure 3.1 B, C), both of which are predicted to interact with ssDNA (71,109). Additionally, Trp-149 and Trp-152 were selected for comparison with Trp-230/406, being located adjacent to the iHDA1 region but much less well conserved across species (Figure 3.1 D).

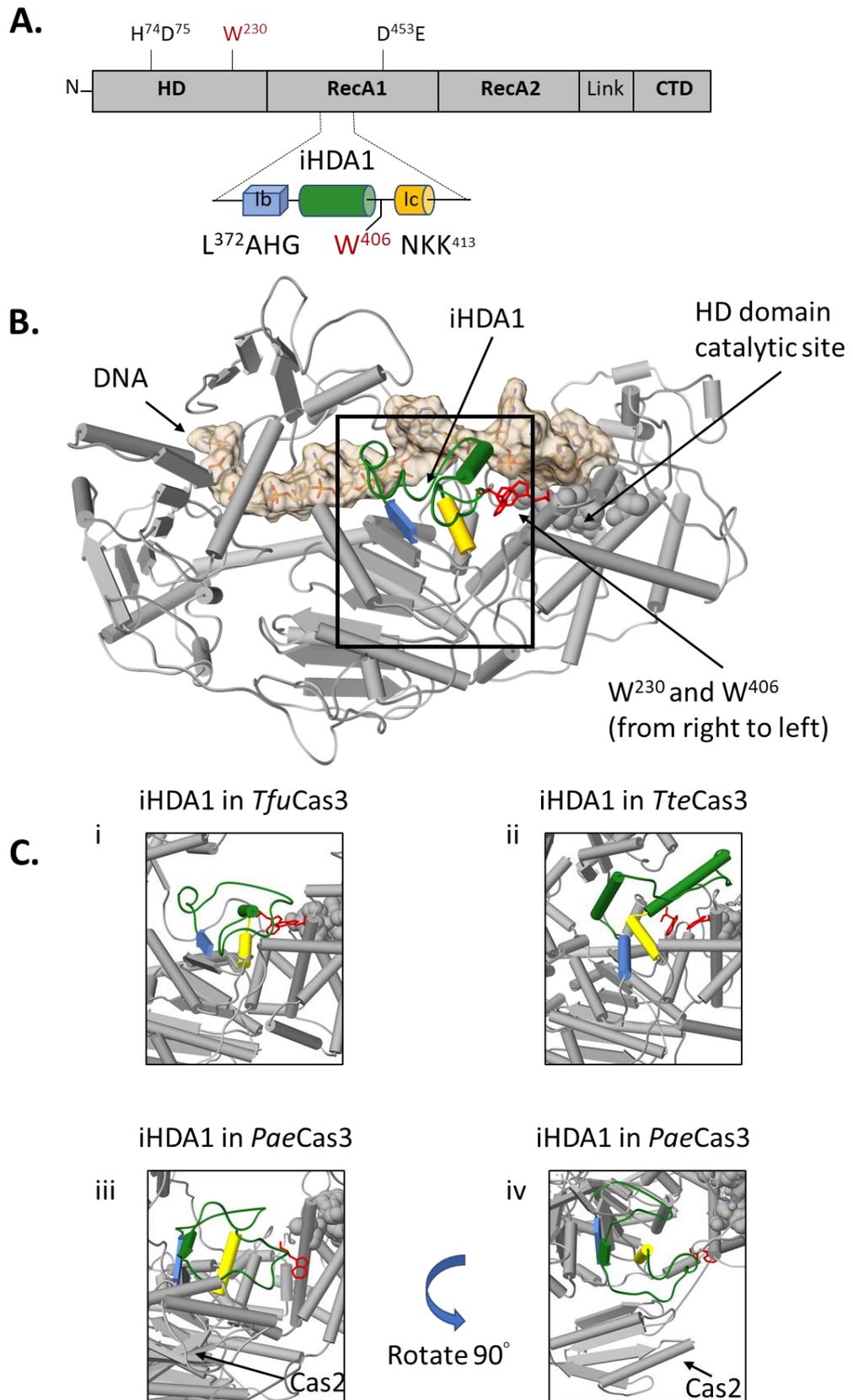


Figure 3.1 Cont.

## D.

## W149 W152

Tfu	A L P S L G Y P N G G L V T G L V A Q M L G G H H G T F H P H P - S F Q S R N P L A E F G F S S P H W E K Q - R H A L L H A V F D A T - G R
Tter	V L A - D V F G L S G R S A R W W A F A V G G H H G F V P S Y D - E V - - R R D L D Q Q A V G W G M W D A A - R E V L L C R L A D A L - G L
Eco	F D A A - P H P Y - E S W F P W V - E A V T G H H G F I L H S O D Q D K S R W E M P A S L A S Y A A Q D K Q A R E E W I S - V L E A L F L T
Sequence Ruler	140 150 160 <b>W230</b> 170 180 190 200
Tfu	P T - - P P D M L D G P T A S V V C G L V I L A D W L V S Q E D F L L E R L T S L P A D G S A S A L R A H F E T S L R R I P - - - S L L D A
Tter	P G S S R P T V E S T P D A F M L A G L V S V A D W I G S N E E Y F P Y A A Q S A L Q V P Q - L D A E A Y L E R A M R Q A E R A M A S L G W
Eco	P A G L S I N D I P P D C S S L L A G F C S L A D W L G S W T T T N T F L F N E D - A P S D I N A L R T Y F Q D R Q Q D A S R V L E L S G L
Sequence Ruler	210 220 230 <b>Ib</b> 240 250 260 270
Tfu	R P G R F L A L P T M A T A D Q M H T R L K E Y A R Y R V E N T D L P R S - - - - - S T L A L L H S M A W L N P D Y A P A D L P G V S K V
Tter	M S G C Y F A L P T M A T S N Q M F G R V T D Y L R H R Y P E D V V V V N L V H G H S D L S A L L Q E L R Q K - - G E E I F L Q G V Y D E
Eco	A D S V I F A L P T Q A T A N A M L T R M E A S A S H L F S S P N L I - - L A H G N S R F N H L F Q S I K S R A - - - - -
Sequence Ruler	340 350 <b>W406 Ic</b> 360 370 380 390
Tfu	L S N L G H R - D P F A A T D W L M G - R K R G L L A P W A V G T I D Q A L M A V L R A K H N A L R L F G L A G K V V V D E A H A V D P Y
Tter	A L G D E Q L - G A V V A G C W F T R - G K R A L L P P Y G E D V T D Q A L L A V L Q V K H V F V R L F A L S T K T V I V D E V H A Y D V Y
Eco	I T E Q G Q E E A W V Q C C W L S Q S N K K Y F L G Q I G V C T I D Q V L I S V L P V K H R F I R G L G I G R S V L I V D E V H A Y D T Y
Sequence Ruler	400 410 420 430 440 450

**Figure 3.1 Structural positioning of two invariant tryptophan residues at the interface of Cas3 HD and RecA1 domains—a key role in Cas3 nuclease function for Trp-406.**

(A). The illustration represents Cas3 protein with amino acid residues indicating the nuclease (HD) and Walker B ATPase-active sites using numbering from the *E. coli* protein. Highlighted in the foreground is the region detailed in the results, located at the interface of the HD and RecA1 domains ('iHDA1'), including the ssDNA binding helicase motifs Ib (blue), Ic (yellow), and an inter-motif sequence in between denoted in green. (B). The *E. coli* Cas3 structure deduced from *T. fusca* Cas3 (PDB: 4QQW-X) and *T. terrenum* (PDB: 4Q2C) highlights the tryptophan residues Trp-230 and Trp-406 (labelled) and the passage of the ssDNA (tan cord). Cas3 regions are denoted as follows in the same way as in Part A: all polypeptides are in grey apart from those which consist of iHDA1. The grey spheres indicate the HD nuclease-active site residues. (C). Showing the iHDA1 region in I-E and I-F CRISPR subtypes. Panel i and ii represent iHDA1 regions in Cas3 from organisms with endogenous I-E subtype CRISPR, which are *T. fusca* Cas3 (denoted as TfuCas3) and *T. terrenum* Cas3 (TteCas3). Panel iii and iv display the iHDA1 region in Cas2-3 from the *P. aeruginosa* PA14 strain (PaeCas2-3) represented in two orientations. (D). Showing protein sequences alignment highlights the conservation of residues Trp-406 and Trp-230 in *Escherichia coli* K-12 (Eco) with the structurally determined Cas3 proteins from *T. fusca* (Tfu) and *T. terrenum* (Tter) proteins, all from the I-E subtype CRISPR. The protein sequences were aligned using Clustal Omega and the results were exported via Lasergene 17.

### 3.3 Protein purification of wild type *EcoCas3* and mutants.

Cell lysate containing either *EcoCas3* or its mutants were prepared as described in the materials and methods section (Chapter 2, Section 2.3.1). The proteins tagged with N-terminus His<sub>6</sub>-MBP were initially isolated from lysates using an Ni-NTA column. The eluted products were loaded onto the heparin column for the removal of contaminants bound to the polynucleotides, while the Cas3 proteins were obtained in the flow-through, and then further concentrated using the MBPTrap column (Figure 3.2 A). The purified proteins were dialysed overnight at 4 °C in a dialysis buffer containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl and 30 % glycerol. The proteins were then stored at -80 °C (Figure 3.2 B).

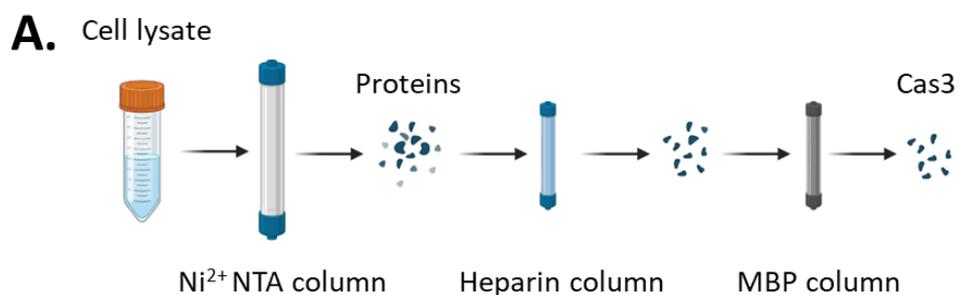
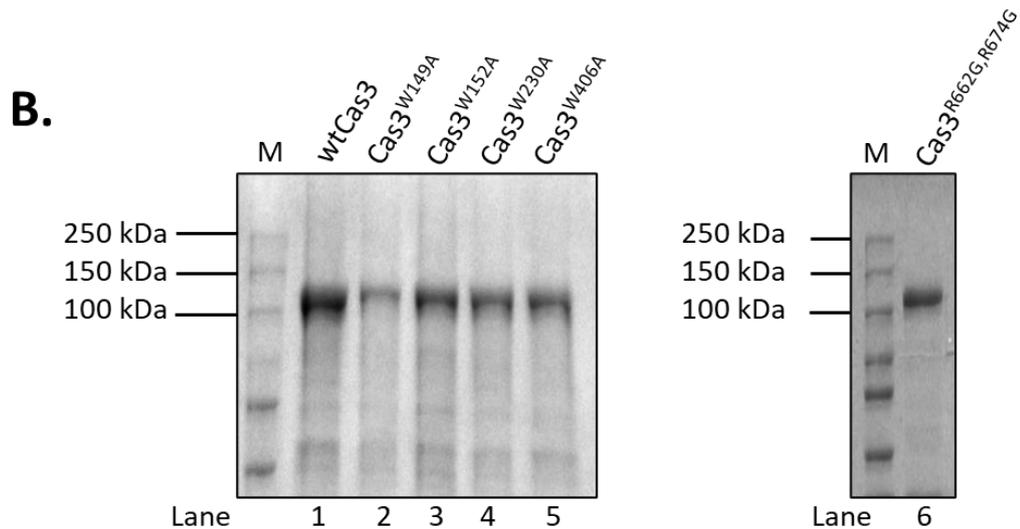


Figure 3.2 Cont.



**Figure 3.2 Purification of EcoCas3 proteins used in this study.**

(A). A schematic diagram of EcoCas3 purification. A cell lysate was loaded onto an Ni-NTA column which has a high affinity to the his<sub>x6</sub>-tag. The proteins bound to the first column were directly eluted using a buffer containing 250 mM imidazole and loaded onto a heparin column for removal of the DNA binding proteins. The EcoCas3 cannot bind to the heparin column. It directly flows through the column and is then concentrated using an MBP column before final dialysis. (B). Purified EcoCas3 and its mutants (145 kDa) are shown on 10 % SDS page gel, post-stained with Coomassie blue. The gels, from left to right, show wild-type EcoCas3 (lane 1), EcoCas3<sup>W149A</sup> (lane 2), EcoCas3<sup>W152A</sup> (lane 3), EcoCas3<sup>W230A</sup> (lane 4), EcoCas3<sup>W406A</sup> (lane 5) and EcoCas3<sup>R662G,R674G</sup> (lane 6), which are the proteins discussed in this chapter.

### 3.4 Selection of DNA substrate used for diagnosing Cas3 nuclease activity.

To analyse if temperature change affects Cas3 function, the nuclease activity of wild-type EcoCas3 was examined on M13ssDNA at different temperatures (Figure 3.3 A). It is apparent that the increased accumulation of cleaved DNA products is concurrent with

the increase in the *EcoCas3* concentration present in the reactions (Figure 3.3 A, lanes 2 to 6). This result suggests purified wild-type *EcoCas3* constantly degrades M13ssDNA at 30 °C and produces cleaved DNA products that are consistent with those in the control reaction (Figure 3.3 A, lane 1), which treated the substrates with 0.01 U DNase I. The *EcoCas3* nuclease activity, however, is not easy to distinguish at 37 °C, and its DNA catalytic function may be repressed with no apparent M13ssDNA loss or any degraded DNA (Figure 3.3 A, lanes 8 to 11). To gain a higher resolution of the products formed by Cas3 nuclease activity, for example, by using denaturing acrylamide gels, a DNA fork substrate was introduced into this study to fulfil this requirement. A Cy5 labelling on the DNA fork significantly increases the accuracy of the tracking nuclease activity on oligonucleotides, including detecting the removal of merely a few nucleotides at the 3' end of the labelled strand. The DNA fork used in this part of the study contains a partially complementary DNA strand of 24 bp with two extended single-stranded flanks of 25 nt each, which includes a 3' ssDNA end for Cas3 to process (Figure 3.3 B). This substrate carries a Cy5 label at the selected DNA end, with maximum absorbance at 646 nm, and it could be applied to trace cleaved DNA products generated by the protein nuclease activity.

Two DNA forks of the same construction but with different Cy5 labelling sites were incubated with the *EcoCas3* separately. One is

## L. He Chapter 3

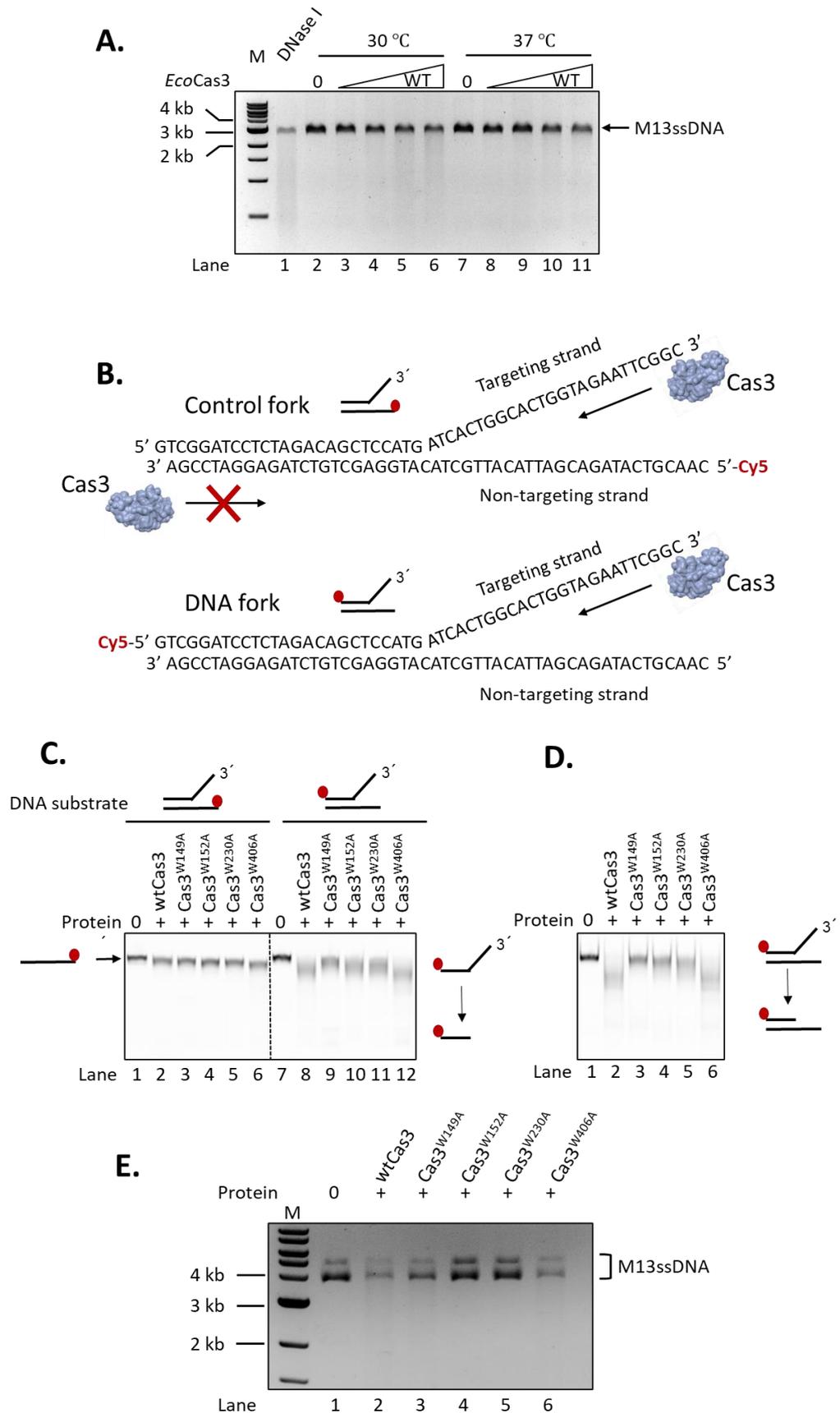
referred to as the *Control fork*, the other one termed the *DNA fork* (Figure 3.3 B). Both fork substrates were made of two partial complementary DNA oligos but with distinct Cy5 labelling sites. The control fork has a labelled 5' end on the non-targeting strand, while the DNA fork containing a label at the 5' end on the targeting strand. Those two types of substrates were introduced for different study purposes. The control fork was used to test for Cas3 nuclease function by providing the necessary 3' to 5' polarity (23) of ssDNA. In theory, the Cas3 is not able to remove any of the nucleotides on the non-targeting strand of the Control fork owing to a lack of a single-stranded 3' end. Whereby the DNA fork substrate was applied to test Cas3 nuclease activity. It was used in the majority of the assays described in the following content, unless mentioned otherwise.

As predicted, the *EcoCas3* did not degrade the labelled strand which was without an exposed 3' ssDNA end on the Control fork (Figure 3.3 C, lanes 2-6); this result agrees with the known features of Cas3 (23,33). In contrast, the DNA fork experienced continuous removal of nucleotides on the labelled DNA strand from its 3' ssDNA end (Figure 3.3 C, lanes 8-12). The wild-type *EcoCas3* and *EcoCas3*<sup>W406A</sup> showed apparent DNA cleavage activity after two hours incubation at 30 °C. The mutants, including *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> gave little or no degraded DNA products. This agrees with the conclusions in the Cas3 nuclease

## L. He Chapter 3

assay using M13ssDNA as the substrate (Figure 3.3 E). *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> are therefore deficient in nuclease activity, consistent with results acquired from those incubated on the DNA fork.

The degraded DNA products by Cas3 nuclease activity on DNA fork could also be analysed while retaining their original 'fork-shape' structure. The Cas3 nuclease assay using wild type protein and mutants was repeated and analysed on a native TBE polyacrylamide gel, instead of on a denaturing TBE polyacrylamide gel. This method revealed results that are consistent with those displayed when utilising denaturing PAGE (lanes 7-12 in Figure 3.3 C, compare to Figure 3.3 D). Therefore, the DNA fork is a beneficial substrate for analysing the DNA catalytic function of Cas3 which could record the DNA products generated by Cas3 nuclease activity, and could be applied to different analysing mediums accordingly.



**Figure 3.3 Cas3 nuclease activity on M13ssDNA and DNA fork substrates.**

## L. He Chapter 3

(A). The *EcoCas3* degrades 200 ng M13ssDNA at 30 °C and 37 °C, shown using 1 % agarose gel. Increasing concentrations of *EcoCas3* were added to the reactions: 0 nM, 14 nM, 28 nM, 56 nM and 112 nM in lanes 2-6 and lanes 7-11. The DNA product generated by the nuclease activity in the samples containing *EcoCas3* is indicated by a control reaction containing endpoint products generated by 0.01 U DNase I degrading M13ssDNA within 12 min at 37 °C. This is shown in lane 1. (B). This shows the oligo sequence and the conformation of the Control fork and DNA fork. Those forks are identical and are made of two oligos, but vary in Cy5 labelling positions, which are denoted in red. The predicted Cas3 cleavage on the fork substrates are represented by the Cas3 cartoon (blue,) and its polarity is denoted by the direction of the arrow. The DNA degradation by Cas3 starts from the 3' ssDNA end and moves towards the 5' end. The schematic diagrams are presented on top of each fork substrate and are applied in the following figures for showing the DNA substrates used in the assays. (C). 56 nM *EcoCas3* and its mutants catalyse the DNA degradation against the 20 nM Control fork and DNA fork separately. This is shown on 10 % TBE polyacrylamide denaturing gel. The *EcoCas3* reveals no cleavage on the non-targeting strand without an exposed 3' end, shown in lanes 2-6, but degrades the targeting strand with a 3' ssDNA end as displayed in lanes 8-12. (D). The analysis of Cas3 (56 nM) nuclease activity against DNA forks (20 nM) using 10 % TBE polyacrylamide native gel reveals results that are consistent with those achieved with denaturing gel (Figure 3.3 C, lanes 7-12). (E). This shows the nuclease activity analysis of *EcoCas3*, *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup>, *EcoCas3*<sup>W230A</sup> and *EcoCas3*<sup>W406A</sup> on M13ssDNA. In each reaction, protein (112 nM) and M13ssDNA (200 ng) were incubated at 30 °C for two hours, before the endpoint products were taken and analysed on 1 % agarose gel. Quantification data in Appendix 9.

### **3.5 *E. coli* Cas3 tryptophan-406 is a determinant for differential nuclease activity at 30 °C and 37 °C.**

The efficacy of CRISPR interference when defending against MGEs is restricted by the environmental temperature (108,167). The  $\Delta hns$  *E. coli* cell strain harbouring CRISPR Cascade and Cas3 is sensitive to phage  $\lambda$  lysis at 37 °C, while it shows robust resistance to the same at 30 °C. However, loss of CRISPR interference was restored

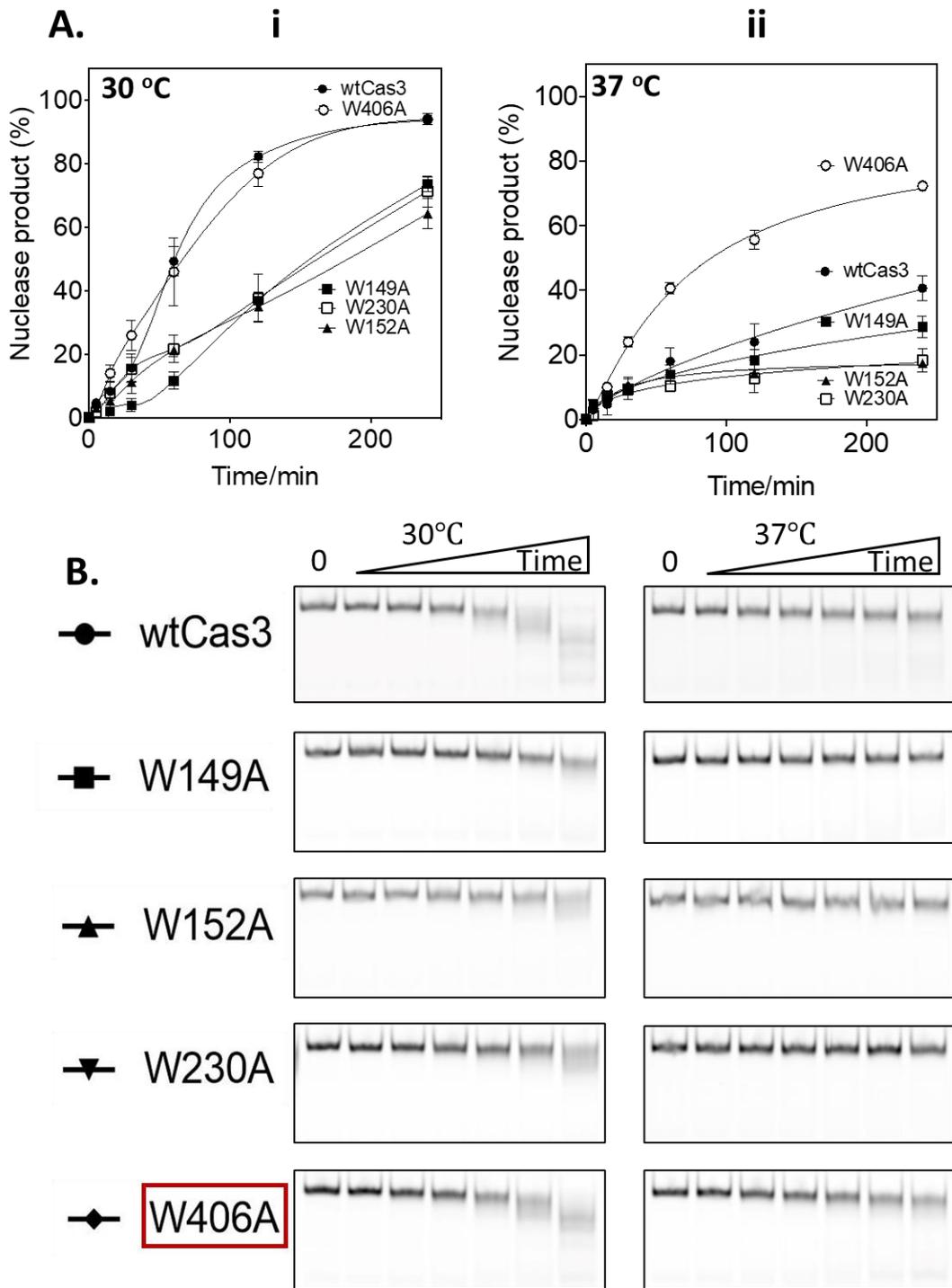
by ectopic overexpression of Cas3, but not other Cas proteins (108,167). This effect directly implicates Cas3 in temperature-dependence of CRISPR immunity in *E. coli*. To investigate the mechanism, *EcoCas3* and tryptophan mutants (each with a disturbed iHDA1 region) were incubated against the DNA fork substrate to analyse the protein nuclease activity at 30 °C and 37 °C.

Nuclease activity against the DNA fork substrate measured as a function of time, showed wild-type *EcoCas3* (56nM) converted maximally 92% of DNA (20nM) into nuclease product at 30 °C, compared to 40% at 37 °C (Figure 3.4 A i and ii). The reduced nuclease activity at 37 °C also occurred in the other three tryptophan mutants *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> (each 56 nM), despite their apparently reduced DNA cleavage function compared with the wild type. However, *EcoCas3*<sup>W406A</sup> (56 nM) revealed similar efficacy to the wild-type *EcoCas3* at 30 °C and was hyperactive compared to the wild-type *EcoCas3* at 37 °C (a mean of 73% fork converted to nuclease product). Therefore, these results suggest *EcoCas3* nuclease activity is repressed at 37 °C but not at 30 °C, which is consistent with the genetic features of CRISPR interference (108). Furthermore, the *EcoCas3* Trp-406 residue may modulate *EcoCas3* nuclease activity because changing it to alanine substantially overcame the inhibitory effect of elevated temperature on nuclease activity. Since the side chain of the Trp-406 is adjacent to Cas3 DNA hydrolysis active sites (Figure 3.1 B), there are two

## L. He Chapter 3

possible explanations for the Cas3<sup>W406A</sup> acquiring temperature-independent nuclease activity. The first one is Trp-406 may cause steric block at 37 °C and prevent DNA accessing to HD active sites, whereas at 30 °C Trp-406 is moved away from DNA channel. This relocation of amino acid may be driven by the conformational change of the polypeptides forming the iHDA1 region. The other possibility is, this tryptophan is necessary for bending DNA phosphate backbone (71,109) thus guide DNA to the catalytic sites in Cas3 HD domain. If so, Cas3<sup>W406A</sup> may have distinct affinity to DNA when comparing with wild type Cas3.

Additionally, reduced nuclease activity from each of *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> indicates that disruption of the iHDA1 region disturbs nuclease function of *EcoCas3*, even though HD-active site residues remain unchanged.



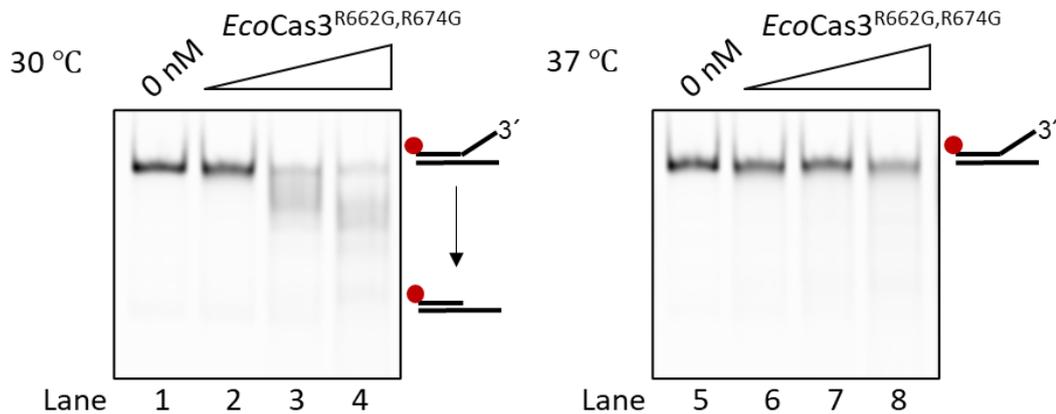
**Figure 3.4 EcoCas3 tryptophan mutants revealed distinct DNA degradation ability, the EcoCas3<sup>W406A</sup> stands out as a hyperactive DNA nuclease.**

(A). This shows the nuclease activity of 56 nM Cas3 and its mutants measured on the DNA fork substrate of 20 nM. The samples were collected at 0, 5, 15, 30, 60, 120 and 240 minutes. The reactions were carried out in triplicate at 30 °C or 37 °C, as indicated, and the data points show standard errors from the mean. (B). This shows how the individual results

## L. He Chapter 3

*of nuclease activity at different temperatures were analysed using Cas3 proteins on the DNA fork substrate, labelled as indicated, and displayed on 10 % native page gels. Each gel on the left shows the results of reactions at 30 °C, and on the right are those at 37 °C. The same analysis was repeated two more times, although the data are not shown.*

The same analysis was carried out on another Cas3 mutant, *EcoCas3*<sup>R662G,R674G</sup>, which has an intact iHDA1 region but has two conserved arginine residues Arg-662 and Arg-674 replaced with Glycine. An increased concentration of *EcoCas3*<sup>R662G,R674G</sup> was incubated against the DNA fork and analysed at 30 °C and 37 °C. We observed no reduced nuclease activity of *EcoCas3*<sup>R662G,R674G</sup> compared to wild-type protein (Figure 3.5 lane 4, compared to Figure 3.3 D lane 2). In addition, *EcoCas3*<sup>R662G,R674G</sup> reveals similar features of wild-type *EcoCas3* in that its nuclease function was repressed at 37 °C but not at 30 °C. This result indicates that mutating conserved amino acids elsewhere rather than in those residing in the iHDA1 region does not affect the DNA degradation function of *EcoCas3*. Therefore, having an intact iHDA1 region is necessary for maintaining *EcoCas3* nuclease function.

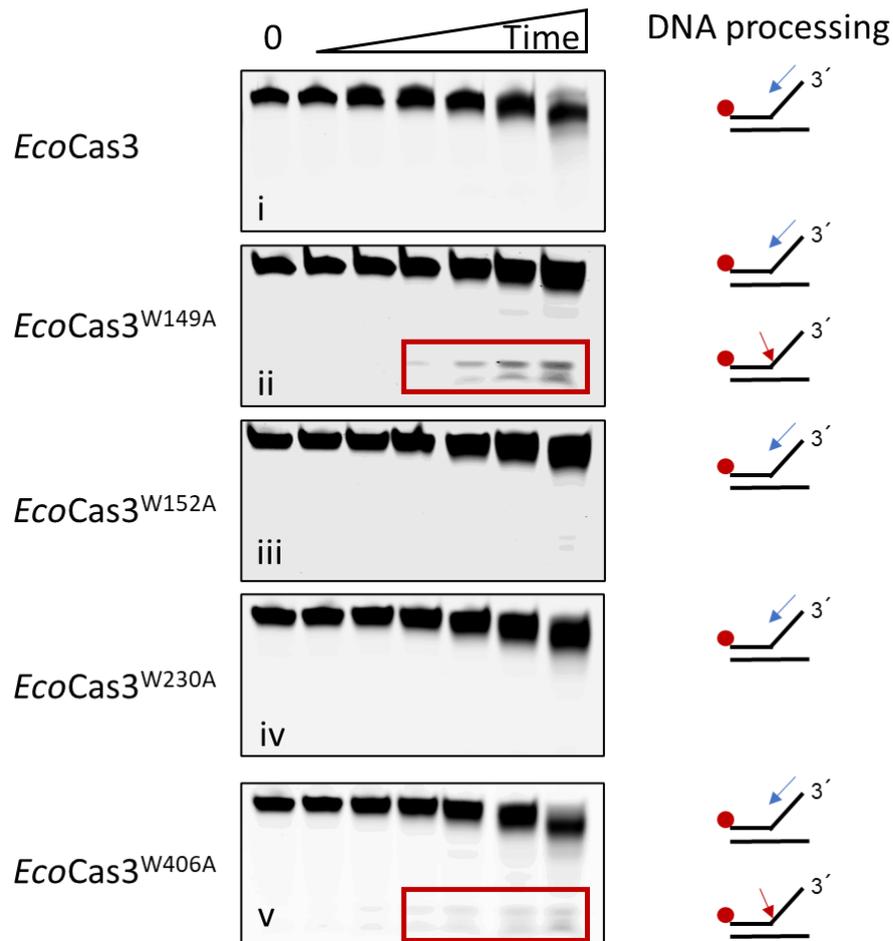


**Figure 3.5. *EcoCas3*<sup>R662G,R674G</sup> degrades DNA fork at 30 °C.**

An increasing concentration of *EcoCas3*<sup>R662G,R674G</sup> was added to reactions containing a 20 nM DNA fork, of 0 nM, 14 nM, 28 nM and 56 nM in lanes 1-4 and lanes 5-8. The reactions were carried out at 30 °C or 37 °C, as indicated. The end point products were analysed on 10 % native TBE polyacrylamide gel. Quantification data in Appendix 9.

### 3.6 Distinct DNA degradation products produced by *EcoCas3* and the mutants.

Two types of DNA products were observed from nuclease activity on DNA fork substrates (Figure 3.6) that were not detected using M13ssDNA (Figure 3.3 E). Most of the degraded DNA was generated by *EcoCas3* through the continuous removal of each nucleotide from the 3' end and towards the 5' end of the targeting strand. This cleavage type was present in wild-type *EcoCas3* and all the tested tryptophan mutants. However, *EcoCas3*<sup>W149A</sup> and *EcoCas3*<sup>W406A</sup> revealed activated nicking activity and were found to be capable of directly truncating DNA at the sites proximal to the complementary base pairs (Figure 3.6, ii and v).



**Figure 3.6 Distinct DNA degradation products are produced by *EcoCas3* and tryptophan mutants.**

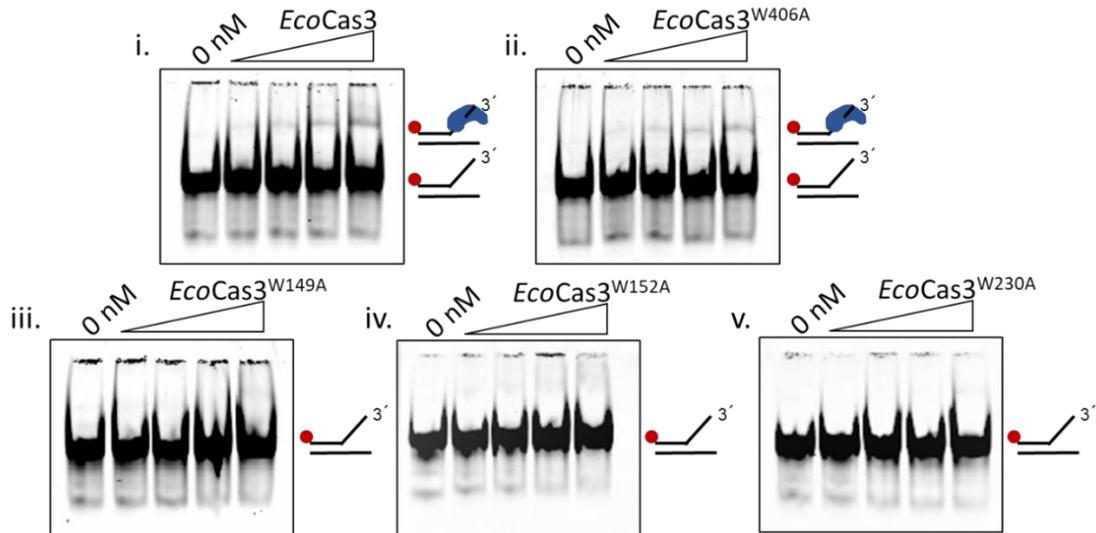
The nuclease activity of *EcoCas3* and its mutants (56 nM) was measured on the DNA fork substrate (20 nM). Samples were collected at 0, 5, 15, 30, 60, 120 and 240 minutes. Reactions were carried out at 30 °C and analysed using 10 % denaturing TBE polyacrylamide gel. The continuous removal of nucleotides from the 3' end is denoted by a blue arrow, while the red arrow represents the nicking activity occurring on the DNA fork. Quantification data in Appendix 9.

### 3.7 DNA binding affinities of Cas3 Trp-mutants.

While none of the four tryptophan residues examined in Section 3.5 directly participated in the DNA hydrolysis by *EcoCas3*, mutating them to alanine more or less impacted *EcoCas3* nuclease activity.

This phenomenon promoted the investigation of the affinity of *EcoCas3* and its mutant to DNA to investigate if introducing a mutant in *EcoCas3* would alter its DNA binding ability.

In the *EcoCas3* EMSA assay, up to 3.3  $\mu\text{M}$  of proteins were pre-incubated with a 20 nM DNA fork at room temperature for one hour, then the free substrates were migrated from the protein-DNA complex onto an 8 % TB acrylamide gel containing 5 mM DTT. In general, the low binding affinity of *EcoCas3* to DNA observed in EMSA assays hinders a precise quantification analysis (Figure 3.7, i). However, we were able to detect loss of Cas3-DNA complex from nuclease defective *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> (Figure 3.7, iii, iv and v). Losing DNA binding capacity may explain the decrease in nuclease activity in those mutants, but *EcoCas3*<sup>W406A</sup> bound to DNA like wild-type *EcoCas3* (Figure 3.7, i and ii). This result led to the hypothesis that the relocation of Trp-406 attributes to the temperature-dependent nuclease activity of Cas3, whereas mutating Trp-406 to Ala results in generation of a nuclease hyperactive Cas3 phenotype. This hypothesis was then confirmed by molecular dynamics (MD) simulation analysis conducted by collaborators in University of Zagreb (167).



**Figure 3.7 EMSAs showing wild-type *EcoCas3* and *EcoCas3*<sup>W406A</sup> forming DNA-protein complexes.**

The EMSAs show that wild-type *EcoCas3* and *EcoCas3*<sup>W406A</sup> form stable DNA-protein complexes (panels i and ii), but other mutant proteins do not (panels iii, iv and v). Increasing concentrations of Cas3 and mutant proteins (0, 0.4 0.8, 1.6 and 3.3  $\mu\text{M}$ ) were incubated with the DNA fork (20 nM). A stable DNA-protein complex is indicated. The *EcoCas3* bound to the DNA fork is marked in blue. Quantification data in Appendix 9.

### **3.8 *EcoCas3* ATPase activity is independent of temperature change.**

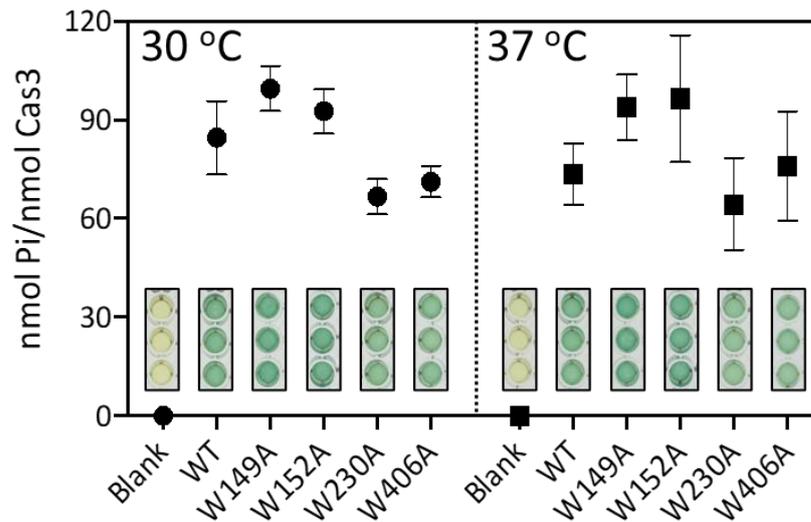
CRISPR interference requires *EcoCas3* nuclease and ATP-dependent DNA translocase activities. It is thought that when Cas3 degrades MGEs, its ATPase activity powers protein conformational movement that promotes the transfer of ssDNA through a hydrophobic DNA tunnel into the HD catalytic site. As discussed previously, Trp-406 in the *EcoCas3* RecA1 domain plays a key role in regulating ssDNA access to the DNA hydrolysis site in the HD domain. Based on this result, a question arises concerning whether

the conformational change of the Cas3 RecA1-RecA2 domains during the ATP binding-releasing cycle causes the relocation of Trp-406. In other words, does mutating tryptophan-406 into alanine have any impact on Cas3 ATPase activity? To answer this, the ATP hydrolysis function of *EcoCas3*<sup>W406A</sup> was tested alongside the other *EcoCas3* proteins, using the malachite green reporter assay (166) at 30 °C and 37 °C. This assay also included a reaction lacking *EcoCas3* as a control.

In contrast to the feature of *EcoCas3* exerting its role as a DNA nuclease, the temperature had no significant effect on the wild-type *EcoCas3* ATPase activity (Figure 3.8). The ATPase activity of each Cas3 mutant was similar, whether measured at 30 °C or 37 °C. This contrasted with the much-reduced nuclease activity of three of the mutants at the higher temperature. *EcoCas3*<sup>W406A</sup> was not hyperactive as ATPase at 37 °C. Therefore, the effect of temperature and the Trp-406 residue are exerted only on *EcoCas3* nuclease activity.

At 30 °C, *EcoCas3*<sup>W149A</sup> and *EcoCas3*<sup>W152A</sup> consumed more ATP compared to the wild-type (Figure 3.8). One possible reason for this effect may be the introduction of the mutant in either Trp-149 or Trp-152, disturbing the position of His-159 in the distal end on the same alpha helix. This Histidine 159 contributes to metal-ion binding in the HD catalytic site. Therefore, relocating His-159 may lead to a

reduction of the magnesium-binding affinity in the HD catalytic site, which consequently increases the concentration of Mg-ATP substrate formed in the sample solution that could be utilised by the *EcoCas3* ATPase motif (Walker B).



