CRISPR Systems – Molecular Mechanisms of the Interaction between Cas1 and DnaK

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

Alison L. Stewart, BSc (Hons)

September 2024

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) systems are a form of adaptive immune system found in a variety of prokaryotes, allowing them to respond to invading mobile genetic elements (MGEs). Short sequences of DNA from an MGE are incorporated into a CRISPR locus of an organism by the Cas1-Cas2 protein complex in a process known as 'adaptation', allowing for degradation of the invading MGE by other CRISPR associated (Cas) proteins upon reinfection of the organism. Work in the Bolt Lab has indicated that Cas1 is restrained by the chaperone protein DnaK to protect self-DNA in *Escherichia coli*. This study attempted to understand more about the mechanism of interaction between the *E. coli* Cas1 and DnaK. Potential sites of Cas1 that may interact with DnaK were identified with amino acids V76 and R41 being selected to study. Cas1 V76L has been shown to be hyperactive in acquisition in previous studies, so it was hypothesised that this may be due to the removal of an interaction with DnaK. R41 appears to interact with DnaK in structural models, so this was also mutagenized to try to remove the interaction with DnaK. Alongside this, the study was expanded the study to determine whether the interaction is conserved across organisms, using *Streptococcus pyogenes* and *Methanothermobacter thermautotrophicus* as model organisms. These organisms were chosen because they have different types of CRISPR systems to *E. coli*.

E. coli Cas1, Cas1 V76L and Cas1 R41G and DnaK, alongside Cas1 and DnaK from *S. pyogenes* and *M. thermautotrophicus* were all purified specifically for use in these assays, and *in vivo* assays were also performed with *E. coli* proteins. Cas1 V76L behaved similarly to wild type Cas1 in DNA binding assays and the interaction with DnaK was not removed, so it is not thought that this amino acid interacts with DnaK. Cas1 R41G did not bind DNA effectively, so it was challenging to study whether it interacts less strongly with DnaK, however it did appear to exhibit a weaker physical interaction with DnaK than wild type Cas1. The results gathered for this study were inconclusive as to whether the interaction between Cas1 and DnaK is conserved across different types of CRISPR system. Further research needs to be done to confirm the interaction between the proteins from *S. pyogenes* and *M. thermautotrophicus*, such as pull down assays and bimolecular fluorescence complementation (BiFC).

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Edward Bolt, for his consistent support and guidance throughout this project. I have developed so much since joining the Bolt Lab, and I am certain I would not be the scientist that I am today had it not been for the opportunities offered to me since I joined this lab.

Secondly, I would like to thank the members of D58, both past and present. Olivia Downs has been a constant source of comfort and laughter throughout this year, along with providing a never-ending stream of advice. Dr Tom Killelea, despite no longer working in D58 has always taken an interest in my work and been there to troubleshoot any problems throughout this project, and I am more grateful to him than I can say. Anna Lou-Hing has also been there for me as a friend, colleague and flatmate, and has been a great source of help for me.

Finally, I would like to thank my family, and in particular my partner James for your love and kindness this year. I could not have done this without you being there for me as my support network, and I appreciate you all so much.

Abbreviations

BCA – Bicinchoninic acid

BiFC - Biomolecular fluorescence complementation

Cas – CRISPR associated

Cascade - CRISPR-associated complex for antiviral defence

CRISPR – Clustered, regularly interspaced short palindromic repeats

CPS – Colour protein standard

crRNA – CRISPR RNA

Ds – Double strand

DSB -Double strand breaks

EMSA – Electrophoretic mobility shift assay

HSP70 – Heat shock protein of 70kDa

IHF – Integration host factor

KLD – Kinase, ligase, digest

LAS – Leader anchoring sequence

LIS – Low ionic salt

MGE – Mobile genetic element

NP – No protein

PAM – Protospacer adjacent motif

PNK – Polynucleotide kinase

SDS PAGE – SDS protein agarose gel electrophoresis

Ss – Single strand

TBE – Tris, boric acid, EDTA

WT – Wild type

Table of Contents

Chapter 1 - Introduction

1.1 CRISPR Systems

Resistance to infection is a fundamental requirement for life and survival, to protect the health of organisms. Clustered regularly interspaced short palindromic repeats (CRISPR) systems provide adaptive immunity and are found in around 40% of bacteria and 90% of archaea (Mitić et al., 2023). They allow for the recognition and destruction of invading mobile genetic elements (MGEs) that could harm the organism, such as transposons, plasmids and bacteriophage viruses (here after referred to as phage viruses) (McGinn & Marraffini, 2019). MGEs lead to horizontal gene transfer in prokaryotes, allowing for the spread of genes from organism to organism. This has great impacts on aspects of human life as well as prokaryotes, as horizontal gene transfer drives the spread of antibiotic resistance and other dangerous capabilities that prokaryotes can possess (Lau et al., 2019). It is therefore essential to understand CRISPR systems to protect against these problems in the future.

The presence of CRISPR systems was first identified in 1987, but their significance was not understood until years later (Ishino et al., 1987). Whilst investigating a gene encoding an alkaline phosphatase enzyme in *Escherichia coli*, it was noted that there were short, repetitive sequences interspersed with non-repeating sequences nearby in the genome (Ishino et al., 1987). In unrelated research, van Soolingen et al. (1993) also identified these clustered repeat sequences, alternating with varied non-repeated sequences in *Mycobacterium tuberculosis*. Mojica et al. (1995) were able to identify and study the function of these short tandem repeats in the archaea *Haloferax mediterranei* and *Haloferax volcanii* and following this Mojica et al. (2000) identified the presence of these systems across a variety of archaeal and bacterial species, and from there the identification of CRISPR systems began. The name CRISPR was first used by Jansen et al., (2002), and the presence of a CRISPR system was determined by 2 main factors; interspaced palindromic repeats with variable non-repeat sequences, and the usage CRISPR associated (Cas) proteins, which a CRISPR response is dependent on (Jansen et al., 2002).