**Figure 3.8 The *EcoCas3* ATPase activity is independent of temperature change.**

The end point measurements of ATP hydrolysis by 790 nM *EcoCas3* proteins were compared at 30 °C and 37 °C, as indicated. The reactions were in triplicate; standard errors from the mean are shown. This gave rise to measurements of phosphate (Pi) released per nM of *EcoCas3* protein. These values were obtained from spectroscopic measurements of malachite green dye intensity, which are shown in the graph, where each measurement was obtained after blanking the spectrophotometer with reactions lacking *EcoCas3*.

### **3.9 Cas3 nicking assay and a Cas3-Cascade in vitro CRISPR interference assay.**

In *E. coli* CRISPR interference, Cas3 is recruited to MGE DNA at an R-loop formed by the Cascade-crRNA complex. Once the Cas3

loads onto the targeting site, its nuclease activity is performed in two steps: first, the magnesium-dependent nicking of supercoiled DNA; then, the ATP-dependent transfer of ssDNA is needed to power the processive nucleolytic digestion. The interplay of Cas3 and Cascade in interference reactions led to the following tests for nuclease activity of *EcoCas3* and its mutants when a plasmid had been targeted by Cascade. This involved analysing the *EcoCas3* nicking activity that was generated on the non-target strand, then the addition of extra ATP to stimulate massive DNA degradation by *EcoCas3*.

### **3.9.1 Introducing mutants in the iHDA1 region promotes the nicking activity of EcoCas3.**

The Cascade apo-protein complex used in the assay described in this section was purified and activated by binding crRNA (79) and the complex target's supercoiled M13dsDNA to form an R-loop, seen as a slower migrating M13dsDNA species compared to reactions lacking Cascade (Figure 3.9 i, compare lanes 2 and 3). At 30 °C, Cas3 converted supercoiled M13dsDNA to nicked and linearized DNA in reactions dependent on magnesium (Figure 3.9 i, compare lanes 4 and 5), and the same effect was observed using the other four tryptophan mutants (Figure 3.9 i, lane 5 compared to lanes 6-9). An interesting result was observed when Cascade R-loops were pre-formed on M13dsDNA—in these reactions, the DNA nicking activity

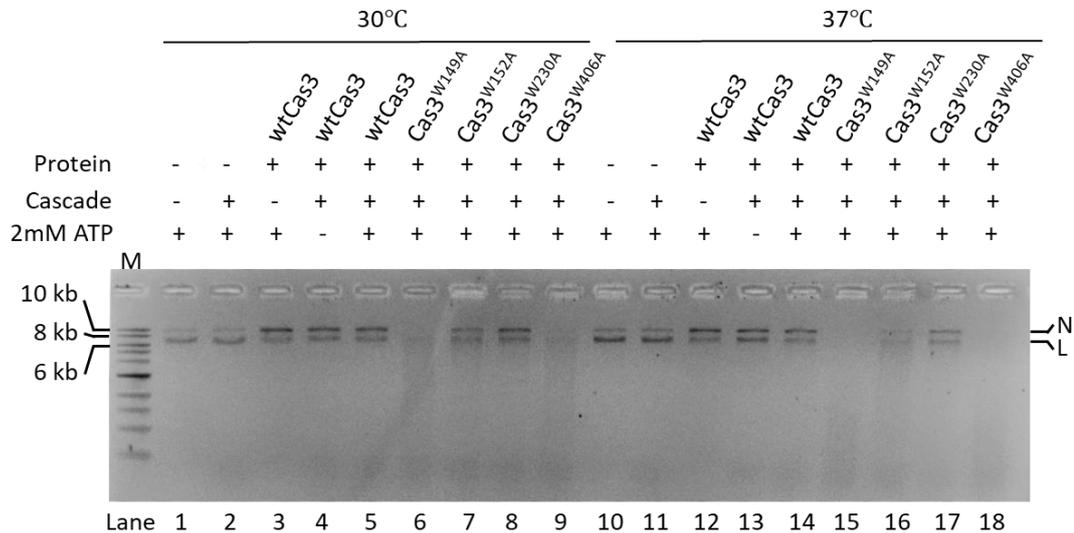


*R-loop substrates, indicated by an \*. The reaction components are, as indicated above, on the ethidium bromide-stained 0.8 % TAE agarose gel panel. In these reactions, no deproteinising 'stop' solution was added, so that the R-loops formed by Cascade could be seen on the gel (e.g., lane 3). The topology of the DNA is highlighted to the right of the gel image by N (nicked) L (linear), \* (R-Loop) and  $\Delta$  (Supercoiled). Quantification data in Appendix 9.*

### **3.9.2 EcoCas3 tryptophan mutants liberate EcoCas3 from functional repression by Cascade, for uncontrolled DNA degradation.**

Assays comparing *EcoCas3* and tryptophan mutants' DNA nicking activity in the presence of Cascade were repeated with the addition of ATP to look for the processive nucleolytic degradation of the M13dsDNA (Figure 3.10). Consistent with a previous analysis of CRISPR interference *in vitro* (122), the reactions containing Cascade, wild-type *EcoCas3* and Mg-ATP only revealed an accumulation of nicked DNA (Figure 3.10, compare lanes 3–5 with lanes 1 and 2). In contrast, all the tryptophan mutants generated products consistent with processive nuclease activity (Figure 3.10, lanes 6-9, 15-18) uncoupled from the constraint of Cascade. This interesting observation indicates that introducing mutants in the iHDA1 region leads to the *EcoCas3* overcoming the down-modulating effect of Cascade *in vitro*. However, those mutants vary in the efficacy of degrading DNA in the presence of Cascade. Another interesting observation was that Cascade recovered the defect of *EcoCas3*

nuclease at 37 °C relative to that at 30 °C as more DNA was degraded by proteins at 37 °C compared to that at 30 °C (Figure 3.10, compare lanes 6-9 and 15-18).



**Figure 3.10 In vitro Cascade-Cas3 interference against M13rfDNA at 30 and 37 °C.**

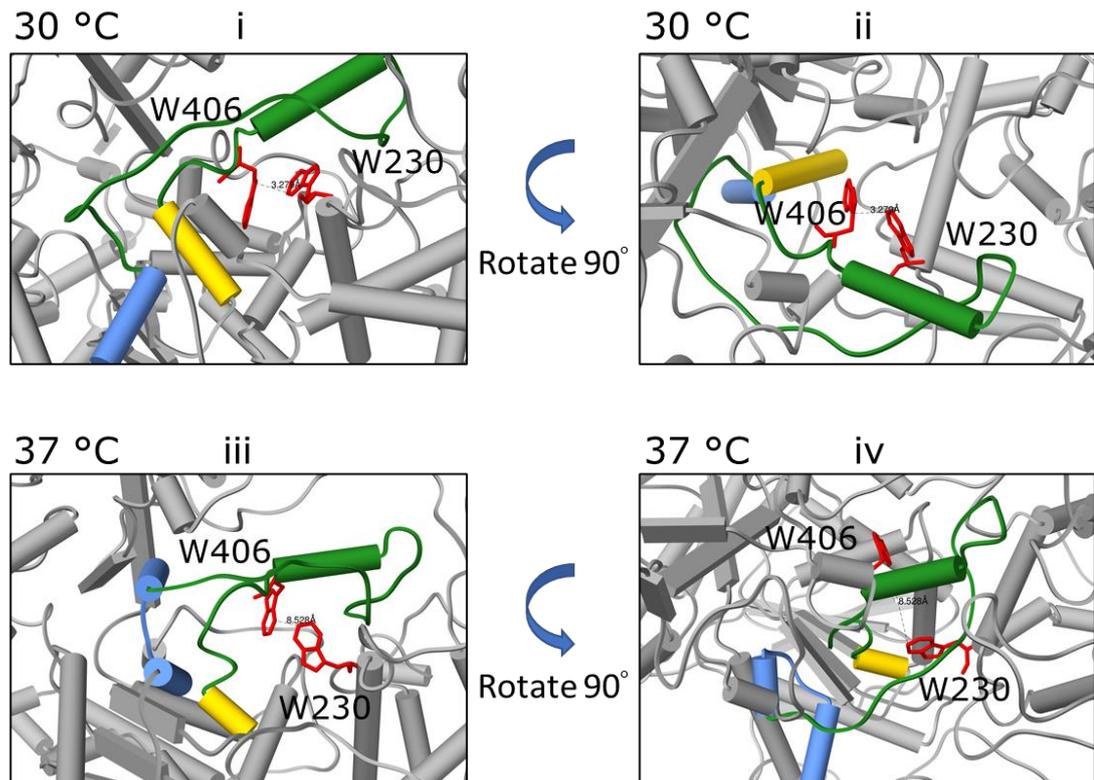
This shows *EcoCas3* nicking and nuclease activity in the presence of  $Mg^{2+}$  and ATP and Cascade (100 nM). The endpoint samples were analysed on ethidium bromide-stained 0.8 % TAE agarose gel. These reactions were stopped by adding a stop solution containing proteinase K and SDS to dissipate the R-loops so that the nuclease products could be more easily discerned. The topology of the DNA is highlighted to the right of the gel image by N (nicked) and L (linear). Quantification data in Appendix 9.

### 3.10 Discussion

The evidence provided in Chapter 3 reveals that the temperature sensitivity of Cas3 nuclease *in vitro* is independent of ATPase activity. Further biophysical analysis conducted by collaborators in University of Zagreb provided direct evidence suggesting this temperature-dependency of Cas3 catalytical function is caused by the

## L. He Chapter 3

repositioning of a highly conserved pair of tryptophan residues close to the nuclease activity site (Figure 3.11). Biochemical data presented in this chapter agrees with previous genetic analysis indicating temperature sensitivity of CRISPR interference in cellular defence against phage  $\lambda$ , which may be relevant to Cas3 (108), and which now can be attributed to the loss of Cas3 nuclease function at 37 °C. The tryptophan residues (Trp-230/406) reside in a region of Cas3 referred to as iHDA1, a domain interface needed to channel ssDNA between the RecA1 helicase domain and the HD nuclease active site. Though the importance of a few amino acids and motifs in iHDA1 has been noted in some structural studies, the biochemical analysis that has been conducted in this study provides the first direct evidence that Trp-406 in the iHDA1 region contributes to the temperature dependency of Cas3 functional status transition.



**Figure 3.11 Repositioning of a highly conserved pair of tryptophan residues Trp-230/406 in *E. coli* Cas3 protein**

(i-iv) The side chains of Trp(W)-230 and Trp(W)-406 are in red. Motifs Ib (blue), Ic (yellow), and an inter-motif sequence in between are denoted in green. Amino acids distances were calculated using UCSF Chimera. (i-ii). Showing the Trp-406 and its adjoining residue Trp-230 at 30 °C, the distance between two amino acids is 3.279 Å. (iii-iv) Showing the Trp-406 and its adjoining residue Trp-230 at 37 °C, the distance between two amino acids is 8.528 Å.

A structural analysis of *TfuCas3* (109) and *TteCas3* (71), which are both *EcoCas3* homologues in I-E subtype CRISPR system, indicate that the DNA phosphate backbone undergoes extraordinary twisting in the iHDA1 region to reach the HD active sites. A conserved tryptophan was predicted to be substantial for DNA interaction in this region. The analysis presented here shows that

Trp-406 in *EcoCas3* may be either 'permissive' or 'inhibitory' to ssDNA accessing the nuclease active site (Figure 3.12) (167). This may explain the temperature-sensitive behaviour of Cas3. Together with CD spectroscopy and the molecular dynamic stimulation data provided by the collaborators from the University of Zagreb, a protein dynamic-function model was proposed which reveals the interplay of Cas3 Trp-230/406 and the Ic motif within the iHDA1 region (167). At 30 °C, Trp-230 and Trp-406 form a stable interaction via their R-groups and retain an open DNA channel to the nuclease catalytic site (DNA permissive state). As the temperature is increased towards 37 °C, the interaction breaks as a consequence of Trp-406 gradually moving away from Trp-230 and stabilizing in a new position through hydrophobic interactions with the surrounding residues residing in the Ic motif. Concurrently, the Trp-406 R-group obstructs the DNA tunnel, which could limit ssDNA access to the nuclease active site (the inhibitory state). Therefore, Trp-406, which is acting as a 'gate', with the assistance of Trp-230 and the Ic motif exerts intrinsic control over the nuclease activity via opening and closing the access of ssDNA to the HD nuclease active site. A summary of this mechanism is shown in Figure 3.12.

Additionally, this steric repression over Cas3 nuclease activity at 37 °C could be substantially overcome by replacing Trp-406 with alanine. This, subsequently, would generate a unique Cas3<sup>W406A</sup> mutant which is nuclease hyperactive. *In vitro* analysis suggests

## L. He Chapter 3

*EcoCas*<sup>W406A</sup> effectively cleaves DNA while it also conquers the functional repression of Cascade over Cas3. Since then, bacterial Cas3 and Cascade have been deployed for genetic editing in human cells, specifically for large scale DNA deletion (127). This identification of the robust mutant *EcoCas*<sup>W406A</sup> may be of use in this respect, particularly if it can be successfully targeted by Cascade. This *EcoCas*<sup>W406A</sup> also produces nuclease activity that is not modulated by Cascade. Together with Cascade, *EcoCas*<sup>W406A</sup> generates productive DNA cleavage at 37 °C, which is the commonly used culture condition for human cells. Therefore, this mutant has the potential to refine the Cascade-Cas3 gene editing system and adapt to the chromosome editing in human cell lines.

Although *EcoCas*<sup>W406A</sup>, along with Cascade, exert efficient target DNA degradation *in vitro*, this mutant seems to be disadvantageous to cells for CRISPR. Based on genetic analysis produced by the collaborators from the University of Zagreb, in the phage sensitivity assay, replacing wild-type *cas3* allele in the *E. coli* chromosome with *Cas3*<sup>W406A</sup> made cells susceptible to phage lysis, which was diagnosed by plaques forming on bacterial lawns relative to those on the control plate (the wild-type strain) (167). Further work is required regarding this study as previous Cas3 research does not wholly explain this phenomenon. It is speculated here that to carry out efficient CRISPR interference requires other factors in addition to Cascade and Cas3, which may include an unknown mechanism.

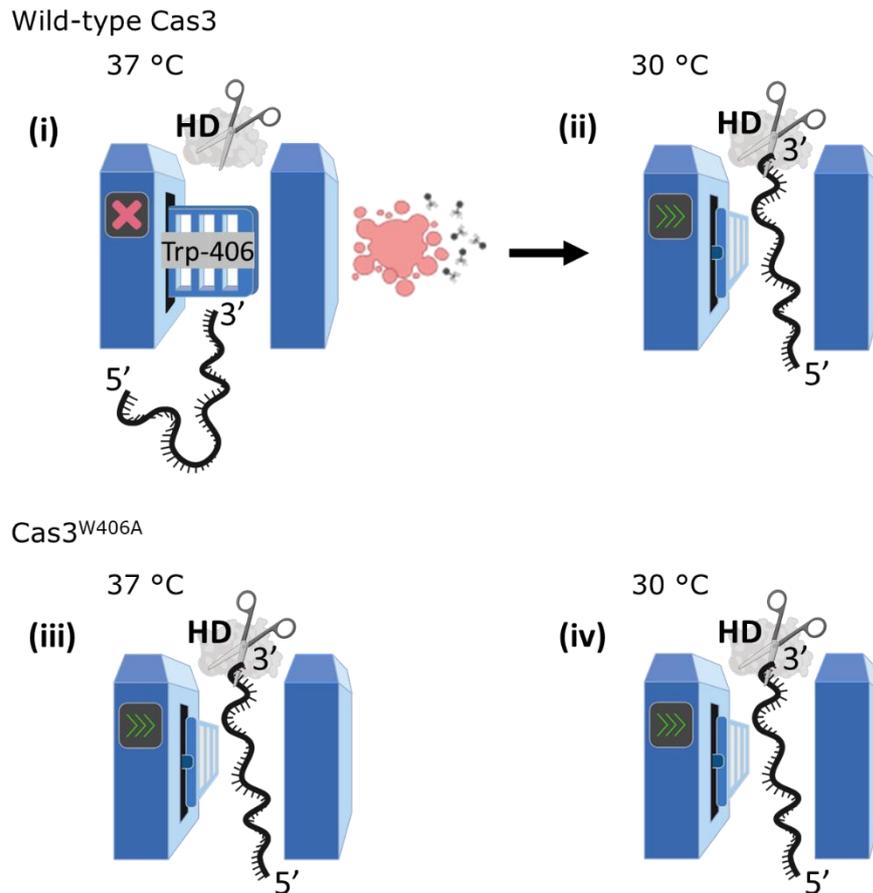
## L. He Chapter 3

Nevertheless, the invariance of Trp-406 across bacterial Cas3 proteins indicates that it holds a crucial role for Cas3 function in the CRISPR systems, one which has not been previously identified; although, a structural study of Cas3 has suggested that tryptophan is likely to be important in Cas3 function.

This study has also highlighted the iHDA1 region that consists of conserved Trp-406 and its competitive binding partners Trp-230, Ib and Ic motif. This region is close to the HD active site; more specifically, the residing Trp-230 is neighbouring the Asp-229, the latter being one of the amino acids that comprise the nuclease activity site in the HD domain. Allosteric regulation inside of the iHDA1 region of the Cas3 nuclease activity ensures that the Cas3 is targeted to destroy only the appropriate substrate nucleic acids at an appropriate time. For Cas3, the interaction with Cascade during CRISPR interference creates functional control of Cascade over it (118,122), but this manipulation, via an extrinsic protein, is abolished when introducing a mutation into the Cas3 iHDA1 region. This indicates that the iHDA1 region may receive external instructions by interacting with foreign factors, and cause a consequent regulation in the Cas3 function by the changing of its conformation. Growth temperature and other environmental factors have a profound effect on bacterial cell physiology; therefore, the effect of temperature that we observe on Cas3 function may reflect the environmental control of the CRISPR systems. Other factors

## L. He Chapter 3

could, instead, be involved, including the CRISPR adaptation of complex Cas1-2 and a chaperone htpG. The Cas1-2 complex has been proved to interact with Cas3 and it together with Cas3 and Cascade forms a prime adaptation complex (PAC) (70,118). A study on I-F subtype CRISPR has revealed that PaeCas1 represses PaeCas2-3 nuclease activity (118). Therefore, it is possible that Cas1-2 physically interacts with Cas3 via the iHDA1 region and induces repression of its function. This could be equivalent to the effect that is brought about by the temperature changing from 30 °C to 37 °C, but further study is needed to investigate this. Additionally, htpG is required for maintaining a functional level of Cas3 in cells at 37 °C (150), indicating it may be a different candidate that exerts a similar regulatory mechanism on Cas3 via contacts in the iHDA1 region. Further analysis has been conducted and is shown in Chapter 4.



**Figure 3.12 Concluding summary of the Trp-406 nuclease 'gate' model.**

(i). When Trp-406 is positioned close to the helicase motif *Ic*, it blocks the ssDNA binding channel to the nuclease (HD) activity site, resulting in a 'closed' gate. In this scenario, CRISPR interference is inhibited because MGE DNA is not destroyed, resulting in cell lysis by phage. (ii). Conformational movement of Cas3 re-positions Trp-406 proximally to Trp-230, with which it interacts to form a lining to the ssDNA channel when the gate is 'open'. In this state, Cas3 can degrade the MGE DNA to fulfil its role in CRISPR adaptation reactions. (iii, iv) For Cas3<sup>W406A</sup> mutant, ssDNA can get access to HD nuclease activity site at both 30 and 37 °C

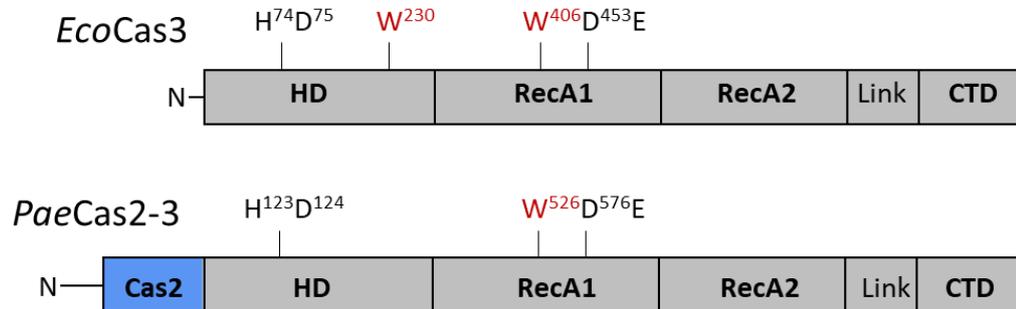
## **Chapter 4**

### **Comparative analysis of two divergent type I Cas3 proteins identified a hypothetical cysteine redox switch, and novel protein-protein interaction.**

In Chapter 3 and in He *et al.* (2021) (167) we provided evidence that the iHDA1 region of *EcoCas3* controls nuclease activity via repositioning the Trp-406 to open and close access of the ssDNA to the nuclease active site. Since that tryptophan is highly conserved across species, two questions arise: Whether the 'tryptophan gate' model discussed in Chapter 3 is also a verifiable biochemical feature of other Cas3 proteins; and, does the conformational change of the Cas3 iHDA1 region is a real environmental control effect over Cas3 function drove by temperature change, or it is accidentally equivalent to protein structural re-arrangement facilitated by protein-protein interaction?

To answer those questions, *Pseudomonas aeruginosa* (*Pae*) Cas2-3 in a CRISPR I-F subtype is introduced into this study to compare with *EcoCas3*, as a unique Cas3 version that has the adaptation protein Cas2 fused to its N- terminus. Its protein structure, along with its biochemical features, has been well characterised (118,120).

It has a conserved Trp-526 that is equivalent to Trp-406 in *EcoCas3* (Figure 4.1).



**Figure 4.1 The structural and amino acid sequence alignment of *PaeCas2-3* and *EcoCas3*.**

The illustration represents *EcoCas3* and *PaeCas2-3* with amino acid residues indicating the nuclease (HD), Walker B ATPase-active sites (DEXD box) and the equivalent tryptophan amino acids.

## Part I

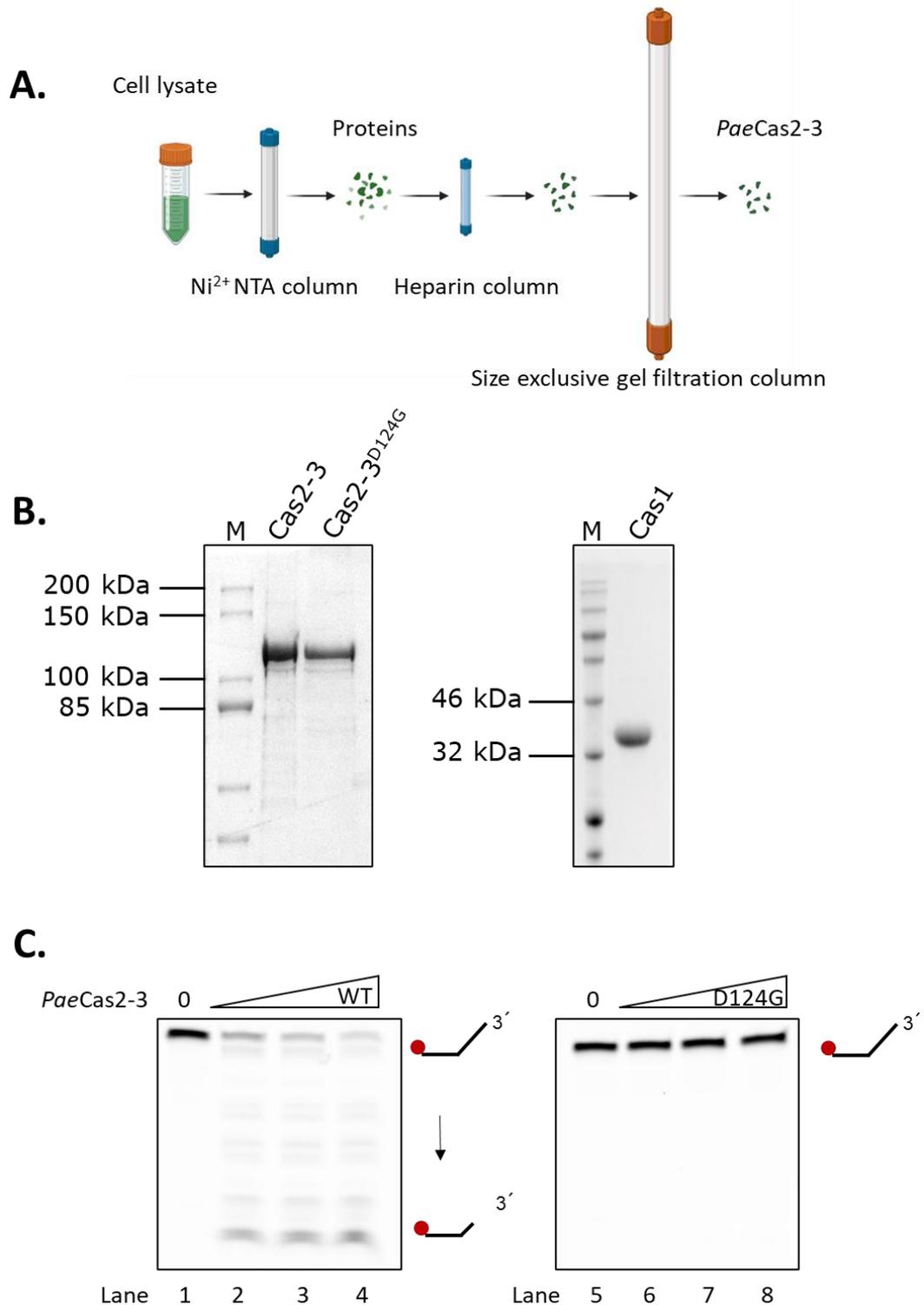
### Comparative analysis of *PaeCas2-3* and *EcoCas3* identified a hypothetical cysteine redox switch in *EcoCas3* iHDA1 region.

#### 4.1 Purified *PaeCas2-3* is a nuclease on the model DNA fork.

*PaeCas2-3* or its mutant proteins were tagged with N-terminus His<sub>6</sub> for isolation from lysates initially using an Ni-NTA column. The eluted products were then loaded onto a heparin column to remove contaminants and concentrating the bound target proteins. The

*PaeCas2-3* eluted from the heparin column was further purified using a Superdex 200 10/300 GL filtration column (Figure 4.2 A). The eluted proteins in the buffer comprise 20 mM Tris-HCl pH 7.5, 500 mM NaCl and 5 % glycerol, which were immediately stored at -80 °C (Figure 4.2 B, left; Chapter 2, Section 2.3.5). The *PaeCas1* fused with a His<sub>x6</sub> tag at its N-terminal was obtained via the same method, but the protein was immediately dialysed after being eluted from the heparin column, and stored (Figure 4.2 B, right; Chapter 2, Section 2.3.6).

Purified *PaeCas2-3* and its nuclease inactive mutant *PaeCas2-3*<sup>D124G</sup> were assayed with 20 nM DNA fork (Figure 4.2 C), with 10 mM MgCl<sub>2</sub> to stimulate the DNA catalytic activity of *PaeCas2-3* reference for this. Wild-type *PaeCas2-3* showed DNA cleavage after one hour incubation at 37 °C (Figure 4.2 C, lanes 2-4), whereas no DNA degradation products were generated by *PaeCas2-3*<sup>D124G</sup> (Figure 4.2 C, lanes 6-8). Therefore, this purification method gave nuclease active *PaeCas2-3*, without any *E. coli* nuclease contaminants.



**Figure 4.2 The nuclease activity of PaeCas2-3 was analysed against DNA fork substrate.**

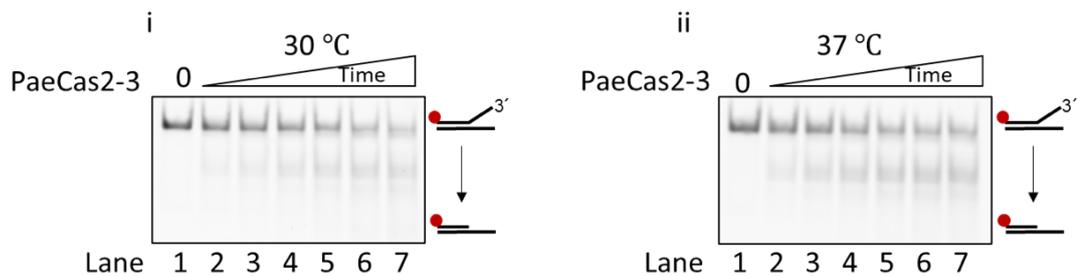
(A). This shows a schematic diagram of the His<sub>x6</sub>-PaeCas2-3 purification method. Cell lysate containing target protein was loaded onto an Ni-NTA column and the proteins bound to the first column were eluted with an increasing imidazole gradient, with the final concentration of imidazole

being 250 mM. The eluted proteins were loaded onto the heparin column to concentrate the target proteins and remove any unnecessary contaminants. The proteins were then concentrated using a spin column and loaded onto a Superdex 200 Increase 10/300 GL column gel filtration column for further purification. The eluted fractions containing the target proteins were immediately frozen and stored at -80 °C. (B). The purified *PaeCas2-3* and *PaeCas2-3<sup>D124G</sup>* (128 kDa) were placed on 10 % SDS page gel and post-stained with Coomassie blue. The gels, from left to right, show wild-type *PaeCas2-3* (lane 1), *PaeCas2-3<sup>D124G</sup>* (lane 2) and purified *PaeCas1* of 40 kDa (lane 3). (C). This shows the nuclease activity analysis of *PaeCas2-3* and *PaeCas3<sup>D124G</sup>* on the DNA fork substrate of 20 nM. The 0 nM, 50 nM, 100 nM and 200 nM of each protein, in lanes 1-4 and lanes 5-8, were incubated at 37 °C for one hour before the endpoint reaction products were taken and analysed on 10 % acrylamide denaturing TBE gel. Quantification data in Appendix 9.

## **4.2 *PaeCas2-3* is nuclease active at both 30 °C and 37 °C.**

It was previously found that *EcoCas3* nuclease activity is repressed at 37 °C (Chapter 3, Section 3.5). To analyse if *PaeCas2-3* has the same biochemical feature as the *EcoCas3* has revealed, the nuclease assay was carried out using *PaeCas2-3* at 30 and 37 °C. The *PaeCas2-3* of 50 nM was incubated with 20 nM DNA fork to examine the impact of temperature on the protein catalytic function. Interestingly, *PaeCas2-3* was similarly active at both temperatures: Almost equal amount of cleaved DNA products were generated by *PaeCas2-3* at 30 °C and 37 °C (Figure 4.3, panel i and ii). This result suggests that the nuclease activity of *PaeCas2-3* is independent of temperature, even though it carries a conserved Trp-526, which is equivalent to the Trp-406 in *EcoCas3*. This phenomenon may be

attributed to that the *PaeCas2-3* containing a Cas2 fusion which is adjacent to the iHDA1 region (Chapter 3, Figure 3.1 C, panel iii and iv). It is possible the existence of Cas2 stabilise the conformation of iHDA1 region in *PaeCas2-3* and lead to no change of protein structure dependent on temperature change.



**Figure 4.3 *PaeCas2-3* is an active nuclease at both 30 °C and 37 °C.**

Individual results of the nuclease assay at different temperatures were carried out using 50 nM *PaeCas2-3* proteins on 20 nM DNA fork substrates. The samples were taken at 0 min, 5 min, 15 min, 30 min, 60 min, 120 min and 240 min, and displayed on 10 % acrylamide native TBE gels. The assays were conducted at 30 °C and 37 °C, as denoted. Quantification data in Appendix 9.

### **4.3 *PaeCas2-3* Trp-526 does not interact with the Ic motif, therefore it may not be able to block ssDNA from accessing its DNA catalytic site in the HD domain.**

Since *PaeCas2-3* nuclease function is independent of temperature change, unlike the *EcoCas3*, a structural analysis was then carried out on a *PaeCas2-3* model (PDB: 5B7I) to understand the interplay

## L. He Chapter 4

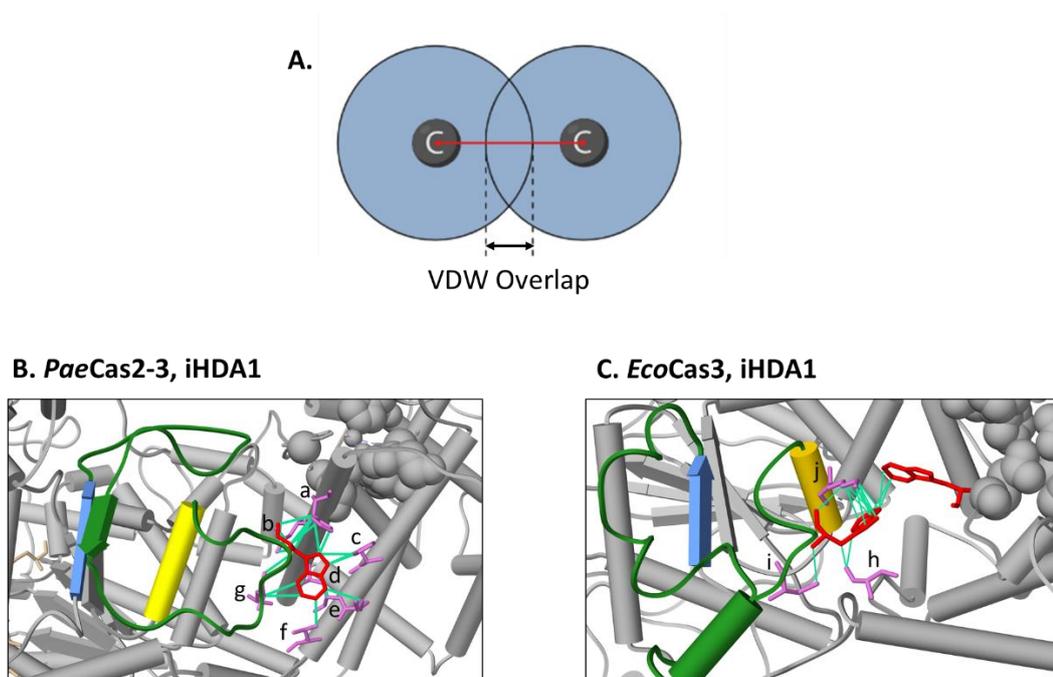
of Trp-526 with its adjoining amino acid residues. This part of the study was done by using UCSF Chimera to analyse existing *PaeCas2-3* (PDB: 5B7I) structural model, in order to predict potential residues that interact with Trp-526. The *E. coli* Cas3 Trp-406 was also included in this in silico prediction as a comparison. Regarding the standard used to select 'possible interactors' to *PaeCas2-3* Trp-526, any amino acids with more than  $-0.4 \text{ \AA}$  Van Der Waals overlap with assigned tryptophan were considered to be candidates. This VDW overlap parameter of two atoms represents the sum of the VDW radii of each atom minus the distance between them (Figure 4.4 A). The formation of an H-bond between the two atoms was not considered because only carbon atoms were used for the analysis.

In *PaeCas2-3*, the Trp-526 was predicted to interact with Leu-313, His-317, His-359, Lys-362, Leu-363 and Thr-366, which are located in the HD nuclease domain. An additional Leu-523 on a linker between the Ib and Ic motifs in RecA1 domain was also highlighted as a possible interactor for Trp-526 (Figure 4.4 B). Most of those interactors, apart from the Leu-523, are on two  $\alpha$ -helices carrying crucial amino acids which form the HD DNA catalytic site in *PaeCas2-3*. These amino acid residues may contribute to the stability of the residue orientation of Trp-526 and prevent it from blocking the ssDNA tunnel. In addition, in this analysis there was no contact between Trp-526 and amino acids in the Ic motif, supporting the lack of temperature-dependent steric regulation of the DNA tunnel

in *PaeCas2-3* that limits nuclease activity in *EcoCas3*. It is possible that other amino acids may exert the 'gate' role in *PaeCas2-3* as the Trp-406 does in *EcoCas3*. Conducting the structural analysis to search for the possible 'gate' amino acids in *PaeCas2-3* is rather difficult when using the available *PaeCas2-3* models. This is because of the lack of detailed structural information from Ala-483 to Ala-502 in the model produced by X-ray crystallography (PDB: 5B7I); and from Gly-452 to Asp-520 in the Cryo-EM model (PDB: 5GQH). This problem may be overcome by matching the incomplete *PaeCas2-3* structure to a complete version which has been predicted by the protein structure-predicting algorithm AlphaFold (168). By doing so, amino acids that may have caused the steric block in the DNA binding channel were narrowed down to 10 candidates, these being Arg-484, Glu-488, Gln-491, Glu-492, Glu-495, Arg-496, Ser-497, Ser-499, Glu-500 and Ser-501 (Figure 4.5 A). The Glu-488 is relatively conserved among Cas3 proteins in different species (Figure 4.5 B), further study may be of value.

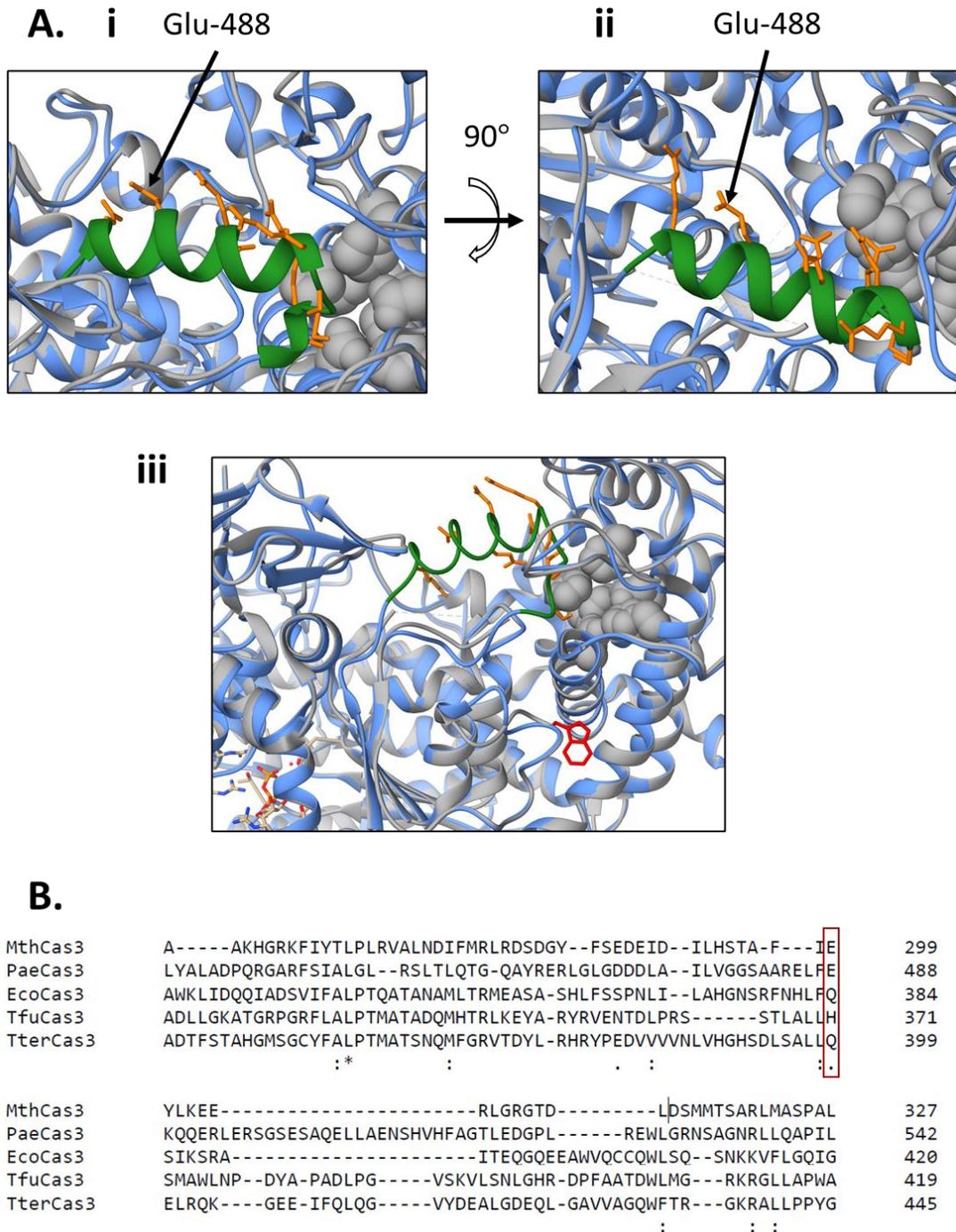
The same interaction predicting method was applied to *EcoCas3* Trp-406, to verify the accuracy of the adopted method (Figure 4.4 C). This was done by comparing the results with the outcome generated previously by MD stimulation conducted by collaborators in the University of Zagreb (167). As expected, Trp-230, plus Lys-413 in the Ic motif, are predicted interactors to Trp-406, which is consistent with previous MD stimulation results. Meanwhile, Ser-271

and Leu-407 may have contact with Trp-406. This Leucine has evolved into phenylalanine in some Cas3 subtypes, but it is still considered to form a highly conserved hydrophobic W-F/L motif with the adjacent tryptophan.



**Figure 4.4 PaeCas2-3 Trp-526 and EcoCas3 interact with adjoining amino acids.**

(A) This shows the definition of 'VDW overlap' that is used to identify possible interactors to assigned tryptophan residues. The red line represents the distance between two carbon atoms. The Van der Waals force is denoted in blue. (B, C) The complete iHDA1 from PaeCas2-3 and the EcoCas3, including the motif Ib (blue), a linker polypeptide (green) and Ic (yellow), are denoted in each panel. The grey spheres represent the nuclease catalytic site in the HD domain. (B). Seven amino acids (a to g) are predicted to interact with PaeCas2-3 Trp-526 (marked in red), denoted as a. His-317, b. Leu-313, c. His-359, d. Leu-363, e. Lys-362, f. Thr-366 and g. Leu 523. (C). EcoCas3 Trp-406, labelled in red on the left, interacted with Trp-230 (red, right) and Lys-413 (j) in the Ic motif. A further two interactors are h. Ser-271 and i. Lys-413.



**Figure 4.5 Potential amino acids in PaeCas2-3 that may cause steric block to DNA accessing HD catalytic site.**

(A). Panel i and ii show amino acid candidates in PaeCas2-3 that may cause steric block; their side chains are coloured in orange. The Glu-488 is denoted. The others are Arg-484, Gln-491, Glu-492, Glu-495, Arg-496, Ser-497, Ser-499, Glu-500 and Ser-501. The PaeCas2-3 is grey (PDB: 5B7I); it is aligned with the model predicted by AlphaFold, which is blue. The missing amino acids from Ala-483 to Ala-502 in the PDB 5B7I model are marked in the AlphaFold model in green. The Trp-526 is denoted in

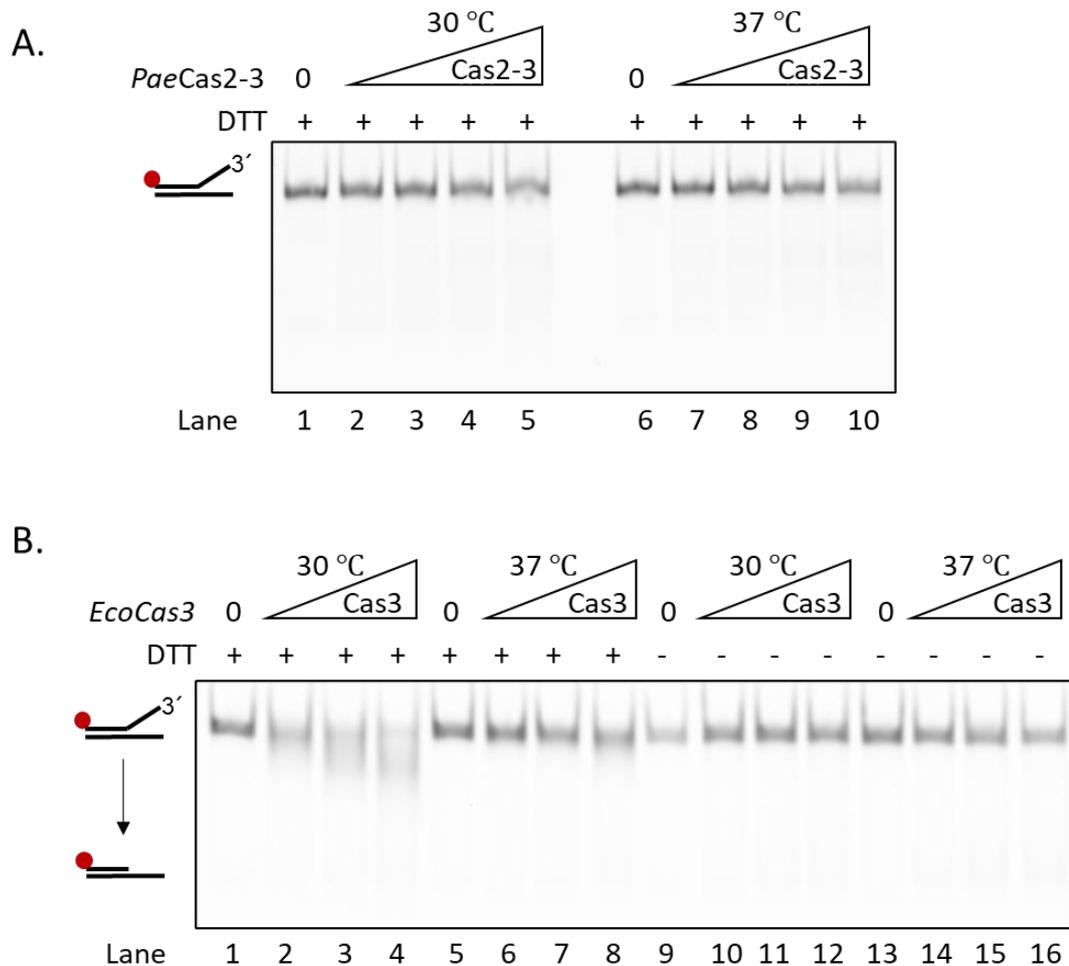
red. Panel iii shows the relative location of the 'steric block' and the Trp-526. (B). This shows the sequence alignment, highlighting the conservation of the residue Glu-488 in *PaeCas2-3* with *Cas3* proteins from *M. ther* (*Mth*), *E. coli* (*Eco*), *T. fusca* (*Tfu*) and *T. terrenum* (*Tter*) proteins. The protein sequences were aligned using Clustal Omega.

#### **4.4 The reducing agent DTT inhibits the *PaeCas2-3* function, but stimulates *EcoCas3*.**

When establishing the nuclease assay conditions for *PaeCas2-3*, we found that DTT inhibited *PaeCas2-3* nuclease (Figure 4.6 A). Adding 20 mM DTT to the nuclease assay using *PaeCas2-3* against the DNA fork, no DNA degradation occurred at 30 °C or 37 °C. Therefore, the DTT was removed in the assays containing *Pae* proteins. Additional tests were then conducted on *EcoCas3* to test if the DTT was necessary for its catalytic function. It was apparent that the DTT is necessary for *EcoCas3* exerting nuclease activity (Figure 4.6 B, lanes 1-8), whereas removal of it led to a significant reduction in DNA degradation by *EcoCas3* (Figure 4.6 B, lanes 9-16).

This preference for DTT may be explained by the existence of cysteine residues in the *EcoCas3* iHDA1 region, that are absent in *PaeCas2-3*. *EcoCas3* contains two adjacent cysteine residues, Cys-403 and Cys-404, on the linker polypeptide connecting Ib and Ic motifs (Figure 4.7 A, panel i and ii). Those two cysteines and the 'gate' residue Trp-406 are one amino acid apart. Their existence in the iHDA1 region may stabilise local polypeptide conformation in

response to DTT, thus reducing the flexibility of this region. This effect may be achieved via disulphide bond formation between the Cys-403 and Cys-404, and the interaction of two cysteines with surrounding amino acid residues. In total, 12 amino acid residues were predicted to interact with the Cys-403 and Cys-404, including: those that reside in the HD domain, Leu-270, Ser271 and Gly-272; the Ib motif's (372-375 aa) adjoining residues Asn-369 and Ile-371; Phe-383 and Lys-387; and Trp-406 and its adjoining residues Val-401, Gln-402, Gln-405 and Ser-408 (Figure 4.7 A, panel iii). Therefore, it is possible that Cys-403 and Cys-404 contribute to conformational stabilisation. Incubating *EcoCas3* with DTT would cause a reduction of the disulfide bond and, consequently, promote the plasticity of the iHDA1 region and restore *EcoCas3* nuclease activity. Interestingly, two adjacent cysteines are often considered to be a redox switch (169,170) (Figure 4.7 B), which plays a key role in enzyme catalytic function regulation by post-translational modifications. So, it is possible that Cys-403 and Cys-404 in *EcoCas3*, form a redox switch that tandemly the Trp-406 'gate' residue. Therefore, a hypothesis is proposed here that Cas3 nuclease activity may be regulated by post-translational modifications, instead of only by Cascade (118) and temperature (167). In contrast to the *EcoCas3*, the iHDA1 region of *PaeCas2-3* contains no cysteine residues. Why the addition of DTT can inhibit DNA hydrolysis by *PaeCas2-3* is still unclear.

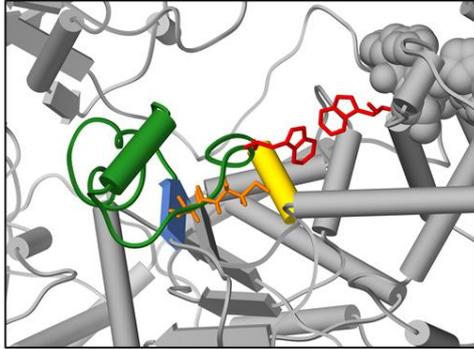


**Figure 4.6 DTT inhibits nuclease activity of *PaeCas2-3* but assists *EcoCas3* nuclease activity.**

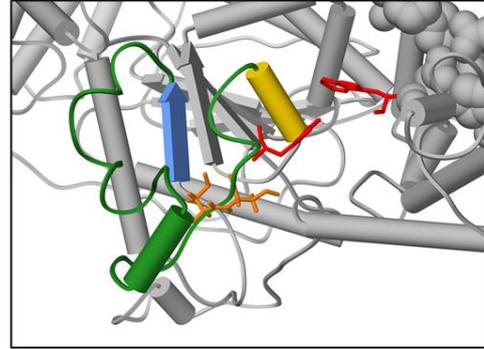
(A). This shows the nuclease activity analysis of *PaeCas2-3* against 20 nM DNA fork in buffer containing 20 mM DTT. The concentration of the *PaeCas2-3* is increased from 0 nM, 50 nM, 100 nM, 200 nM to 400 nM in lanes 1-5 and lanes 6-10. The reactions were carried at 30 °C or 37 °C two hours in advance, before taking the endpoint reaction products and analysing them on 10 % acrylamide TBE native gel. (B). This shows the nuclease activity analysis of *PaeCas2-3* against 20 nM DNA fork in the presence of 20 nM DTT, and in a buffer without DTT. The concentration of *EcoCas3* was increased from 0 nM, 28 nM, 56 nM to 112 nM in lanes 1-4, lanes 5-8, lanes 9-12 and lanes 12-16. The reactions were conducted at 30 °C or 37 °C for two hours, before the endpoint reaction products were taken and analysed on 10 % acrylamide TBE native gel. Quantification data in Appendix 9.

### EcoCas3

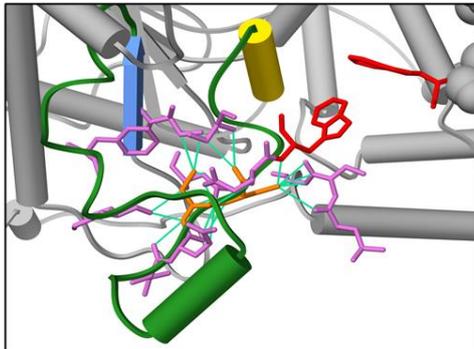
A.i



A.ii

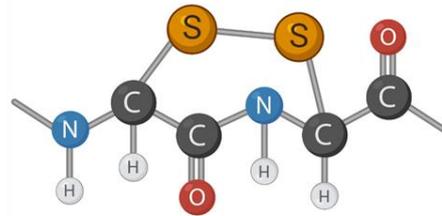


A.iii



B.

Cysteine Redox Switch



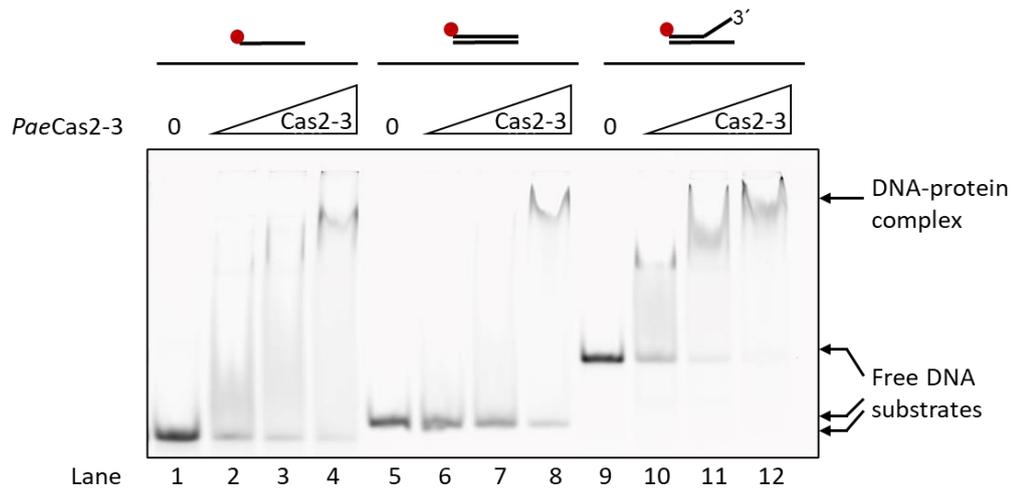
**Figure 4.7 Two adjacent cysteine residues Cys-403 and Cys-404 tandem the Trp-406 'gate' residue.**

(A). Panel i and ii show the side chains of Cys-403 and Cys-404, which are denoted in orange, as well as their location relative to Trp-406 (red, left), Trp-230 (red, right), Ib motif (blue) and Ic motif (yellow). Panel iii shows the predicted interactions of Cys-403 and Cys-404, with the surrounding amino acid residues (coloured in orchid), including Leu-270, Ser271, Gly-272, Asn-369, Ile-371, Phe-383, Lys-387, Val-401, Gln-402, Gln-405, Trp-406 and Ser-408. The lines in light green denote the interactions between atoms. (B). This shows the conformation of a redox switch formed by two adjacent cysteine residues.

#### **4.5 EMSA shows *PaeCas2-3* binds to DNA substrates of different types.**

Previous data have shown that *EcoCas3* has low affinity to DNA (Section 3.7). However, this has not yet been tested on *PaeCas2-3* of its DNA binding capacity. To this end, *PaeCas2-3* was incubated with three different DNA substrates separately, all with a Cy5 label at its 5' end. Those selected substrates, 50 nt ssDNA, 50-mer dsDNA and DNA fork, varied in conformation to compare the affinity of *PaeCas2-3* to different DNA substrates.

*PaeCas2-3* forms stable protein-DNA complex with all three selected DNA substrates (Figure 4.8). *PaeCas2-3* may perform the highest affinity to the DNA fork as there are no detectable free substrates in lane 12 compared to lanes 4 and 8. This result suggests that *PaeCas2-3* binds to DNA better than *EcoCas3* (Figure 3.7). This higher affinity to DNA in *PaeCas2-3* may be attributable to its Cas2 fusion. However, whether *PaeCas2-3* preferentially interacts with the DNA fork needs further quantification to be validated.



**Figure 4.8 EMSA assay showing wild-type *PaeCas2-3* forming a stable protein-DNA complex with different substrates.**

The EMSA shows that wild-type *PaeCas2-3* forms stable protein-DNA complexes with 20 nM substrates, including ssDNA, dsDNA and the DNA fork. An increasing concentration of *PaeCas2-3* from 0 nM, 50 nM, 100 nM to 200 nM is shown in lanes 1-4, lanes 5-8 and lanes 9-12. The reactions took place at room temperature (20°C) for 15 min before the products were taken and analysed on 8 % acrylamide TBE native gel. Quantification data in Appendix 9.

## 4.6 Discussion for Part I

This part of the study analysed the biochemical features for *PaeCas2-3*, including its nuclease activity at different temperatures, for comparison with *EcoCas3*. Interestingly, *PaeCas2-3* exhibited temperature-independent nuclease activity in contrast to what has been revealed in previous work using *EcoCas3*. Unlike *EcoCas3*, this Cas2-Cas3 fusion protein is an active nuclease at both 30 °C and 37 °C. Notably, even though *PaeCas2-3* contains an iHDA1 region which includes a Trp-526 that is equivalent to the Trp-406 in *EcoCas3*, it reveals a distinct biochemical feature. This phenomenon

may be explained by the subsequent structural analysis whereby it was seen that *PaeCas2-3* Trp-526 does not interact with the Ic motif; instead, it forms a stable interaction with surrounding amino acids in the HD domain. Therefore, Trp-526 is predicted to lose its role of constructing a steric block in the DNA tunnel with the aid of the Ic motif as an effect of iHDA1 structural alteration driven by the changing of the temperature.

This difference in *PaeCas2-3* nuclease activity compared to *EcoCas3* suggests that *PaeCas2-3* contains a more stable iHDA1 region, one in which there is a fused Cas2 in the vicinity of the iHDA1 (Chapter 3, Figure 3.1 C, iii and iv). This would deprive structural flexibility in this region and ultimately result in a temperature-independent effect on *PaeCas2-3* nuclease activity. Furthermore, Cas2 may be capable of operating Cas3 via interaction with the iHDA1 region. Additional *in vitro* studies using Cas3 and Cas2 would offer insight into this topic.

## **Part II**

### **Analysis of *EcoCas3* with proteins that potentially target the *EcoCas3* iHDA1 region**

We suggested that Cas3 undergoes a switch-like functional change generated by the positioning of Trp-406 in the iHDA1 region

(Chapter 3 and (167)). Through further analysis of *EcoCas3* and *PaeCas2-3*, a hypothesis was proposed that Cas3 iHDA1 may be an interaction zone for allosteric factors (protein or non-protein) that modulate Cas3 functions in response to specific conditions. Therefore, three putative interactor proteins (Cas1, Cas2 and HtpG) to the *EcoCas3* iHDA1 region were tested for modulation of Cas3 function, using as targets the wild type *EcoCas3* and hyperactive nuclease *EcoCas3*<sup>W406</sup>. We predicted that any modulation of the wild type iHDA1 region would not be effective in *EcoCas3*<sup>W406A</sup>.

Cas1 and Cas2 were selected as possible interactors because they form the PAC complex with Cas3 and Cascade in I-E and I-F CRISPR subtypes (70,118). In I-F CRISPR, Cas1 alone has been shown to repress Cas2-3 nuclease activity (118). Meanwhile, Cas2 fusion, which is located next to the iHDA1 region in *PaeCas2-3* may reduce the flexibility of this region, details of which have been discussed earlier in this chapter. The chaperone protein HtpG, was also tested for its crucial role during CRISPR interference in maintaining functional Cas3 protein levels (150).

#### **4.7 *EcoCas1* represses *EcoCas3* nuclease activity but not hyperactive nuclease *EcoCas3*<sup>W406A</sup>.**

An increasing concentration of an inactivated *EcoCas2*<sup>E9A</sup> and wild-type *EcoCas1* were incubated separately with 550 nM *EcoCas3*

at 30 °C. Those two adaptation proteins were used individually to preclude the competitive binding effect which may block Cas3 accesses substrates when utilising the Cas1-2 complex. No obvious variation of DNA product appeared in the reactions that involved *EcoCas2*<sup>E9A</sup> compared to those without it (Figure 4.9 A, compare lane 8 and lanes 9-14). This suggests *EcoCas2*<sup>E9A</sup> has no impact on *EcoCas3* nuclease activity at 30 °C. Meanwhile, apparent nuclease inhibition occurred in the reaction which included 1.2 uM *EcoCas1* and 550 nM *EcoCas3* (Figure 4.9 B, compare lanes 8 and 14). It is clear that in the reactions which included a lower amount of *EcoCas3* of 300 nM, increasing the *EcoCas1*, gradually repressed the *EcoCas3* nuclease (figure 4.9 C, lanes 3-5). These results indicate that the role of Cas1 in repressing the nuclease activity of Cas3 may be a conserved feature of the CRISPR-Cas system because this mechanism has been observed both in I-E (in this study) and I-F CRISPR subtypes. The latter has been previously shown in Rollins *et al* (118). using *PaeCas1* and *PaeCas2-3*. Interestingly, *EcoCas1* could not exert similar suppression on hyperactive *EcoCas3*<sup>W406A</sup> as no convincing repression of the production of degraded DNA could be observed (Figure, 4.9 D).

The data suggest *EcoCas1*, but not *EcoCas2*<sup>E9A</sup>, represses *EcoCas3* nuclease activity at 30 °C, while nuclease hyperactive *EcoCas3*<sup>W406A</sup> overcomes this regulatory mechanism. This is apparent as no obvious *EcoCas3*<sup>W406A</sup> nuclease activity decay

occurred in the presence of *EcoCas1*. Another possibility is that the iHDA1 region exerts a regulatory role over the *EcoCas3* function when it interacts with other proteins; the Trp-406 has been identified as one of a few crucial amino acids located in it. It is, therefore, hypothesised that by interacting with Cas3 in the iHDA1 region, Cas1 promotes relocation of the core amino acid residue Trp-406, consequently reducing Cas3 nuclease activity.

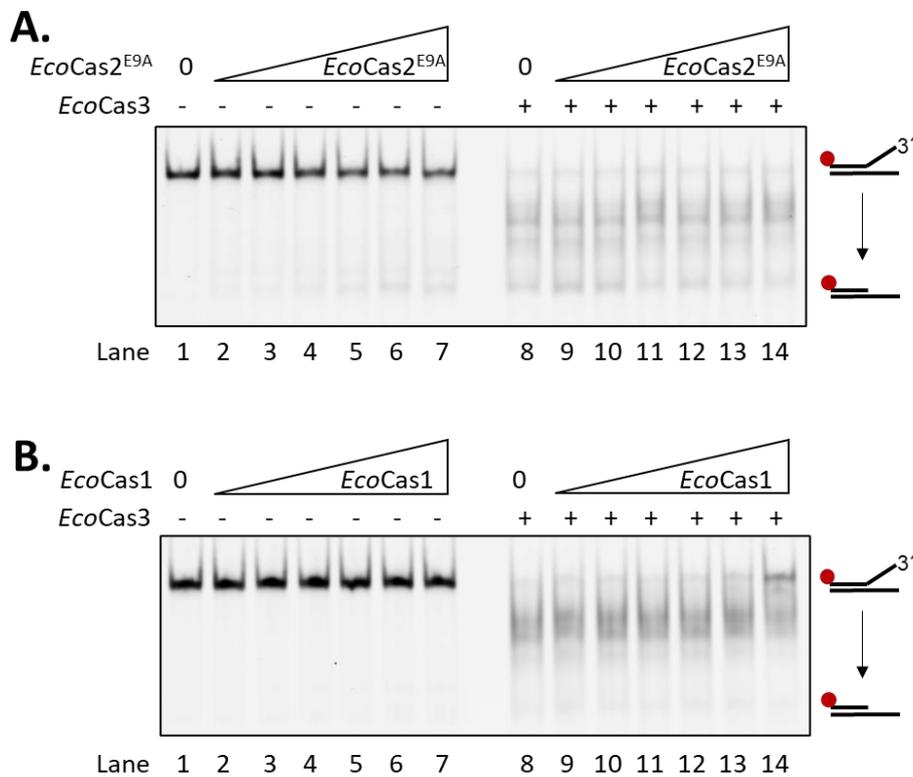
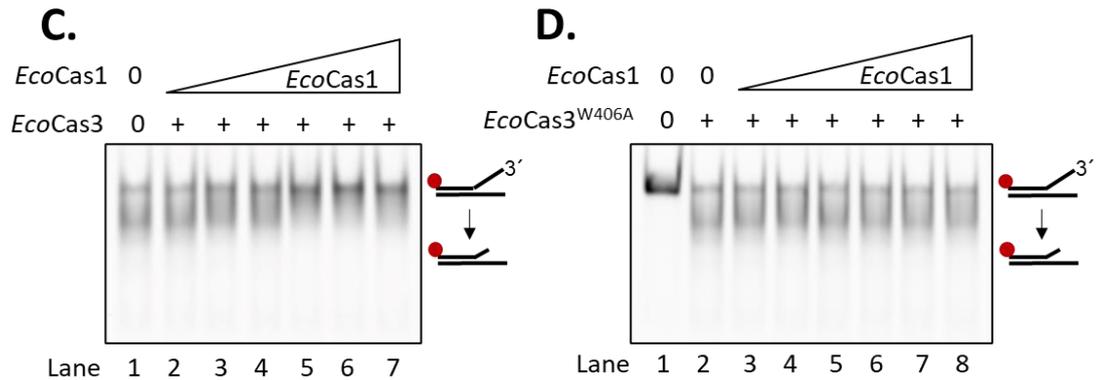


Figure 4.9 Cont.



**Figure 4.9 EcoCas3 nuclease activity in the presence of EcoCas2<sup>E9A</sup> and EcoCas1.**

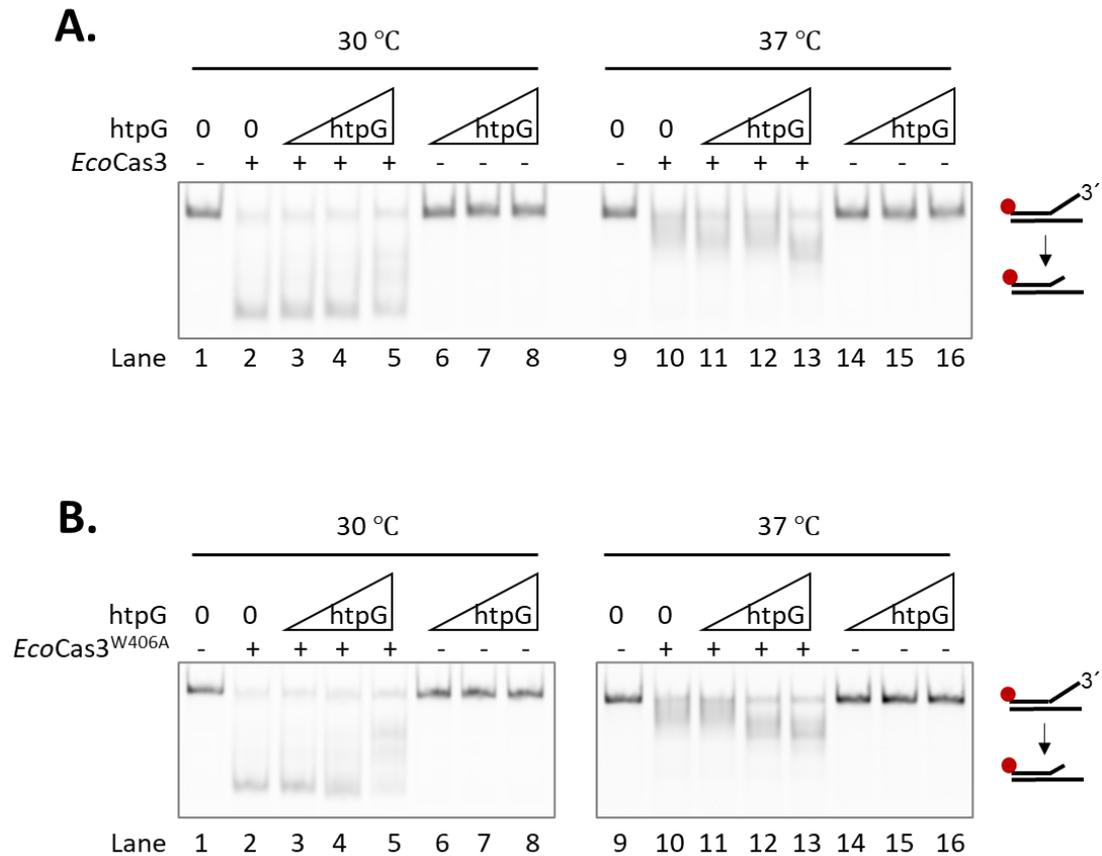
(A, B). This shows the nuclease activity analysis of 550 nM EcoCas3 against 20 nM DNA fork with additional EcoCas2<sup>E9A</sup>. The concentration of EcoCas2<sup>E9A</sup> (in A) or Cas1 (in B) increases from 0 nM, 50 nM, 100 nM, 200nM, 400 nM, 800 nM to 1.2  $\mu$ M in lanes 1-7 and lanes 8-14. The reactions took place at 30 °C two hours in advance, before the endpoint reaction products were taken and analysed on 10 % acrylamide TBE native gel. Only a small amount of degraded DNA can be seen in lanes 4-7 containing EcoCas2<sup>E9A</sup> only. This has been generated by star activity, which usually occurs when glycerol is present up to 5 % (v/v) of the total reaction volume, and most of the glycerol is from the EcoCas2<sup>E9A</sup> protein sample, as is the case here. (C, D). This shows the nuclease activity analysis of 330 nM EcoCas3 (in C) or EcoCas3<sup>W406A</sup> (in D), against 20 nM DNA fork with additional EcoCas1. The concentration of Cas1 increases from 0 nM, 50 nM, 100 nM, 200nM, 400 nM, 800 nM to 1.2  $\mu$ M in lanes 1-7 (in C) and lanes 2-8 (in D). The reactions took place at 30 °C two hours in advance, before the endpoint reaction products were taken and analysed on 10 % acrylamide TBE native page gel. Quantification data in Appendix 9.

#### 4.8 HtpG promotes the nuclease activity of EcoCas3 and EcoCas3<sup>W406A</sup> at 37 °C.

Apart from Cas proteins such as Cas1 and Cas2, there is one non-CRISPR protein HtpG which may also have an impact on Cas3 function. This chaperone protein has been proved as a necessary factor for promoting CRISPR efficacy against  $\lambda$  phage and maintains a functioning level of Cas3 in host cells (150) (108). Therefore, it is

possible that HtpG interacts in the iHDA1 region and, concurrently, alters Cas3 activities. To address this, DNA cleavage on a 20 nM DNA fork by 800 nM *EcoCas3* was analysed at increasing HtpG concentrations of 10 nM, 20 nM and 40 nM. The samples were analysed at 30 °C and 37 °C for two hours before comparing the endpoint products for each reaction.

In an assay taken at 30 °C, it is difficult to compare the total amount of substrate loss when comparing the intact DNA fork left in each reaction after being treated with *EcoCas3*. Nevertheless, those DNA products in reactions with extra 40 nM HtpG revealed significant differences compared with those containing no or less HtpG, regarding their variation in length (Figure 4.10 A, compare lanes 2 and 5). Whereas at 37 °C, HtpG promoted *EcoCas3* nuclease activity and the smaller DNA products which were generated migrated further away from the intact DNA substrates (Figure 4.10 A, compare lanes 10 and 13). Similar results were observed in the repeat assay utilising *EcoCas3*<sup>W406A</sup> not wild-type protein. These data indicate that HtpG could promote *EcoCas3* nuclease activity at 37 °C, and the addition of it to *EcoCas3* at 30 °C could result in a change in the DNA product length, though the mechanism is still unknown.



**Figure 4.10 EcoCas3 nuclease activity in the presence of chaperon protein HtpG.**

(A, B). This shows the nuclease activity analysis of 800 nM EcoCas3 against 20 nM DNA fork, with additional htpG at 10 nM, 20 nM and 40 nM in lanes 3-5 and lanes 11-13. The reactions took place at 30 °C or 37 °C two hours in advance, before the endpoint reaction products were taken and analysed on 10 % acrylamide TBE native gel. Quantification data in Appendix 9.

## 4.9 Discussion.

Chapters 3 and 4 provide three biochemical analyses of Cas3 mechanism: a. Identifying function of the Trp-406/iHDA1 region in EcoCas3; b. Differential features of EcoCas3 and PaeCas2-3, revealing a likely redox-switch that contributes to the iHDR function

of *EcoCas3*; and c. Evidence for interaction of *EcoCas3* - *EcoCas1* that suppresses *EcoCas3* nuclease activity.

In this study, a regulatory structure of Cas3, referred to as the iHDA1 region, was identified as being necessary for the change in Cas3 nuclease activity. By reconstructing the relative position of Trp-230/406 and the Ic motif, the Trp-406 within the iHDA1 region was able to block/open the ssDNA tunnel (Chapter 3). This exerts control over the DNA accessing the nuclease active site and, thus, regulates Cas3 nuclease activity on DNA. At first, it was thought that the structural reorganisation of the iHDA1 region was stimulated only by temperature change and that this may reflect environmental control of the Cas3 function and the CRISPR-Cas systems. However, it was then found that the iHDA1 region may be impacted by other factors: for instance, proteins. Previous studies have shown that Cas3 is tightly restricted by Cascade (26,118), over Cas3 functioning on the elimination of MGEs. We showed that this extrinsic regulation could be overcome by mutating amino acids in and near the iHDA1 region, and could initiate robust Cas3 nuclease activity, leading to uncontrolled DNA degradation, as revealed in Chapter 3, Section 3.9. This effect suggests that an intact iHDA1 region is necessary for Cas3 to be governed by Cascade, though, unfortunately, it does not match with what has been observed where the Cas3 nuclease was repressed at 37 °C but not at 30 °C. Nevertheless, it can also be said that this conclusion elicited an idea of structural change of the

## L. He Chapter 4

iHDA1 which may have been stimulated by other factors rather than only by environmental temperatures.

When assuming the identity of probable interactors with the iHDA1 region that results in the rearrangement of its structure, Cas2 was primarily suspected. This speculation relied on published material concerning *PaeCas2-3* structure (PDB 5GQH, 5B7I) because the Cas2 fusion is relatively close to the iHDA1 in *PaeCas2-3* compared to Csy (crRNA-guided surveillance complex in I-F CRISPR). However, it was then proved that in *PaeCas2-3* fused Cas2, there is perhaps stabilisation of the iHDA1 region. Since the nuclease activity of *PaeCas2-3* is independent of temperature change, this suggests that this distinct Cas3 version does not undergo similar structural rearrangement as the temperature changes, unlike *EcoCas3* (Sections 4.2 and 4.3). In addition, by comparing the preference to DTT of *PaeCas2-3* and *EcoCas3* (Section 4.4), a cysteine redox switch in *EcoCas3* was identified, which is absent in *PaeCas2-3*. This discovery leads to another possibility that the regulation over *EcoCas3* nuclease activity may involve post-translational modification.

Interestingly, subsequent analysis deploying *EcoCas1*, which is known to be (70,76), suggests *EcoCas1* is likely to interact with the *EcoCas3* iHDA1 region and, consequently, suppress *EcoCas3* nuclease activity. At 30 °C, the *EcoCas3* nuclease active state is

inactivated in the presence of *EcoCas1*, at an approximate molar ratio of 1:1 of *EcoCas1* to *EcoCas3*, and higher (Figure 4.9 C). In addition, no inhibition of *EcoCas1* over the hyperactive *EcoCas3*<sup>W406A</sup> mutant was observed (Figure 4.9 D). These results imply the interaction between *EcoCas1* and *EcoCas3* occurs in the iHDA1 region, with a consequent stimulation of structural change which arises within this targeting district and causes a repression effect on *EcoCas3* nuclease function. This was sustained by the loss of inhibition of *EcoCas1* on the *EcoCas3* Trp-406 mutant, which carries a disrupted iHDA1 region. In contrast, *EcoCas2*, the close companion of *EcoCas1*, does not exert the same impact on the *EcoCas3* catalytic reaction on DNA. This indicates that *EcoCas1*, rather than *EcoCas2*, retains intrinsic interaction specificity for *EcoCas3* and, very likely, this is via the iHDA1 region. The importance of this conclusion will be discussed further.

*EcoCas3* nuclease activity was also examined alongside a chaperone protein, HtpG. Genetic studies (108,150) have suggested it is a key factor that helps maintain a functional intracellular level of *EcoCas3* protein to meet the requirements during CRISPR interference. In this study, for the first time, direct biochemical evidence has been proposed to prove that HtpG impacts *EcoCas3* DNA cleavage as well as assisting those *EcoCas3* proteins in a defective state at 37 °C to complete effective DNA degradation (Figure 4.10). This robust chaperone not only promotes the function

of wild-type *EcoCas3* as a nuclease but also increases DNA cleavage which has been generated by the hyperactive *EcoCas3*<sup>W406A</sup>. Though the mechanism is still unknown, it is speculated that HtpG may help with the restoration of partially miss-folded polypeptides in *EcoCas3*, which were caused by the intense buffer conditions during purification. Further study is required to uncover more details.

During CRISPR interference, Cas3 would be expected to efficiently degrade the MGEs which are fully or partially recognised by Cascade-crRNA (67,126,171). This would avoid invaders evading the CRISPR immunity by rapid translation of heterogeneous proteins and/or replication to preserve the polynucleotides from being undermined by the interference effectors of Cas3. However, the CRISPR-Cas system often adopts a distinct strategy of repressing Cas3 efficacy during CRISPR interference, especially when Cas1-2-mediated primed adaptation is involved. *In vivo* analysis has reported that decayed plasmid loss usually occurs in those individuals that carry ineffective PAM sequences or mutant protospacers (67,70,107,172). During this inefficient CRISPR interference, concurrently, the CRISPR-Cas system generates productive primed adaptation to update new spacers to host the CRISPR array (67). The consensus is that the defectivity of Cas3 during CRISPR interference may be caused by an imperfect Cascade R-loop formation on a mutant protospacer or a fully complementary protospacer with a disadvantaged PAM (126,173). However, this statement has been

challenged since some studies have proposed evidence for a different theory. Structural analysis has found Cas1-2 and Cas3 form a stable complex, referred to as PAC (primed adaptation complex) (70), and that this complex exists throughout CRISPR interference and in primed adaptation. Additional biochemical analysis using *PaeCas1* in the I-F subtype CRISPR has displayed a suppression regulation over *PaeCas2-3* nuclease activity via PAC formation, regardless of the existence of Cascade (Csy) (118). In this study, *EcoCas1* exerted repression on *EcoCas3* nuclease activity, consistent with that revealed in the I-F system. It was further identified that this regulatory mechanism may be involved in the physical interaction between *EcoCas1* and the iHDA1 region in *EcoCas3* (Section 4.7). These data indicate a possibility that a stagnant Cas3 in the PAC complex is required for CRISPR adaptation: through interacting with the iHDA1 region of Cas3, Cas1 can engage Cas3 to cut the substrate at an appropriate time or position.

Additionally, the Cas3 iHDA1 region itself is an intriguing protein construction. The discussions above are unlikely to be wholly representative. Considering its structural flexibility, alongside the evidence indicating that the iHDA1 region may interact with other proteins, for instance, Cas1 proposed in this study enables insight into the characteristics of the Cas3 iHDA1 region, which closely fit the description of the 'intrinsically disordered proteins or protein regions' (IDPs/IDPRs). IDPs/IDPRs represent proteins or regions

## L. He Chapter 4

lacking structural stability, or those that reveal an obvious transition from folding to unfolding states, and *vice versa*. They are distributed broadly as eukaryotes, archaea and bacteria, and they sustain biological activities, including protein-protein interaction, accommodating post-translational modification sites, stimulating chemical reactions and promoting protein phase separation (174). Although IDPs/IDPRs are unstructured, unlike conformational constrained protein motifs and domains, their contribution to biological activities is indispensable. Because of their mobility, they are often missed in electron density when analysing protein crystal structures, resulting in a lack of 15 amino acid residues (in Trypsinogen (175), PDB: 1TGB) and up to 273 residues missing (in Fibrinogen (174,176,177), PDB: 1M1J).

In the case of the Cas3 iHDA1 region, this study has proved that the structural malleability of this region impacts protein function from a biochemical perspective (Chapter 3 and 4). This complements the MD stimulation data generated by collaborators (167). Further analysis suggests the iHDA1 region may exert a role in interacting with other proteins; and accommodate a redox switch which tandoms the Trp-406 'gate' residue. Lacking electron density in the iHDA1 region can be found in existing structural data from I-E subtype *TfuCas3* (109) and I-F subtype *PaeCas2-3* (119). These Cas3 proteins exhibit the absence of structured segments at different levels. *TfuCas3* reveals residues missing in the iHDA1

region from Asp-384 to Gly-396 (PDB: 4QQX), whereas expanded signal loss between Arg-353 and Leu-407 occurs when it interacts with Cascade (PDB: 6C66). Clearly the *Tfu*Cascade does not contribute to stabilising the *Tfu*Cas3 iHDA1 region; what it does is the opposite. The same happens in *Pae*Cas2-3 (119), as electron density decreases from Ala-483 to Ala-502 (PDB: 5B7I). Despite the lack of a structural model of *Eco*Cas3, an emerging deep learning technique termed AlphaFold2 (168) has produced a predicted model that carries an unstructured segment of His-374 to Val-414, which resided in the iHDA1 region proposed in this study (Figure 3.1 A, Leu-372 to Lysine-413). This predicted model is consistent with the conclusions generated in this study; that the iHDA1 region may be an inherently flexible construction.

Regarding the Cysteines-403/404 redox switch spotted in the *Eco*Cas3 iHDA1 region, which tandems the Trp-406 'gate' residues. This discovery has led to the possibility that Cas3 DNA hydrolysis efficacy may be regulated by post-translational modification, rather than only by Cas proteins and protein chaperones. Though the redox switch is not a conserved feature among Cas3 proteins in various species, this could be explained if the iHDA1 region is considered to be an IDPR. From an evolutionary perspective, the IDPRs are much less conserved compared with those structured protein motifs and domains (178). The former's mutation rate is significantly higher than the latter's. The reason why the equivalent IDPRs vary in

## L. He Chapter 4

different species is still unknown; one possibility is that their plasticity may increase their tolerance to mutation (178). For the Cas3 iHDA1 region, it is not even conserved among the I-E subtype Cas3 proteins, as revealed in Chapter 3, Figure 3.1 D.

A comprehensive study of the iHDA1 region may be necessary to define its role in coordinating the Cas3 function and its involvement in CRISPR interference and primed adaptation.

## **Chapter 5**

# **Biochemical analysis of the Cas3 oligomeric state and novel protein interactions**

### **5.1 Summary for Part I**

This chapter describes observations for two functional forms of *EcoCas3*, a nuclease activated oligomeric state Cas3 and a nuclease-depleted monomeric Cas3, isolated unexpectedly. The original study purpose was to design a purification protocol for MBP-tagged Cas3, meanwhile trying to avoid the occurrence of instability of the fusion protein that resulted previously in protein degradation (24), and observed during the purification of the His<sub>X6</sub>-MBP-Cas3 used in Chapter 3 (Figure 3.2 B). A modified Cas3 purification protocol revealed that Cas3 proteins can exist in two states: the Cas3 oligomer is more abundant, with much less monomeric Cas3. Biochemical analysis of these two Cas3 species revealed distinct biochemical characteristics. One hypothetical explanation for this is that Cas3 oligomer is formed via liquid-liquid phase separation (LLPS). Proteins undergoing LLPS are often found to be intrinsically disordered. In the case of Cas3, the structurally flexible iHDA1 region (Chapter 3 and 4) may be an intrinsically disordered protein region (IDPR) and facilitate Cas3 forming LLPS. Bioinformatic analysis was carried out to investigate this hypothesis.

## **Part I**

### **Identifying Cas3 in oligomer and monomer states.**

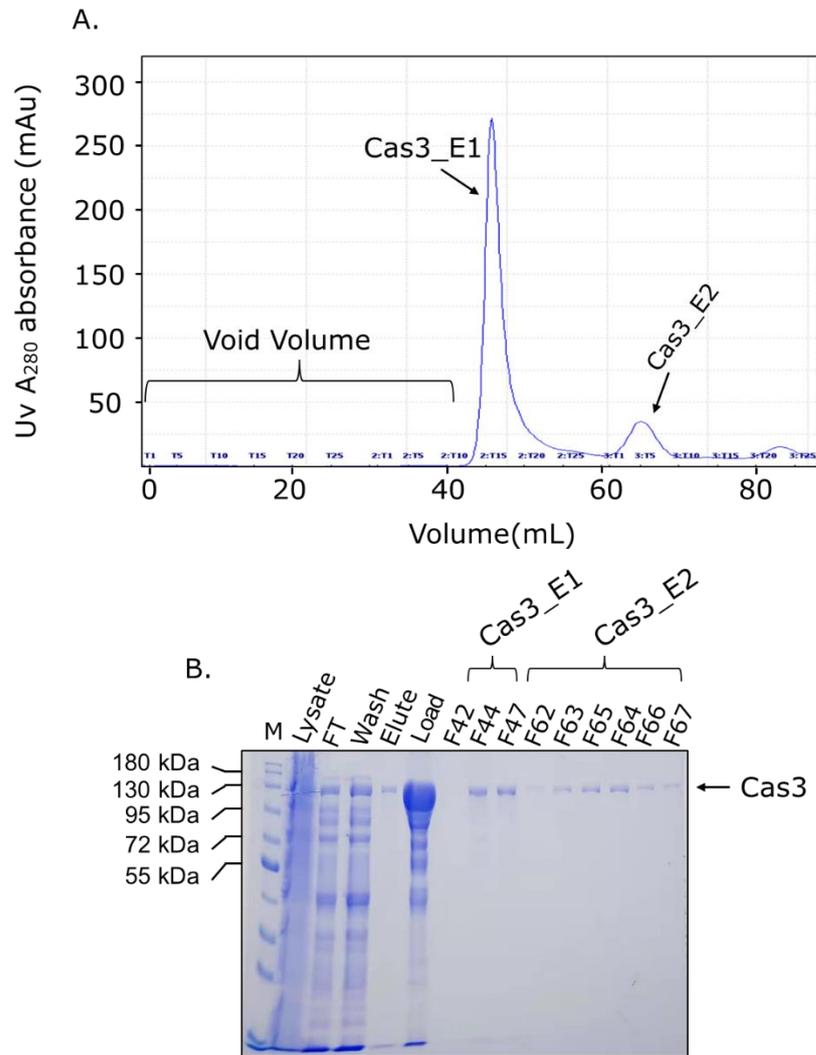
#### **5.2 Cas3 purified in two forms, distinguished in size exclusion chromatography.**

In Chapter 3, wild type Cas3 and its specific mutants were purified. However, those samples underwent massive protein degradation. To acquire more intact Cas3 and remove truncated proteins, a new purification protocol was designed and tested. This protocol utilised an initial affinity-chromatograph step, followed by a second step involving size exclusion chromatography (SEC). By applying SEC, the intact Cas3 proteins (fused to an N-terminus MBP-tag) could be separated from degraded products according to their distinct molecular weights, thus the MBP-Cas3 can be further purified.

This new purification protocol began by separating MBP-Cas3 (145 kDa) from cell lysate by affinity chromatography using amylose resin. The bound proteins were then processed by filtration through a 0.22  $\mu\text{m}$  filter before being loaded onto a pre-equilibrated HiLoad 16/600 Superdex 200 pg preparative SEC column. Eluted proteins were measured using UV absorbance at a fixed wavelength of 280 nM, with UV peaks corresponding to protein fractions collected. During SEC, two UV absorbance  $A_{280}$  peaks were observed. These

curves suggested that two batches of proteins with different molecular masses were eluted separately from the gel filtration column (Figure 5.1 A, Appendix 8).