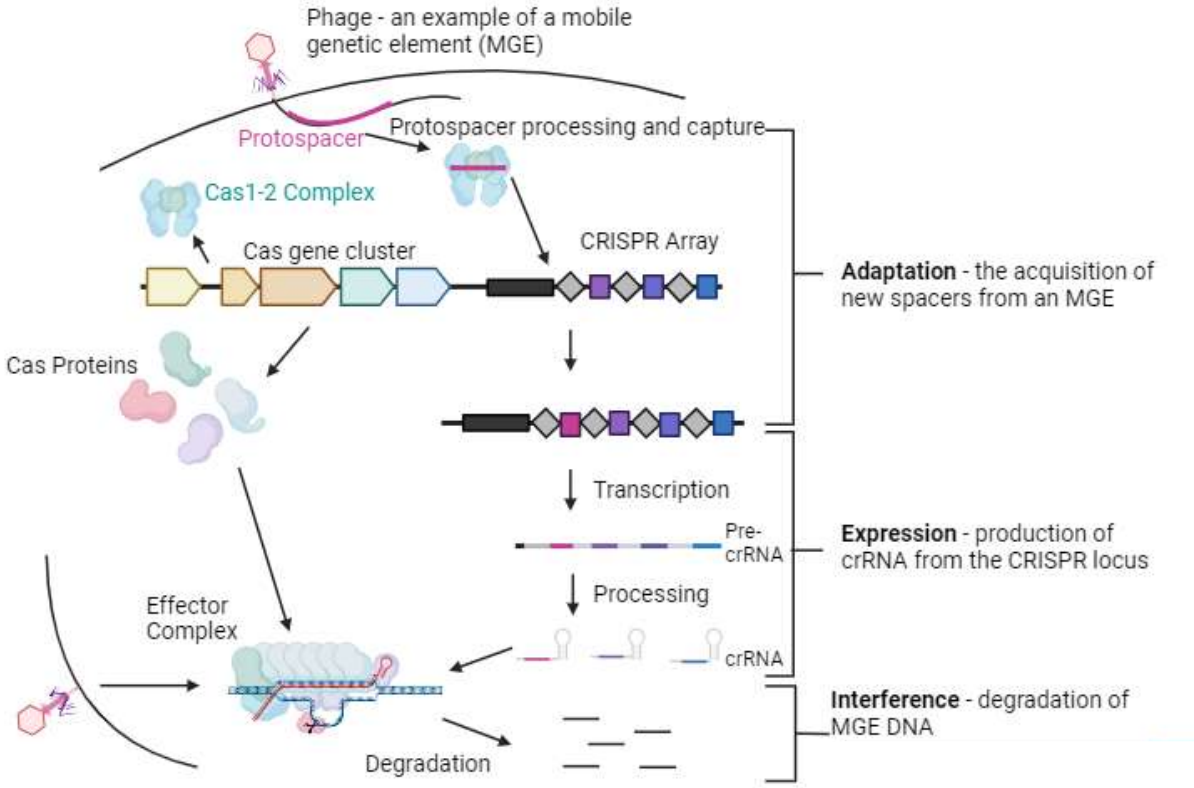


***Figure 1.1* - Overview of CRISPR Systems**. *A CRISPR response is divided up into 3 sections, adaptation, expression and interference. These rely on Cas proteins, which are transcribed from cas genes in the CRISPR locus. These Cas proteins have varied functions, and lead to the recognition and degradation of invading MGE DNA, alongside long-term immunological memory. Created with Biorender.com*

A CRISPR response can be divided up into 3 sections, Adaptation, Expression and Interference, an overview of which can be seen in Figure 1.1. Adaptation is the recognition of an MGE and incorporation of its genetic sequence into the CRISPR locus (Nuñez et al., 2014). Expression produces and processes CRISPR RNA (crRNA) molecules from the CRISPR locus (Brouns et al., 2008). Interference requires the combination of crRNA with an effector complex of Cas proteins that can then target the complementary sequence to this crRNA in an MGE (Lau et al., 2019). This leads to the formation of an R-loop in the MGE DNA, leaving an unpaired DNA strand which can then be targeted for nucleolytic degradation (Killelea & Bolt, 2017). As the work in this study relates to adaptation, this is the only stage that will be discussed in detail.

CRISPR systems rely on the presence of at least one CRISPR locus in the genome, the structure of which can be seen in Figure 1.2. CRISPR loci consist of cas genes, encoding Cas proteins, followed by an AT rich leader sequence (Hille & Charpentier, 2016). Following the leader sequence there are alternating spacers (derived from MGEs which have previously infected the cell) from which crRNA molecules are transcribed to be used in the interference step, and palindromic repeats (usually 20-30 nucleotides) (Barrangou & Marraffini, 2014). CRISPR systems contain diverse Cas proteins, so a method of classification has been established. There are 2 classes of CRISPR system, Class 1 and Class 2. These vary based on their effector proteins, as Class 1 CRISPR Systems rely on a complex of Cas proteins known as Cascade to degrade the invading MGE, whereas Class 2 CRISPR Systems have a single protein that performs this role (Makarova & Koonin, 2015). These two classes can be broken down further into 6 types based on the different proteins used and the target genetic material, as seen in Figure 1.3. Class 1 CRISPR Systems encompasses Type I, III and IV, whereas Class 2 encompasses Type II, V and VI (Makarova et al., 2020).