The first protein sample eluted at 36.7-41.7 % wash volume was collected and denoted as Cas3\_E1 (Elution 1); this corresponded to a high molecular mass protein, protein complex or protein aggresome. The second protein sample eluted at 51.6-56.6 % wash volume and was termed Cas3\_E2 (Elution 2); this corresponded to the monomeric Cas3 (Appendix 8). Both Cas3\_E1 and Cas3\_E2 were over 95 % pure based on the validation by SDS PAGE, each with a single band at the size expected for MBP-Cas3 (Figure 5.1 B). Since the Cas3\_E1 eluted at an early stage during SEC, it may exist in the large protein complex formed by multiple Cas3 molecules, which are often considered as protein aggresome.



**Figure 5.1 Size exclusion chromatography (SEC) analysis showing Cas3 polymer (Cas3\_E1) and monomeric Cas3 (Cas3\_E2).**

(A). UV A<sub>280</sub> absorbance graph produced during protein purification showing two eluting Cas3 species. Cas3\_E1 is the Cas3 in a large complex state eluted earlier than the Cas3 in monomer state (Cas3\_E2) through the gel filtration column. (B) 10 % (v/v) SDS PAGE gel showing Cas3 proteins in different purification stages, including the Cas3\_E1 and Cas3\_E2. On the gel from left to right are exhibited cell lysate (Lysate), unbound proteins that flowed-through amylose resin (FT), unbound proteins that rinsed off during resin washing step (Wash), proteins eluted from amylose resin (Elute), concentrated Cas3 proteins that were loaded onto gel filtration column (Load) and Cas3 proteins eluted from gel filtration column. Cas3\_E1 was collected in fraction F44-47 and Cas3\_E2 existed in fraction F62-67, showing partially collected samples.

### **5.3 The Cas3 oligomer (Cas3\_E1) revealed features biochemically distinct from monomeric Cas3 (Cas3\_E2).**

Previous published data about *E. coli* Cas3 had tested only the monomer form (24,74,122). Those Cas3 proteins with a larger molecular mass than monomer Cas3 were considered as aggresome and precluded from biochemical analysis on the grounds that protein aggregates are unlikely to be functionally active. However, the nuclease assay using purified monomeric Cas3, as standard procedure after each Cas3 purification, revealed inactive nuclease (Figure 5.2, panel ii). This result prompted us to test nuclease activity for the Cas3 'aggresome/oligomer' fractions.

The nuclease activity of Cas3 oligomer (Cas3\_E1) was analysed using the nuclease assay condition developed in Chapter 3 and compared with the same assay carried out using the Cas3 monomer (Cas3\_E2). Increasing Cas3 concentrations were incubated against DNA fork (20nM), at 30 °C and 37 °C. The temperature-dependent Cas3 nuclease activities were exhibited in reactions containing Cas3 oligomer (Figure 5.2, Panel i, lanes 1-6), but not in those containing Cas3 monomer (Figure 5.2, Panel ii). This result suggested under the same reaction condition, the Cas3 oligomer and Cas3 monomer revealed distinct biochemical features.

## L. He Chapter 5

The same assay was repeated using the nuclease inactive Cas3 mutant Cas3<sup>D75G</sup> in an oligomeric state. Dramatically reduced DNA degradation was produced by Cas3<sup>D75G</sup> oligomer (Figure 5.3 A) compared to wild type Cas3 oligomer, therefore providing more convincing evidence that the wild-type Cas3 in oligomeric state is nuclease active, rather than formed by dysfunctional misfolded proteins.

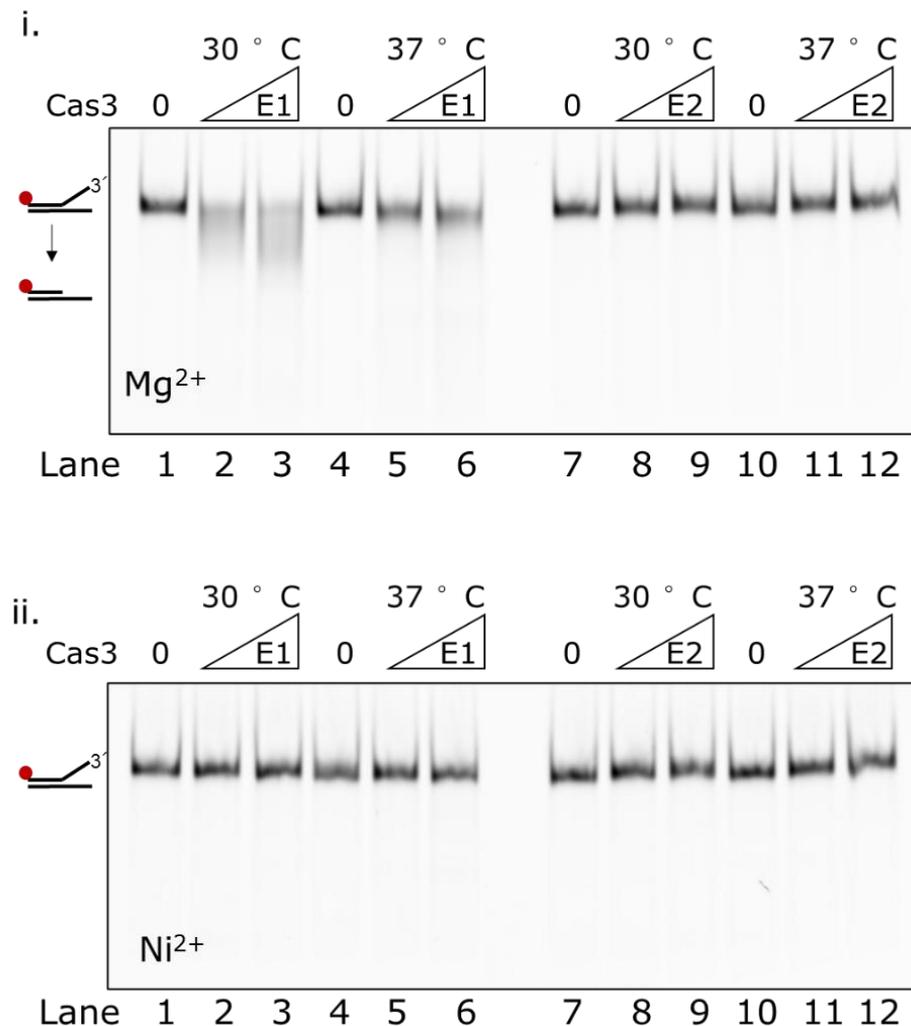
Another Cas3 mutant, Cas3<sup>D452A</sup>, was included in this part of the study. This mutant has an intact HD nuclease domain but carries a disrupted ATP catalytic site (Walker B motif). Unexpectedly, this mutant in an oligomeric polymer state was found to be nuclease inactive (Figure 5.3 B). This result could be attributed to the reduced DNA binding affinity of Cas3<sup>D452A</sup> (Figure 5.3 C) compared to the wild type Cas3 (Chapter 3, Section 3.7). One explanation of this phenomenon is that the conformation of the DNA tunnel situated within Cas3 was disturbed by introducing a mutation at its adjoining Walker B motif. This structural change disables Cas3 for binding and transferring DNA substrates to the HD nuclease site, thus generating a nuclease inactive outcome. Notably, when purifying Cas3<sup>D75G</sup> and Cas3<sup>D452A</sup>, only proteins in the oligomeric state were acquired, because no monomer mutants could be detected by either SEC or SDS PAGE.

## L. He Chapter 5

In contrast to the nuclease-active oligomeric Cas3, reactions with the addition of Cas3 monomer showed no apparent accumulation of degraded DNA products (Figure 5.2, Panel i, lanes 7-12). Since this monomeric Cas3 was a poor nuclease when against DNA fork substrates, its DNA cleaving capacity was then tested using 150  $\mu\text{M}$   $\text{Ni}^{2+}$  instead of  $\text{Mg}^{2+}$ . This change of metal ions used in the reaction was because nickel chloride was adopted in a previously published Cas3 nuclease assay (122) to stimulate DNA degradation by Cas3 monomer on M13ssDNA. However, use of  $\text{Ni}^{2+}$  had no stimulatory effect on Cas3 monomer nuclease activity (Figure 5.2, Panel ii). The same conclusion was obtained using Cas3 oligomer and  $\text{Ni}^{2+}$ , which has revealed that its nuclease activity was stimulated when incubated with  $\text{Mg}^{2+}$ . This may be attributed to the absence of DTT in reactions containing  $\text{Ni}^{2+}$ , while Cas3 requires DTT for nuclease activity (Chapter 4, Section 4.4). The removal of DTT is necessary because it could reduce the nickel ions, this will lead to nickel ions precipitate from reaction solution thus may have impact on protein function.

In general, these results suggested that Cas3 oligomer and Cas3 monomer have different DNA catalytic capacities when against DNA forks. The Cas3 oligomer that was thought to be the misfolded Cas3 in fact exhibited temperature-dependent DNA nuclease activity. In contrast, the monomer Cas3 did not exert DNA degradation on DNA fork substrates. Concerning the formation of the Cas3 in oligomer

state, from a protein-function perspective, this large protein complex was distinct from the typical protein aggregates. Unlike the inactive and insoluble protein oligomer, this aggregated Cas3 was revealed to be functionally active and soluble.

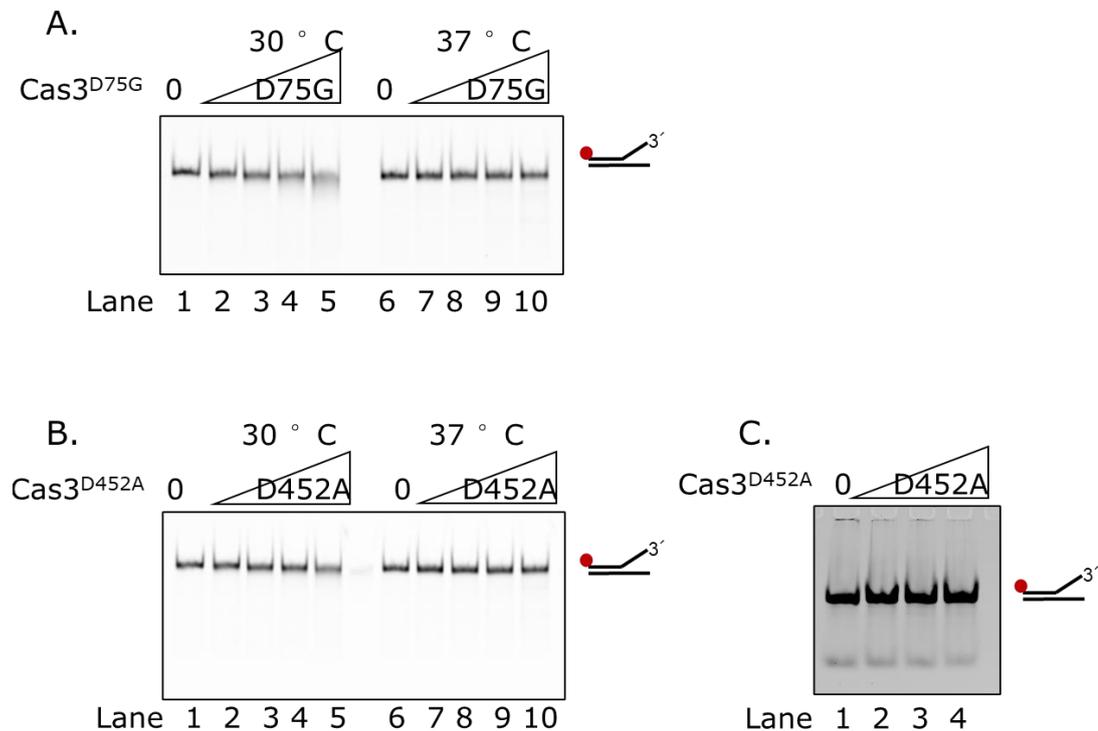


**Figure 5.2 Cas3\_E1 and Cas3\_E2 revealed distinct DNA hydrolysis capacity, which was stimulated by MgCl<sub>2</sub> rather than NiCl<sub>2</sub>.**

(i and ii) Showing nuclease assays using increasing concentrations of Cas3\_E1 and Cas3\_E2 against 20 nM DNA fork. Each Cas3 species of 0 nM, 200 nM and 400 nM were added in reactions, shown in Lanes 1-3, Lanes 4-6, Lanes 7-9 and Lanes 10-12. Reactions were carried out at 30°C or 37 °C as indicated, and endpoint products were analysed on 10 % (v/v) polyacrylamide TBE native gel. (i). The assay was carried out in a buffer containing 50 mM Tris-HCl pH7.5, 10 mM MgCl<sub>2</sub>, 100 mM NaCl,

L. He Chapter 5

0.1 mg/mL BSA and 20 mM DTT. (ii). The assay was carried out in a buffer containing 50 mM Tris-HCl pH7.5, 150  $\mu$ M NiCl<sub>2</sub>, 100 mM NaCl and 0.1 mg/mL BSA. Quantification data in Appendix 9.



**Figure 5.3 The Cas3<sup>D75G</sup> oligomer and Cas3<sup>D452A</sup> oligomer exhibited reduced nuclease activity on the DNA fork substrate.**

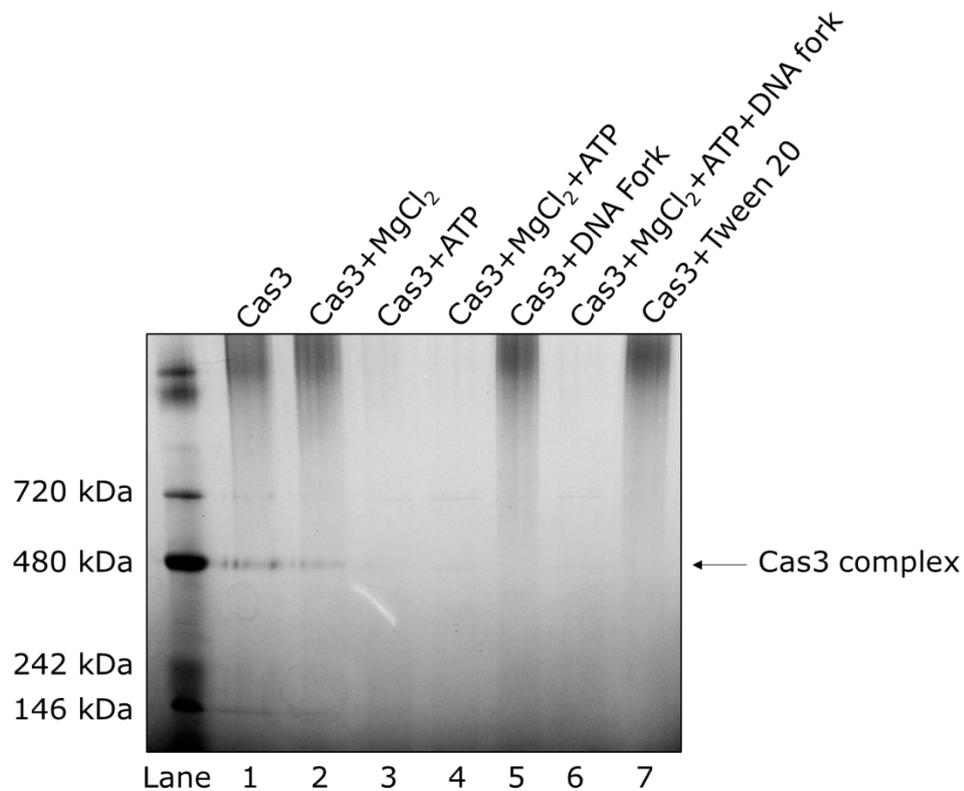
(A, B). Showing nuclease activity of Cas3 mutants Cas3<sup>D75G</sup> and Cas3<sup>D452A</sup> on 20 nM DNA fork, both mutants were in oligomeric state. Protein concentrations of 50 nM, 100 nM, 200 nM and 400 nM were used in Lanes 2-5 and Lanes 7-10. Reactions were carried out at 30 °C or 37 °C as indicated, and endpoint products were analysed on 10 % (v/v) polyacrylamide TBE native gel. (C). The EMSA assay shows no stable protein-DNA binding complex formed by Cas3<sup>D452A</sup> oligomer. 20 nM DNA forks were treated with Cas3<sup>D452A</sup> oligomer of 0, 150, 300 and 600 nM in Lanes 1-4. Quantification data in Appendix 9.

## **5.4 Different affinity tags fused to Cas3 have no impact on the formation of Cas3 oligomer.**

Is the formation of the MBP-Cas3 oligomer state an inherent tendency of Cas3, or is it due to the fusion of certain affinity tags? To answer this question, Cas3 attached with a relatively smaller His<sub>x6</sub>-tag (27 aa) at its N-terminus was investigated utilising Blue Native polyacrylamide gel electrophoresis (BN-PAGE) analysis. Purified Cas3 was compared alongside the same batch of Cas3 but was pre-incubated separately with known cofactors and binding targets of Cas3 such as DNA, Mg<sup>2+</sup> and ATP. These samples were prepared for the validation of whether the interaction between Cas3 and selected cofactors have any impact on the formation or dissociation of Cas3 oligomer.

Similar to MBP-Cas3, His<sub>x6</sub>-Cas3 (104.9 kDa) was observed to form a large complex with a proximal molecular mass of about 480 kDa (Figure 5.4). However, this Cas3 oligomer formation could be disrupted when protein was pre-incubated with ATP (Figure 5.4, Lanes 3, 4 and 6). The addition of selected factors that interact with Cas3, including DNA fork and MgCl<sub>2</sub>, had no impact on dissociating Cas3 from its oligomeric state. In addition, adding Tween 20, a detergent often used during protein purification to prevent protein precipitation, did not disassemble Cas3 oligomer. This result suggested the formation of Cas3 oligomer may be an inherent

feature of Cas3, one which can be reversed by adding ATP to dissociate the large complex formed by Cas3 proteins.



**Figure 5.4 Blue Native PAGE showing dissociation of His<sub>x6</sub>-Cas3 oligomer by adding ATP.**

Purified His<sub>x6</sub>-Cas3 on 3 to 12%(v/v) BN-PAGE gel, post-stained with Coomassie blue G-250. On the gel from left to right are shown 1 ug His<sub>x6</sub>-Cas3 in Lanes 1-7. In Lanes 2-7, Cas3 was pre-incubated with cofactors (Lanes 2-6) and detergent (Lane 7). MgCl<sub>2</sub> used in Lanes 2, 4 and 6 was 47 pmol. ATP used in Lanes 3, 4 and 6 was 47 pmol. DNA fork used in Lane 5 and Lane 6 was 94 pmol. Tween 20 was added to Cas3 with a final concentration of 0.2%(v/v) (Lane 7).

## 5.5 Bioinformatics analyses of potential intrinsically disordered protein regions within *E. coli* Cas3.

The atypical Cas3 oligomer remained confusing until a publication (179) was released discussing LLPS in *E. coli*, which attracted attention to LLPS formation in bacterial proteins. That study revealed

that the *E. coli* RNA polymerase adopts a strategy termed LLPS to enhance its synthetical function. Protein LLPS is a phenomenon of homogeneous solutions of proteins that separate into a dense phase and a dilute phase; the former is enriched in proteins in the liquid condensate format, whereas the latter lacks proteins (180). This protein LLPS used to be considered as the classic feature of eukaryotic proteins, but the emerging research on LLPS in bacteria suggests otherwise (174,178). Before protein LLPS was defined and proved to assist with biological processes, proteins undergoing LLPS were thought to form protein aggresomes with no biochemical function. It has now been revealed that several proteins adopt LLPS to enhance their inherent functions (179,181). A typical example is the nucleocapsid protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (182), which can adopt LLPS to infect human cells. By interacting with RNA, SARS-CoV-2 nucleocapsid protein forms a protein-RNA condensate, so it can effectively track RNA-binding proteins in human cells and consequently cause infection (182). Protein LLPS is a strategy bacteria cells have adopted to survive from the living pressure caused by multiple factors, for instance absence of ATP (183), whereby the addition of ATP to proteins in the LLPS stage could disassociate this protein-enriched phase (183-185).

A way to primarily distinguish a protein that may undergo LLPS is to analyse whether it is an *i*ntrinsically *d*isordered *p*rotein or carrying

any protein regions (IDP, IDPR) (174,178). IDP and IDPR represent proteins and polypeptides that are unable to maintain stable conformations and are often under structural transitions from folding to unfolding and *vice versa*. Owing to their malleability, IDP and IDPR fulfil the activity of folded protein domains, including their involvement during protein-protein interaction, accommodation of post-transcriptional modification sites and regulation function over functional motifs and domains (174,178).

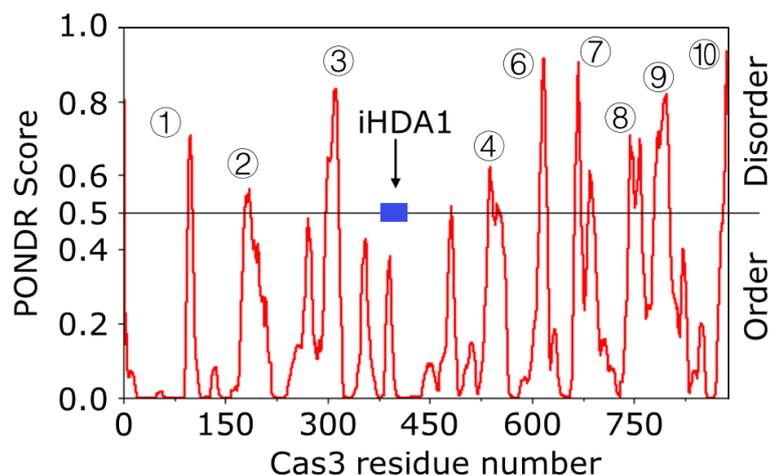
The roles that IDPR plays in biochemical processes and its structural plasticity are reminiscent of the iHDA1 region in Cas3. In Chapter 3, a comprehensive analysis of the Cas3 iHDA1 region suggested it is a structural mobility region and might participate in protein-protein interaction with Cas1. In addition, the iHDA1 also accommodates a post-transcriptional modification site: a redox switch formed by two adjacent cysteines (Chapter 4). Therefore, the iHDA1 region may be an IDPR, which could facilitate the formation of protein LLPS. Based on this, it is hypothesised here that the nuclease-active Cas3 oligomer is the Cas3 in the LLPS stage.

It is possible for a protein to have more than one IDPR within its structure. Apart from the iHDA1 region, which was likely intrinsically disordered as discussed in Chapter 3, it was necessary to validate whether *E. coli* Cas3 contains any other IDPRs. To this end, the amino acid sequence of Cas3 was analysed using the online

prediction programme PONDR VL-XT (186,187). This algorithm was designed to analyse amino acids in a given protein sequence and determine how likely it is that they are intrinsically disordered and structured. By tracking the distribution of a few amino acids preferentially enriched in known IDP/IDPRs, this predictor could score polypeptides based on previous studies on proteins carrying IDPRs to help diagnose protein regions in the target protein, which tend to be intrinsically disordered. Notably, this predictor utilises the eukaryotic proteins database to compare with inputted proteins, so it may not be ideal for applying to protein from prokaryotes, such as the *E. coli* Cas3 this study focused on. Therefore, results generated by PONDR VL-XT were cross-compared with results from the deep-learning algorithm AlphaFold.

*E. coli* Cas3 was diagnosed with an overall disorder percentage of 14.08%, with a total of ten disordered regions of >5 amino acids identified (Figure 5.5, Table 5.1). Most were located on the solvent-exposed surface of Cas3, indicating those regions may be the protentional interaction sites of Cas3 with other factors (e.g., proteins). However, two of the polypeptide sequences highlighted by PONDR VL-XT, amino acids 298-317 (Region ③) and 684-691 (Region ⑦) are both internalised within Cas3. Region ③ has no known function, but it is adjacent to a crucial amino acid Lysine-320, which plays an important role in R-loop formation by Cas3 revealed via *in vitro* analysis (24); and spacer acquisition from downstream

of the protospacer during primed adaptation, suggested by *in vivo* study (unpublished, conference poster) conducted by Nikita Vaulin from Peter the Great St. Petersburg Polytechnic University. Therefore, Region ③ could be involved in multiple unrevealed biochemical mechanisms that may be crucial for linking together the CRISPR interference and primed adaptation. Region ⑦ overlapped the Cas3 Q/RGR motif (665-674aa), which is a structured region. This bias could be explained because the Q/RGR motif consists of multiple arginine amino acids, so its amino acid composition is similar to a structural disordered motif distributed in eukaryotic proteins termed Arginine Rich Motif (ARM). Surprisingly, the iHDA1 region was not predicted as IDPR by PONDR VL-XT (186), in contrast to the result produced by AlphaFold. This discrepancy may be related to the distinction of protein databases used to train these two programmes.



**Figure 5.5 The possible IDPRs in *E. coli* Cas3 predicted by PONDR VL-XT.**

The PONDR score value 1.0 denoted disordered segments, while 0.0 represented ordered sequence. The score represents the tendency of each amino acid in Cas3

*protein to be intrinsically disordered, shown as red curves. The location of the Cas3 iHDA1 region is labelled in blue.*

**Table 9. Disordered regions predicted by PONDR VL-XT.**

Region Number	Predicted disorder segment (>5 amino acids)	Domains
①	96-101	HD
②	179-187	HD
③	298-317	RecA1
④	536-542	RecA2
⑤	612-622	RecA2
⑥	664-673	RecA2
⑦	684-691	RecA2
⑧	742-762	Linker sequence
⑨	781-802	CTD
⑩	882-888	CTD

*Showing amino acid numbers and residential domains of each possible IDPRs in Cas3, the prediction was carried out using PONDR VL-XT*

## 5.6 Discussion for Part I.

This study revisited the Cas3 oligomer that could be acquired during a standard Cas3 overexpression and purification but had been precluded in previous biochemical analyses of Cas3. By analysing the nuclease activity of the Cas3 oligomer and comparing it with that of the Cas3 monomer, the former revealed a temperature-dependent nuclease activity on DNA fork substrate, whereas the Cas3 monomer did not generate any DNA cleavage products. This result indicates the Cas3 oligomer may not be due to typical protein degradations.

## L. He Chapter 5

The nuclease assay using two states of Cas3 was repeated using nickel ions instead of magnesium. Nickel was adopted to assist with Cas3 DNA catalytic activity as published in Mulepati, S., *et al.* (122) and Loeff, L., *et al.* (74). Interestingly, no evidence suggested Cas3 oligomer and monomer were able to exert nuclease activity on DNA fork in the presence of nickel. This may be attributed to the DNA fork substrates used in the nuclease assay in this study. The fork substrates were much smaller than the M13ssDNA used in the Cas3 nuclease assay conducted by Mulepati, S., *et al.* and Loeff, L., *et al.* (74,122).

The discovery of Cas3 forms oligomer led to the possibility that this large soluble Cas3 complex may be the Cas3 in LLPS, and the formation of the enriched Cas3 phase may be driven by the iHDA1 region that is possibly the IDPR in Cas3.

Bioinformatic analysis was introduced in this part of the study to assist with identifying potential IDPR in Cas3. The protein sequence of *E. coli* Cas3 was analysed by an IDP/IDPR online predicting programme, PONDR VL-XT (186,187). In general, Cas3 was suggested by the PONDR VL-XT to have ten regions predicted to be unstructured, with a 14.08 % overall disorder percentage. However, the iHDA1 region identified in Chapter 3 that fitted the profile of an IDPR was not highlighted by the PONDR VL-XT. The emerging protein structure predicting algorithm AlphaFold gave an opposing

result; it suggested the iHDA1 region was likely to be disordered because of its low per-residue confidence score. This bias could be attributed to the different protein databases that PONDR VL-XT and AlphaFold use (168). The former was designed to compare new protein sequences with known eukaryotic proteins, while the latter uses a broader protein database including eukaryotic and prokaryotic proteins. Additional studies would offer insight into this topic.

## **Part II**

### **Cas3 in oligomer state reveals novel protein interactions**

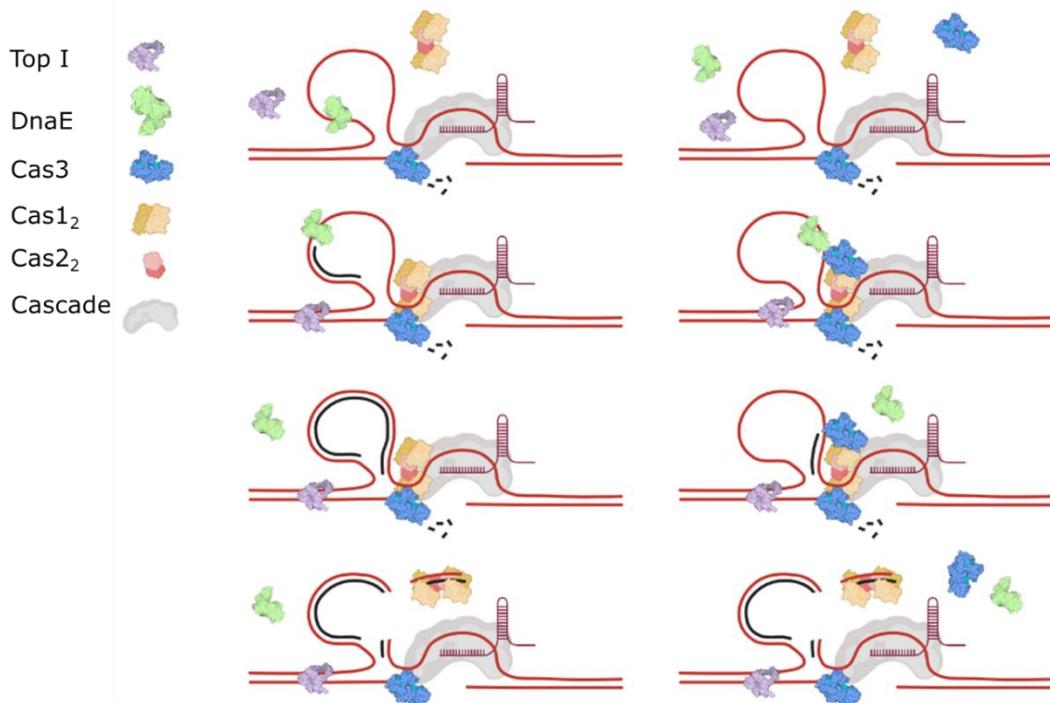
#### **5.7 Validation of DNA polymerase III alpha and DNA polymerase I activities in CRISPR-Cas primed adaptation.**

Type I CRISPR systems utilize Cascade (that is CasA-E bound to crRNA) and Cas3 to form an interference R-loop with MGE DNA leading to DNA degradation. Cas3 is recruited by Cascade and loaded onto the non-targeting strand where it reels DNA for cleaving at least 10 kb (70,122,127).

However, this model leaves an open question: How does the nuclease activity of Cas3 result in DNA capture and integration required of primed adaptation? I considered this regarding DNA topoisomerases and DNA polymerases, based on unpublished data produced by Dr Jamieson Howard and Dr Edward Bolt. Their data indicating that both DNA topoisomerases and DNA polymerases physically interact with CRISPR Cas3. Topoisomerases can relieve DNA tension ahead of helicase-DNA fork unwinding complexes in many contexts (188). The DNA polymerase catalytic core protein DnaE/Pol alpha/Pol III core (herein called DnaE) was identified to interact physically with Cas3 via a comprehensive protein interaction study using over 4000 *E. coli* proteins (151). Because no further analysis was carried out the significance of this interaction is unknown. Both DnaE and topoisomerase enzymes are essential for DNA replication in *E. coli*. A hypothesis is proposed that interaction of Cas3 with DnaE may be part of Cas3-DnaE coupled interference-replication that facilitates primed adaptation by generating nascent duplex DNA from Cas3 nuclease DNA products. In the model (Figure 5.6), Cas3 recruits DnaE to an MGE recognised by Cascade-crRNA and then facilitates DNA synthesis using the targeting strand as a template. Those synthesised DNA fragments are then protospacers and pre-spacers for Cas1-Cas2 catalysed adaptation.

In this scenario Cas1-2 could participate in two ways. The first potential assembly involves Cas1, Cas2 and Cas3 in a 4:2:1 ratio

(Figure 5.6, left). The second describes a Cas1-Cas2-Cas3 complex in a 4:2:2 ratio (Figure 5.6, right). The latter was considered because of the existing Cas1-Cas2-3 model in the I-F CRISPR-Cas system, which has suggested Cas2-3 assembled with Cas1 in a 2:4 ratio (118,120).

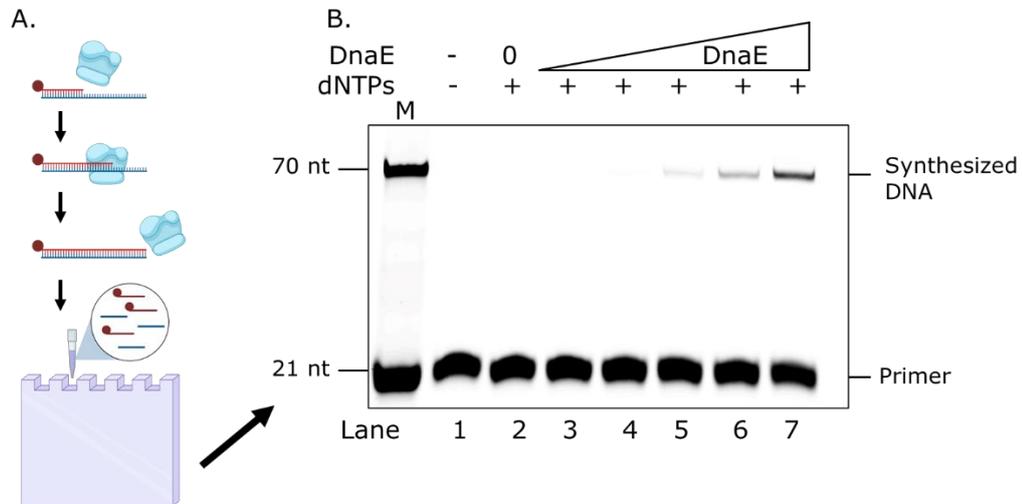


**Figure 5.6 A model of CRISPR interference in which Cas3 recruits DnaE to synthesise DNA fragments, which are then processed by Cas1-2 into a pre-spacer.**

*During CRISPR interference, the Cas3-Cascade complex anchors to the protospacer sequence on MGEs via R-loop formation. When Cas3 exerts DNA unwinding and degradation activities on the non-targeting strand, this process will lead to the accumulation of ssDNA on the targeting strand. By interacting with Cas3, DnaE may utilise the available ssDNA as a template to synthesise short DNA fragments (black). This self-nonspecific DNA hybrid (red-black) is then processed by Cas1-2 as a pre-spacer. At the same time, Top I may assist with Cas3 unwinding.*

## **5.8 A primer extension assay to analyse DNA synthesis by DnaE.**

The model proposed above involves the possible coordination of four different proteins. First, DnaE activity was analysed alongside Cas3, to determine if Cas3 has any impact on DNA synthesis by DnaE. For this purpose, a primer extension assay was adopted for DnaE. The substrate used was a partial duplex DNA comprising two DNA oligonucleotides with different lengths: A 70 nt DNA oligo (ELB41) was annealed with a 21 nt DNA (ELB40Primer), which was a 21 nt DNA oligo with a Cy5 labelling its 5' end (Figure 5.7 A). In theory, when incubating DnaE or any other DNA polymerases with ELB40Primer/41 (ELB40P/41) substrate in the presence of dNTPs, DnaE can elongate the Cy5 labelled primer while using the unlabelled DNA strand as a template. Any primer extension that has occurred could be detected by denaturing polyacrylamide TBE gel electrophoresis and shown as DNA products that are longer than the original primer fragments (Figure 5.7 A). This developed assay was first tested using an increasing amount of DnaE on 10 nM ELB40P/41 substrate, with additional 200  $\mu$ M dNTPs (Figure 5.7 B). Increased accumulation of synthesised DNA was concurrent with the increase in DnaE concentration added to the reactions (Figure 5.7 B, Lanes 3-7). This result suggested that DnaE exerted its DNA synthetic function on the designed ELB40P/41 substrate.



**Figure 5.7 DnaE exhibited DNA synthetic activity on ELB40P/41 substrate, as revealed by primer extension assay.**

(A) A diagram showing how the primer extension assay is carried out in this study. The primer oligo in the DNA substrate designed for doing primer extension is 21 nt (red) with a Cy5 labelled 5' end (red dot). It is pre-annealed to a 70 nt template DNA (blue). After incubating the substrate with DNA polymerase, the endpoint products are denatured and analysed using 10 % polyacrylamide denaturing TBE gel. The sample processing methods are discussed in Chapter 2, Section 2.6 and 2.8.2. (B) A primer extension assay using 10 nM ELB40P/41 substrate incubates with an increasing amount of DnaE, of 5 nM, 10 nM, 25 nM, 50 nM and 100 nM in Lanes 3-7. Reactions were incubated at 37 °C for 30 minutes. Endpoint products were analysed using 10 % denaturing acrylamide TBE gel. Primer extension was taken in a buffer containing 10 mM Magnesium Acetate, 40 mM HEPES-NaOH pH 8.0, 0.1 mg/mL BSA and 200 mM dNTPs. Unless mentioned otherwise, primer extension assays mentioned in the following content used the same buffer condition. Quantification data in Appendix 9.

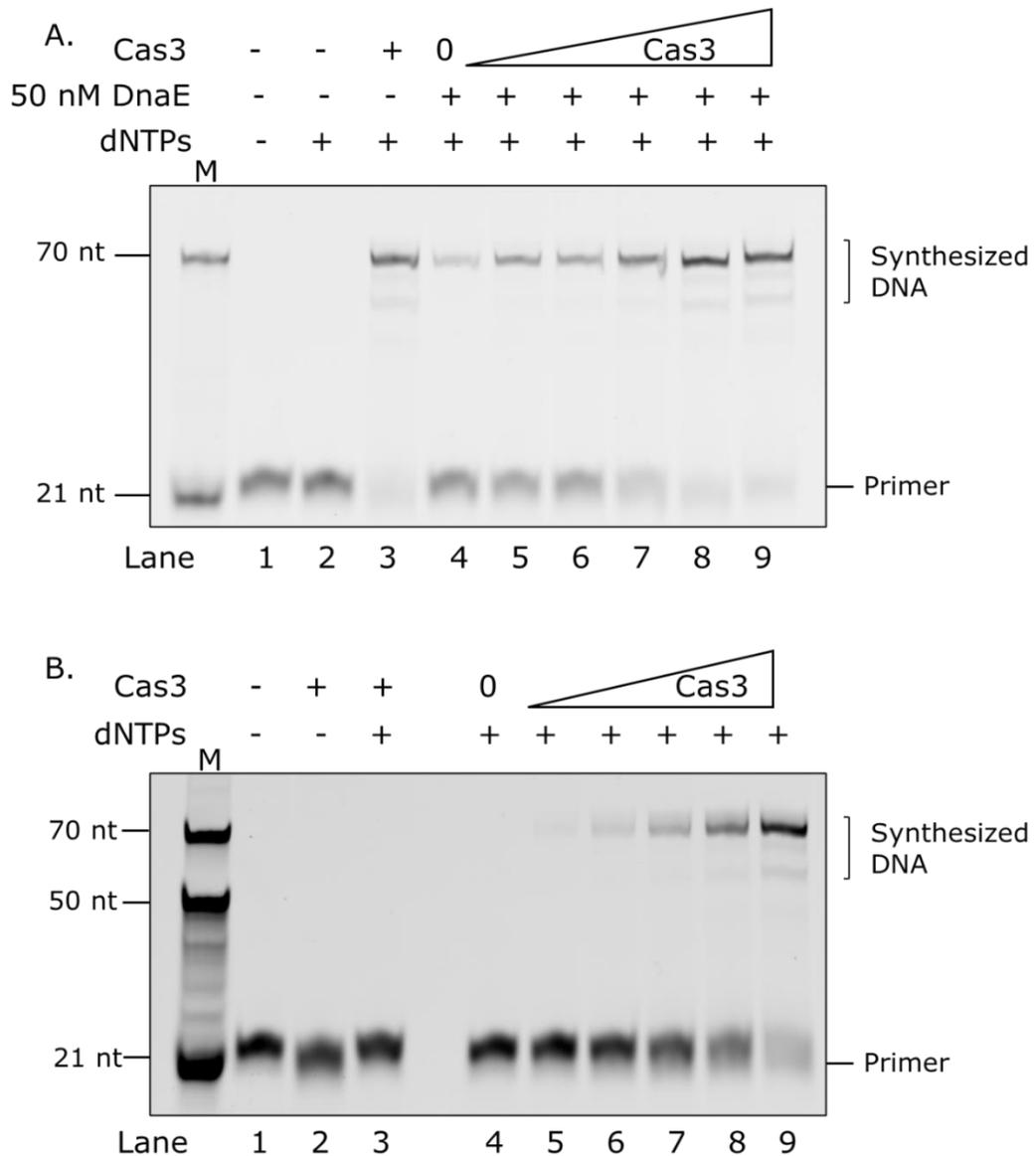
### 5.9 DNA polymerase I co-purified with Cas3.

Since DnaE could extend DNA substrate (ELB40P/41), the next test repeated DnaE primer extension assay with the addition of Cas3 to gain an initial assessment of whether Cas3 impacts DnaE function. To this end, 50 nM DnaE was incubated against 10 nM ELB40P/41

## L. He Chapter 5

with an increasing concentration of Cas3. Unless mentioned otherwise, all Cas3 proteins used in primer extension assays presented in this part of study were the MBP-Cas3 monomer (Section 5.2).

In the course of this assay, unexpected primer elongation was observed in the Cas3-only control reaction (Figure 5.8 A, Lane 3). The same phenomenon was then observed repeatedly, as exhibited in Figure 5.8 B of a primer extension assay using an increasing amount of Cas3 as the only protein source. The accumulation of extended primers was concurrent with the increasing amount of Cas3 added to reactions. This result suggested that Cas3 or trace amounts of DNA polymerases in purified Cas3 preparation carried out DNA synthesis.



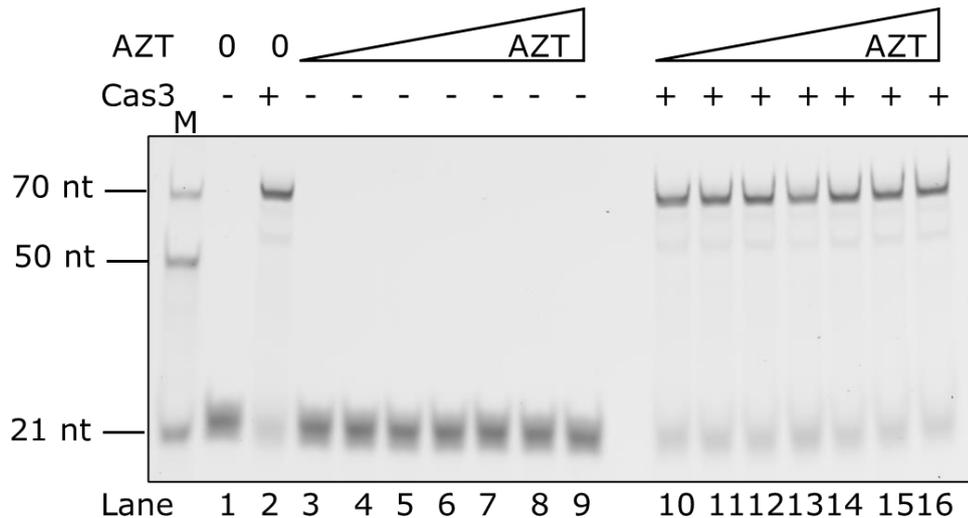
**Figure 5.8 Primer extension occurred in samples containing Cas3.**

(A). When analysing the impact of Cas3 on DnaE activity, extended primers were generated in a sample that contained Cas3 but not DnaE. This control reaction is shown in Lane 3. In Lanes 5-9, increasing concentrations of Cas3 from 10 nM, 25 nM, 50 nM, 100 nM and 200 nM were added to reactions, each containing 50 nM DnaE and 200 mM dNTPs. Reactions were carried out at 37 °C for 30 minutes. Endpoint products were analysed using 10 % denaturing acrylamide TBE gel. (B). Showing DNA synthesis was co-occurrent with adding purified Cas3. The primer extension assay showed in (A) was repeated but only used an increasing amount of Cas3 in each reaction, without any DnaE. The sample in Lane 2 contained 200 nM Cas3 in a buffer condition omitting dNTPs. Quantification data in Appendix 9.

## L. He Chapter 5

The absence of known polymerase motifs or domains in Cas3 led to the only possibility that the observed Cas3-dependent primer elongation occurring was attributable to co-purifying traces of one or more DNA polymerases (Pols). Six DNA polymerases are known in *E. coli*: five DNA-dependent polymerases Pol I, Pol II, Pol III catalytic subunit DnaE, Pol IV, Pol V subunit UmuD; and an RNA-dependent polymerase, also known as reverse transcriptase, Retron EC67 (189,190). Since DnaE was previously suggested to interact with Cas3 (151), it would be the most likely candidate.

Retron EC67 is the easiest to diagnose because its DNA synthetic activity could be inhibited by an antiviral compound Azidothymidine (AZT), which specifically targets reverse transcriptases and stalls their activity. Including AZT in reactions containing Cas3 solution and carrying out a primer extension test could reveal whether the contaminated polymerase is Retron EC67 or not. As shown in Figure 5.9, AZT did not inhibit primer extension in reactions (10 uL) containing 2 uL purified Cas3 (Figure 5.9, Lanes 10-16). This result suggested that polymerase contamination in purified Cas3 was not due to reverse transcriptase, precluding Retron EC67 from listed polymerases.



**Figure 5.9 Reverse transcriptase inhibitor AZT did not block DNA synthetic activity generated by the DNA polymerase co-purified with Cas3.**

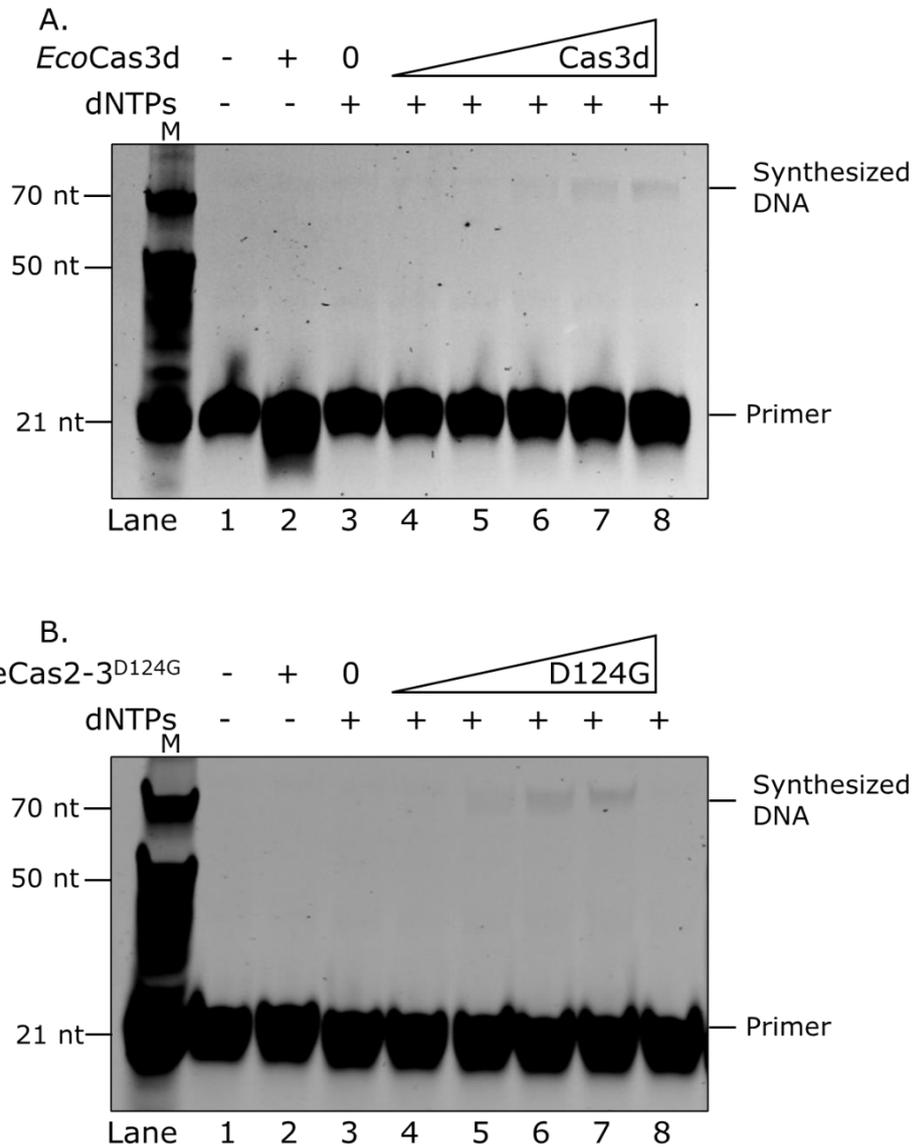
Primer extension reactions using DNA Pol contaminated Cas3 were treated with an increasing amount of AZT, from 5 nM, 50 nM, 250 nM, 500 nM, 1 uM, 2 uM and 5 uM in Lanes 3-9 and Lanes 10-16. Reactions were incubated at 37 °C for 30 minutes before adding a stop solution containing proteinase K and SDS. Endpoint products were analysed using 10% denaturing acrylamide TBE gel. Quantification data in Appendix 9.

A primer extension assay is not capable of distinguishing the rest of the DNA polymerases that may co-purify with Cas3. In this circumstance, the most effective way to identify the contaminating polymerase is mass spectrometry. Therefore, the purified Cas3 was processed (method see Chapter 2, Section 2.13) and submitted to the Cambridge Centre for Proteomics (University of Cambridge) for liquid chromatography-mass spectrometry (LC-MS) analysis. Unexpectedly, the LC-MS result (see Appendix 7) suggested *E. coli* polymerase I (Pol I) is the only DNA polymerase diagnosed in

purified Cas3. Pol I had not been reported before to form any interaction with Cas3; however, it is known to be a factor required for CRISPR adaptation, probably for its gap-filling reactions in pre-spacer DNA capture and/or protospacer integration (10). Although DnaE was not observed in the submitted Cas3 sample, it had previously been shown to interact with Cas3 (151). One explanation for this is that the Cas3 (88 kDa) used in this work is fused to a relatively large MBP tag (43 kDa), which possibly disturbs the protein interaction between Cas3 and DnaE. Because MBP-Cas3 (143 kDa) and DnaE (130 kDa) have a similar mass molar, the interaction between them cannot be tested via a protein pull-down assay.

As the further investigation was based on the LC-MS result, the prevalence of the Pol I-Cas3 co-purification was then conducted to get a sense of whether it is a consistent effect and not limited to one batch of purified Cas3: The former may suggest a meaningful physical interaction between Cas3 and Pol I; the latter would not support this. To probe this idea, I returned to previously purified *EcoCas3* and *PaeCas2-3* samples to validate if they were contaminated by polymerases. The inclusion of *PaeCas2-3* proteins in this analysis was important because of its very different purification strategy to the *E. coli* MBP-Cas3, and it is highly homologous to *E. coli* Cas3, but its fusion with Cas2 may influence any interaction with polymerase I.

Interestingly, various samples of *EcoCas3* (Chapter 3 and 4) and *PaeCas2-3*<sup>D124G</sup> (Chapter 4) were observed to contain polymerase activity (Figure 5.10, A and B). In the reaction containing only Cas3 without dNTPs, shortened primers were produced (Figure 5.10 A, Lane 2). This result represented 5' to 3' endonuclease activities on ELB40P/41 substrates and a high likelihood of production by polymerases in the Cas3 sample rather than Cas3. For primer extension analysis using *PaeCas2-3*, the addition of 200 nM *PaeCas2-3* inhibited DNA synthesis by a contaminated polymerase. This could be explained by the high binding affinity of *PaeCas2-3* to DNA. Excessive *PaeCas2-3* (200 nM) may titrate free DNA substrates (10 nM) via protein-DNA, binding and consequently blocking polymerases from accessing DNA substrates. This was discussed in Chapter 4, Section 4.5. Altogether, polymerase contaminations existed in different batches of purified *EcoCas3* and *PaeCas2-3* that were purified using different methods. This result suggests that co-purification of Cas3 with polymerase may be a significant and physiologically relevant protein-protein interaction. Apart from that, the Pol Is in *E. coli* and *P. ae* are highly conserved, so the Pol I-Cas3 interaction might exist between *PaePol I* and *PaeCas2-3*. This interaction may be a conserved feature in all Cas3 subtypes.

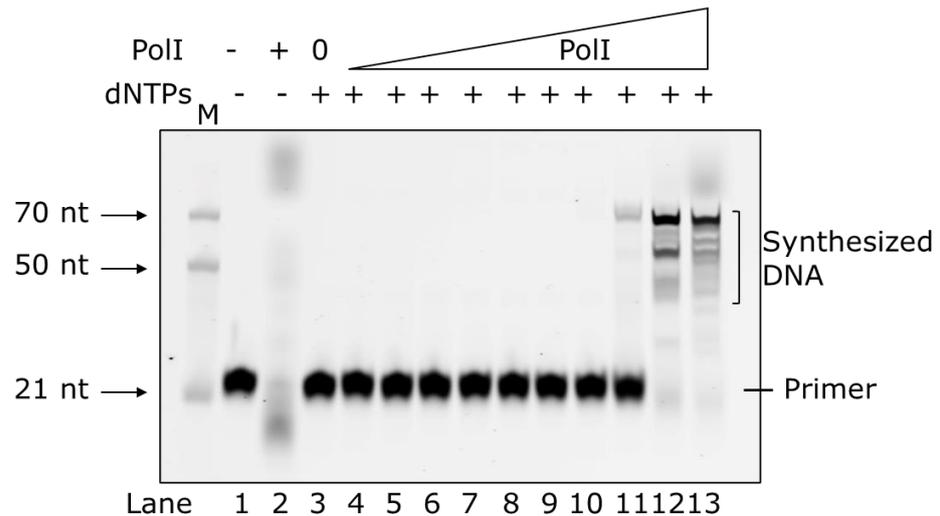


**Figure 5.10 DNA polymerase was co-purified with *EcoCas3* and *PaeCas2-3*.**

(A and B) Primer extension assays showing DNA polymerase contamination existed in *EcoCas3* polymer (*Cas3d*) and *PaeCas2-3<sup>D124G</sup>*. Increasing concentrations of *EcoCas3* and *PaeCas2-3<sup>D124G</sup>* from 10 nM, 25 nM, 50 nM, 100 nM and 200 nM were examined, shown in Lanes 4-8. Reactions were carried out at 37 °C for 30 minutes before adding a stop solution containing proteinase K and SDS. Endpoint products were analysed using 10 % denaturing acrylamide TBE gel. Quantification data in Appendix 9.

### **5.10 Removal of Pol I during Cas3 purification.**

We needed Cas3 without Pol I contamination to be able to unequivocally assess Cas3 with polymerisation from DnaE. To assist with this, we first tested the minimal amount of Pol I that could be diagnosed by doing a primer extension assay. An increasing amount of Pol I from 0.6 up to 60 nM was incubated with 10 nM ELB40P/41 in the presence of dNTPs and Mg<sup>2+</sup>. The endpoint products were then analysed on a 10 % denaturing TBE polyacrylamide gel (Figure 5.11). The increased accumulation of synthesised DNA was concurrent with the increase in Pol I activity units added to each reaction (Figure 5.11, Lanes 10-13). Extended primers could be detected in reactions with a minimum 0.6 nM Pol I added to a 10 uL reaction (Figure 5.11, Lane 11). Meanwhile, in the absence of dNTPs, Pol I exhibited 5' to 3' exonuclease activity and generated DNA degradation products that migrated much slower than the Cy5 labelled primer (Figure 5.6 B, Lane 2). This occurrence of anomalous migration was because the positively charged Cy5 label could not sustain electrophoretic migration of DNA oligos less than 8 nt (191). In general, these results suggested that the primer extension assay was capable of detecting at least 6 femtomoles of Pol I present in the reaction solution, feasible for detecting Pol I in purified Cas3.



**Figure 5.11 Primer extension analysis could detect a minimum of 0.6 nM Pol I added to the reaction.**

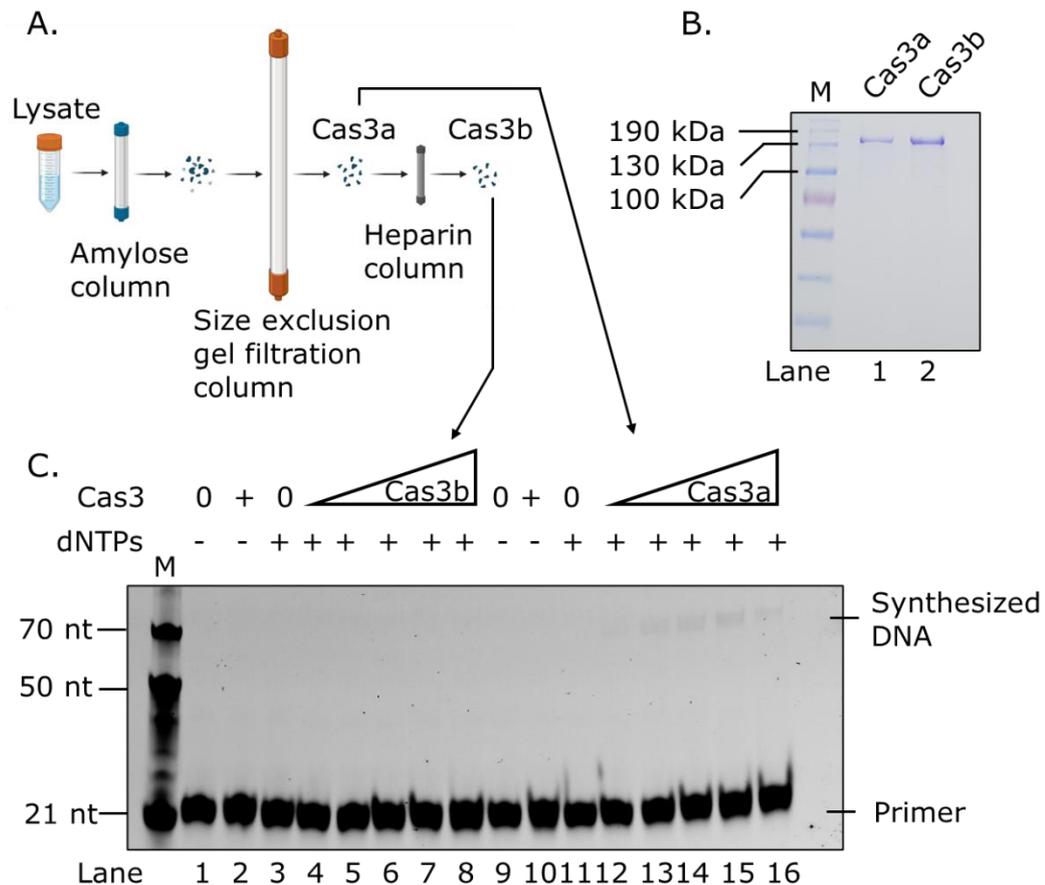
Increasing concentrations of Pol I from 0.6 fM to 6 fM, 60 fM, 600 fM, 6 pM, 60 pM, 600 pM, 0.6 nM, 6 nM and 60 nM were treated to 10 nM ELB40P/41, shown in Lanes 4-13. Reactions were incubated at 37 °C for 30 minutes before adding a stop solution containing proteinase K and SDS. Endpoint products were analysed using 10 % denaturing acrylamide TBE gel. Lane 2 contained 60 nM Pol I but no dNTPs. Quantification data in Appendix 9.

To obtain Cas3 protein without Pol I, the current purification protocol was optimised via two strategies. The first is increasing NaCl in purification buffers from 200 mM to 500 mM, which may potentially reduce electrostatic interactions between Cas3 and any contaminating proteins during amylose affinity purification. The second is the inclusion of a heparin column in the purification protocol, to separate any DNA binding protein (e.g., Pol I) from Cas3. Though Cas3 is a DNA binding protein, it does not bind to the heparin column during purification (Chapter 2, Section 2.3.1) (24).

## L. He Chapter 5

During a purification optimisation test, two Cas3 samples were taken at different elution stages (Figure 5.12 A) to trace the remaining Pol I in the purified sample. The Cas3a was the Cas3 monomer eluted from HiLoad 16/600 Superdex 200 pg gel filtration column. Half of the Cas3a sample was then dialysed into a buffer containing 100 nM NaCl and then loaded onto a heparin column. The Cas3 protein collected in flow-through was denoted as Cas3b.

Primer extension assays for Cas3a and Cas3b were used to detect the presence of Pol I in each stage of the Cas3 purification. Synthesised DNA products could be detected in reactions containing increasing amounts of Cas3a solution (Figure 5.11 C, Lanes 13-16). However, no detectable DNA replication occurred in reactions containing Cas3b (Figure 5.11 C, Lanes 4-8). These results suggested that Pol I still co-purified with Cas3 in a high salt buffer condition. It could be separated from Cas3 by exploiting the heparin column, to which Pol I bound, but Cas3 did not.



**Figure 5.12 Separating Pol I from Cas3.**

(A). Schematic diagram of optimised EcoCas3 purification and protein samples that were collected in different purification stages. Proteins were purified using the amylose column, HiLoad 16/600 Superdex 200 pg preparative SEC column and heparin column. Monomer Cas3 acquired at different purification stages was taken to test the remaining Pol I by doing primer extension assay, samples were denoted as Cas3a and Cas3b, and they were both dialysed into a buffer containing 20 mM Tris pH 7.5, 200 nM NaCl and 30 % glycerol for storage purpose. (B). Cas3a (Lane 1) and Cas3b (Lane 2) on a 10 % SDS page gel, post-stained with Coomassie blue. The migration of protein fits the MBP-Cas3 protein mass of 143 kDa. (C). Co-purified Pol I was diagnosed in Cas3a but not in Cas3b. Increasing amounts of Cas3a of 0.1  $\mu$ L, 0.25  $\mu$ L, 0.5  $\mu$ L, 1  $\mu$ L, 2  $\mu$ L were added to reactions, each of which is 10  $\mu$ L in total, shown in Lanes 4-8. Cas3b of 0.23  $\mu$ L, 0.6  $\mu$ L, 1.2  $\mu$ L, 2.3  $\mu$ L and 4.6  $\mu$ L were added to reactions to make up a 10  $\mu$ L sample, shown in Lanes 12-16. Lanes 2 and 10 show 2  $\mu$ L Cas3a in a 10  $\mu$ L reaction and 4.6  $\mu$ L of Cas3b added to 10  $\mu$ L reaction. All samples were incubated at 37°C for one hour. These reactions were stopped by adding a stop solution containing proteinase K and SDS. DNA products were then mixed with Orange G loading dye (78 % Formamide, 22 %

*Glycerol and Orange G) before analysis on a 10 % acrylamide denaturing TBE gel. Quantification data in Appendix 9.*

### **5.11 Less DNA elongation is produced by DnaE when it is incubated with an excessive amount of Cas3.**

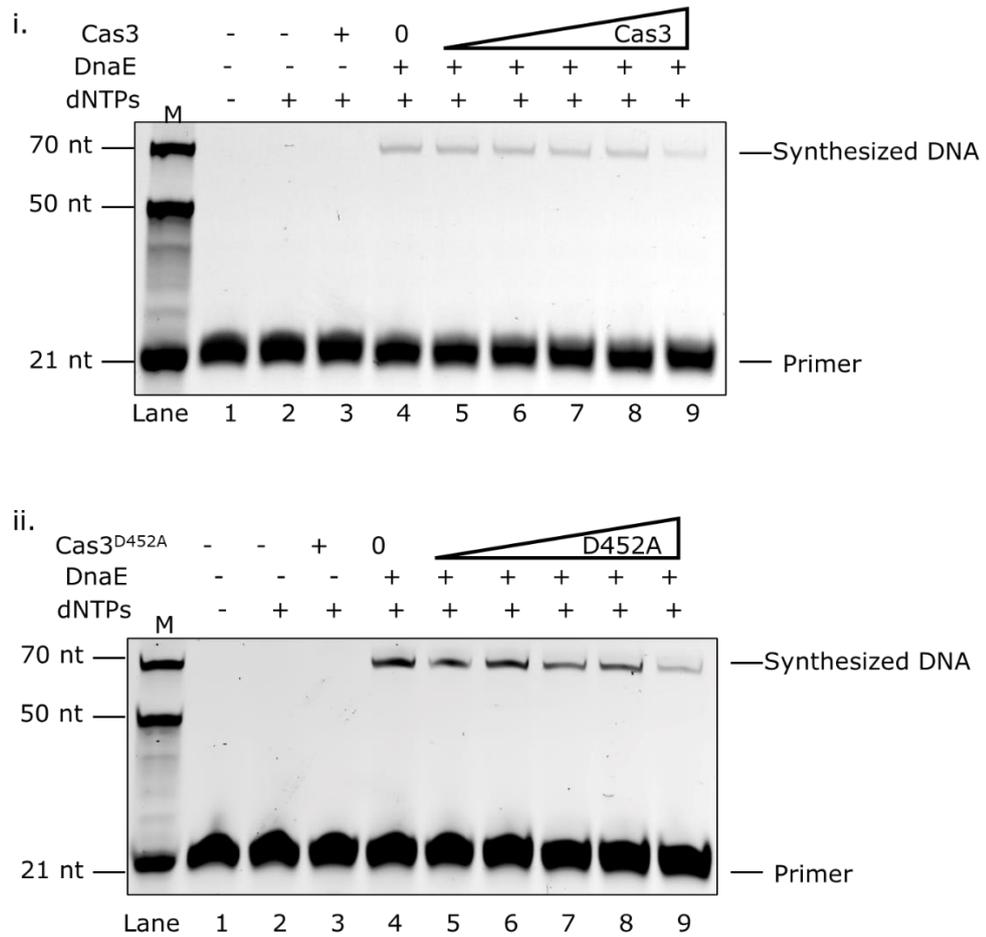
The newly purified Cas3 with no detectable Pol I contamination was used to investigate the potential protein-protein interaction with DnaE. Though published research suggests Cas3 physically interact with DnaE diagnosed by protein pull-down assay (151), this study revisited the previous report but focused mostly on whether this interaction had a functional effect. To this end, increasing Cas3 was titrated into 50 nM DnaE to assess primer extension of the ELB40P/41 substrate. The total amount of primer elongation generated by DnaE revealed a Cas3 concentration-dependent reduction (Figure 5.13, Panel i; compare Lane 4 and Lane 9). Notably, no convincing evidence suggests Cas3 inhibits DnaE function because of the excessive amount of Cas3 used against DnaE, but clearly, the DnaE function has altered when Cas3 exists.

Because DnaE alone is an inefficient DNA polymerase that frequently falls off from the DNA strand without the aid of Pol III subunit  $\epsilon$  and  $\theta$  (192)(Figure 5.14), one explanation for what was observed in the DnaE-Cas3 PE assay is that the Cas3 may compete with DnaE for binding to DNA, thus disturbing DnaE from accessing DNA substrates. However, other possibilities cannot be ignored,

because Cas3 reveals a low DNA binding affinity (Chapter 3, Section 3.7), which suggests the inhibitory effect may not be solely explained by DNA binding.

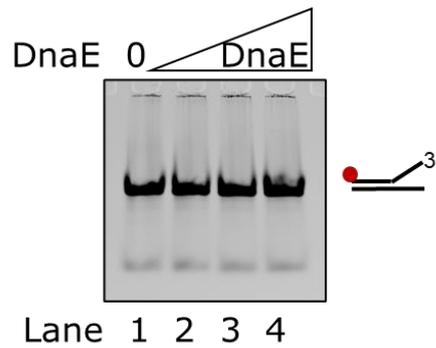
To test this further, Cas3<sup>D452A</sup> was used in a primer extension assay containing DnaE. This Cas3 mutant cannot form a stable DNA-protein binding complex (Section 5.3), so no inhibitory effect on DnaE would be expected if the DNA synthetic product loss is caused only by Cas3-DNA binding. However, the same reduction of DNA synthesis was observed when using Cas3<sup>D452A</sup> in DnaE PE assay (Figure 5.13, panel ii; compare Lane 9 to Lane 4). This result rules out the competitive binding inference; meanwhile, it indicates the change of DNA synthetic function of DnaE may arise from the physical interaction between DnaE and Cas3.

L. He Chapter 5



**Figure 5.13 DNA synthesis by DnaE was moderately reduced by Cas3.**

Primer extension assays using 50 nM DnaE and 10 nM ELB40P/41 substrate, with increasing concentration of Cas3 proteins. In each panel, reactions exhibited in Lanes 5-9 contained 10 nM, 25 nM, 50 nM, 100 nM and 200 nM Cas3 and Cas3<sup>D452A</sup> as denoted above. All reactions were incubated at 37 °C for 30 minutes before adding the proteinase K stop solution, and endpoint reaction products were then taken and analysed on 10 % denaturing page gel. Quantification data in Appendix 9.



**Figure 5.14 EMSA showing DnaE did not form a stable protein-DNA complex.**

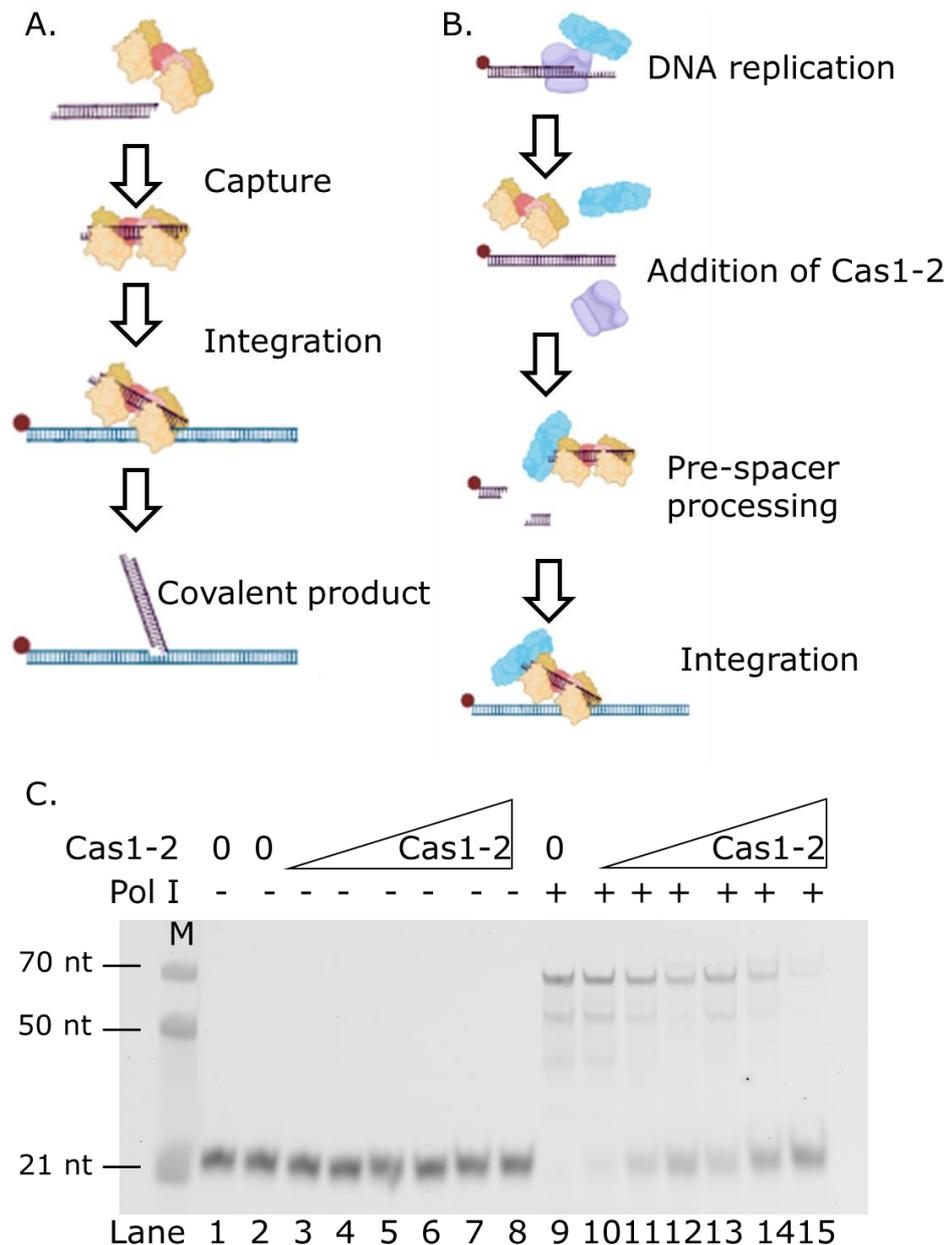
Increasing concentrations of DnaE were incubated with 10 nM ELB40P/41 substrate. Lanes 2-4 contained 150 nM, 300 nM and 600 nM DnaE. Samples were analysed on 8 % acrylamide TB gel which contained 5 mM DTT.

## 5.12 DnaE influences spacer integration by Cas1-2.

The proposed hypothesis described in Section 5.7 (Figure 5.6) that Cas1-2 may utilise newly synthesised DNA by DnaE as the source of pre-spacer was tested in an *in vitro* Spacer Integration (SPIN) assay adapted for use with DnaE. A standard *in vitro* SPIN assay utilises Cas1-2 to capture a model DNA fragment pre/protospacer (donor DNA), which is half-site integrated into a synthetic CRISPR locus (acceptor DNA) termed CRISPR Is, as a spacer. This reaction is stimulated by the integration host factor (IHF) (Figure 5.15 A). In this study, a SPIN assay was combined with a prior primer extension (PE) assay (PE-SPIN) that was designed to provide donor DNA of different lengths for spacer integration by Cas1-Cas2 (Figure 5.15 B).

## L. He Chapter 5

In the primary stage of a PE-SPIN assay, DNA products of varying lengths are produced by a defined DNA polymerase (e.g., DnaE) using ELB40P/41 substrate in the presence of Cas3. Then, the premixed Cas1-2, IHF and the labelled acceptor sequence CRISPR1s (integration mix) is added to the reaction to initiate the second stage SPIN (Figure 5.15 B). The reason for adding DNA polymerase and Cas1-2 to the reaction separately is that Cas1-2 can bind to ELB40P/41 and consequently block DNA synthesis by DNA Pols. This effect was revealed by carrying out a prior test using Cas1-2 and Pol I for assay designing purposes. The result suggested that adding Cas1-2 and DNA polymerase at the same time would dramatically reduce the amount of synthesised DNA (Figure 5.15 C).



**Figure 5.15 The PE reaction using Pol I alongside Cas1-2, for the development of a combined PE-SPIN assay that coupled DNA synthesis to spacer integration.**

(A, B). Illustration of the SPIN assay and its modification into a PE-SPIN assay. Cas1-2 is in yellow (Cas1) and red (Cas2). Spacer DNA (black) and CRISPR1s acceptor (blue) were labelled with Cy5 (red dot). (A). Showing different processes carried out during a standard SPIN assay. (B). In a PE-SPIN assay, DNA polymerase and Cas3 were firstly incubated with ELB40P/41 to synthesise various DNA products. Cas1-2 and other integration factors (not shown) are then added to the reaction mix. If spacer processing by Cas1-2 occurred, the covalent product could then be generated. (C) The addition of Cas1-2 inhibited DNA synthesis by Pol I.

## L. He Chapter 5

*Each reaction contained 10 nM ELB40P/41 substrate with increasing concentration of Cas1-2 of 0 nM, 10 nM, 25 nM, 50 nM, 100 nM, 200 nM and 400 nM in Lanes 2-8 and Lanes 9-15. In each sample, 3 nM Pol I was added, as shown in Lanes 9-15. Lanes 1 and 2 were identical. Quantification data in Appendix 9.*

DnaE was introduced into the PE-SPIN assay to produce DNA fragments that could be utilised by Cas1-2 as pre/proto-spacers. At the first stage of the PE-SPIN assay, 200 nM DnaE was pre-incubated with 10 nM ELB40P/41 at 37 °C for 30 minutes, to generate synthesised DNA products varying in lengths. After pre-incubation, integration activity was initiated by adding integration machinery, including Cas1-2, IHF and acceptor sequence CRISPR1s. Reactions were further incubated at 37 °C for one hour before analysing endpoint products. This assay was carried out with the addition of Cas3 because it was found to interact physically with Cas1-2 during primed adaptation (76). In addition, two standard SPIN reactions were introduced in PE-SPIN assay as controls, which could exhibit the formation of covalent products as the sign of a successful SPIN process being completed. Those control reactions utilised pre-made donor DNA substrates TK24/25 and ELB40/41 for Cas1-2 to capture (11) (Figure 5.11 A). The former is a 23 bp dsDNA with extended 5 nt overhangs at both of its 3' ends, while the latter is a DNA fragment of 70 bp. The oligos used to construct those two spacer DNAs is mentioned in Chapter 2, Section 2.4, Table 8.

## L. He Chapter 5

Intriguingly, multiple integration intermediates of varying sizes could be observed in PE-SPIN assay containing DnaE (Figure 5.16 B, Lane 6 to Lanes 3-5). To begin with, spacer integration by Cas1-2 was observed in control reactions containing different types of pre-made spacers. In Figure 5.16 B, the appearance of covalent products in Lanes 3-5 suggested that Cas1-2 utilised ELB40P/41, TK24/25 and ELB40/41 as pre-spacers and attempted to complete spacer integration. These results suggested Cas1-2 revealed no specificity on substrate conformation when it captured pre-spacers, whereas, in the presence of DnaE in reaction, two distinct integration intermediates were generated when compared to the control reaction that did not contain DnaE (Figure 5.16 B, compare Lane 6 to Lane 3). This indicated the presence of DnaE exerted a certain impact on the formation of covalent products. However, the role of DnaE during that process was different from what was expected in that it synthesised DNA products for Cas1-2 to capture and process (compare Lane 6 to Lane 5).

On the other hand, the addition of Cas3 inhibited spacer integration by Cas1-2. In the presence of Cas3, fewer or even no integration intermediates could be generated by Cas1-2 utilising any kind of DNA substrates available in these reaction solutions (Figure 5.16 B, compare Lane 6 with Lanes 7-9). These results revealed a repression function of Cas3 over Cas1-2, which consequently inhibited Cas1-2 from integrating DNA fragments into CRISPR Is.



and 250 nM Cas1-2 were added to reactions exhibited in Lanes 3-10, to initiate a SPIN assay. Reactions were further incubated at 37 °C for one hour, before adding a stop solution containing proteinase K and SDS. Endpoint products were analysed on 10 % acrylamide TBE gels. The right side of the gel panel shows the ELB40P/41, synthesised DNA substrates and spacer (black). CRISPR Is acceptor sequence is coloured in blue. A red dot represents Cy5 labelling. Quantification data in Appendix 9.

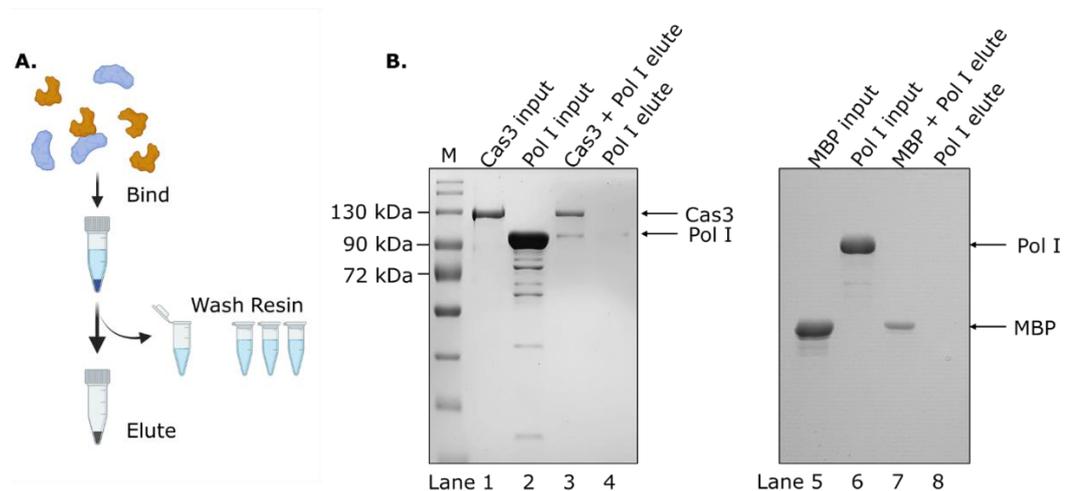
### **5.13 Cas3 physically interacts with Pol I, but it has no impact on DNA synthesis by Pol I.**

In Section 5.9, Pol I was found to be a persistent contaminant that co-purified with Cas3. The same phenomenon was repeatedly observed during purification conditions utilising both low and high concentrations of salt. This indicated the possibility that Pol I may be physically interacting with Cas3. To address this, an *in vitro* pull-down assay was applied using purified Cas3 fused with MBP tag as the bait to validate whether un-tagged Pol I could interact with it. If Cas3 and Pol I formed a stable protein complex, then two proteins would elute together from amylose resin.

An illustration explaining how a Cas3-Pol I pull-down assay was carried out is shown in Figure 5. 17 A. In brief, purified Cas3 and Pol I were premixed on ice before loading onto pre-equilibrated amylose resin. After rinsing out proteins that did not bind to the resin, those proteins that were bound to the resin were directly eluted by adding a buffer containing 10 mM Maltose. The same pull-down assay was repeated using Pol I and a mixture of Pol I and purified MBP

separately. This was to compare with the Cas3-Pol I pull-down analysis to gain direct evidence of whether Pol I interacts with Cas3 or MBP.

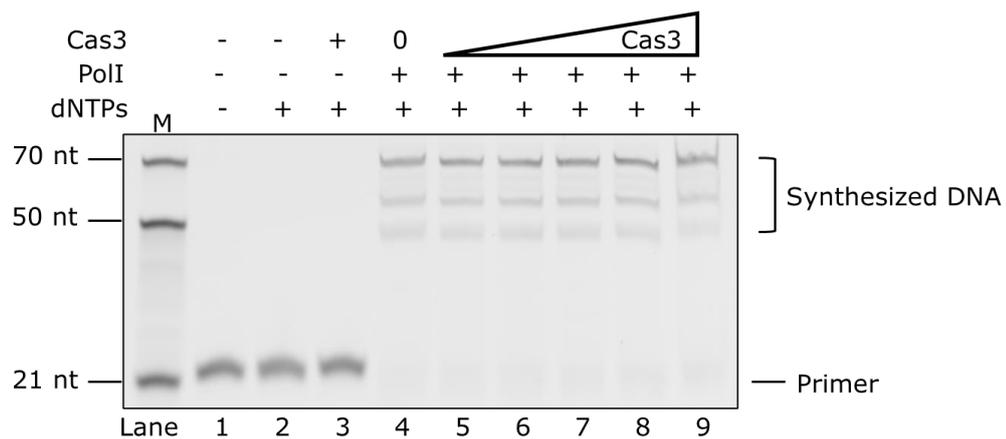
Pol I was eluted from amylose resin together with MBP-Cas3 but not with MBP (Figure 5.17 B, compare Lane 3 and Lane 7). This result proved that Cas3 and Pol I could form a stable physical interaction *in vitro*. It explained the phenomenon observed previously that Pol I co-purified with Cas3.



**Figure 5.17 Pull-down assay using MBP-Cas3 and Pol I indicated Cas3 and Pol I formed a physical interaction *in vitro*.**

(A). An illustration showing the process of a pull-down assay. Cas3 coloured in blue, Pol I coloured in yellow. (B). Showing MBP-Cas3 and Pol I (input) that were used in the pull-down assay and proteins that eluted from rinsed amylose resin on 10 % SDS page gel, post-stained with Coomassie blue. Lanes 1, 2, 5 and 6 represented proteins that were used to prepare protein mixtures before loading onto amylose resin, each of which were denoted on the top of gel images. Lanes 3 and 7 showed eluted proteins from the resin after washing out proteins that did not bind to amylose resin. Lanes 4 and 8 showed that Pol I used in this pull-down experiment did not have an affinity to amylose resin.

Since Cas3 has been proved to interact with Pol I, this gave rise to the question of whether Cas3 has any impact on the catalytic function of Pol I. To validate this, increasing concentrations of Cas3 were incubated with 3 nM Pol I and treated with 10 nM ELB40P/41 in the presence of dNTPs. However, no evidence suggested that Cas3 had any impact on the catalytic function of Pol I. No apparent change of total accumulation of synthesised DNA by Pol I was concurrent with the addition of Cas3 (Figure 5.18, compare Lane 4 and Lanes 5-9).



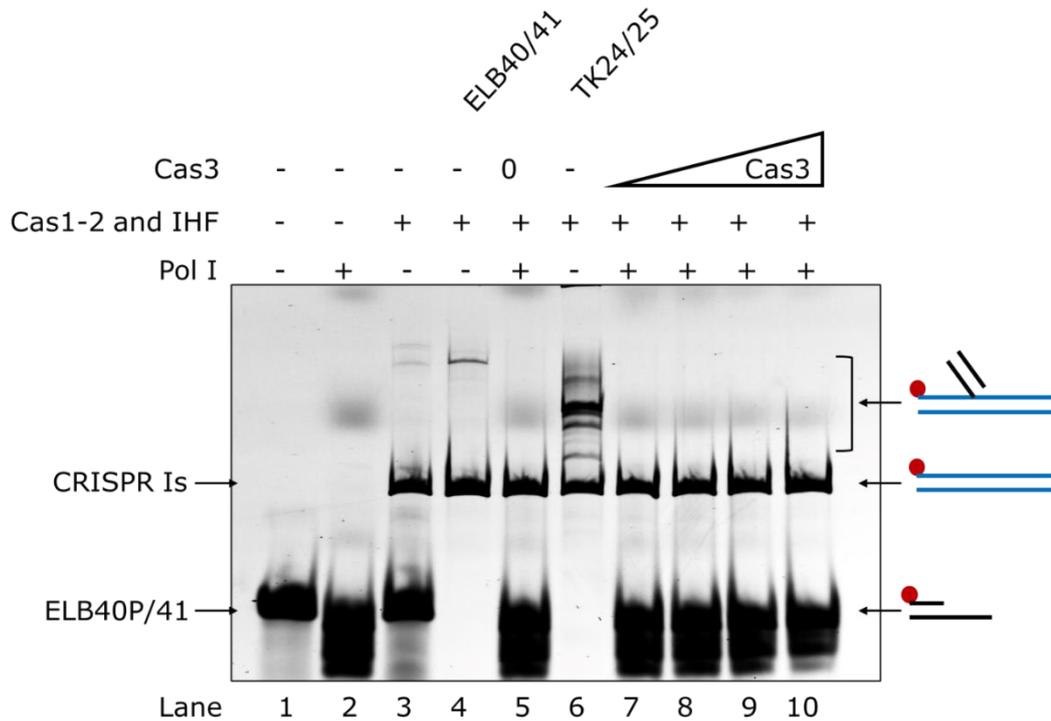
**Figure 5.18 Cas3 did not affect primer elongation by Pol I.**

*Primer extension using Pol I on ELB40P/41 substrate, with increasing concentration of Cas3. Reactions exhibited in Lanes 4-9 contain 3 nM of Pol I. The extra additions of Cas3 in Lanes 4-9 were 0 nM, 10 nM, 25 nM, 50 nM, 100 nM and 200 nM. Lane 3 contained 200 nM Cas3. All reactions were incubated at 37 °C for 30 minutes, before adding the proteinase K stop solution, and then taking the endpoint reaction products for analysis on 10 % polyacrylamide denaturing TBE gel. Quantification data in Appendix 9.*

### **5.14 Cas1-2 was not capable of using DNA fragments generated by Pol I as pre-spacers.**

To test if Cas1-2 could process synthesised DNA by Pol I into the spacer, the PE-SPIN assay was repeated using Pol I as a parallel test to the PE-SPIN assay using DnaE (Section 5.12). In this assay, 3 nM Pol I was pre-incubated with 10 nM ELB40P/41 at 37 °C for 30 minutes. Increasing concentrations of Cas3 proteins were added at the beginning of the PE-SPIN assay because it did not disturb the catalytic function of Pol I (Figure 5.18). After pre-incubation, integration activity was initiated by adding Cas1-2, IHF and acceptor sequence CRISPR1s. Reactions were further incubated at 37 °C for one hour before analysing endpoint products.