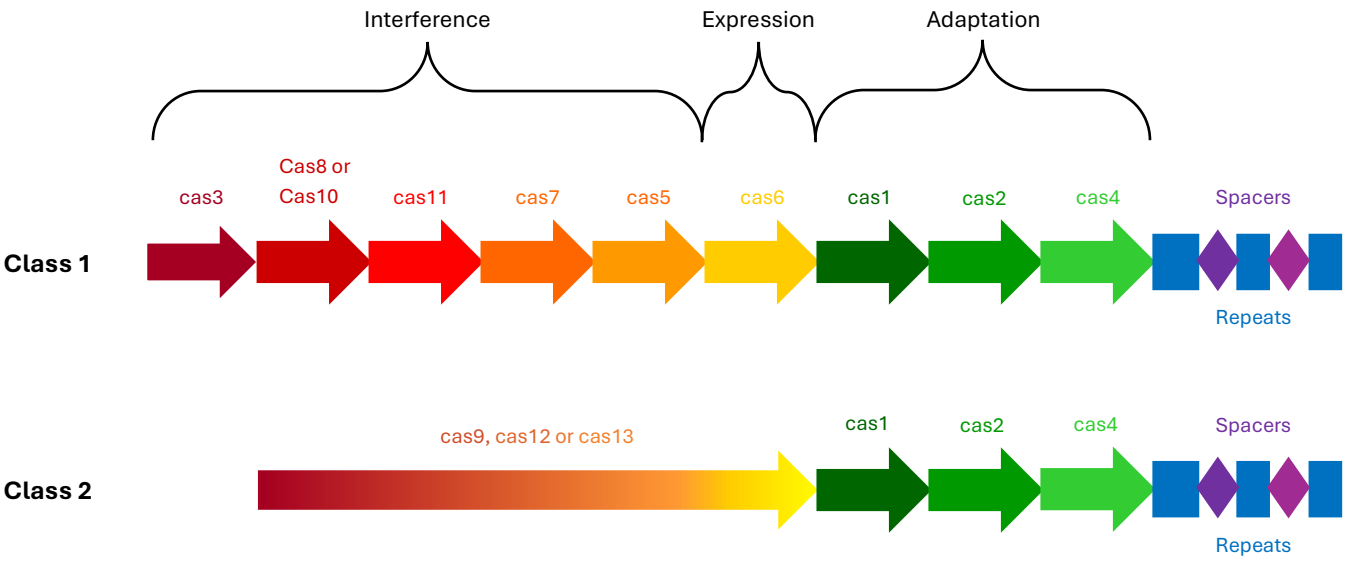


Figure 1.2 - Structure of the CRISPR locus in Class 1 and Class 2 CRISPR systems, showing the different effector proteins used. *Class 1 CRISPR systems rely on a protein complex called Cascade to perform interference, and Cas6 to process crRNA molecules, whereas Class 2 CRISPR systems have a single protein which can perform these rolled. Following this, the CRISPR locus consists of short palindromic DNA repeats, interspersed with DNA sequences from MGEs that have previously infected to the cell known as spacers. Adapted from (Makarova et al., 2020)*

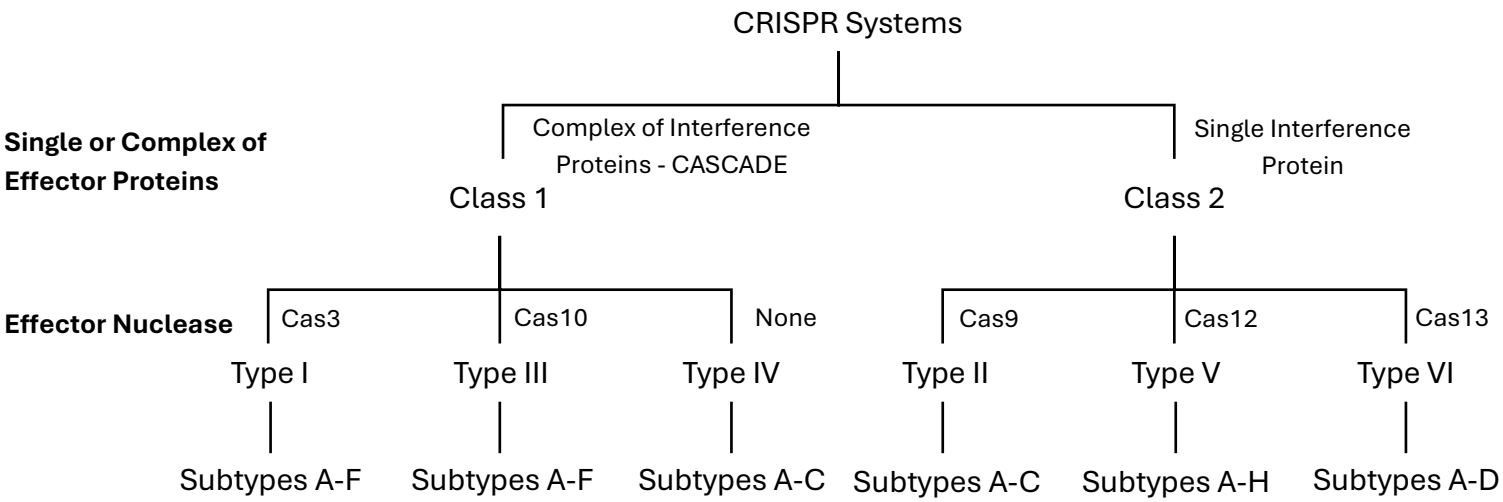


Figure 1.3 - Classification of CRISPR systems. *CRISPR systems are divided up into 2 classes depending on whether they use a single effector protein or a complex of several cas proteins known as Cascade. These are then divided up into 6 types depending on the protein that acts as a nuclease to degrade MGE DNA, and from there a further 33 subtypes based on a variety of factors. Note that this figure does not indicate phylogenetic relationship between the different groups, merely the classification used.*

The division of CRISPR systems into different types relies on the differences in the effector nuclease used to degrade MGE DNA, an overview of which can be seen in Figure 1.3. Class I encompasses types I, III and IV, which use Cas3, Cas10 and no effector nuclease respectively (Makarova et al., 2018). Class II encompasses types II, V and VI., which use Cas9, Cas12 and Cas13 nucleases respectively (Shmakov et al., 2017). These types of CRISPR systems can be further classified into 33 different subtypes based on variation in gene sequences, whether the system recognises DNA or RNA based MGEs, organisation of the CRISPR locus and proteins used (Koonin et al., 2017).

1.2 Adaptation

Most of the variation between CRISPR systems is found in the method of interference, whereas the process of adaptation is universal amongst CRISPR systems. Adaptation can be divided up into two stages; selection and processing of MGE DNA to form a new protospacer, and its integration into the CRISPR locus (Rollie et al., 2015). Adaptation can be 'primed' (because a cell has before encountered the MGE) or 'naïve' (when a cell is faced with a novel MGE).

1.2.1 Selection of MGE DNA to form a Protospacer

Primed adaptation occurs when a MGE that has previously infected a cell re-infects the cell. MGEs are subjected to a variety of selection pressures that can impact their survival, including the immunity that CRISPR systems provide to their host organisms (Wang et al., 2016). There is therefore a drive for evolution within MGEs, and as a consequence of this evolution the DNA sequence of an infecting MGE can undergo a mutation known as an escape mutation (Mitić et al., 2023). This escape mutation means that the MGE is no longer directly complementary to the crRNA transcribed from the CRISPR locus, and this mismatch can then be detected by Cascade complex which will then recruit the nuclease Cas3 in Class 1 CRISPR systems (Wang et al., 2016). Cas3 will then degrade the MGE DNA to form a new protospacer (Ivančić-Baće et al., 2015). Equivalent nucleases perform this role across CRISPR systems, and the active nuclease used in each type of CRISPR system can be seen in Figure 1.3. The protospacer generated via this nuclease activity can then be incorporated into the CRISPR locus as described in Chapter 1.2.2, thus re-establishing optimal immunity to the invading MGE.