However, no integration intermediates were produced in reactions containing Pol I (Figure 5.19, compare Lane 5). The integration activity of Cas1-2 was stalled in the presence of Pol I, although in the reaction mix DNA substrates were available for it to capture (compare Lane 5 to Lane 3 and Lane 4). In addition, the addition of Cas3 did not assist or disturb Cas1-2 to integrate any type of DNA substrates that existed in reactions (Figure 5.19, compare Lane 5 to Lanes 7-9).



**Figure 5.19 No spacer integration by Cas1-2 occurred when using Pol I to produce DNA fragments as a pre-spacer source.**

*Pol I of 3 nM was first added to reactions exhibited in Lane 2, Lane 5 and Lanes 7-10, to generate DNA substrates as pre-spacers for subsequent spacer processing by Cas1-2. Lanes 7-10 contained Cas3 of 50 nM, 100 nM, 250 nM and 435 nM. Cas3 was then added to reactions together with Pol I when setting up reaction mixes. After incubating reactions at 37 °C for 30 minutes, integration factors including premixed IHF of 1  $\mu$ M, 25 nM CRISPR1s, and 250 nM Cas1-2 were added to reactions exhibited in Lanes 3-10, to initiate the SPIN assay. Reactions were further incubated at 37 °C for one hour, before adding a stop solution containing proteinase K and SDS. Quantification data in Appendix 9.*

## 5.15 Discussion for part II.

This part of the study proposed a new CRISPR interference-primed adaptation model which involves Cascade, Cas3, Cas1-2, DnaE and topoisomerases, and attempted to validate this model via biochemical approaches.

The hypothesis proposed in this study is that Cas3 may recruit DnaE on MGE to synthesise DNA fragments. This DNA product is then processed by Cas1-2 into pre-spacers. To validate this model, Cas3 and DnaE were primarily analysed via primer extension assay to test if adding Cas3 affects DNA synthesis by DnaE. In a control reaction containing only Cas3, extended primers were produced (Figure 5.8). This phenomenon was then further revealed that it was caused by Pol I that was co-purified with Cas3. In addition, DNA Pol I that existed in purified Cas3 was not a random case that happened in only one batch of Cas3 purified so far. Previously yielded *EcoCas3* in condensate state were revealed to be contaminated by DNA polymerase (Figure 5.10). These results indicated a possible occurrence of *EcoCas3*-Pol I interaction, which was then verified by *in vitro* pull-down assay (Figure 5.17). Interestingly, purified *PaeCas2-3* was also contaminated by *EcoPol I*. Considering that Pol I is a highly conserved protein among *E. coli* and *P. ae*, and any other species, Cas3-Pol I interaction could be a common feature for all Cas3 subtypes.

To continue the study on Cas3 and DnaE, the issue of Pol I persistently co-purifying with Cas3 needed to be resolved. This was achieved by optimising the Cas3 purification protocol. By increasing the salt concentration in purification buffers, less Pol I remained in the purified Cas3 solution. This small amount of Pol I could be further separated from Cas3 using a heparin column (Figure 5.12). This

## L. He Chapter 5

removal of Pol I was attributed to the different DNA binding affinities of Cas3 and Pol I.

After acquiring purer Cas3 with no Pol I, the impact of Cas3 on DnaE synthetic function was tested. An exclusive amount of Cas3 added to PE reactions containing DnaE could cause reduced accumulation of synthesised DNA products, but it is less convincing that Cas3 exerts an inhibition function over DnaE. What would cause the phenomenon observed? It was originally suspected that the occurrence of this phenomenon was driven by the competitive binding of Cas3 to DNA because DnaE itself frequently falls off from the DNA strand in the absence of its participants, including Pol III subunits  $\theta$  and  $\epsilon$ . However, examination using a Cas3 mutant Cas3<sup>D452A</sup> precluded this possibility. Cas3<sup>D452A</sup> did not interact with DNA (Section 5.3), while it still exhibited a slight repression effect on DNA synthesis by DnaE. Altogether, exclusive amount of Cas3 slightly disturbs the catalytic activity of DnaE, which was independent of the competitive binding of protein to DNA substrate. Additional studies are necessary for revealing further insight into these phenomena.

To test whether Cas1-2 can capture DNA fragments produced by DnaE as a pre-spacer, a PE-SPIN assay was designed and adopted. The PE-SPIN assay combining primer extension assay using defined DNA Pols with consequent SPIN assay involves Cas1-2, IHF and

receptor DNA sequence CRISPR 1s. Though no evidence showed Cas1-2 utilised the DNA products generated by DnaE and formed integration intermediates, those covalent products generated by Cas1-2 in the presence of DnaE were distinct from those acquired in a condition without DnaE (Figure 5.16, compare Lane 6 to Lanes 3 and 5). Intriguingly, the addition of Cas3 in the PE-SPIN assay repressed Cas1-2 integration. In reactions containing Cas3, the accumulation of integration intermediates generated by Cas1-2 was dramatically reduced (Figure 5.16, compare Lane 7 to Lane 6). This phenomenon was distinct from Cas3 blocking DNA synthesis by DnaE because adding Cas3 enabled Cas1-2 to integrate ELB40P/41 substrates into CRISPR 1s (Figure 5.16, compare Lanes 3 and 6 to Lanes 7-10).

Regarding Pol I, though it physically interacted with Cas3, the latter does not disturb the DNA extension activity of Pol I (Figure 5.12 and 5.13). This crucial factor for CRISPR primed adaptation (10) was analysed via PE-SPIN assay (Figure 5.14). However, no Cas1-2 integration occurred in the presence of Pol I. This result indicated that the role of Pol I in primed adaptation may be relying on its function of gap-filling DNA synthesis, rather than producing DNA fragments as pre-spacers for Cas1-2 to capture. The underlying mechanism lies in the interaction between Cas3 and Pol I and requires further studies to be fully understood.

## Chapter 6

### Conclusion and future work

The works conducted discussed in Chapters 3-5 included: Identification of 1. An iHDA1 region which exerts regulation over Cas3 nuclease activity against DNA; 2. Cas1, instead of Cas2, represses Cas3 nuclease activity. This effect of Cas1 over Cas3 is possibly generated via protein interaction of Cas1 at the Cas3 iHDA1 region; 3. A cysteine redox switch accommodated in the iHDA1 region that may facilitate allosteric regulation on Cas3 function via post-translational modification; 4. Cas3 may form a protein-dense phase by liquid-liquid separation for maintaining a nuclease activated state; 5. Cas3 physically interacts with DNA polymerase I; 6. Cas3 represses spacer integration by Cas1-2 *in vitro*; 7. DnaE, one of the proteins that interacts with Cas3, participates in Cas1-2 spacer integration. In the following content, the findings listed above will be summarized to discuss future works that could be undertaken based on current understanding of Cas3.

#### **6.1 The iHDA1 region accommodates a Trp-406 'gate' amino acid and a cysteine redox switch.**

Allosteric regulation is a common mechanism that can impact enzyme activities, but for the first time it is identified to have

occurred in *E. coli* CRISPR Cas3. By biochemical analysis of Cas3 proteins (Chapter 3 and 4), together with biophysical data provided by collaborators, we reported an allosteric regulation effect on Cas3 nuclease activity which is attributed to the structural rearrangement of an iHDA1 region located in the interface of the Cas3 HD and the RecA1 domain. Notably, this region was firstly hypothesised to be important for Cas3 functioning (109), in the I-E subtype Cas3 found in *T. fusca*. Based on a structural analysis using *TfuCas3*, Huo (109) suggested the DNA substrate undergoes extraordinary geometry twists to avoid steric collision with the Cas3 peptides adjacent to the DNA catalytic sites in the HD domain. However, no follow up study was conducted to further reveal the role of those peptides and their underlying mechanism during Cas3 functioning as a CRISPR interference effector. In this thesis, those peptides are further characterised because they revealed allosteric regulation over Cas3 function, we defined those peptides forming a region termed iHDA1.

The structural change of iHDA1 region leads to Cas3 to form two conformational states: the nuclease active state (at 30 °C) and nuclease defective state (at 37 °C). Stimulating by temperature change, as has proved in this study, Cas3 transitions from one state to another and eventually leads to a phenomenon where CRISPR interference in *E. coli* CRISPR-Cas system is temperature dependent. However, temperature may not be the only factor that causes the

amino acids to shift in the Cas3 iHDA1 region. Regarding this aspect, this thesis has concluded two relevant factors that impact the iHDA1, apart from the environmental temperature (Chapter 3), is the CRISPR Cas1 that may interact with Cas3 at the iHDA1 region and facilitate the iHDA1 region to re-construct (Chapter 4). In addition, this thesis proposed a hypothesis that the post-translational modification may also lead to a change in iHDA1 structure by targeting the redox switch accommodated in this region (Chapter 4). To validate this theory, biochemical and genetic analysis using a Cas3 mutant with a removed redox switch is necessary.

## **6.2 Cas3 phase separation-formation and dissociation.**

During Cas3 purification, two types of Cas3 proteins were acquired. The most abundant Cas3 proteins existed in soluble large protein complexes, with a small amount of monomeric state. The first assumption regarding the Cas3 complex was that it was a protein aggresome consisting of misfolded Cas3 proteins. Surprisingly, biochemical analysis using this Cas3 'aggresome' revealed that it was nuclease active, whilst under the same reaction condition, the Cas3 monomer was nuclease defective. This phenomenon, the formation of oligomeric Cas3, leads to a possibility that Cas3 may undergo protein phase separation.

## L. He Chapter 6

Protein phase separation is often found to have occurred in eukaryotic proteins which have intrinsically disordered protein regions, or themselves are intrinsically disordered proteins. However, it is a fairly novel topic in research on prokaryotic proteins. Why and how proteins can form phase separation is still unclear. But several factors involve the formation and dissociation of protein phase separation, including RNA (formation) and the existence of intrinsically disordered polypeptide in protein (formation). In addition, ATP is also an important factor that determines whether or not protein undergoes phase separation. It has been observed that protein phase separation is adopted as a survival strategy for cells when lacking ATP. In other words, adding ATP can dissociate protein phase separation.

Cas3 is hypothesised to form phase separation based on the following facts: 1. Cas3 forms large soluble complex meanwhile its catalytical function is intact; 2. The Cas3 complex/aggresome can be dissociated by incubating protein with ATP; 3. Cas3 has a DEXD box (also termed DEAD box) in the RecA1 region, which has been proven to facilitate phase separation in other proteins, for instance, in DDX RNA helicases (185); 4. Cas3 has an iHDA1 region which revealed structural plasticity, therefore it is likely to be an intrinsically disordered protein region. However, it can be challenging when conducting a further study on Cas3 and protein phase separation, because defining protein phase separation in

bacteria requires the usage of multiple methods, so comprehensive analysis using various approaches is necessary. For example, the protein droplet formed by protein phase separation (PPS) can be directly observed using confocal microscope, if adding 1,6-Hexanediol in tested sample, this could potentially stimulate disassociation of protein droplet. These could provide direct evidence to prove the target protein undergoes phase separation.

In addition, adapting study methods used in investigating phase separation in eukaryotic cells into that can be used in prokaryotic cells may be required.

### **6.3 Detecting the interaction between Cas3 and DNA polymerases using Bimolecular fluorescence complementation (BiFC) assay.**

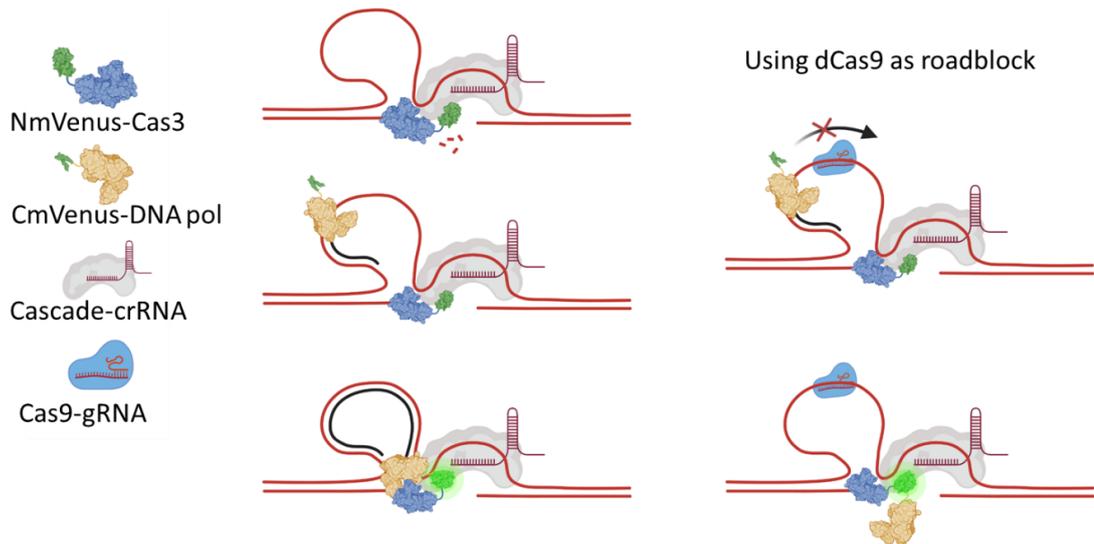
Cas3 is found to interact with DNA polymerase I (Chapter 5). Based on this, a BiFC assay has been designed and set up to detect its interaction *in vivo* (genetic study). Briefly, the BiFC assay denoted protein-protein interaction by reconstitution of a complete yellow fluorescent protein (YFP). An illustration of the specific BiFC assay designed for Cas3-polymerase is shown in Figure 6.1. In this system, a YFP variation termed mVenus is split into two complementary portions termed NmVenus by N terminus mVenus and CmVenus by

## L. He Chapter 6

C terminus mVenus. NmVenus and CmVenus are fused with Cas3 and target DNA polymerase separately. During CRISPR interference, the interaction of Cas3 with DNA polymerase will bring NmVenus and CmVenus together and consequently those two portions reconstitute an intact YFP. This formation could be detected using 514 nm argon-laser lines for excitation, under this laser wavelength, the complete YFP can be 'lighted-up' but not those individual YFP portions. If Cas3 does not interact with DNA polymerase, no signal could be detected.

During BiFC assay, dCas9 can also be introduced for distinguishing the interaction between Cas3 and the DNA polymerase loaded on the targeting strand or DNA pol from elsewhere. A dCas9 targeting the upstream of the protospacer is a roadblock for DNA replication and potentially prevent the interaction of Cas3 and DNA polymerase if the DNA polymerase is approaching Cas3 from the targeting strand of MGE (Figure 6.1).

This whole BiFC assay system has been set up and tested fusion protein expression (Appendix 3). Owing to time limits, this assay has not been conducted.



**Figure 6.1. BiFC assay for analysing Cas3-DNA polymerase interaction during CRISPR interference.**

The interaction between NmVenus-Cas3 fusion and CmVenus-DNA pol fusion will lead to reconstitution of mVenus (YFP). Intact YFP has maximum excitation at 514 nm laser wavelength-showing in the figure as 'light-up'. dCas9 targets the upstream of protospacer on the targeting strand can be used in BiFC assay.

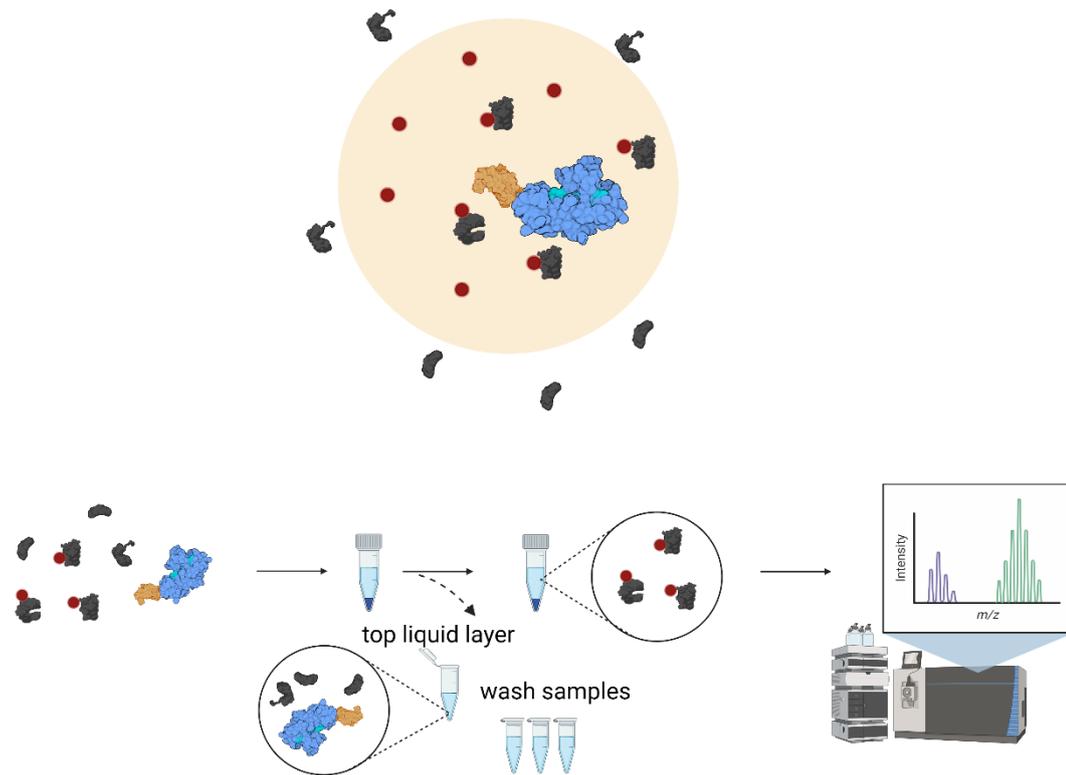
#### **6.4. BioID could be adopted to screen broader participators of Cas3.**

Apart from Cas proteins, Cas3 also interacts with other non-Cas proteins. By using pull-down assay (151), Cas3 was found to interact with GroL, DnaE, GyrA, AceE, IbpA, CrfC, and MdoD. However, the disadvantage of pull-down assay is it can only detect the stable protein-protein interaction, for detecting those weak interactions Cas3 participated in, pull-down assay is not feasible. To analyse broader participators of Cas3, the BioID assay can be used for this

purpose and to identify those proteins forming weak interaction with Cas3.

BioID is a technique that exploited biotin ligase to label proximal proteins using biotin via biotinylation, with approximately 10 nm radius when conduct labelling. Labelled proteins are then isolated from cell via biotin affinity pull-down assay and can be further analysed by mass spectrometry. Because protein biotinylation is rare in bacteria cells, so the endogenous modification in cells is less likely to disturb BioID biotinylation (Figure 6.2).

Cas3 fused with second-generation BioID, the biotin-protein ligase<sup>R118G</sup> from *Aquifex aeolicus*, is constructed and tested expression (Appendix 4). Due to time limits, this assay was not completed before the submission of this thesis.



**Figure 6.2. BioID assay for detecting broader participators of Cas3.**

*Cas3* is labelled in blue. Biotin-protein ligase is yellow and biotin molecule is in red. Proximal proteins in the detectable range (light yellow) are biotinylated by Biotin-protein ligase and isolated from cell lysate by doing pull down assay. Protein sample is further analysed by mass spectrometry.

## 6.5 Cas3 and beyond

This thesis displayed research conducted on Cas3 in Dr Edward Bolt's lab. More features of Cas3 have been revealed, including a protein function regulatory portion, the iHDA1 region, and its interaction with Pol I. Observations during biochemical analysis on Cas3 have indicated there are still mysterious features of Cas3 that are yet to be elucidated. For instance, a putative redox switch in *E.*

*coli* Cas3 iHDA1 region may link the catalytical function of Cas3 with post-translational modification or cell signalling systems.

Though both *in vitro* and *in vivo* studies have been done on different Cas3 subtypes by numerous research groups, Cas3 is still far from being a protein that is fully understood. Such a protein, with several known functions, yet still full of mysteries is reminiscent of a concept often used in the computer science field, termed 'black box'. This phrase is to describe a system with a known input and output, but which does not reveal its internal workings. For Cas3, the known 'input' is that Cascade mediated R-loop formation on MGEs stimulates its nuclease-helicase functions; and the 'output' is elimination of MGEs, and primed adaptation if Cas1-Cas2 is present. This thesis has further revealed that the helicase and nuclease domains of Cas3 are joined by a conserved iHDA1 region; with biochemical analysis suggests this region may interact with other factors, e.g., Cas1. However, the underlying mechanism of Cas3 function and the effect that other factors have on this function requires further study.

Considering that Cas3 is far from being fully understood, the emerging application of Cascade-Cas3 system as an applied tool in gene editing and phage therapy is controversial. Regardless of unrevealed risks, such as the possible off-target effect when applying Cascade-Cas3 system, using Cas3 as a tool may provide

less value when so little is known about it. It is unfortunate that the importance of understanding the fundamental biology of Cas3 has been overlooked compared of other aspects of CRISPR-Cas research. However, this situation may change in the near future. One of the signs is increasing publications that provide direct and indirect evidence suggesting Cas3 is the key factor that coordinates the CRISPR interference and primed adaptation. These outcomes will draw more attention to Cas3 as an individual protein, along with its roles in CRISPR and non-CRISPR systems.

## References

1. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P. *et al.* (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*, **18**, 67-83.
2. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, **9**, 467-477.
3. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
4. Mojica, F.J., Juez, G. and Rodríguez-Valera, F. (1993) Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Mol Microbiol*, **9**, 613-621.
5. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, **169**, 5429-5433.
6. Jansen, R., Embden, J.D., Gastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, **43**, 1565-1575.
7. Jansen, R., van Embden, J.D., Gastra, W. and Schouls, L.M. (2002) Identification of a novel family of sequence repeats among prokaryotes. *Omic : a journal of integrative biology*, **6**, 23-33.
8. 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 (Reading, England)*, **151**, 653-663.
9. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (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.
10. Ivancic-Bace, I., Cass, S.D., Wearne, S.J. and Bolt, E.L. (2015) Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res*, **43**, 10821-10830.

## References

11. Radovicic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E.L. and Ivancic-Bace, I. (2018) CRISPR-Cas adaptation in *Escherichia coli* requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic Acids Res*, **46**, 10173-10183.
12. Bernheim, A., Bikard, D., Touchon, M. and Rocha, E.P.C. (2019) A matter of background: DNA repair pathways as a possible cause for the sparse distribution of CRISPR-Cas systems in bacteria. *Philos Trans R Soc Lond B Biol Sci*, **374**, 20180088.
13. Bernheim, A., Calvo-Villamanan, A., Basier, C., Cui, L., Rocha, E.P.C., Touchon, M. and Bikard, D. (2017) Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat Commun*, **8**, 2094.
14. Depardieu, F., Didier, J.P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B. and Bikard, D. (2016) A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. *Cell Host Microbe*, **20**, 471-481.
15. Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, **151**, 2551-2561.
16. Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67-71.
17. Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843-1845.
18. Andersson, A.F. and Banfield, J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, **320**, 1047-1050.
19. Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C. and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of bacteriology*, **190**, 1401-1412.
20. Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P. and Moineau, S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology*, **190**, 1390-1400.
21. Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*, **109**, E2579-2586.

## References

22. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602-607.
23. Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A. and Yakunin, A.F. (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J*, **30**, 4616-4627.
24. Howard, J.A., Delmas, S., Ivancic-Bace, I. and Bolt, E.L. (2011) Helicase dissociation and annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem J*, **439**, 85-95.
25. Pul, U., 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. *Mol Microbiol*, **75**, 1495-1512.
26. Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K. *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell*, **46**, 595-605.
27. Plagens, A., Tjaden, B., Hagemann, A., Randau, L. and Hensel, R. (2012) Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *Journal of bacteriology*, **194**, 2491-2500.
28. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
29. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
30. DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*, **41**, 4336-4343.
31. Waltz, E. (2016) Gene-edited CRISPR mushroom escapes US regulation. *Nature*, **532**, 293.
32. Koonin, E.V. and Makarova, K.S. (2019) Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci*, **374**, 20180087.
33. 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. *EMBO J*, **30**, 1335-1342.
34. Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M. and Doudna, J.A. (2018) CRISPR-Cas12a target

## References

- binding unleashes indiscriminate single-stranded DNase activity. *Science*, **360**, 436-439.
35. O'Connell, M.R. (2019) Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *Journal of molecular biology*, **431**, 66-87.
  36. Kolesnik, M.V., Fedorova, I., Karneyeva, K.A., Artamonova, D.N. and Severinov, K.V. (2021) Type III CRISPR-Cas Systems: Deciphering the Most Complex Prokaryotic Immune System. *Biochemistry (Mosc)*, **86**, 1301-1314.
  37. Zhou, Y., Bravo, J.P.K., Taylor, H.N., Steens, J.A., Jackson, R.N., Staals, R.H.J. and Taylor, D.W. (2021) Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience*, **24**, 102201.
  38. Moya-Beltran, A., Makarova, K.S., Acuna, L.G., Wolf, Y.I., Covarrubias, P.C., Shmakov, S.A., Silva, C., Tolstoy, I., Johnson, D.B., Koonin, E.V. *et al.* (2021) Evolution of Type IV CRISPR-Cas Systems: Insights from CRISPR Loci in Integrative Conjugative Elements of Acidithiobacillia. *CRISPR J*, **4**, 656-672.
  39. Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol*, **21**, 528-534.
  40. He, L., St John James, M., Radovicic, M., Ivancic-Bace, I. and Bolt, E.L. (2020) Cas3 Protein—A Review of a Multi-Tasking Machine. *Genes*, **11**, 208.
  41. Rollie, C., Graham, S., Rouillon, C. and White, M.F. (2018) Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Res*, **46**, 1007-1020.
  42. Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N. and Doudna, J.A. (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*, **527**, 535-538.
  43. Xiao, Y., Ng, S., Nam, K.H. and Ke, A. (2017) How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*, **550**, 137-141.
  44. Nussenzweig, P.M. and Marraffini, L.A. (2020) Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annu Rev Genet*, **54**, 93-120.
  45. Yoganand, K.N., Sivathanu, R., Nimkar, S. and Anand, B. (2017) Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res*, **45**, 367-381.
  46. Sun, C.L., Thomas, B.C., Barrangou, R. and Banfield, J.F. (2016) Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J*, **10**, 858-870.

## References

47. Nicholson, T.J., Jackson, S.A., Croft, B.I., Staals, R.H.J., Fineran, P.C. and Brown, C.M. (2019) Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol*, **16**, 566-576.
48. Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*, **8**, R61.
49. Hatoum-Aslan, A., Maniv, I. and Marraffini, L.A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci U S A*, **108**, 21218-21222.
50. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K. *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*, **60**, 385-397.
51. Chylinski, K., Le Rhun, A. and Charpentier, E. (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol*, **10**, 726-737.
52. Nickel, L., Ulbricht, A., Alkhnbashi, O.S., Forstner, K.U., Cassidy, L., Weidenbach, K., Backofen, R. and Schmitz, R.A. (2019) Cross-cleavage activity of Cas6b in crRNA processing of two different CRISPR-Cas systems in *Methanosarcina mazei* Go1. *RNA Biol*, **16**, 492-503.
53. Wei, W., Zhang, S., Fleming, J., Chen, Y., Li, Z., Fan, S., Liu, Y., Wang, W., Wang, T., Liu, Y. *et al.* (2019) Mycobacterium tuberculosis type III-A CRISPR/Cas system crRNA and its maturation have atypical features. *FASEB J*, **33**, 1496-1509.
54. Sokolowski, R.D., Graham, S. and White, M.F. (2014) Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic Acids Res*, **42**, 6532-6541.
55. Taylor, H.N., Warner, E.E., Armbrust, M.J., Crowley, V.M., Olsen, K.J. and Jackson, R.N. (2019) Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biol*, **16**, 1438-1447.
56. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. and Charpentier, E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517-521.
57. Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P. and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/D/vulg CRISPR-Cas system. *Structure*, **20**, 1574-1584.
58. Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.E., Maier, L.K., Richter, H., Nickel, L., Schmitz, R.A. *et al.* (2014) A complex of Cas proteins 5, 6, and 7 is required

## References

- for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. *J Biol Chem*, **289**, 7164-7177.
59. Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V. *et al.* (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem*, **286**, 21643-21656.
  60. Punetha, A., Sivathanu, R. and Anand, B. (2014) Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system. *Nucleic Acids Res*, **42**, 3846-3856.
  61. Liu, Q., Wang, S., Long, J., Chen, Z., Yang, B. and Lin, F. (2021) Functional Identification of the *Xanthomonas oryzae* pv. *oryzae* Type I-C CRISPR-Cas System and Its Potential in Gene Editing Application. *Front Microbiol*, **12**, 686715.
  62. Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P. and Siksnys, V. (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol*, **10**, 841-851.
  63. Peng, W., Li, H., Hallstrom, S., Peng, N., Liang, Y.X. and She, Q. (2013) Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol*, **10**, 738-748.
  64. Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P. and Siksnys, V. (2013) In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J*, **32**, 385-394.
  65. Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E. and Doudna, J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A*, **111**, 6618-6623.
  66. Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B. and Ke, A. (2016) Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature*, **530**, 499-503.
  67. Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K.A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S.J.J. and Severinov, K. (2019) Systematic analysis of Type I-E *Escherichia coli* CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol Microbiol*, **111**, 1558-1570.
  68. O'Brien, R.E., Santos, I.C., Wrapp, D., Bravo, J.P.K., Schwartz, E.A., Brodbelt, J.S. and Taylor, D.W. (2020) Structural basis

## References

- for assembly of non-canonical small subunits into type I-C Cascade. *Nat Commun*, **11**, 5931.
69. Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M. and Ke, A. (2017) Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell*, **170**, 48-60 e11.
  70. Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V. *et al.* (2018) Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell*, **175**, 934-946 e915.
  71. Gong, B., Shin, M., Sun, J., Jung, C.H., Bolt, E.L., van der Oost, J. and Kim, J.S. (2014) Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc Natl Acad Sci U S A*, **111**, 16359-16364.
  72. van Erp, P.B.G., Patterson, A., Kant, R., Berry, L., Golden, S.M., Forsman, B.L., Carter, J., Jackson, R.N., Bothner, B. and Wiedenheft, B. (2018) Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. *ACS Chem Biol*, **13**, 481-490.
  73. Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A. and Greene, E.C. (2015) Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, **163**, 854-865.
  74. Loeff, L., Brouns, S.J.J. and Joo, C. (2018) Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. *Mol Cell*, **70**, 385-394 e383.
  75. Dolan, A.E., Hou, Z., Xiao, Y., Gramelspacher, M.J., Heo, J., Howden, S.E., Freddolino, P.L., Ke, A. and Zhang, Y. (2019) Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Mol Cell*, **74**, 936-950 e935.
  76. Musharova, O., Medvedeva, S., Klimuk, E., Guzman, N.M., Titova, D., Zgoda, V., Shiriaeva, A., Semenova, E., Severinov, K. and Savitskaya, E. (2021) Prespacers formed during primed adaptation associate with the Cas1-Cas2 adaptation complex and the Cas3 interference nuclease-helicase. *Proc Natl Acad Sci U S A*, **118**.
  77. Diancourt, L., Passet, V., Chervaux, C., Garault, P., Smokvina, T. and Brisse, S. (2007) Multilocus sequence typing of Lactobacillus casei reveals a clonal population structure with low levels of homologous recombination. *Appl Environ Microbiol*, **73**, 6601-6611.
  78. Lloyd, R.G. and Rudolph, C.J. (2016) 25 years on and no end in sight: a perspective on the role of RecG protein. *Curr Genet*, **62**, 827-840.
  79. Killelea, T., Hawkins, M., Howard, J.L., McGlynn, P. and Bolt, E.L. (2019) DNA replication roadblocks caused by Cascade

## References

- interference complexes are alleviated by RecG DNA repair helicase. *RNA Biol*, **16**, 543-548.
80. Heussler, G.E., Miller, J.L., Price, C.E., Collins, A.J. and O'Toole, G.A. (2016) Requirements for *Pseudomonas aeruginosa* Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm Enrichment Assay. *J Bacteriol*, **198**, 3080-3090.
  81. Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., Peng, W., Liang, Y., She, Q. and Peng, N. (2017) Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in *Sulfolobus islandicus*. *Nucleic Acids Res*, **45**, 8978-8992.
  82. Liu, Z., Sun, M., Liu, J., Liu, T., Ye, Q., Li, Y. and Peng, N. (2020) A CRISPR-associated factor Csa3a regulates DNA damage repair in Crenarchaeon *Sulfolobus islandicus*. *Nucleic Acids Res*, **48**, 9681-9693.
  83. Lee, K.H., Lee, S.G., Eun Lee, K., Jeon, H., Robinson, H. and Oh, B.H. (2012) Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins*, **80**, 2573-2582.
  84. Koo, Y., Jung, D.K. and Bae, E. (2012) Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS One*, **7**, e33401.
  85. Nam, K.H., Kurinov, I. and Ke, A. (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca<sup>2+</sup>-dependent double-stranded DNA binding activity. *J Biol Chem*, **286**, 30759-30768.
  86. Ellinger, P., Arslan, Z., Wurm, R., Tschapek, B., MacKenzie, C., Pfeiffer, K., Panjekar, S., Wagner, R., Schmitt, L., Gohlke, H. *et al.* (2012) The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J Struct Biol*, **178**, 350-362.
  87. Zhang, J., Kasciukovic, T. and White, M.F. (2012) The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One*, **7**, e47232.
  88. Rocha, E.P., Cornet, E. and Michel, B. (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet*, **1**, e15.
  89. Pickar-Oliver, A. and Gersbach, C.A. (2019) The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol*, **20**, 490-507.
  90. Barman, A., Deb, B. and Chakraborty, S. (2020) A glance at genome editing with CRISPR-Cas9 technology. *Curr Genet*, **66**, 447-462.

## References

91. Li, G., Wang, H., Zhang, X., Wu, Z. and Yang, H. (2021) A Cas9-transcription factor fusion protein enhances homology-directed repair efficiency. *J Biol Chem*, **296**, 100525.
92. Reint, G., Li, Z., Labun, K., Keskitalo, S., Soppa, I., Mamia, K., Tolo, E., Szymanska, M., Meza-Zepeda, L.A., Lorenz, S. *et al.* (2021) Rapid genome editing by CRISPR-Cas9-POLD3 fusion. *Elife*, **10**.
93. Swarts, D.C. and Jinek, M. (2018) Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA*, **9**, e1481.
94. Meliawati, M., Schilling, C. and Schmid, J. (2021) Recent advances of Cas12a applications in bacteria. *Appl Microbiol Biotechnol*, **105**, 2981-2990.
95. Shi, Y., Fu, X., Yin, Y., Peng, F., Yin, X., Ke, G. and Zhang, X. (2021) CRISPR-Cas12a System for Biosensing and Gene Regulation. *Chem Asian J*, **16**, 857-867.
96. Aregger, M., Xing, K. and Gonatopoulos-Pournatzis, T. (2021) Application of CHyMErA Cas9-Cas12a combinatorial genome-editing platform for genetic interaction mapping and gene fragment deletion screening. *Nat Protoc*, **16**, 4722-4765.
97. Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R. and Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, **5**, e00928-00913.
98. Csörgő, B., León, L.M., Chau-Ly, I.J., Vasquez-Rifo, A., Berry, J.D., Mahendra, C., Crawford, E.D., Lewis, J.D. and Bondy-Denomy, J. (2020) A compact Cascade-Cas3 system for targeted genome engineering. *Nature Methods*.
99. Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., Joung, J., Collins, J.J. and Zhang, F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, **360**, 439-444.
100. Broughton, J.P., Deng, X., Yu, G., Fasching, C.L., Servellita, V., Singh, J., Miao, X., Streithorst, J.A., Granados, A., Sotomayor-Gonzalez, A. *et al.* (2020) CRISPR-Cas12-based detection of SARS-CoV-2. *Nat Biotechnol*, **38**, 870-874.
101. Azhar, M., Phutela, R., Kumar, M., Ansari, A.H., Rauthan, R., Gulati, S., Sharma, N., Sinha, D., Sharma, S. and Singh, S. (2020) Rapid, accurate, nucleobase detection using FnCas9. *MedRxiv*.
102. Gupta, R., Kazi, T.A., Dey, D., Ghosh, A., Ravichandiran, V., Swarnakar, S., Roy, S., Biswas, S.R. and Ghosh, D. (2021) CRISPR detectives against SARS-CoV-2: a major setback against COVID-19 blowout. *Appl Microbiol Biotechnol*, **105**, 7593-7605.
103. Whinn, K.S., Kaur, G., Lewis, J.S., Schauer, G.D., Mueller, S.H., Jergic, S., Maynard, H., Gan, Z.Y., Naganbabu, M.,

## References

- Bruchez, M.P. *et al.* (2019) Nuclease dead Cas9 is a programmable roadblock for DNA replication. *Sci Rep*, **9**, 13292.
104. Alvarez, B., Mencia, M., de Lorenzo, V. and Fernandez, L.A. (2020) In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat Commun*, **11**, 6436.
105. Mojica, F.J. and Rodriguez-Valera, F. (2016) The discovery of CRISPR in archaea and bacteria. *Febs j*, **283**, 3162-3169.
106. Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell*, **172**, 1239-1259.
107. Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Mielliet, W.R., Klein, M., Depken, M., Suarez-Diez, M. and Brouns, S.J. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol Cell*, **63**, 852-864.
108. Majsec, K., Bolt, E.L. and Ivancic-Bace, I. (2016) Cas3 is a limiting factor for CRISPR-Cas immunity in Escherichia coli cells lacking H-NS. *BMC Microbiol*, **16**, 28.
109. Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I. *et al.* (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol*, **21**, 771-777.
110. Jackson, R.N., Lavin, M., Carter, J. and Wiedenheft, B. (2014) Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr Opin Struct Biol*, **24**, 106-114.
111. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*, **20**, 313-324.
112. Rho, J., Choi, S., Seong, Y.R., Choi, J. and Im, D.S. (2001) The arginine-1493 residue in QRRGRTGR1493G motif IV of the hepatitis C virus NS3 helicase domain is essential for NS3 protein methylation by the protein arginine methyltransferase 1. *J Virol*, **75**, 8031-8044.
113. Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B. and Koonin, E.V. (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic acids research*, **30**, 482-496.
114. Swarts, D.C., Mosterd, C., van Passel, M.W. and Brouns, S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**, e35888.
115. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.

## References

116. Nimkar, S. and Anand, B. (2020) Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. *Nucleic Acids Res*, **48**, 2486-2501.
117. Stoll, B., Maier, L.K., Lange, S.J., Brendel, J., Fischer, S., Backofen, R. and Marchfelder, A. (2013) Requirements for a successful defence reaction by the CRISPR-Cas subtype I-B system. *Biochem Soc Trans*, **41**, 1444-1448.
118. Rollins, M.F., Chowdhury, S., Carter, J., Golden, S.M., Wilkinson, R.A., Bondy-Denomy, J., Lander, G.C. and Wiedenheft, B. (2017) Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc Natl Acad Sci U S A*, **114**, E5113-E5121.
119. Wang, X., Yao, D., Xu, J.G., Li, A.R., Xu, J., Fu, P., Zhou, Y. and Zhu, Y. (2016) Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat Struct Mol Biol*, **23**, 868-870.
120. Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L. *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**, E5122-E5128.
121. Sinkunas, T., Gasiunas, G. and Siksnys, V. (2015) In Lundgren, M., Charpentier, E. and Fineran, P. C. (eds.), *CRISPR: Methods and Protocols*. Springer New York, New York, NY, pp. 277-291.
122. 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. *J Biol Chem*, **288**, 22184-22192.
123. 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**, 31896-31903.
124. Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V. and Seidel, R. (2015) Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep*, **10**, 1534-1543.
125. Sashital, D.G., Wiedenheft, B. and Doudna, J.A. (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell*, **46**, 606-615.
126. Tuminauskaite, D., Norkunaite, D., Fiodorovaite, M., Tumas, S., Songailiene, I., Tamulaitiene, G. and Sinkunas, T. (2020) DNA interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system. *BMC Biol*, **18**, 65.
127. Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., Xu, H., Sasakawa, N., Naito, Y., Nakada, S. *et al.*

## References

- (2019) CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat Commun*, **10**, 5302.
128. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annual Review of Biochemistry*, **76**, 23-50.
129. Awate, S. and Brosh, R.M., Jr. (2017) Interactive Roles of DNA Helicases and Translocases with the Single-Stranded DNA Binding Protein RPA in Nucleic Acid Metabolism. *Int J Mol Sci*, **18**.
130. Fagerburg, M.V., Schauer, G.D., Thickman, K.R., Bianco, P.R., Khan, S.A., Leuba, S.H. and Anand, S.P. (2012) PcrA-mediated disruption of RecA nucleoprotein filaments--essential role of the ATPase activity of RecA. *Nucleic Acids Res*, **40**, 8416-8424.
131. Terakawa, T., Redding, S., Silverstein, T.D. and Greene, E.C. (2017) Sequential eviction of crowded nucleoprotein complexes by the exonuclease RecBCD molecular motor. *Proc Natl Acad Sci U S A*, **114**, E6322-E6331.
132. Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem Soc Trans*, **37**, 343-348.
133. Cui, T.J. and Joo, C. (2019) Facilitated diffusion of Argonaute-mediated target search. *RNA Biol*, **16**, 1093-1107.
134. Xue, C., Zhu, Y., Zhang, X., Shin, Y.K. and Sashital, D.G. (2017) Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Rep*, **21**, 3717-3727.
135. Park, J., Myong, S., Niedziela-Majka, A., Lee, K.S., Yu, J., Lohman, T.M. and Ha, T. (2010) PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell*, **142**, 544-555.
136. Gupta, M.K., Guy, C.P., Yeeles, J.T., Atkinson, J., Bell, H., Lloyd, R.G., Marians, K.J. and McGlynn, P. (2013) Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli. *Proc Natl Acad Sci U S A*, **110**, 7252-7257.
137. Sternberg, S.H., Richter, H., Charpentier, E. and Qimron, U. (2016) Adaptation in CRISPR-Cas Systems. *Mol Cell*, **61**, 797-808.
138. Shiriaeva, A., Fedorov, I., Vyhovskyi, D. and Severinov, K. (2020) Detection of CRISPR adaptation. *Biochem Soc Trans*, **48**, 257-269.
139. Lau, C.H., Reeves, R. and Bolt, E.L. (2019) Adaptation processes that build CRISPR immunity: creative destruction, updated. *Essays in biochemistry*, **63**, 227-235.
140. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res*, **42**, 8516-8526.

## References

141. Fineran, P.C. and Charpentier, E. (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*, **434**, 202-209.
142. Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M. and Fineran, P.C. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun*, **7**, 12853.
143. Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun*, **3**, 945.
144. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. and Brouns, S.J. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet*, **9**, e1003742.
145. Westra, E.R., Swarts, D.C., Staals, R.H.J., Jore, M.M., Brouns, S.J.J. and Oost, J.v.d. (2012) The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. *Annual Review of Genetics*, **46**, 311-339.
146. 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.
147. Shiriaeva, A.A., Savitskaya, E., Datsenko, K.A., Vvedenskaya, I.O., Fedorova, I., Morozova, N., Metlitskaya, A., Sabantsev, A., Nickels, B.E., Severinov, K. *et al.* (2019) Detection of spacer precursors formed in vivo during primed CRISPR adaptation. *Nat Commun*, **10**, 4603.
148. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A. *et al.* (2011) A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol*, **79**, 484-502.
149. Ka, D., Hong, S., Jeong, U., Jeong, M., Suh, N., Suh, J.Y. and Bae, E. (2017) Structural and dynamic insights into the role of conformational switching in the nuclease activity of the *Xanthomonas albilineans* Cas2 in CRISPR-mediated adaptive immunity. *Struct Dyn*, **4**, 054701.
150. Yosef, I., Goren, M.G., Kiro, R., Edgar, R. and Qimron, U. (2011) High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc Natl Acad Sci U S A*, **108**, 20136-20141.
151. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.C., Hirai, A. *et al.* (2006) Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res*, **16**, 686-691.

## References

152. Stachler, A.E., Turgeman-Grott, I., Shtifman-Segal, E., Allers, T., Marchfelder, A. and Gophna, U. (2017) High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res*, **45**, 5208-5216.
153. Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U. and Sorek, R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505-510.
154. Ivancic-Bace, I., Radovicic, M., Bockor, L., Howard, J.L. and Bolt, E.L. (2013) Cas3 stimulates runaway replication of a ColE1 plasmid in Escherichia coli and antagonises RNaseHI. *RNA Biol*, **10**, 770-778.
155. Radovicic, M., Culo, A. and Ivancic-Bace, I. (2019) Cas3-stimulated runaway replication of modified ColE1 plasmids in Escherichia coli is temperature dependent. *FEMS Microbiol Lett*, **366**.
156. Cheok, C.F., Wu, L., Garcia, P.L., Janscak, P. and Hickson, I.D. (2005) The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic Acids Res*, **33**, 3932-3941.
157. Muftuoglu, M., Sharma, S., Thorslund, T., Stevnsner, T., Soerensen, M.M., Brosh, R.M., Jr. and Bohr, V.A. (2006) Cockayne syndrome group B protein has novel strand annealing and exchange activities. *Nucleic Acids Res*, **34**, 295-304.
158. Machwe, A., Xiao, L., Groden, J., Matson, S.W. and Orren, D.K. (2005) RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange. *J Biol Chem*, **280**, 23397-23407.
159. 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.
160. Louwen, R., Horst-Kreft, D., de Boer, A.G., van der Graaf, L., de Knegt, G., Hamersma, M., Heikema, A.P., Timms, A.R., Jacobs, B.C., Wagenaar, J.A. *et al.* (2013) A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain-Barre syndrome. *Eur J Clin Microbiol Infect Dis*, **32**, 207-226.
161. Viswanathan, P., Murphy, K., Julien, B., Garza, A.G. and Kroos, L. (2007) Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. *J Bacteriol*, **189**, 3738-3750.
162. Bozic, B., Repac, J. and Djordjevic, M. (2019) Endogenous Gene Regulation as a Predicted Main Function of Type I-E CRISPR/Cas System in E. coli. *Molecules*, **24**.
163. He, F., Vestergaard, G., Peng, W., She, Q. and Peng, X. (2017) CRISPR-Cas type I-A Cascade complex couples viral infection

## References

- surveillance to host transcriptional regulation in the dependence of Csa3b. *Nucleic Acids Res*, **45**, 1902-1913.
164. Cady, K.C. and O'Toole, G.A. (2011) Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J Bacteriol*, **193**, 3433-3445.
165. Heussler, G.E., Cady, K.C., Koeppen, K., Bhujju, S., Stanton, B.A. and O'Toole, G.A. (2015) Clustered Regularly Interspaced Short Palindromic Repeat-Dependent, Biofilm-Specific Death of *Pseudomonas aeruginosa* Mediated by Increased Expression of Phage-Related Genes. *mBio*, **6**, e00129-00115.
166. Bird, L.E., Håkansson, K., Pan, H. and Wigley, D.B. (1997) Characterization and crystallization of the helicase domain of bacteriophage T7 gene 4 protein. *Nucleic acids research*, **25**, 2620-2626.
167. He, L., Matosevic, Z.J., Mitic, D., Markulin, D., Killelea, T., Matkovic, M., Bertosa, B., Ivancic-Bace, I. and Bolt, E.L. (2021) A Tryptophan 'Gate' in the CRISPR-Cas3 Nuclease Controls ssDNA Entry into the Nuclease Site, That When Removed Results in Nuclease Hyperactivity. *Int J Mol Sci*, **22**.
168. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
169. Ryu, S.E. (2012) Structural mechanism of disulphide bond-mediated redox switches. *Journal of biochemistry*, **151**, 579-588.
170. Klomsiri, C., Karplus, P.A. and Poole, L.B. (2011) Cysteine-based redox switches in enzymes. *Antioxidants & redox signaling*, **14**, 1065-1077.
171. Cooper, L.A., Stringer, A.M. and Wade, J.T. (2018) Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation in the Type I-E CRISPR-Cas System. *mBio*, **9**.
172. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A*, **111**, E1629-1638.
173. Garrett, S., Shiimori, M., Watts, E.A., Clark, L., Graveley, B.R. and Terns, M.P. (2020) Primed CRISPR DNA uptake in *Pyrococcus furiosus*. *Nucleic Acids Res*.
174. Uversky, V.N. (2019) Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Frontiers in Physics*, **7**.
175. Bode, W., Fehlhämmer, H. and Huber, R. (1976) Crystal structure of bovine trypsinogen at 1-8 Å resolution. I. Data collection, application of Patterson search techniques and preliminary structural interpretation. *Journal of molecular biology*, **106**, 325-335.

## References

176. Yang, Z., Kollman, J.M., Pandi, L. and Doolittle, R.F. (2001) Crystal structure of native chicken fibrinogen at 2.7 Å resolution. *Biochemistry*, **40**, 12515-12523.
177. Oldfield, C.J., Xue, B., Van, Y.Y., Ulrich, E.L., Markley, J.L., Dunker, A.K. and Uversky, V.N. (2013) Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochimica et biophysica acta*, **1834**, 487-498.
178. Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*, **83**, 553-584.
179. Ladouceur, A.M., Parmar, B.S., Biedzinski, S., Wall, J., Tope, S.G., Cohn, D., Kim, A., Soubry, N., Reyes-Lamothe, R. and Weber, S.C. (2020) Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid-liquid phase separation. *Proc Natl Acad Sci U S A*, **117**, 18540-18549.
180. Jiang, S., Fagman, J.B., Chen, C., Alberti, S. and Liu, B. (2020) Protein phase separation and its role in tumorigenesis. *Elife*, **9**.
181. Du, M. and Chen, Z.J. (2018) DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science*, **361**, 704-709.
182. Perdikari, T.M., Murthy, A.C., Ryan, V.H., Watters, S., Naik, M.T. and Fawzi, N.L. (2020) SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *Embo j*, **39**, e106478.
183. Wright, R.H.G., Le Dily, F. and Beato, M. (2019) ATP, Mg(2+), Nuclear Phase Separation, and Genome Accessibility. *Trends Biochem Sci*, **44**, 565-574.
184. Weis, K. (2021) Dead or alive: DEAD-box ATPases as regulators of ribonucleoprotein complex condensation. *Biological chemistry*, **402**, 653-661.
185. Hondele, M., Sachdev, R., Heinrich, S., Wang, J., Vallotton, P., Fontoura, B.M.A. and Weis, K. (2019) DEAD-box ATPases are global regulators of phase-separated organelles. *Nature*, **573**, 144-148.
186. Li, X., Romero, P., Rani, M., Dunker, A.K. and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics*, **10**, 30-40.
187. Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. and Dunker, A.K. (2001) Sequence complexity of disordered protein. *Proteins*, **42**, 38-48.
188. Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D. and Cozzarelli, N.R. (2001) Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci U S A*, **98**, 8219-8226.
189. (2005), *DNA Repair and Mutagenesis*, pp. 9-69.

## References

190. Lampson, B.C., Viswanathan, M., Inouye, M. and Inouye, S. (1990) Reverse transcriptase from *Escherichia coli* exists as a complex with msDNA and is able to synthesize double-stranded DNA. *J Biol Chem*, **265**, 8490-8496.
191. Killelea, T., Saint-Pierre, C., Ralec, C., Gasparutto, D. and Henneke, G. (2014) Anomalous electrophoretic migration of short oligodeoxynucleotides labelled with 5'-terminal Cy5 dyes. *Electrophoresis*, **35**, 1938-1946.
192. Lewis, J.S., Jergic, S. and Dixon, N.E. (2016) The *E. coli* DNA Replication Fork. *The Enzymes*, **39**, 31-88.

# Appendices

## Appendix 1

### I-E subtype Cas3 sequences alignment

Consensus	x x P x - - x x F W A K x x x G L x - - x x x H P L x C H x L D V A A V A L x x W x x x L x x x L x x T I x x x x x x x x E x A x x W L A
Tfu	I P P L - - D L R F W A K E R - G L R - - G K T Y P L V C H S L D A A A A A L V L W N E Y L S P G L R D T I A S S M E T D E E H A G H C I A
Tter	W R V D - - P W I F W A K W G S G P D - - L G W H P L L C H M L D V A A V T L Q M W R R V L P A A W K A R I S G V L G V G Q E D A E R W L A
Eco	M E P F K Y I C H Y W G K S S K S L T K G N D I H L L I Y H C L D V A A V A D C W W D Q S V V - - L Q N T F C R N E M L S K Q R V K A W L L
Sequence Ruler	1 10 20 30 40 50 60
Sequence Logo	
Consensus	F F A G L H D I G K x x x x F Q x Q x A x x x x x - - - - - x x G x E L x x x - - - x x x x A G L x x F N x x x x H x x x x G x x L x F
Tfu	F W A G L H D I G K L T R E F Q Q I A I D L S A - - - - - Y P G E E L S G E - - - Q R - - - - - S H A A T G K W L P F
Tter	F F A G G H D I G K A S P A F Q L Q L R P E Q - - - - - G R E L V A R - - - R L R D A G L P L F N A R A P H G T I S A N V L E T
Eco	F F I A L H D I G K F D I R F Q Y K S A E S W L K L N P A T P S L N G P S T Q M C R K F N H G A A G L Y W F N Q D S L S E Q S L G D F F T
Sequence Ruler	70 80 90 100 110 120 130
Sequence Logo	
Consensus	x L A x x x x P x x G x x x x W V A x A V G G H H G F x x x x x - x x x S R x x L x x x x x S x x x x W D K Q - R E x L L x x V x D A L - G x
Tfu	A L P S L G Y P N G G L V T G L V A Q M L G G H H G T F H P H P - S F Q S R N P L A E F G F S S P H W E K Q - R H A L L H A V F D A T - G R
Tter	V L A - D V F G L S G R S A R W V A F A V G G H H G F V P S Y D - E V - - R R D L D Q Q A V G W G M W D A A - R E V L L C R L A D A L - G L
Eco	F D A A - P H P Y - E S W F P W V - E A V T G H H G F I L H S Q D Q D K S R W E M P A S L A S Y A A Q D K Q A R E E W I S - V L E A L F L T
Sequence Ruler	140 150 160 170 180 190 200
Sequence Logo	
Consensus	P x x x x P x x x x P x A S x L A G L V S L A D W L G S x E x x x x x x x x S x x A x x x x x A L R A Y F E x x x R x A x R x x x L L G x
Tfu	P T - - P P D M L D G P T A S V V C G L V I L A D W L V S Q E D F L L E R L T S L P A D G S A S A L R A H F E T S L R R I P - - S L L D A
Tter	P G S S R P T V E S T P D A F M L A G L V S V A D W I G S N E E Y F P Y A A Q S A L Q V P Q - L D A E A Y L E R A M R Q A E R A M A S L G W
Eco	P A G L S I N D I P P D C S S L L A G F C S L A D W L G S W T T T N T F L F N E D - A P S D I N A L R T Y F Q D R Q Q D A S R V L E L S G L
Sequence Ruler	210 220 230 240 250 260 270
Sequence Logo	
Consensus	V G x R P x x x G x x x x T E L F P x x x Q P x x L Q A x - A x x L x x L x x x P G L T x I E A P M G E G K T E A A x x x A D x L x x A x G
Tfu	A G L R P I T V P P A T F T E S F P H L S K P N G L Q A S L A K H L P C L C T G P G L V I T A P M G E G K T E A A Y H V A D L L G K A T G
Tter	V G W R P A S - G S M R L T E L F P Y I R Q P T T V Q A A - A E E L A G E V K S P S I T I I E A P M G E G K T E A A M L L A D T F S T A H G
Eco	V S N K R C Y E G - - V H A L L D N G Y Q P R Q L Q V - - - L V D A L P V A P G L T V I E A P T G S G K T E T A L A Y A W K L I D Q Q I
Sequence Ruler	280 290 300 310 320 330
Sequence Logo	
Consensus	x x G x x F A L P T M A T A N Q M x T R x x x Y A R H R x x x x L x x x L x H G x S x x x A L L Q S x x x x x - x x x x x L x G V x x x
Tfu	R P G R F L A L P T M A T A D Q M H T R L K E Y A R Y R V E N T D L P R S - - - - S T L A L L H S M A W L N P D Y A P A D L P G V S K V
Tter	M S G C Y F A L P T M A T S N Q M F G R V T D Y L R H R Y P E D V V V V N L V H G H S D L S A L L Q E L R Q K - - G E E I F Q L Q G V Y D E
Eco	A D S V I F A L P T Q A T A N A M L T R M E A S A S H L F S S P N L I - - L A H G N S R F N H L F Q S I K S R A - - - - -
Sequence Ruler	340 350 360 370 380 390
Sequence Logo	

# Appendices

Consensus	x x x G Q x - x x V x A x Q W L x x - x K R x L L x P x G V G T I D Q A L x A V L x V K H x F x R L F G L x x K V V I V D E V H A Y D x Y
Tfu	L S N L G H R - D P F A A T D W L M G - R K R G L L A P W A V G T I D Q A L M A V L R A K H N A L R L F G L A G K V V V V D E A H A V D P Y
Tter	A L G D E Q L - G A V V A G Q W F T R - G K R A L L P P Y G V G T V D Q A L L A V L Q V K H V F V R L F A L S T K T V I V D E V H A Y D V Y
Eco	I T E Q G Q E E A W V Q C C Q W L S Q S N K K V F L G Q I G V C T I D Q V L I S V L P V K H R F I R G L G I G R S V I V D E V H A Y D T Y
Sequence Ruler	400 410 420 430 440 450
Sequence Logo	ALEPQGEAAVAACQWIMSGGNKRALAPVGVGTIDQALAVIPVKHNEFLRLEGLAGKVVIVDEVHAYDXY LISGLTNGEHRREGWVFCDFIFTRSRNRKGFPGWVACVQALMSVGRARAVAVRGLAISTSTLVVAHAHAYDXY
Consensus	M x x L L E x L L x W L G x L x V P V V L L S A T L P x x x x x L V K A Y x x G A x x x x W x x x x x P x x x x x Y P x I x x x x x x x
Tfu	M Q V L L E Q L L R W L G T L D V P V V L L S A T L H H S I A N S L V K A Y L E G A R G R R W N R S E P Q P V S E V S Y P G W L H V D A R I
Tter	M T T L L H R L L E W L G A L S V P V V V L S A T L P S A R R R L V K A Y A R G A - - - G W Q A E R D L P P A - - G Y P R I T Y A - - - A
Eco	M N G L L E A V L K A Q A D V G S V I L L S A T L P M K Q K L L D T Y G L H T D P V - - - - - E N N S A Y P L I N W R G V N G
Sequence Ruler	470 480 490 500 510 520
Sequence Logo	MNGVTLLEALLEWLGALDVPVVLISATLPHALANEKLVKAYAEGLADGRGWNAEEDLPEANESGYPGLLHARNDANGA MOTVLEHRVRAQATVSGSVVLSATLHMKQKOKKLVKAYLRHTRPVRVQRSRPQVSNVSRWTVYVGRIT
Consensus	A - - - - - x D V D x L x x - P x x x x R x x L x x R x V x x P E - - x x x x x x x x x L x x x L x x A x x Q G G C A A I I C N T V x x A
Tfu	G K V T R S S D V D P L P I - - A T T P R K P L E V R L V D V P V - - K E G A L N R S T V L A K E L T P L V K Q G G C A A I I C T T V A E A
Tter	A - - - - - E D V R G I H F A P S E A S R R K V A L R W V S A P E - - - - - H E A L G Q L L A E A L S Q G G C A A I I C N T V P R A
Eco	A - - - - - Q R F D L L A H - P E - - - - - Q L P P R F S I Q P E P I C L A D M L P D L T M L E R M I A A A N A G A Q V C L I C N L V D V A
Sequence Ruler	530 540 550 560 570
Sequence Logo	AKVTRSSRDVDGLHFADEAPRKHLEPVEVDAPEPIKCEAALLPHEALAEELAAAKQGGCAAIICTVVAEVA GKVTSSSRFRFPPIAISTTSRKRQVPPVWSVVPVPIKLGDMNRSTVMLQRMTPLVSAQAQVCLICTLVVAEVA
Consensus	Q x x Y x x L x E x F x x L x E D x x P D L D L L H A R F P x x x R x E x E A R x x x x F G K N G x x - - R R P x x G A I L V A T Q V V E Q
Tfu	Q G V Y D L L S Q W F A T L G E D - A P D L Y L L H S R F P N R Q R T E I T A T I V D L F G K E G A Q S G R R P T R G A V L V A T Q V V E Q
Tter	Q A L Y S A L R E V F P G L A E D G M P E L D L L H A R Y P Y E E R E V R E A R T L G R F S R N G - - - - R R P - H R A I L V A T Q V I E Q
Eco	Q V C Y Q R L K E L N N - - - - T Q V D I D L F H A R F T L N D R R E K E N R V I S N F G K N G K R - - - - N V G R I L V A T Q V V E Q
Sequence Ruler	590 600 610 620 630
Sequence Logo	QAGVDALEKELFANGLAEDGMPDIDLHARFPLENDRERLEKEARLIDLGNFGKNGAQRRRPNRGAII VATQVVEQ QVLSRQSNWPTLGEDQVEVDFSSVTRRQVVRINTVVSRSREQRRSRRPVRVRRV
Consensus	S L D L D F D L M I T D L A P V D L L Q R x G R L H R H x x x x x - x R P x x x x x P E L x V L x P x x x G D x x x x P I F x R x x x S V
Tfu	S L D L D V D L M I S D L A P V S L L L Q R A G R C W R H E H L G I I N R P Q W A K Q P E L V V L T P E Q N G D A D R A P W F P R S W T S V
Tter	S L D L D F D L M V T D L A P V D L V L Q R M G R L H R H P V H D P - L R P E R L R S P E L W V V S P Q V M G D - - - V P I F D R G S A S V
Eco	S L D V D F D W L I T Q H C P A D L L F Q R L G R L H R H R K - - - - Y R P A G F E I P V A T I L L P D G E G Y G R H E H I Y S - - - - -
Sequence Ruler	650 660 670 680 690 700
Sequence Logo	SLDQDFDLMITDLAPVDLLQRAGRLHRHEHHRKDIYRPPAGAEIPELTVLSDGEEQMGGADHAPLFPGRGSASV VWLVVSHCASVFCWRCWRHHPVLPGLIYRPPQWLRQVAVIWTQVNGRRRHHVYSRWSV
Consensus	Y x x x L x R T x x x L R x x N G A x L Q x P E D x x Q L V D x V Y D D x x x x E x x x - - x M x x x - - - x E x x F K A x x x V L R
Tfu	Y P L A L L Q R T Y T L L R R R N G A P V Q I P E D V Q L V D D V Y D D S L A E D L E - - A D M E R M - - - G E E L A Q - - R G L A R
Tter	Y D E H T L L R S W L A L R D - - R D T L Q L P E D I E E L V E Q V Y S D G R V P Q G A S - - E E L R S L - - - W E R T F K A Q Q K V L R
Eco	- N V R V M W R T Q Q H I E E L N G A S L F F P D A Y R Q W L D S I Y D D A E M D E P E W V G N G M D K F E S A E C E K R F K A R K - V L -
Sequence Ruler	710 720 730 740 750 760
Sequence Logo	YREALLRRTQLALRLPENGAPLQFPEDIEQLVDDVYDDAELAFDAEAEAPMDFKFEELRFAKQKGLR VNLHTGQWGHLELERRRDTVFLIDAYREWLESISDGRMDLGESEERLRSMSAEWVRTAQAARRKRLR
Consensus	x x - - - - - x I x x P x - - x x x x W G x x E x S x x x D x x x x x x x x A x T R x G E
Tfu	N A - - - - - V I P D P D A E D N L N G L T E F S F D V D E H - - - - - V L A T R F G A
Tter	E D S L Q A K Y R Y I K G P G - - Y N S I W G I V T A S V E E D A P E L H P A L Q A L T R L A E
Eco	- - - - - Q W - A E E Y S L Q - - - - D N D E T I L A V T R D G E
Sequence Ruler	770 780 790
Sequence Logo	EANSLQAKRYIKDPPEDNLEWGFEEAFEDDAHDLDDEAIGLACTRDFGE NDSLQAKRYIKDPPEDNLEWGFEEAFEDDAHDLDDEAIGLACTRDFGE

Note: The protein sequences were aligned using Clustal Omega and the results were exported via Lasergene 17.

## Appendices

### Appendix 2

#### Cas3 sequence alignment from five different class I subtypes

CLUSTAL O(1.2.4) multiple sequence alignment

MthCas3	-----	0
PaeCas3	MNILLVSQCEKRALSETRRILDQFAERRGERTWQTPITQAGLDTLRRLL--KKS-----	53
EcoCas3	-----MEPFKYICHYWGKSSKSLT	19
TfuCas3	-----MPEHDSDDKHGIPPL--DLRFWAKER-GLR	28
TterCas3	MN-----GGPGMDGTSEV-----DLSGAGSPVGAAGWRVD--PWIFWAKWGS	43
MthCas3	-----	0
PaeCas3	--RRNTAVACHWIRGRDHSELLWIVGDASRFNAQGAVPTNRTCRDILRKEDE	111
EcoCas3	KGNDIHLLIYHCLDVAAVADCWWDQSVV-----LQNTFCRNEM-----	62
TfuCas3	--GKTYPLVCHSLDAAAAALVLWNEYLSPG-----LRDTIASSME-----	71
TterCas3	--LGWHPLLCHMLDVAAVTLQMWRRVLPAA-----WKARISGVLG-----	86
MthCas3	-----MASIIHDLGKIDYNFQKLMGDDSEAWEVLEEFLSPLKPLKRSPRHEILSIIW-	53
PaeCas3	IRLLTVMAALFHDIGKASQAFQAKLRNRGKPM-----ADAYRHEWVSLRLF	157
EcoCas3	VKAWLFFIALHDIGKFDIRFQYKSAESWLKLNPATPSLNGPSTQMCRFNHGAAGLYWF	122
TfuCas3	AGHCIAFWAGLHDIGKLTREFQQQIAIDLSA-----YPGEELSGE---QR-----	113
TterCas3	AERWLAFFAGGHDIKASPAFQLQLRPEQ-----GRELVAR---RLRDAGLPLF	132
	: **:** ** :	.
MthCas3	-----STFLGNDLDAM-----M-----	67
PaeCas3	EAFVGGSSDEDWLRRLADKRETGDAWLSQLARDDRQSAPPGPQKSRPLPLAQAVGWL	217
EcoCas3	NQDSLSEQLGDFFSFFDAA-----PHPY-----ESWFPWV-	153
TfuCas3	----SHAAATGKWLFPALPS-----LGYPNG-----GLVTGLVA	143
TterCas3	NARAPHGTISANVLETVLA-----DVFGLS-----GRSARWVA	165
MthCas3	-----RTAILLHHYNEYFLNDK-----DLM-EIIFTYRD---A-	96
PaeCas3	VSHHRLPNGDHRGSASLARLPAPIQSQWCGARDADAKEKAACWQFPHGLPFASAHWRART	277
EcoCas3	E----AVTGHGFFIL--HSQDQDKSRW-----EMP-ASLASAAQDKQA	190
TfuCas3	Q----MLGGHHGTFH--PHP-SFQSRN-----PLA-EFGFSSPHWEKQ-	178
TterCas3	F----AVGGHHGFVP--SYD-EV--RR-----DLD-QQAVGWGMWDA--	198
	:	:
MthCas3	FEIYLN----FIEKKETLRIFIEDILNYIQNSLESIDLVISAADEIRSNMDFEKPELLLE	152
PaeCas3	---ALC--AQSMLEPGLLAR-GPALLHDSYVMHVSRLILMLADH----H-----	319
EcoCas3	REEWIS-VLEALFLTPAGLS---INDIPPDCSSLLAG-FCSLADWLGSWTTNTFLFN--	243
TfuCas3	RHALLHAVFDAT-GRPT--P---PDMLDGPTASVCG-LVILADWLVSQEDFLLERLT--	229
TterCas3	REVLLCRLADAL-GLPGSSR---PTVESTPDFAFLAG-LVSVADWIGSNEEYFPYAAQ--	251
	:	. . **
MthCas3	KIKEYDDDISEFAAFYEPEE-RSADI-----LILSGIL-RRADYSASAGVDIELFSEEV	204
PaeCas3	-SLPADSRLGDPNFP LHANTDRDSGKLGKQLRDEHLLGVALHSRKL--AGTLPR--	373
EcoCas3	-ED-APSDINALRTRYFQDRQ-QDAS---RV-LELSGLVSNKRCY---EG-----	286
TfuCas3	-SLPADGSASALRAHFETSL-RRIP-----SLLDAAGLRPIT--VP-PAT--	273
TterCas3	-SALQVPQ-LDAEAYLERAM-RQAE---RA-MASLGWGWWRPAS-----G-SMR--	296
	.	.
MthCas3	FRDIDE----KITSKIGGAPWQIRLMGE-----LGGPKKMVLVAPTSGSKTEFSILW	252
PaeCas3	LPRLARHKGFTRRVEQPRFRWQDKAYDCAMACREQAMEHGFFGLNLASTGCGKTLANGRI	433
EcoCas3	LDNGYQPRQLQV-----LVDAL--PVA-----PGLTVEAPTSGSKTETALAY	327
TfuCas3	FPHLSKPNGLQASL-----AKHLPCL--CTG-----PGLVLITAPMGEGKTEAAYHV	318
TterCas3	FPYIRQPTTVQAA-----AEELAGE--VKS-----PSITIEEAPMGEGKTEAAML	340
	:	: * * **

## Appendices

MthCas3	A-----AKHGRKFIYTLPLRVALNDIFMRLRSDGY--FSEDEID--ILHSTA-F---IE	299
PaeCas3	LYALADPQRGARFSIALGL--RSLTLQTG-QAYRERLGLGDDDLA--ILVGGSAARELFE	488
EcoCas3	AWKLIDQQIADSVIFALPTQATANAMLTRMEASA-SHLFSSPNLI--LAHGNSRFNHLFQ	384
TfuCas3	ADLLGKATGRPGRF LALPTMATADQMTRLKEYA-RYRVENTDLPRS-----STLALLH	371
TterCas3	ADTFSTAHGMSGCYFALPTMATSNQMFRVTDYL-RHRYPEDVVVVNLVHGSDLSALLQ	399
	:* : . :	
MthCas3	YLKEE-----RLGRGTD-----LDSMMSARLMASPAL	327
PaeCas3	KQQERLERSGSESAQELLAENSHVHFAGTLEDGPL-----REWLGRNSAGNRLQLAPIL	542
EcoCas3	SIKSRA-----ITEQGQEEAWVQCCQWLSQ--SNKKVFLGQIG	420
TfuCas3	SMAWLN--DYA-PADLPG----VSKVLSNLGHR-DPFAATDWLMG--RKRGLLAPWA	419
TterCas3	ELRQK----GEE-IFQLQG----VYDEALGDEQL-GAVVAGQWFR---GKRALLPPYG	445
	: :	
MthCas3	LTPDQVLITSLNYFGSDKVISVYPF--ASMILDEIQTYNEEMAAVIKTLLELVNEVDGN	385
PaeCas3	ACTIDHLMPAESLRGGHQIAPLLRLMTSDLVLDEVDDFDIDDLPALSRVLVHWAGLFGSR	602
EcoCas3	VCTIDQVLIS--VLPVKHRFIRGLGIGRSVLI VDEVHAYDTYMNGLLEAVLKAQADVGG	478
TfuCas3	VTIDQALMA--VLRKHNALRFLGLAGKVVVVEAHAVDPYMQVLLLEQLLRWGLDVP	477
TterCas3	VGTVDQALLA--VLQVKHVFVRLFALSTKTVIVDEVHAYDVYMTLLHRLLEWLGAHSV	503
	* * : : . : : : * * . : : : . .	
MthCas3	ILVMTATLPPYFRSFLD-AMNFEVMDV-----AAIPGA-----	417
PaeCas3	VLLSSATLPPALVQGLFEAYRSGREIFQRHRGAPGRATEI---RCAWFDEFSSQSSAHGA	659
EcoCas3	VILLSATLPMKQKQLLDTYGLHTDPV-----ENNSAYPLINWRGVNGA-----	522
TfuCas3	VVLLSATLHHSIANSLVKAYLEGARRRNRSEPPVSEVSPGWLHVDARIG-----K	531
TterCas3	VVLSATLPSARRRELVKAYARGA---GWQAERDLPPA--GYPRITYA---AA-----	548
	: : : * * * * :	
MthCas3	----HDIKNLNL-----KRHVPQLIEE--PLFNDELEVS DKLKILDE-----	454
PaeCas3	VTSFSEAHATFVAQRLAKLEQLPPRRQAQLCTVHAAG-EARPALCRELAGQMNTWMADLH	718
EcoCas3	----QRFDLLAH-PE-----QLPPRFSIQPEPICLADMLPDLTMLERMIA-----	562
TfuCas3	VTRSSDVPDLPPI--ATTPRKPLEVRLVDVPV--KEGALNRSTVLAKELTP-----	577
TterCas3	----EDVRGIHFAPSEASRRKVALRWVSAPE-----HEALGQLLAE-----	585
	: : :	
MthCas3	-----NSEKNV-----LIVVNVQKAIELYR---EY-----QDDPDVYLLHSRLL	492
PaeCas3	RCHHTEHQRRISFGLLRLANIEPLIELAQAI-----LAQGAPEGLHVHLCVYHSRHP	772
EcoCas3	----AANAGAQC-----LICNLVDVAQVCYQRLKELNN-----TQVDIDLHARFTL	606
TfuCas3	----LVKQGGCAA-----IICTTVAEAQGVYDLSQWFATLGED-APDLYLLHSRFPN	625
TterCas3	----ALSQGGCAA-----IICNTVPRAQALYSALREVFPGLAEDGMPELDLLHARYPY	634
	: : * : : * * *	
MthCas3	KVKSQR---IGEVKKRSQ-----DERGLTVISTQIIIEASVDIDF	528
PaeCas3	LVRSAIERQLDELLKRSDDDAALFARPTLAKALQASTERDHLFVVLASPAEVEGRDHDY	832
EcoCas3	NDRREKENRVISNFGKNGKR-----NVGRILVATQVVEQSLDVDF	646
TfuCas3	RQREITATIVDLFGKEGAQ-----SGRRPTRGAVLVATQVVEQSLDLVDV	670
TterCas3	EEREVREARTLGRFSRNG-----RRP-HRAILVATQVIEQSLDLDF	674
	: : . : : : * . * *	
MthCas3	DLMITEISTIDSQIRWGRIRNRDADY-----DSGDPNIIIFTSDR---RTSLIYDKK	580
PaeCas3	DWAIVEPSSMRSIIQLAGRIRRRHSFGS-----GEANLYLLSRNIRSLEGQNPQAFQRP	885
EcoCas3	DWLITQHCPADLLFQRLGRLHRHRK---YRPAGFEIPVATILLPDGEGYGRHEHIYS--	701
TfuCas3	DLMISDLAPVSLLLQRAGRCWRHEHLGIINRPQWAKQPELVVLTPEQNGADRAPWFFRS	730
TterCas3	DLMVTDLAPVDLVLQRMGRLLHRHPVHDP-LRPERLRSPELWVWVSPQVMGD---VPIFDRG	730
	* : : . : * * * * : . . :	
MthCas3	VLDATRAILERYDQGILDYNLERSMIEEVFQEEIDGSTLKEIYENQIRETISDLDYFTVE	640
PaeCas3	GFET-----PDFPLDSDL-----HDLDP-----A	906
EcoCas3	-----NV-----R	704
TfuCas3	W-----TS-----VYPL-----A	738
TterCas3	S-----AS-----VYDE-----H	738

## Appendices

MthCas3	KRTQAQRLFRNMAGYKVFIPDAVLRyseSEIERT -FAE-----LIKGDYRLWKDILGEIER	695
PaeCas3	LLARIDASPRIVEPFPLFPRSRLVDLEHRRRLRALMLADDPSSLLGVPLWQTPAS -L--	963
EcoCas3	VMWRTQQHIEELNGASLFFPDAYR-----QWLDSIYDD-----AEMDEPEWVGNMGDKFES	755
TfuCas3	LLQRTYTLRRRRNGAPVQIPEDVQ-----QLVDDVYDD-----DSLAEDE--ADMERM--	785
TterCas3	TLLRSWLALRD--RDTLQLPEDIE-----ELVEQVYSD-----GRVPQGAS--EELRSL--	783
	: . : . . :	
MthCas3	I-----TGEKTTMWDLKKVLYEYSVNVV -----VF	720
PaeCas3	---SGALQT-SQPFRAGAKERCYALLPDEDDEER-----LHF	996
EcoCas3	AECEKRFKARK-VL-----QW-AEEYSLQ-----DNDETILAV	786
TfuCas3	--GEELAQ--RGLARNA-----VIPDPDDAEDNLNGLTEFSFD--VDEH-----VLA	826
TterCas3	--WERTFKAQQKVLREDSLQAKYRYIKGPG--YNSIWGIVTASVE--EDAPELHPALQAL	837
MthCas3	YEEKSDFWSRTTGE-----FKGFYVWGSMEDDD-----	748
PaeCas3	_SRYEETWSNODNLLRNLDLTYGPRIOTW-----GTVNYREELVAMAGREDLDLROCA	1049
EcoCas3	TRDGEMSLPL-----LPYVQTSSGKQLLDGQVYEDLSHEQQY-----	823
TfuCas3	TRFGAGSVRV-----LCYYVDTAGNRWLDPECTVEFPEQGTGREGRFTMADCRDLVARTIP	882
TterCas3	TRLAEPSVSA----VCLVAGSGGPCLPD-GTPVDLDTPPDA-----AMAERLLRRSVA	885
	.	
MthCas3	-----VELLEELGLDSIFGETESSLIV-----	770
PaeCas3	MRYGEVRL-----RENTQ	1062
EcoCas3	-----EALALNRVNVVFTWKRSEV-----DE-----DGLLWLEGKQNL	860
TfuCas3	VRMGPWASQLTEDNHPPPEAWRESFYLRDLVLIQQRVTDEGAVLPTETGGREW--LDPCK	940
TterCas3	ITDARV-LDPLLDVVPKGVWERSLLRGRPLVFDAS--GRA-----MVGRWIVRIDPEL	937
MthCas3	-----	770
PaeCas3	GWSYHPYLGFKKYN-----	1076
EcoCas3	GWVWQGNSIVITYTGDEGMTRVIPANPK	888
TfuCas3	GLIF-----	944
TterCas3	GIVVES-----P-	944

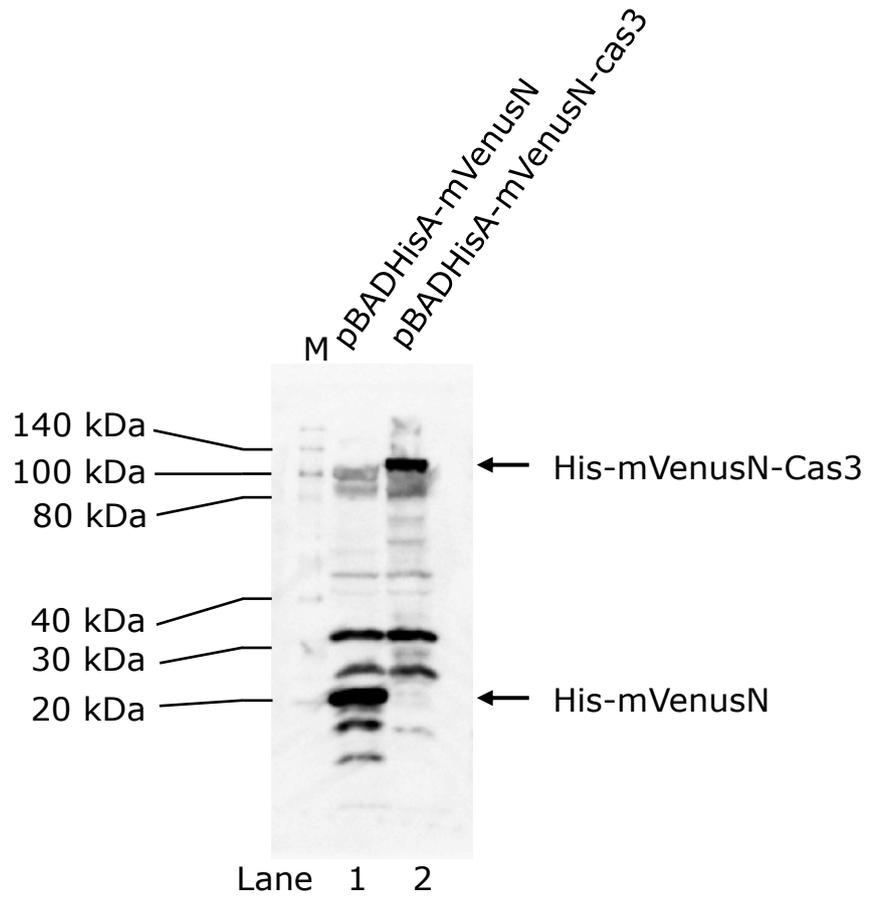
*Note: The protein sequences were aligned using Clustal Omega.*

[https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20191118-144133-0028-56802634-p1m/aln-clustal\\_num](https://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20191118-144133-0028-56802634-p1m/aln-clustal_num)

Appendices

Appendix 3

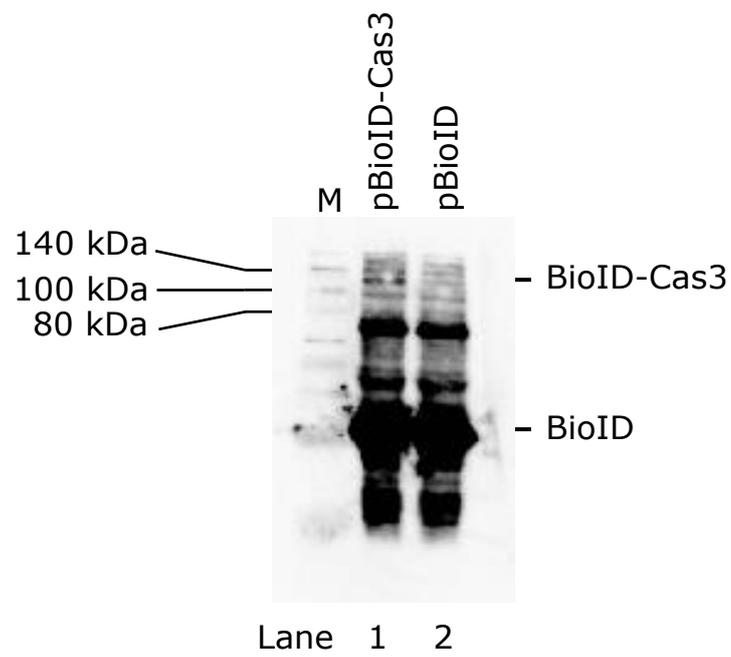
Expression of mVenus-N-Cas3 in BL21AI, result was validated by western blot



Appendices

Appendix 4

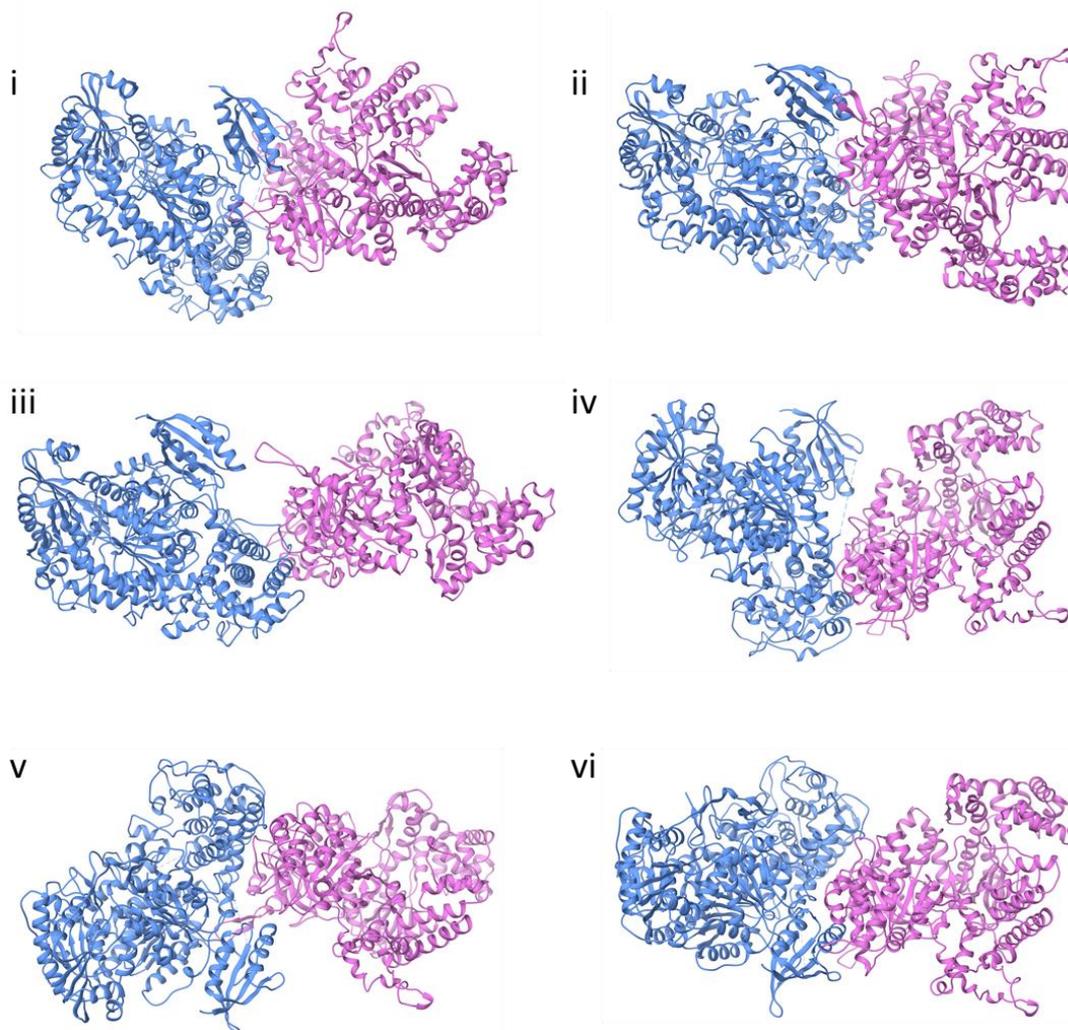
Expression of BioID-Cas3 in BL21AI, result was validated by western blot