In naïve adaptation, non-Cas proteins are required for the generation of protospacers for incorporation into the CRISPR locus (Mitić et al., 2023). Protospacers are preferentially generated from DNA fragments produced by the degradation of DNA during the repair of double strand (ds) DNA breaks (DSB) due to the involvement of RecBCD (Ivančić-Baće et al., 2015). RecBCD is a key enzyme involved in the repair of DSBs, as it recognises dsDNA ends that are created by the breakage of DNA at replication forks during homologous DNA replication (Dillingham & Kowalczykowski, 2008). Once it has bound to the end of a dsDNA break, RecBCD will use its dual helicase-nuclease activity to unwind DNA and degrade DNA, in particular the 3’ DNA strand, into small, single stranded DNA fragments (Dixon & Kowalczykowski, 1993). DNA degradation is halted once the RecBCD molecule reaches a short, protective DNA sequence known as a Chi-sequence on the 3’ strand, which causes the RecBCD molecule to undergo a conformational change and causes the RecBCD molecule to pause and initiate loading of the protein RecA onto the 3’ DNA strand, allowing for the repair of the double strand break to continue (Anderson & Kowalczykowski, 1997). It was previously theorised that protospacer fragments were produced via the nuclease activity of RecBCD and harnessed by Cas1-Cas2 before

incorporation of the DNA into the CRISPR locus as a spacer (McGinn & Marraffini, 2019). However, it has since been shown that the helicase activity of the RecBCD protein is required for CRISPR adaptation, and not the nuclease activity (Radovcic et al., 2018).

It is now thought that RecBCD helicase activity separates DNA into two single strands, and Cas1-Cas2, or alternative host exonucleases such as Dna pol III, DnaQ, and ExoT (Ramachandran et al., 2020) degrade the ssDNA into short fragments of the appropriate size, which can then be processed and used to generate a new DNA spacer (Mitić et al., 2023). In certain types of CRISPR system, Cas4 is also a required protein for spacer acquisition (Lee et al., 2018). Cas4 is a RecB-like nuclease that forms a complex with the Cas1-Cas2 complex and is required for the selection of protospacers and their processing to the correct size (Shiimori et al., 2018). However, unlike Cas1-Cas2, Cas4 is not universally conserved, so some CRISPR systems rely on only Cas1-Cas2 (Lee et al., 2018).

Both primed and naïve adaptation result in the presence of the Cas1-Cas2 complex bound to a protospacer. In *Escherichia coli*, this protospacer is 33bp long, but this is variable depending on the species, as the length is dictated by the 3D spacing of the Cas1 active sites by the Cas2 homodimer (Nuñez, Lee et al., 2015). The protospacer is held in a manner that can be seen in Figure 1.4, whereby there is a double stranded stretch between two Cas1 molecules on adjacent homodimers, with flayed ends at either end of the molecule (Nuñez et al., 2014). The flayed ends of the protospacer then must be cleaved to form the final spacer that will be integrated into the CRISPR locus. In *E. coli,* the flayed ends are 5bp in length, leaving the size of the final spacer in the CRISPR locus as 23bp spacer following cleavage of these sequences (Nuñez, Lee, et al., 2015).

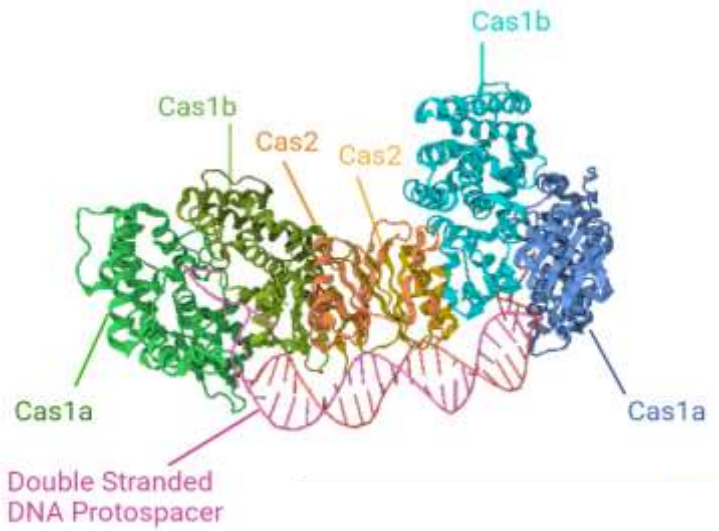


Figure 1.4 – Structural model of the Cas1-Cas2 complex in E. coli. *Two Cas1 homodimers join with a single Cas2 homodimer to form the Cas1-Cas2 complex. Protein sequences taken from Uniprot.com, PDB number 4P6I. Protein structure and annotations created in Biorender.com*

Cas1 has been shown to exhibit nuclease activity *in-vivo* when in the Cas1-Cas2 complex (Lau et al., 2019), so it is likely that in most CRISPR systems Cas1 is directly responsible for the processing of protospacers into spacers of the correct length (Rollie et al., 2018), thus generating a 3'-OH end which is required for the incorporation of the spacer into the CRISPR locus (Radovcic et al., 2018). There is also some evidence that in species where Cas4 is found, it can also be responsible for the cleavage of protospacer DNA to prepare it for insertion into the CRISPR locus (Lee et al., 2019), however as Cas4 is only found in certain types of CRISPR system it is most likely the nuclease activity of Cas1-Cas2 that is responsible for the cleavage of these flayed ends.

1.2.2 Insertion of a New Spacer into the CRISPR Locus

Insertion of new spacers into the CRISPR locus occurs preferentially directly following the AT-rich leader sequence. It also must be polarised to ensure that the spacer is inserted in a manner whereby crRNA can be transcribed from the template and be complementary to the MGE from which it was derived (McGinn & Marraffini, 2016). Direction of the spacer to the leader sequence utilises different mechanisms in Class I and Class II CRISPR systems (McGinn & Marraffini, 2019). In CRISPR systems found in Gram-negative organisms, a protein called integration host factor (IHF) is required, as the

leader sequences of these CRISPR loci contain an IHF-binding site (Nuñez et al., 2016). IHF interacts with Cas1 and the IHF binding site in the leader sequence, bending DNA and guiding Cas1 to the leader sequence for integration (McGinn & Marraffini, 2019). Gram – positive bacteria however lack IHF, therefore there must be another as yet unknown host DNA bending protein that is able to fulfil this role (Nuñez et al., 2016). In Class II CRISPR systems however, the leader sequence contains a specific sequence known as the leader anchoring sequence (LAS) (Xiao et al., 2017). This binds to Cas1 with sufficient affinity that no additional factors are required to direct spacer acquisition to the leader sequence (McGinn & Marraffini, 2019).

Once the leader sequence of the CRISPR locus has been identified, and the integration complex has been oriented, spacer integration will occur, an overview of which can be seen in Figure 1.5. Cas1 exhibits nickase activity and will cleave a single strand of DNA adjacent to the leader sequence (McGinn & Marraffini, 2019). The 3'-OH end of one of the DNA strands of the spacer will then attack the phosphodiester backbone, and Cas1 catalyses a transesterification reaction between the spacer DNA, and the start of the first repeat in the CRISPR locus (McGinn & Marraffini, 2019). Cas1 then catalyses a second transesterification reaction between the 3'-OH end of the other strand of spacer DNA and the end of the first repeat in the CRISPR locus (Rollie et al., 2018). This leaves gaps in the DNA sequence, which are filled by host enzymes. A DNA polymerase and a ligase enzyme are required to complete the process of spacer integration, and it has been identified that DNA Polymerase I is required for both primed and naïve adaptation (Ivančić-Baće et al., 2015).

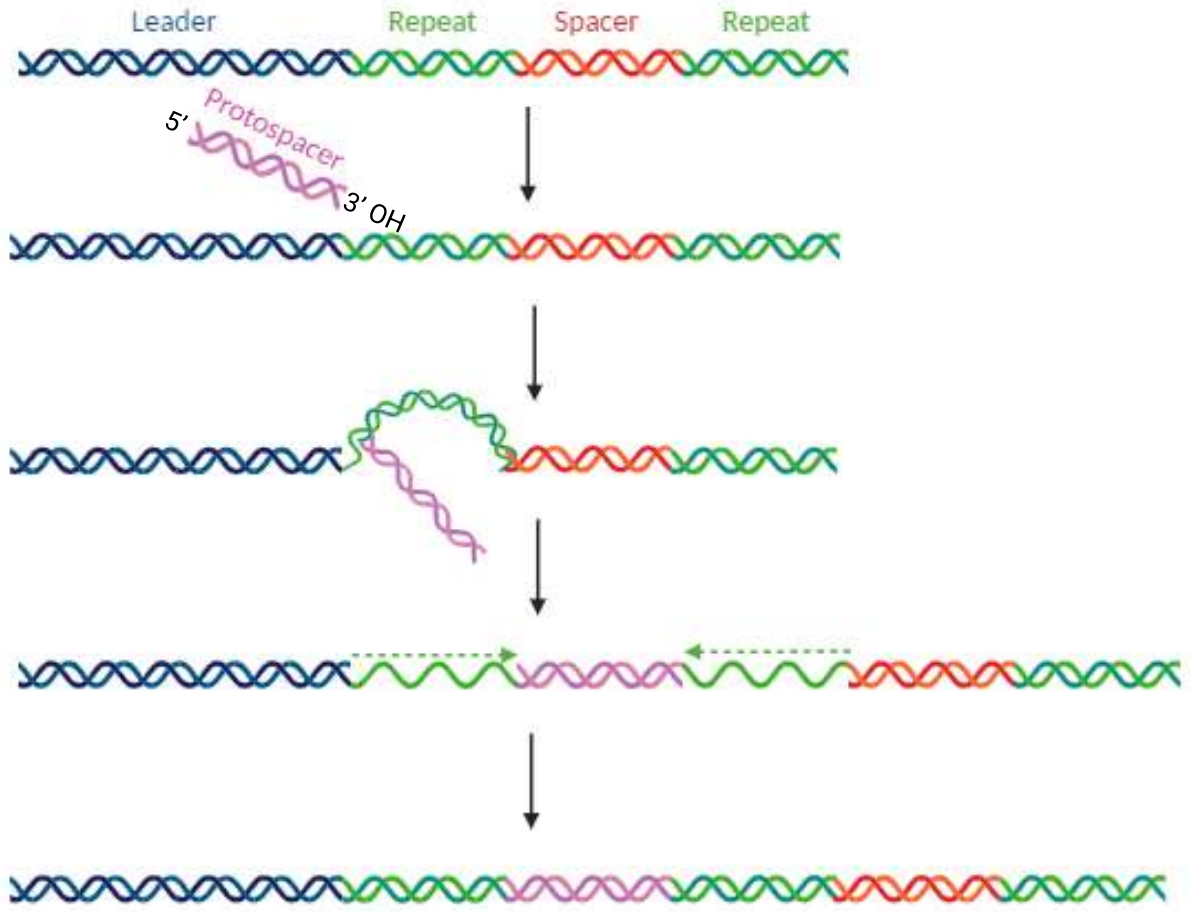


***Figure 1.5** - **An Overview of spacer integration during adaptation.** The 3'- OH end of a protospacer bound by Cas1-Cas2 attacks the end of the leader sequence in the CRISPR locus, and a transesterification reaction occurs catalysed by Cas1.A second transesterification reaction, catalysed by Cas1 then occurs to join the other strand of protospacer DNA to the repeat sequence in the CRISPR locus. DNA repair machinery including DNA polymerase I and DNA ligase join the gaps in the DNA to leave a newly incorporated spacer. Adapted from (McGinn & Marraffini, 2019), created using Biorender.com*

The way the new spacer is inserted into the CRISPR locus next to the leader sequence is beneficial to the organism, as spacers close to the leader sequence are more readily transcribed into crRNA, therefore the organism ensures that it is protecting itself against the pathogens which are most likely to infect the cell (Liao et al., 2022). Because of this, cells can perform spacer deletion upon infection to ensure that the relevant spacers are in proximity to the leader sequence, thus ensuring effective expression of the crRNA and interference of the MGE (Garrett, 2021).

1.3 Cas1-Cas2 in Adaptation

Cas1 and Cas2 are the only universally conserved proteins amongst CRISPR systems (Nuñez et al., 2014). Cas2 exhibits non-specific catalytic activity, but this has been proven not to be required for spacer integration, instead it acts as a scaffolding protein to ensure appropriate positioning of the two Cas1 homodimers that make up the integration complex (Nuñez et al., 2014). The Cas1-Cas2 complex is essential for spacer integration across all CRISPR systems, and there are no alternative proteins that can perform this role (Nuñez et al., 2014) (He et al., 2018).