## Appendices

### Appendix 5

Images representing the *PaeCas2-3* and *PaePolI* docking results

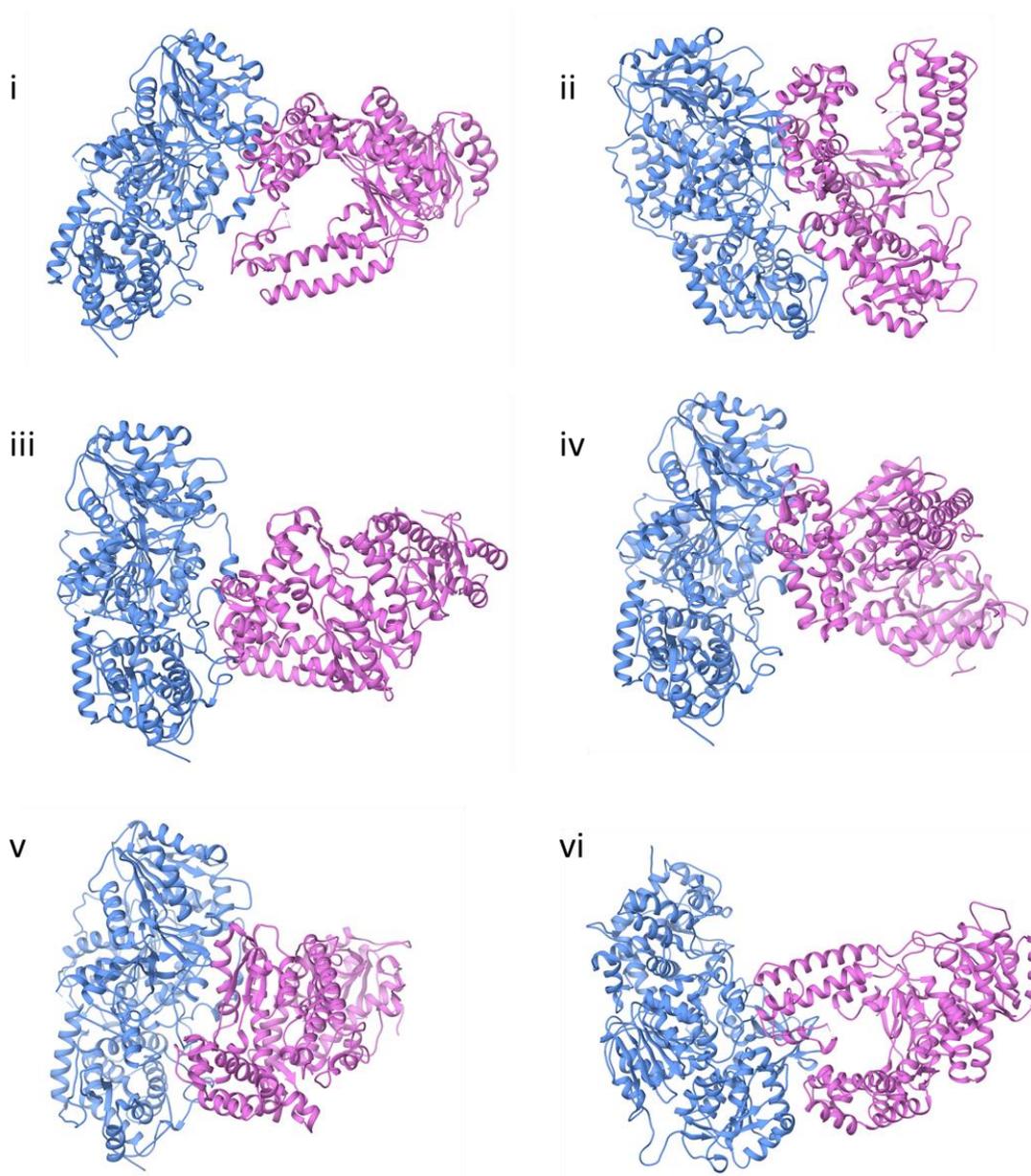


Note: *PaeCas2-3* is denoted in blue and *PaePolI* is coloured in orchid. Protein docking was conducted using *PaeCas2-3* structure (PDB: 5GQH) and predicted *PaePolI* structure by AlphaFold. Results were generated by online docking service ClusPro2. <https://cluspro.bu.edu/publications.php>

## Appendices

### Appendix 6

Images representing the *EcoCas3* and the Klenow fragment of *EcoPolI* docking results



*Note: EcoCas3 is denoted in blue and EcoPolI Klenow fragment is coloured in orchid. Protein docking was conducted using EcoCas33 structure predicted by AlphaFold and EcoPolI Klenow fragment structure (PDB: 1D8Y). Results were generated by online docking service ClusPro2.*

<https://cluspro.bu.edu/publications.php>

## Appendices

### Appendix 7

#### Mass spectrometry result

Identified Proteins	Accession Number	Alternate ID	Molecular Weight
CRISPR-associated endonuclease/helicase Cas3 OS=Escherichia coli (strain K12) OX=83333 GN=ygcB PE=1 SV=2	P38036	ygcB	101 kDa
Maltose/maltodextrin-binding periplasmic protein OS=Escherichia coli (strain K12) OX=83333 GN=malE PE=1 SV=1	P0AEX9	malE	43 kDa
Pyruvate dehydrogenase E1 component OS=Escherichia coli (strain K12) OX=83333 GN=aceE PE=1 SV=2	P0AFG8	aceE	100 kDa
Chaperone protein DnaK OS=Escherichia coli (strain K12) OX=83333 GN=dnaK PE=1 SV=2	P0A6Y8	dnaK	69 kDa
30S ribosomal protein S1 OS=Escherichia coli (strain K12) OX=83333 GN=rpsA PE=1 SV=1	P0AG67	rpsA	61 kDa
Lon protease OS=Escherichia coli (strain K12) OX=83333 GN=lon PE=1 SV=1	P0A9M0	lon	87 kDa
NADH-quinone oxidoreductase subunit G OS=Escherichia coli (strain K12) OX=83333 GN=nuoG PE=1 SV=4	P33602	nuoG	100 kDa
Glyceraldehyde-3-phosphate dehydrogenase A OS=Escherichia coli (strain K12) OX=83333 GN=gapA PE=1 SV=2	P0A9B2	gapA	36 kDa
Bifunctional protein PutA OS=Escherichia coli (strain K12) OX=83333 GN=putA PE=1 SV=3	P09546	putA	144 kDa
Cluster of Formate acetyltransferase 1 OS=Escherichia coli (strain K12) OX=83333 GN=pflB PE=1 SV=2 (P09373)	P09373 [2]	pflB	85 kDa
Formate acetyltransferase 1 OS=Escherichia coli (strain K12) OX=83333 GN=pflB PE=1 SV=2	P09373	pflB	85 kDa
PFL-like enzyme TdcE OS=Escherichia coli (strain K12) OX=83333 GN=tdcE PE=1 SV=2	P42632	tdcE	86 kDa
Keratin, type I cytoskeletal 15 OS=Ovis aries GN=KRT15 PE=2 SV=1	cRAP040	KRT15	49 kDa
Glutamate synthase [NADPH] large chain OS=Escherichia coli (strain K12) OX=83333 GN=gltB PE=1 SV=3	P09831	gltB	163 kDa
Pyruvate kinase I OS=Escherichia coli (strain K12) OX=83333 GN=pykF PE=1 SV=1	P0AD61	pykF	51 kDa
50S ribosomal protein L2 OS=Escherichia coli (strain K12) OX=83333 GN=rplB PE=1 SV=2	P60422	rplB	30 kDa
Lactose operon repressor OS=Escherichia coli (strain K12) OX=83333 GN=lacI PE=1 SV=3	P03023	lacI	39 kDa

## Appendices

Translation initiation factor IF-2 OS=Escherichia coli (strain K12) OX=83333 GN=infB PE=1 SV=1	P0A705	infB	97 kDa
Alanine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=alaS PE=1 SV=2	P00957	alaS	96 kDa
Cluster of Trypsin OS=Sus scrofa PE=1 SV=1 (cRAP112)	cRAP112 [2]		24 kDa
Trypsin OS=Sus scrofa PE=1 SV=1	cRAP112		24 kDa
Cationic trypsin OS=Bos taurus PE=1 SV=3	cRAP110		26 kDa
Glycine dehydrogenase (decarboxylating) OS=Escherichia coli (strain K12) OX=83333 GN=gcvP PE=1 SV=3	P33195	gcvP	104 kDa
Chaperone protein ClpB OS=Escherichia coli (strain K12) OX=83333 GN=clpB PE=1 SV=1	P63284	clpB	96 kDa
50S ribosomal protein L1 OS=Escherichia coli (strain K12) OX=83333 GN=rplA PE=1 SV=2	P0A7L0	rplA	25 kDa
50S ribosomal protein L9 OS=Escherichia coli (strain K12) OX=83333 GN=rplI PE=1 SV=1	P0A7R1	rplI	16 kDa
Elongation factor Tu 1 OS=Escherichia coli (strain K12) OX=83333 GN=tufA PE=1 SV=1	P0CE47 (+1)	tufA	43 kDa
60 kDa chaperonin OS=Escherichia coli (strain K12) OX=83333 GN=groL PE=1 SV=2	P0A6F5	groL	57 kDa
DNA protection during starvation protein OS=Escherichia coli (strain K12) OX=83333 GN=dps PE=1 SV=2	P0ABT2	dps	19 kDa
Tryptophanase OS=Escherichia coli (strain K12) OX=83333 GN=tnaA PE=1 SV=1	P0A853	tnaA	53 kDa
NADH-quinone oxidoreductase subunit C/D OS=Escherichia coli (strain K12) OX=83333 GN=nuoC PE=1 SV=3	P33599	nuoC	68 kDa
Periplasmic pH-dependent serine endoprotease DegQ OS=Escherichia coli (strain K12) OX=83333 GN=degQ PE=1 SV=1	P39099	degQ	47 kDa
4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase OS=Escherichia coli (strain K12) OX=83333 GN=kdul PE=1 SV=1	Q46938	kdul	31 kDa
Succinate--CoA ligase [ADP-forming] subunit beta OS=Escherichia coli (strain K12) OX=83333 GN=sucC PE=1 SV=1	P0A836	sucC	41 kDa
DNA gyrase subunit B OS=Escherichia coli (strain K12) OX=83333 GN=gyrB PE=1 SV=2	P0AES6	gyrB	90 kDa
Maltodextrin phosphorylase OS=Escherichia coli (strain K12) OX=83333 GN=malP PE=1 SV=7	P00490	malP	91 kDa
Methionine synthase OS=Escherichia coli (strain K12) OX=83333 GN=metH PE=1 SV=5	P13009	metH	136 kDa
30S ribosomal protein S4 OS=Escherichia coli (strain K12) OX=83333 GN=rpsD PE=1 SV=2	P0A7V8	rpsD	23 kDa
Chaperone protein DnaJ OS=Escherichia coli (strain K12) OX=83333 GN=dnaJ PE=1 SV=3	P08622	dnaJ	41 kDa
Dihydrolipoyl dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=lpdA PE=1 SV=2	P0A9P0	lpdA	51 kDa

## Appendices

Isoleucine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=ileS PE=1 SV=5	P00956	ileS	104 kDa
ATP-dependent Clp protease ATP-binding subunit ClpX OS=Escherichia coli (strain K12) OX=83333 GN=clpX PE=1 SV=2	P0A6H1	clpX	46 kDa
Trigger factor OS=Escherichia coli (strain K12) OX=83333 GN=tig PE=1 SV=1	P0A850	tig	48 kDa
P-protein OS=Escherichia coli (strain K12) OX=83333 GN=pheA PE=1 SV=1	P0A9J8	pheA	43 kDa
ATP-dependent Clp protease ATP-binding subunit ClpA OS=Escherichia coli (strain K12) OX=83333 GN=clpA PE=1 SV=1	P0ABH9	clpA	84 kDa
DNA polymerase I OS=Escherichia coli (strain K12) OX=83333 GN=polA PE=1 SV=1	P00582	polA	103 kDa
Pyruvate dehydrogenase [ubiquinone] OS=Escherichia coli (strain K12) OX=83333 GN=poxB PE=1 SV=1	P07003	poxB	62 kDa
NADH-quinone oxidoreductase subunit F OS=Escherichia coli (strain K12) OX=83333 GN=nuoF PE=1 SV=3	P31979	nuoF	49 kDa
50S ribosomal protein L7/L12 OS=Escherichia coli (strain K12) OX=83333 GN=rplL PE=1 SV=2	P0A7K2	rplL	12 kDa
Glycine--tRNA ligase beta subunit OS=Escherichia coli (strain K12) OX=83333 GN=glyS PE=1 SV=4	P00961	glyS	77 kDa
Polyribonucleotide nucleotidyltransferase OS=Escherichia coli (strain K12) OX=83333 GN=pnp PE=1 SV=3	P05055	pnp	77 kDa
2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=kduD PE=1 SV=2	P37769	kduD	27 kDa
Threonine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=thrS PE=1 SV=1	P0A8M3	thrS	74 kDa
ADP-L-glycero-D-manno-heptose-6-epimerase OS=Escherichia coli (strain K12) OX=83333 GN=hldD PE=1 SV=1	P67910	hldD	35 kDa
50S ribosomal protein L21 OS=Escherichia coli (strain K12) OX=83333 GN=rplU PE=1 SV=1	P0AG48	rplU	12 kDa
50S ribosomal protein L10 OS=Escherichia coli (strain K12) OX=83333 GN=rplJ PE=1 SV=2	P0A7J3	rplJ	18 kDa
50S ribosomal protein L20 OS=Escherichia coli (strain K12) OX=83333 GN=rplT PE=1 SV=2	P0A7L3	rplT	13 kDa
50S ribosomal protein L17 OS=Escherichia coli (strain K12) OX=83333 GN=rplQ PE=1 SV=1	P0AG44	rplQ	14 kDa
Proline--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=proS PE=1 SV=4	P16659	proS	64 kDa
50S ribosomal protein L15 OS=Escherichia coli (strain K12) OX=83333 GN=rplO PE=1 SV=1	P02413	rplO	15 kDa
Lysine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=lysS PE=1 SV=2	P0A8N3	lysS	58 kDa
50S ribosomal protein L13 OS=Escherichia coli (strain K12) OX=83333 GN=rplM PE=1 SV=1	P0AA10	rplM	16 kDa

## Appendices

Aspartate--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=aspS PE=1 SV=1	P21889	aspS	66 kDa
D-3-phosphoglycerate dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=serA PE=1 SV=2	P0A9T0	serA	44 kDa
Purine nucleoside phosphorylase DeoD-type OS=Escherichia coli (strain K12) OX=83333 GN=deoD PE=1 SV=2	P0ABP8	deoD	26 kDa
Chaperone protein Skp OS=Escherichia coli (strain K12) OX=83333 GN=skp PE=1 SV=1	P0AEU7	skp	18 kDa
Pyruvate kinase II OS=Escherichia coli (strain K12) OX=83333 GN=pykA PE=1 SV=3	P21599	pykA	51 kDa
Protein GrpE OS=Escherichia coli (strain K12) OX=83333 GN=grpE PE=1 SV=1	P09372	grpE	22 kDa
Small heat shock protein IbpB OS=Escherichia coli (strain K12) OX=83333 GN=ibpB PE=1 SV=1	P0C058	ibpB	16 kDa
Isocitrate lyase OS=Escherichia coli (strain K12) OX=83333 GN=aceA PE=1 SV=1	P0A9G6	aceA	48 kDa
Protein-lysine deacetylase OS=Escherichia coli (strain K12) OX=83333 GN=dhaM PE=1 SV=3	P37349	dhaM	51 kDa
50S ribosomal protein L3 OS=Escherichia coli (strain K12) OX=83333 GN=rplC PE=1 SV=1	P60438	rplC	22 kDa
30S ribosomal protein S18 OS=Escherichia coli (strain K12) OX=83333 GN=rpsR PE=1 SV=2	P0A7T7	rpsR	9 kDa
Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha OS=Escherichia coli (strain K12) OX=83333 GN=accA PE=1 SV=2	P0ABD5	accA	35 kDa
Methionine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=metG PE=1 SV=2	P00959	metG	76 kDa
Valine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=valS PE=1 SV=2	P07118	valS	108 kDa
DNA topoisomerase 1 OS=Escherichia coli (strain K12) OX=83333 GN=topA PE=1 SV=2	P06612	topA	97 kDa
30S ribosomal protein S15 OS=Escherichia coli (strain K12) OX=83333 GN=rpsO PE=1 SV=2	P0ADZ4	rpsO	10 kDa
Small heat shock protein IbpA OS=Escherichia coli (strain K12) OX=83333 GN=ibpA PE=1 SV=1	P0C054	ibpA	16 kDa
Catalase HP11 OS=Escherichia coli (strain K12) OX=83333 GN=katE PE=1 SV=1	P21179	katE	84 kDa
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex OS=Escherichia coli (strain K12) OX=83333 GN=aceF PE=1 SV=3	P06959	aceF	66 kDa
30S ribosomal protein S6 OS=Escherichia coli (strain K12) OX=83333 GN=rpsF PE=1 SV=1	P02358	rpsF	16 kDa
NADH-quinone oxidoreductase subunit E OS=Escherichia coli (strain K12) OX=83333 GN=nuoE PE=1 SV=1	P0AFD1	nuoE	19 kDa
Bifunctional protein GlmU OS=Escherichia coli (strain K12) OX=83333 GN=glmU PE=1 SV=1	P0ACC7	glmU	49 kDa

## Appendices

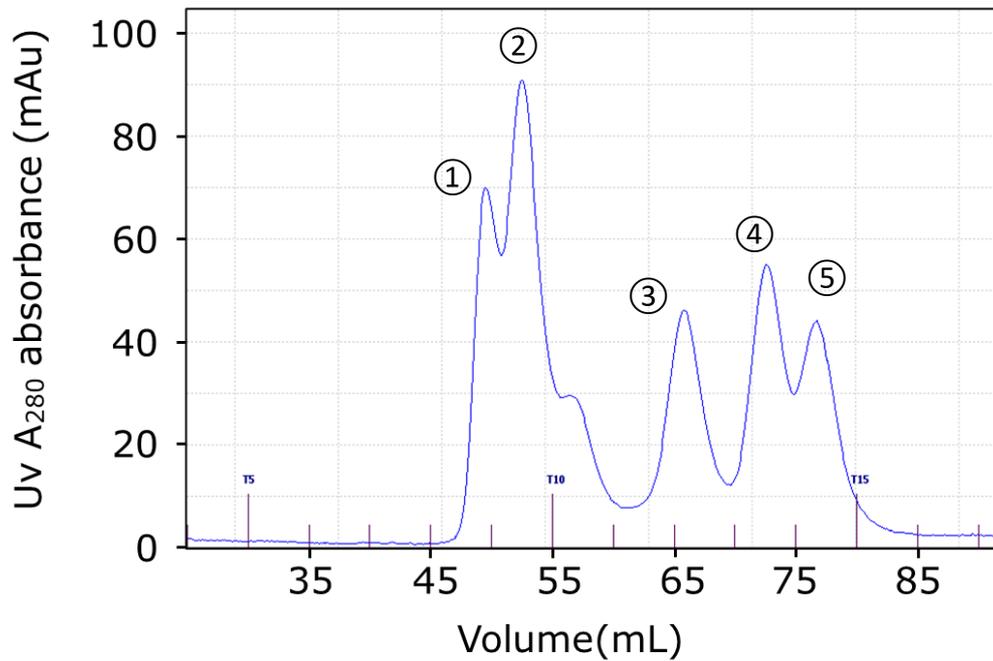
DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) OX=83333 GN=rpoB PE=1 SV=1	P0A8V2	rpoB	151 kDa
Protein-export protein SecB OS=Escherichia coli (strain K12) OX=83333 GN=secB PE=1 SV=1	P0AG86	secB	17 kDa
NADP-specific glutamate dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=gdhA PE=1 SV=1	P00370	gdhA	49 kDa
30S ribosomal protein S7 OS=Escherichia coli (strain K12) OX=83333 GN=rpsG PE=1 SV=3	P02359	rpsG	20 kDa
Glutamate synthase [NADPH] small chain OS=Escherichia coli (strain K12) OX=83333 GN=gltD PE=1 SV=3	P09832	gltD	52 kDa
Chaperone protein HtpG OS=Escherichia coli (strain K12) OX=83333 GN=htpG PE=1 SV=1	P0A6Z3	htpG	71 kDa
Pyridoxine 5'-phosphate synthase OS=Escherichia coli (strain K12) OX=83333 GN=pdxJ PE=1 SV=2	P0A794	pdxJ	26 kDa
50S ribosomal protein L11 OS=Escherichia coli (strain K12) OX=83333 GN=rplK PE=1 SV=2	P0A7J7	rplK	15 kDa
30S ribosomal protein S9 OS=Escherichia coli (strain K12) OX=83333 GN=rpsI PE=1 SV=2	P0A7X3	rpsI	15 kDa
Glucose-6-phosphate 1-dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=zwf PE=1 SV=1	P0AC53	zwf	56 kDa
Glycogen phosphorylase OS=Escherichia coli (strain K12) OX=83333 GN=glgP PE=3 SV=1	P0AC86	glgP	93 kDa
GTP cyclohydrolase 1 type 2 homolog OS=Escherichia coli (strain K12) OX=83333 GN=ybgl PE=1 SV=1	P0AFP6	ybgl	27 kDa
Stringent starvation protein B OS=Escherichia coli (strain K12) OX=83333 GN=sspB PE=1 SV=1	P0AFZ3	sspB	18 kDa
Succinate--CoA ligase [ADP-forming] subunit alpha OS=Escherichia coli (strain K12) OX=83333 GN=sucD PE=1 SV=2	P0AGE9	sucD	30 kDa
Tyrosine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=tyrS PE=1 SV=2	P0AGJ9	tyrS	48 kDa
Selenocysteine-specific elongation factor OS=Escherichia coli (strain K12) OX=83333 GN=selB PE=1 SV=3	P14081	selB	69 kDa
PTS system mannose-specific EIIAB component OS=Escherichia coli (strain K12) OX=83333 GN=manX PE=1 SV=2	P69797	manX	35 kDa
PKHD-type hydroxylase YbiX OS=Escherichia coli (strain K12) OX=83333 GN=ybiX PE=3 SV=2	P75779	ybiX	26 kDa
tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG OS=Escherichia coli (strain K12) OX=83333 GN=mnmG PE=1 SV=1	P0A6U3	mnmG	70 kDa
30S ribosomal protein S13 OS=Escherichia coli (strain K12) OX=83333 GN=rpsM PE=1 SV=2	P0A7S9	rpsM	13 kDa
30S ribosomal protein S5 OS=Escherichia coli (strain K12) OX=83333 GN=rpsE PE=1 SV=2	P0A7W1	rpsE	18 kDa

## Appendices

UDP-3-O-(3-hydroxymyristoyl)glucosamine N-acyltransferase OS=Escherichia coli (strain K12) OX=83333 GN=lpxD PE=1 SV=2	P21645	lpxD	36 kDa
Lactaldehyde dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=aldA PE=1 SV=2	P25553	aldA	52 kDa
Ribosomal RNA large subunit methyltransferase K/L OS=Escherichia coli (strain K12) OX=83333 GN=rlmL PE=1 SV=1	P75864	rlmL	79 kDa
S-adenosylmethionine synthase OS=Escherichia coli (strain K12) OX=83333 GN=metK PE=1 SV=2	P0A817	metK	42 kDa
4-aminobutyrate aminotransferase GabT OS=Escherichia coli (strain K12) OX=83333 GN=gabT PE=1 SV=1	P22256	gabT	46 kDa
ATP-dependent protease subunit HslV OS=Escherichia coli (strain K12) OX=83333 GN=hslV PE=1 SV=2	P0A7B8	hslV	19 kDa
Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta OS=Escherichia coli (strain K12) OX=83333 GN=accD PE=1 SV=1	P0A9Q5	accD	33 kDa
RNA chaperone ProQ OS=Escherichia coli (strain K12) OX=83333 GN=proQ PE=1 SV=2	P45577	proQ	26 kDa
Branched-chain-amino-acid aminotransferase OS=Escherichia coli (strain K12) OX=83333 GN=ilvE PE=1 SV=2	P0AB80	ilvE	34 kDa
Acyl carrier protein OS=Escherichia coli (strain K12) OX=83333 GN=acpP PE=1 SV=2	P0A6A8	acpP	9 kDa
ATP-dependent RNA helicase SrmB OS=Escherichia coli (strain K12) OX=83333 GN=srmB PE=1 SV=1	P21507	srmB	50 kDa
ATP synthase subunit alpha OS=Escherichia coli (strain K12) OX=83333 GN=atpA PE=1 SV=1	P0ABB0	atpA	55 kDa
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4	cRAP087	ALB	69 kDa
Bifunctional glutamine synthetase adenylyltransferase/adenylyl-removing enzyme OS=Escherichia coli (strain K12) OX=83333 GN=glnE PE=1 SV=2	P30870	glnE	108 kDa
Alpha-S2-casein OS=Bos taurus GN=CSN1S2 PE=1 SV=2	cRAP014	CSN1S2	26 kDa
Kappa-casein OS=Bos taurus GN=CSN3 PE=1 SV=1	cRAP016	CSN3	21 kDa
Bifunctional aspartokinase/homoserine dehydrogenase 2 OS=Escherichia coli (strain K12) OX=83333 GN=metL PE=1 SV=3	P00562	metL	89 kDa
Lysine-sensitive aspartokinase 3 OS=Escherichia coli (strain K12) OX=83333 GN=lysC PE=1 SV=2	P08660	lysC	49 kDa
Alpha-S1-casein OS=Bos taurus GN=CSN1S1 PE=1 SV=2	cRAP013	CSN1S1	25 kDa

Appendix 8

Size exclusion chromatography showing a calibration standard (Cytiva 28-4038-42) including Thyroglobulin (669 kDa, ①), Ferritin (440 kDa, ②), Aldolase (158 kDa, ③), Conalbumin (75 kDa, ④) and Ovalbumin (43 kDa, ⑤), eluted separately from the HiLoad 16/600 Superdex 200 pg preparative SEC column used in this study.



## Appendices

### Appendix 9

Quantification was done by using GelAnalyzer v19.1.

**Figure 3.3(A).** In lanes 3-6 showing 14 nM, 28 nM, 56 nM and 112 nM of wild-type EcoCas3 converted 15.5 %, 23.3 %, 38 % and 43.6 % of 200 ng M13ssDNA into degradation products at 30 °C. In lanes 8-11 showing 14 nM, 28 nM, 56 nM and 112 nM of wild-type Eco Cas3 converted 0.5 %, 6.2 %, 17.9 % and 23.7 % of 200 ng M13ssDNA into DNA products at 37 °C. In lane 1 showing 0.01 U DNase I degraded 59.3 % of M13ssDNA (200 ng).

**Figure 3.3(C).** In lanes 8-12 showing 56 nM of wild-type EcoCas3, EcoCas3<sup>W149A</sup>, EcoCas3<sup>W152A</sup>, EcoCas3<sup>W230A</sup> and EcoCas3<sup>W406A</sup> converted 81.9 %, 50 %, 71.9 %, 65.3 % and 88.5 % of 20 nM DNA fork into DNA products, reactions were incubated at 30 °C. This is shown on 10 % polyacrylamide TBE denaturing gel.

**Figure 3.3(D).** In lanes 2-6 showing 56 nM of wild-type EcoCas3, EcoCas3<sup>W149A</sup>, EcoCas3<sup>W152A</sup>, EcoCas3<sup>W230A</sup> and EcoCas3<sup>W406A</sup> converted 86.2 %, 34.6 %, 55.9 %, 65.6 % and 86.6 % of 20 nM DNA fork into DNA products, reactions were incubated at 30 °C. This is shown on 10 % polyacrylamide TBE gel.

**Figure 3.3(E).** In lanes 2-6 showing 112 nM of wild-type EcoCas3, EcoCas3<sup>W149A</sup>, EcoCas3<sup>W152A</sup>, EcoCas3<sup>W230A</sup> and EcoCas3<sup>W406A</sup> converted 86.2 %, 34.6 %, 55.9 %, 65.6 % and 86.6 % of 200 ng M13ssDNA into DNA products, reactions were incubated at 30 °C. This is shown on 1 % TBE agarose gel.

## Appendices

**Figure 3.5.** In lanes 2-4 showing at 30 °C, 14 nM, 28 nM and 56 nM of *EcoCas3*<sup>R662G,R674G</sup> converted 4.7 %, 66.5 % and 79.4 % of 20 nM DNA fork into DNA products; in lanes 6-8 showing 14 nM, 28 nM and 56 nM of *EcoCas3*<sup>R662G,R674G</sup> converted 2.9 %, 5.2 % and 42.3 % of 20 nM DNA fork into DNA products when incubate proteins with DNA substrates at 37 °C. The endpoint products were analysed on 10 % polyacrylamide TBE native gel.

**Figure 3.6.** Showing the nuclease activity of *EcoCas3* and its mutants (56 nM) was measured on the DNA fork substrate (20 nM). Samples were collected at 0, 5, 15, 30, 60, 120 and 240 minutes. Reactions were carried out at 30 °C and analysed using 10 % denaturing TBE polyacrylamide gel. *EcoCas3* converted 4.3 %, 6.1 %, 13.5 %, 47.9 %, 83.3 % and 92 % of 20 nM DNA fork at each time points. *EcoCas3*<sup>W149A</sup> converted 1.6 %, 7 %, 10.3 %, 14.2 %, 32.7 % and 51 % of 20 nM DNA fork at each time points. *EcoCas3*<sup>W152A</sup> converted 4.4 %, 6.2 %, 13 %, 27.2 %, 42 % and 70.8 % of 20 nM DNA fork at each time points. *EcoCas3*<sup>W230A</sup> converted 1 %, 3.3 %, 15.1 %, 32 %, 63.7 % and 85.5 % of 20 nM DNA fork at each time points. *EcoCas3*<sup>W406A</sup> converted 1.9 %, 10 %, 19.4 %, 48 %, 77.1 % and 93.2 % of 20 nM DNA fork at each time points.

**Figure 3.7.** The EMSAs show that wild-type *EcoCas3* of 0.4 0.8, 1.6 and 3.3 µM could bind to 2.7 %, 4.1 % 6.6 % and 8.9 % of 20 nM DNA fork (in panel i). *EcoCas3*<sup>W406A</sup> of 0.4 0.8, 1.6 and 3.3 µM could bind to 2.2 %, 3.7 %, 5.9 % and 7.3 % of 20 nM DNA fork (in panel i). *EcoCas3*<sup>W149A</sup>, *EcoCas3*<sup>W152A</sup> and *EcoCas3*<sup>W230A</sup> are not able to form stable protein-DNA complex.

## Appendices

**Figure 3.9.** Each lane contains 50 ng M13dsDNA. Lane 1: At 30 °C, 9.4 % nicked DNA and 90.6 % supercoiled DNA; At 37 °C, 9.4 % 12.6 % nicked DNA and 87.4 % supercoiled DNA. Lane 2: At 30 °C, 14 % nicked DNA and 86 % supercoiled DNA; At 37 °C, 16.3 % nicked DNA and 83.7 % supercoiled DNA. Lane 3: At 30 °C, 19.7 % nicked DNA and 80.3 % R-loop; At 37 °C, 26.6 % nicked DNA, 45.1 % linear DNA and 28.3 % R-loop. Lane 4: At 30 °C, 17.7 % nicked DNA and 82.3 % supercoiled DNA; At 37 °C, 15.9 % nicked DNA, 32.7 % linear DNA and 51.4 % R-loop. Lane 5: At 30 °C, 74.8 % nicked DNA and 25.2 % linear DNA; At 37 °C, 77.9 % nicked DNA and 22.1 % linear DNA. Lane 6: At 30 °C, 59 % nicked DNA and 41 % linear DNA; At 37 °C, 35.4 % nicked DNA and 64.6 % linear DNA. Lane 7: At 30 °C, 82.5 % nicked DNA and 17.5 % linear DNA; At 37 °C, 86.9 % nicked DNA and 13.1 % linear DNA. Lane 8: At 30 °C, 87.9 % nicked DNA and 12.1 % linear DNA; At 37 °C 91.8 % nicked DNA and 8.2 % linear DNA. Lane 9: At 30 °C, 38.2 % nicked DNA and 61.8 % linear DNA; At 37 °C, 41 % nicked DNA and 59 % linear DNA. Lane 10: At 30 °C, 22 % nicked DNA and 78 % R-loop; At 37 °C 29.4 % nicked DNA and 70.6 % linear DNA and R-loop. Lane 11: At 30 °C, 64.4 % nicked DNA and 35.6 % linear DNA; At 37 °C, 55.4 % nicked DNA and 44.6 % linear DNA. Lane 12: At 30 °C, 79.2 % nicked DNA and 20.8 % linear DNA; At 37 °C, 88 % nicked DNA and 12 % linear DNA. Lane 13: At 30 °C, 84.2 % nicked DNA and 15.8 % linear DNA; At 37 °C, 82.9 % nicked DNA and 17.1 % linear DNA. Lane 14: At 30 °C, 56.6 % nicked DNA and 43.4 % linear DNA; At 37 °C, 49.3 % nicked DNA and 50.7 % linear DNA

**Figure 3.10.** In the presence of  $Mg^{2+}$  and ATP and Cascade (100 nM), 850 nM of EcoCas3 converted 9.7 % and 13.6 % of M13dsDNA into DNA

## Appendices

products at 30 °C and 37 °C. 850 nM of EcoCas3<sup>W149A</sup> converted 93.3 % and 100 % of M13dsDNA into DNA products at 30 °C and 37 °C. 850 nM of EcoCas3<sup>W152A</sup> converted 37.1 % and 69.7 % of M13dsDNA into DNA products at 30 °C and 37 °C. 850 nM of EcoCas3<sup>W230A</sup> converted 28.3 % and 47.1 % of M13dsDNA into DNA products at 30 °C and 37 °C. 850 nM of EcoCas3<sup>W406A</sup> converted 88.4 % and 100 % of M13dsDNA into DNA products at 30 °C and 37 °C.

**Figure 4.2(C).** In lanes 2-4 showing at 37 °C, 50 nM, 100 nM and 200 nM of PaeCas2-3 converted 68 %, 84 % and 91 % of 20 nM DNA fork into DNA products.

**Figure 4.3.** Showing nuclease activity of 50 nM PaeCas2-3 proteins on 20 nM DNA fork substrates. The samples were taken at 0 min, 5 min, 15 min, 30 min, 60 min, 120 min and 240 min. At 30 °C (panel i), PaeCas2-3 converted 0 %, 13.1 %, 20.9 %, 25.7 %, 37.3 %, 51.6 % and 61.2 % of 20 nM DNA fork into DNA products at each time point. At 37 °C (panel ii), PaeCas2-3 converted 0 %, 21.1 %, 21.3 %, 36.5 %, 44.7 %, 50 % and 53.6 % of 20 nM DNA fork into DNA products at each time point.

**Figure 4.6 (A).** In the presence of 20 mM DTT, PaeCas2-3 of 50 nM, 100 nM, 200 nM to 400 nM (lanes 2-5) converted 7.6 %, 9.4 %, 16.1 % and 24.6 % of 20 nM DNA fork into DNA products at 30 °C; at 37 °C PaeCas2-3 of 50 nM, 100 nM, 200 nM to 400 nM (lanes 7-10) converted 7.1 %, 11.6 %, 19 % and 29.3 % of 20 nM DNA fork into DNA products.

**Figure 4.6 (B).** In the presence of 20 mM DTT, EcoCas2-3 of 28 nM, 56 nM to 112nM (lanes 2-3) converted 33 %, 43.7 %, 66.4 % of 20 nM DNA fork into DNA products at 30 °C; at 37 °C EcoCas2-3 of 28 nM, 56 nM to

## Appendices

112nM (lanes 6-8) converted 3.6 %, 15.9 % and 34.2 % of 20 nM DNA fork into DNA products. In the absence of DTT, EcoCas2-3 of 28 nM, 56 nM to 112nM (lanes 10-12) converted 17 %, 19.6 %, 26.5 % of 20 nM DNA fork into DNA products at 30 °C; at 37 °C EcoCas2-3 of 28 nM, 56 nM to 112nM (lanes 14-16) converted 7.4 %, 24.8 % and 31.5 % of 20 nM DNA fork into DNA products.

**Figure 4.8.** PaeCas2-3 of 50 nM, 100 nM to 200 nM binds to 72.5 %, 80.6 % and 89.4 % of 20 nM single-stranded DNA and forms protein-DNA complexes. PaeCas2-3 of 50 nM, 100 nM to 200 nM binds to 19.4 %, 52 % and 78.6 % of 20 nM double-stranded DNA and forms protein-DNA complexes. PaeCas2-3 of 50 nM, 100 nM to 200 nM binds to 61.6 %, 87.5 % and 91.9 % of 20 nM DNA fork and forms protein-DNA complexes.

**Figure 4.9 (A).** 550 nM EcoCas3 converted about 81.3 % of 20 nM DNA fork into DNA products in the absence and presence of EcoCas2<sup>E9A</sup> (lanes 9-14)

**Figure 4.9 (B).** 550 nM EcoCas3 converted about 79.2 % of 20 nM DNA fork into DNA products in the absence of EcoCas1; 550 nM EcoCas3 converted about 78.9 % of 20 nM DNA fork into DNA products in the presence of 50 nM EcoCas1; 550 nM EcoCas3 converted about 79.3 % of 20 nM DNA fork into DNA products in the presence of 100 nM EcoCas1; 550 nM EcoCas3 converted about 77.8 % of 20 nM DNA fork into DNA products in the presence of 200 nM EcoCas1; 550 nM EcoCas3 converted about 79.1 % of 20 nM DNA fork into DNA products in the presence of 400 nM EcoCas1; 550 nM EcoCas3 converted about 79.1 % of 20 nM DNA fork into DNA products in the presence of 800 nM EcoCas1; 550 nM EcoCas3

## Appendices

converted about 47.2 % of 20 nM DNA fork into DNA products in the presence of 1.2  $\mu$ M EcoCas1.

**Figure 4.9 (C).** 330 nM EcoCas3 converted about 76.1 % of 20 nM DNA fork into DNA products in the absence of EcoCas1; 330 nM EcoCas3 converted about 72.6 % of 20 nM DNA fork into DNA products in the presence of 50 nM EcoCas1; 330 nM EcoCas3 converted about 64 % of 20 nM DNA fork into DNA products in the presence of 100 nM EcoCas1; 330 nM EcoCas3 converted about 67.5 % of 20 nM DNA fork into DNA products in the presence of 200 nM EcoCas1; 330 nM EcoCas3 converted about 50.5 % of 20 nM DNA fork into DNA products in the presence of 400 nM EcoCas1; 330 nM EcoCas3 converted about 50.3 % of 20 nM DNA fork into DNA products in the presence of 800 nM EcoCas1; 330 nM EcoCas3 converted about 60 % of 20 nM DNA fork into DNA products in the presence of 1.2  $\mu$ M EcoCas1.

**Figure 4.9 (D).** EcoCas3<sup>W406A</sup> converted about 78.3 % of 20 nM DNA fork into DNA products in the absence and presence of EcoCas1.

**Figure 4.10.** 800 nM EcoCas3 converted about 85.1 % and 59.3 % of 20 nM DNA fork into DNA products at 30 and 37 °C. Additional HtpG in reaction does not impact the total DNA amount EcoCas3 consumed, but it has an impact on forming DNA products of different lengths.

**Figure 5.2.** Panel i: Cas3\_E1 of 200 nM converted 57 % and 32.1 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E1 of 400 nM converted 73.5 % and 48.2 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E2 of 200 nM converted 1.6 % and 2.1 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E2 of 400 nM

## Appendices

converted 2.7 % and 1.9 % of 20 nM DNA fork into DNA products at 30°C and 37 °C. Panel ii: Cas3\_E1 of 200 nM converted 2.2 % and 1.7 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E1 of 400 nM converted 1.9 % and 1.1 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E2 of 200 nM converted 2.9 % and 2.3 % of 20 nM DNA fork into DNA products at 30°C and 37 °C; Cas3\_E2 of 400 nM converted 2.7 % and 1.5 % of 20 nM DNA fork into DNA products at 30°C and 37 °C

**Figure 5.3 (A).** Cas3<sup>D75G</sup> of 50 nM, 100 nM, 200 nM and 400 nM converted 9 %, 11.1 %, 14 % and 38.5 % of 20 nM DNA fork into DNA products at 30 °C. Cas3<sup>D75G</sup> of 50 nM, 100 nM, 200 nM and 400 nM converted 4.5 %, 4.4 %, 0 % and 16 % of 20 nM DNA fork into DNA products at 37 °C.

**Figure 5.3 (B).** Cas3<sup>D452A</sup> of 50 nM, 100 nM, 200 nM and 400 nM converted 4.3 %, 7.2 %, 11.2 % and 14.8 % of 20 nM DNA fork into DNA products at 30 °C. Cas3<sup>D452A</sup> of 50 nM, 100 nM, 200 nM and 400 nM converted 10.6 %, 5.5 %, 9.1 % and 12.4 % of 20 nM DNA fork into DNA products at 37 °C.

**Figure 5.3 (C).** Cas3<sup>D452A</sup> shows no DNA-protein interaction.

**Figure 5.7 (B).** DnaE of 5 nM, 10 nM, 25 nM, 50 nM and 100 nM (Lanes 3-7) extended 0 %, 0 %, 2.6 %, 15.1 % and 29.5 % of 10 nM DNA substrates.

**Figure 5.8 (A).** In lanes 3-9 showing 72.7 %, 49.6 %, 64.5 %, 67.1 %, 71.1 %, 78 % and 81.5 % of 10 nM DNA substrates were extended by an unknown DNA polymerase in purified Cas3 sample.

## Appendices

**Figure 5.8 (B).** In lanes 5-9 showing 11.8 %, 24.4 %, 38.1 %, 48.7 %, and 68.1 % of 10 nM DNA substrates were extended by an unknown DNA polymerase in the purified Cas3 sample.

**Figure 5.9.** Lane 2 and lanes 10-16 showing about 75.1- 78.9 % of 10 nM DNA substrates were extended by an unknown DNA polymerase in the purified Cas3 sample.

**Figure 5.10 (A).** Lanes 4-8 showed 0 %, 4.4 %, 10 %, 14.6 % and 22.1 % of 10 nM DNA substrates were extended by an unknown DNA polymerase in the purified EcoCas3 in the polymeric state.

**Figure 5.10 (B).** In lanes 4-8 showing 0 %, 0 %, 8.2 %, 10.7 % and 3.3 % of 10 nM DNA substrates were extended by an unknown DNA polymerase in the purified PaeCas2-3 in the polymeric state.

**Figure 5.11.** In lane 2 showing 60 nM PolI cleaved 77 % of 10 nM DNA substrates by its exonuclease function in the absence of dNTPs. In lanes 4-13 showing PolI of 0.6 fM to 6 fM, 60 fM, 600 fM, 6 pM, 60 pM, 600 pM, 0.6 nM, 6 nM and 60 nM extended 0 %, 0 %, 0 %, 0 %, 0 %, 3 %, 9.1 %, 27.3 %, 82.4 % and 91.4 % of 10 nM DNA substrates.

**Figure 5.12.** In lane 12-16 showing 0 %, 9 %, 12.2 %, 12.4 % and 13.9 % of 10 nM DNA substrates were extended by PolI co-purified with Cas3 (in Cas3b sample).

**Figure 5.13.** Panel i: In lane 4-9 showing 21.3 %, 22.7 %, 20 %, 17.5 %, 16.1 % and 13.1 % of 10 nM DNA substrates were extended by 50 nM DnaE, with increasing concentration of wild-type Cas3 added in reactions (0 nM, 10 nM, 25 nM, 50 nM, 100 nM and 200 nM Cas3). Panel ii: In lane

## Appendices

4-9 showing 17.3 %, 14.4 %, 16.2 %, 13.7 %, 12.8 % and 9.2 % of 10 nM DNA substrates were extended by 50 nM DnaE, with increasing concentration of Cas3<sup>D452A</sup> added in reactions (0 nM, 10 nM, 25 nM, 50 nM, 100 nM and 200 nM Cas3<sup>D452A</sup>)

**Figure 5.15 (C).** PolI of 3 nM extended 92.9 %, 87 %, 59 %, 35 %, 53.4 %, 42.9 % and 10 % of 10 nM DNA substrates in the presence of Cas1-2 of 0 nM, 10 nM, 25 nM, 50 nM, 100 nM, 200 nM and 400 nM in lanes 9-15.

**Figure 5.16 (B).** In lanes 3-10 showing 250 nM Cas1-2 generated integration intermediates using 22.7 %, 72 %, 26 %, 15 %, 7.1 %, 4.6 %, 4.7 % and 0.8 % of 25 ng CRISPR Is acceptor sequence.

**Figure 5.18.** In lanes 4-9 showing 3 nM PolI extended 93.6-97.5 % of 10 nM DNA substrates, with increasing concentration of Cas3 (0 nM, 10 nM, 25 nM, 50 nM, 100 nM and 200 nM) added in reactions.

**Figure 5.19.** In lanes 3-10 showing 250 nM Cas1-2 generated integration intermediates using 11.5 %, 17.4 %, 6 %, 73.5 %, 3.2 %, 1.6 %, 0.5 % and 1.1 % of 25 ng CRISPR Is acceptor sequence.