When overexpressed in cells, Cas1 will attack the host chromosome, and yet *in vivo*, it exhibits sequence selectivity whereby it will only target stretches of DNA from MGEs. This is particularly interesting, as there is no molecular difference between the DNA in the host chromosome compared to that of the MGE, and yet organisms protect their own host chromosome from destruction by Cas1 (Wimmer & Beisel, 2020). Protospacer adjacent motifs (PAMs) have long been recognised as assisting with the selectivity of Cas1. PAMs are short (2-5 base pairs) sequences which vary from species to species and Cas1 has been found to preferentially target DNA sequences that are flanked by PAMs (Shah et al., 2013). The sequences of these PAMs are very simple, for example in *E. coli* the consensus sequence is 5'-AAG-3', meaning they are sufficiently conserved to be found across MGEs (Mitić et al., 2023). PAMs are cleaved from the protospacer by Cas1 before insertion into the CRISPR locus, thus preventing autoimmunity and self-targeting at the CRISPR locus (Nuñez et al., 2014). The existence of PAMs gives some insight into why Cas1 preferentially targets MGE DNA, but as PAMs are only short sequences of DNA, they do appear in the host chromosome, so these sites would still be potential targets for Cas1 (Liao et al., 2022). Therefore, there must be another mechanism that exists to prevent Cas1-Cas2 from targeting host DNA, and Killelea et al. (2023) indicated that the chaperone protein DnaK is responsible for restraining Cas1 to prevent host chromosome targeting in *E. coli*.

1.4 DnaK and its role as a Protein Chaperone

Cas1 was identified as interacting with DnaK by Calloni et al. (2012), yet it was not until Killelea et al. (2023) investigated this interaction further that it was found that DnaK

restrains Cas1 in *E. coli* to prevent targeting of genomic DNA for spacer acquisition. It is theorised that DnaK binds Cas1 in cells, until the presence of a MGE element protein prompts DnaK to release Cas1, thus freeing Cas1 to perform spacer acquisition (Killelea et al., 2023). To understand this better, the following summarises the role of DnaK as a protein chaperone.

DnaK is a member of the 70 kDa heat shock protein (Hsp70) family of molecular chaperone proteins, a group of proteins which are essential for protein folding and the prevention of protein aggregates (W. Wang et al., 2021). They are named heat shock proteins as their expression is induced in times of cellular distress, such as extreme changes in heat, to aid correct protein folding (Kim et al., 2013), however they are constitutively expressed at a lower level in cells regardless of cellular conditions. By expressing molecular chaperones such as DnaK, cells can ensure proper structure, and therefore function, of cellular proteins (Calloni et al., 2012).

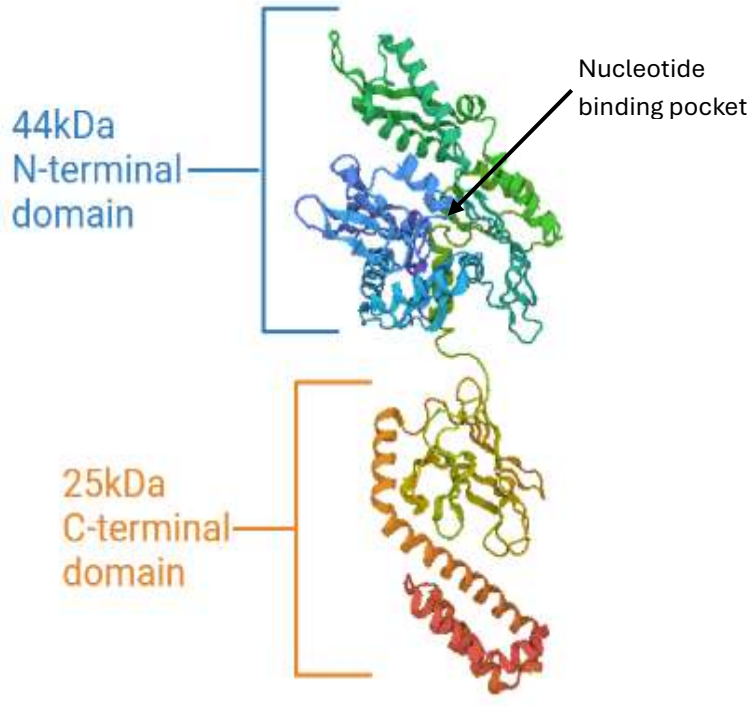
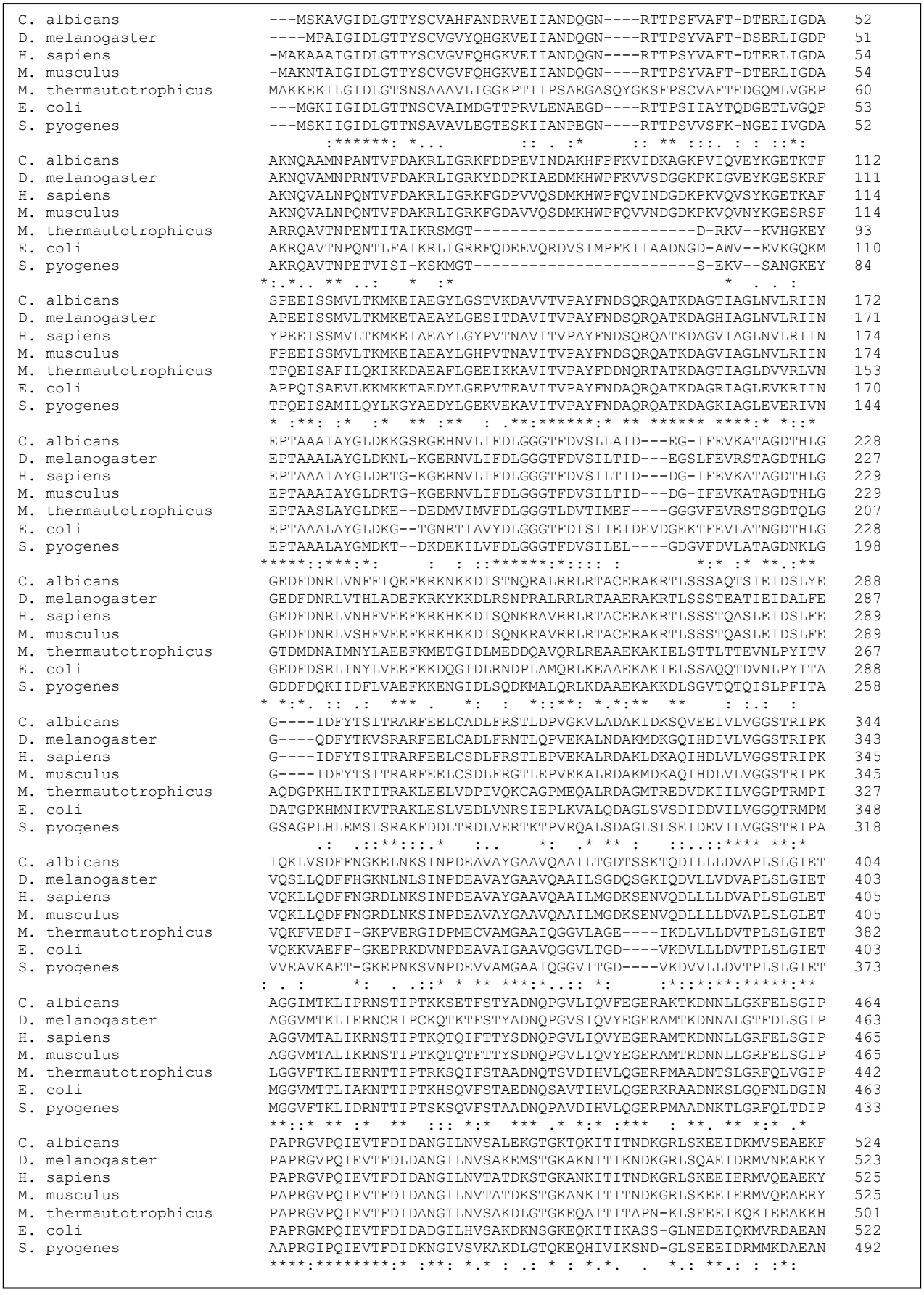


***Figure 1.6 – E. coli DnaK structure when in its ADP bound state**. Protein sequence taken from UniProt.com, PDB number 2KHO, image created in Biorender.com*

DnaK is comprised of 2 domains: a 44-kDa N-terminal nucleotide-binding domain and a 25-kDa C-terminal substrate-binding domain, as seen in Figure 1.6. The N-terminal domain is capable of ATP-hydrolysis to allow for the binding and release of substrates which are bound to the substrate binding domain (Bertelsen et al., 2009). When DnaK is in its ADP bound state, it will bind substrate proteins via hydrophobic residues, however when in its ATP bound state, a conformational change occurs in the substrate binding domain, meaning it will release any bound substrates (Bertelsen et al., 2009).

The binding motif of DnaK and other Hsp70 proteins is conserved, consisting of a 4-5 residue sequence composed of hydrophobic residues. Leu is the most conserved amino acid in DnaK binding sequences (Rüdiger et al., 1997), as this amino acid sits in the central pocket of the binding cleft of the substrate binding domain (Zhu et al., 1996), but Ile, Val, Phe and Tyr are also common in the binding cleft (Rüdiger et al., 1997). This consensus DNA binding motif occurs regularly in protein sequences, in fact in *E. coli*, it was found to occur every 36 residues (Jansen et al., 2002). The frequency of occurrence of this consensus sequence enables DnaK to be an incredibly effective protein chaperone, as the substrate binding domain of DnaK is complementary to almost all other proteins (Rüdiger et al., 1997).

The protection of hydrophobic residues is essential for proper protein folding in the cluttered environment of the cell. As the cell cytoplasm is a hydrated environment, hydrophobic residues will tend to cluster together when exposed, leading to improper folding and protein aggregation (Kim et al., 2013). In addition to this, the levels of protein crowding in the cell cytoplasm also increase the propensity of proteins for aggregation, and protein aggregates are non-functional and often harmful to the cell (Kim et al., 2013). Therefore, DnaK and other Hsp70 family proteins are essential for proper functionality of the cell, including the functionality of CRISPR systems. The Hsp70 family of proteins is one of the most highly conserved family of proteins across all species with very little sequence variation between species, as seen in Figure 1.7. The highly conserved nature of this protein is indicative of how essential its function is. With it being almost universally conserved amongst all organisms, not just those with CRISPR systems, it would provide an explanation for the spatiotemporal control of adaptation across organisms that rely on Cas1-Cas2 for adaptation if this mechanism was shown to be conserved.


C. albicans ---MSKAVGIDLGTTYSCVAHFANDRVEIIANDQGN----RTTPSFVAFT-DTERLIGDA 52
D. melanogaster ----MPAIGIDLGTTYSCVGVYQHGKVEIIANDQGN----RTTPSYVAFT-DSERLIGDP 51
H. sapiens -MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGN----RTTPSYVAFT-DTERLIGDA 54
M. musculus -MAKNTAIGIDLGTTYSCVGVFQHGKVEIIANDQGN----RTTPSYVAFT-DTERLIGDA 54
M. thermautotrophicus MAKKEKILGIDLGTSNSAAAVLIGGKPTIIPSAEGASQYGKSFPSCVAFTEDGQMLVGEP 60
E. coli ---MGKIIGIDLGTTNSCVAIMDGTTPRVLENAEGD----RTTPSIIAYTQDGETLVGQP 53
S. pyogenes ---MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGN----RTTPSVVSFK-NGEIIVGDA 52
C. albicans AKNQAAMNPANTVFDAKRLIGRKFDDPEVINDAKHFPFKVIDKAGKPVIQVEYKGETKTF 112
D. melanogaster AKNQVAMNPRNTVFDAKRLIGRKYDDPKIAEDMKHWPFKVVSDGGKPKIGVEYKGESKRF 111
H. sapiens AKNQVALNPQNTVFDAKRLIGRKFGDPVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAF 114
M. musculus AKNQVALNPQNTVFDAKRLIGRKFGDAVVQSDMKHWPFQVVNDGDKPKVQVNYKGESRSF 114
M. thermautotrophicus ARRQAVTNPENTITAIKRSMGT----------D-RKV--KVHGKEY 93
E. coli AKRQAVTNPQNTLFAIKRLIGRRFQDEEVQRDVSIMPFKIIAADNGD-AWV--EVKGQKM 110
S. pyogenes AKRQAVTNPETVISI-KSKMGT----------S-EKV--SANGKEY 84
C. albicans SPEEISSMVLTKMKEIAEGYLGSTVKDAVVTVPAYFNDSQRQATKDAGTIAGLNVLRIIN 172
D. melanogaster APEEISSMVLTKMKETAEAYLGESITDAVITVPAYFNDSQRQATKDAGHIAGLNVLRIIN 171
H. sapiens YPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQRQATKDAGVIAGLNVLRIIN 174
M. musculus FPEEISSMVLTKMKEIAEAYLGHPVTNAVITVPAYFNDSQRQATKDAGVIAGLNVLRIIN 174
M. thermautotrophicus TPQEISAFILQKIKKDAEAFLGEEIKKAVITVPAYFDDNQRTATKDAGTIAGLDVVRLVN 153
E. coli APPQISAEVLKKMKKTAEDYLGEPVTEAVITVPAYFNDAQRQATKDAGRIAGLEVKRIIN 170
S. pyogenes TPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYFNDAQRQATKDAGKIAGLEVERIVN 144
C. albicans EPTAAAIAYGLDKKGSRGEHNVLIFDLGGGTFDVSLLAID---EG-IFEVKATAGDTHLG 228
D. melanogaster EPTAAALAYGLDKNL-KGERNVLIFDLGGGTFDVSILTID---EGSLFEVRSTAGDTHLG 227
H. sapiens EPTAAAIAYGLDRTG-KGERNVLIFDLGGGTFDVSILTID---DG-IFEVKATAGDTHLG 229
M. musculus EPTAAAIAYGLDRTG-KGERNVLIFDLGGGTFDVSILTID---DG-IFEVKATAGDTHLG 229
M. thermautotrophicus EPTAASLAYGLDKE--DEDMVIMVFDLGGGTLDVTIMEF----GGGVFEVRSTSGDTQLG 207
E. coli EPTAAALAYGLDKG--TGNRTIAVYDLGGGTFDISIIEIDEVDGEKTFEVLATNGDTHLG 228
S. pyogenes EPTAAALAYGMDKT--DKDEKILVFDLGGGTFDVSILEL----GDGVFDVLATAGDNKLG 198
C. albicans GEDFDNRLVNFFIQEFKRKNKKDISTNQRALRRLRTACERAKRTLSSSAQTSIEIDSLYE 288
D. melanogaster GEDFDNRLVTHLADEFKRKYKKDLRSNPRALRRLRTAAERAKRTLSSSTEATIEIDALFE 287
H. sapiens GEDFDNRLVNHFVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFE 289
M. musculus GEDFDNRLVSHFVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFE 289
M. thermautotrophicus GTDMDNAIMNYLAEEFKMETGIDLMEDDQAVQRLREAAEKAKIELSTTLTTEVNLPYITV 267
E. coli GEDFDSRLINYLVEEFKKDQGIDLRNDPLAMQRLKEAAEKAKIELSSAQQTDVNLPYITA 288
S. pyogenes GDDFDQKIIDFLVAEFKKENGIDLSQDKMALQRLKDAAEKAKKDLSGVTQTQISLPFITA 258
C. albicans G----IDFYTSITRARFEELCADLFRSTLDPVGKVLADAKIDKSQVEEIVLVGGSTRIPK 344
D. melanogaster G----QDFYTKVSRARFEELCADLFRNTLQPVEKALNDAKMDKGQIHDIVLVGGSTRIPK 343
H. sapiens G----IDFYTSITRARFEELCSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPK 345
M. musculus G----IDFYTSITRARFEELCSDLFRGTLEPVEKALRDAKMDKAQIHDLVLVGGSTRIPK 345
M. thermautotrophicus AQDGPKHLIKTITRAKLEELVDPIVQKCAGPMEQALRDAGMTREDVDKIILVGGPTRMPI 327
E. coli DATGPKHMNIKVTRAKLESLVEDLVNRSIEPLKVALQDAGLSVSDIDDVILVGGQTRMPM 348
S. pyogenes GSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLSLSEIDEVILVGGSTRIPA 318
C. albicans IQKLVSDFFNGKELNKSINPDEAVAYGAAVQAAILTGDTSSKTQDILLLDVAPLSLGIET 404
D. melanogaster VQSLLQDFFHGKNLNLSINPDEAVAYGAAVQAAILSGDQSGKIQDVLLVDVAPLSLGIET 403
H. sapiens VQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLSLGLET 405
M. musculus VQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLSLGLET 405
M. thermautotrophicus VQKFVEDFI-GKPVERGIDPMECVAMGAAIQGGVLAGE----IKDLVLLDVTPLSLGIET 382
E. coli VQKKVAEFF-GKEPRKDVNPDEAVAIGAAVQGGVLTGD----VKDVLLLDVTPLSLGIET 403
S. pyogenes VVEAVKAET-GKEPNKSVNPDEVVAMGAAIQGGVITGD----VKDVVLLDVTPLSLGIET 373
C. albicans AGGIMTKLIPRNSTIPTKKSETFSTYADNQPGVLIQVFEGERAKTKDNNLLGKFELSGIP 464
D. melanogaster AGGVMTKLIERNCRIPCKQTKTFSTYADNQPGVSIQVYEGERAMTKDNNALGTFDLSGIP 463
H. sapiens AGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLGRFELSGIP 465
M. musculus AGGVMTALIKRNSTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTRDNNLLGRFELSGIP 465
M. thermautotrophicus LGGVFTKLIERNTTIPTRKSQIFSTAADNQTSVDIHVLQGERPMAADNTSLGRFQLVGIP 442
E. coli MGGVMTTLIAKNTTIPTKHSQVFSTAEDNQSAVTIHVLQGERKRAADNKSLGQFNLDGIN 463
S. pyogenes MGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAADNKTLGRFQLTDIP 433
C. albicans PAPRGVPQIEVTFDIDANGILNVSALEKGTGKTQKITITNDKGRLSKEEIDKMVSEAEKF 524
D. melanogaster PAPRGVPQIEVTFDLDANGILNVSAKEMSTGKAKNITIKNDKGRLSQAEIDRMVNEAEKY 523
H. sapiens PAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRLSKEEIERMVQEAEKY 525
M. musculus PAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRLSKEEIERMVQEAERY 525
M. thermautotrophicus PAPRGVPQIEVTFDIDANGILNVSAKDLGTGKEQAITITAPN-KLSEEEIKQKIEEAKKH 501
E. coli PAPRGMPQIEVTFDIDADGILHVSAKDKNSGKEQKITIKASS-GLNEDEIQKMVRDAEAN 522
S. pyogenes AAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIKSND-GLSEEEIDRMMKDAEAN 492

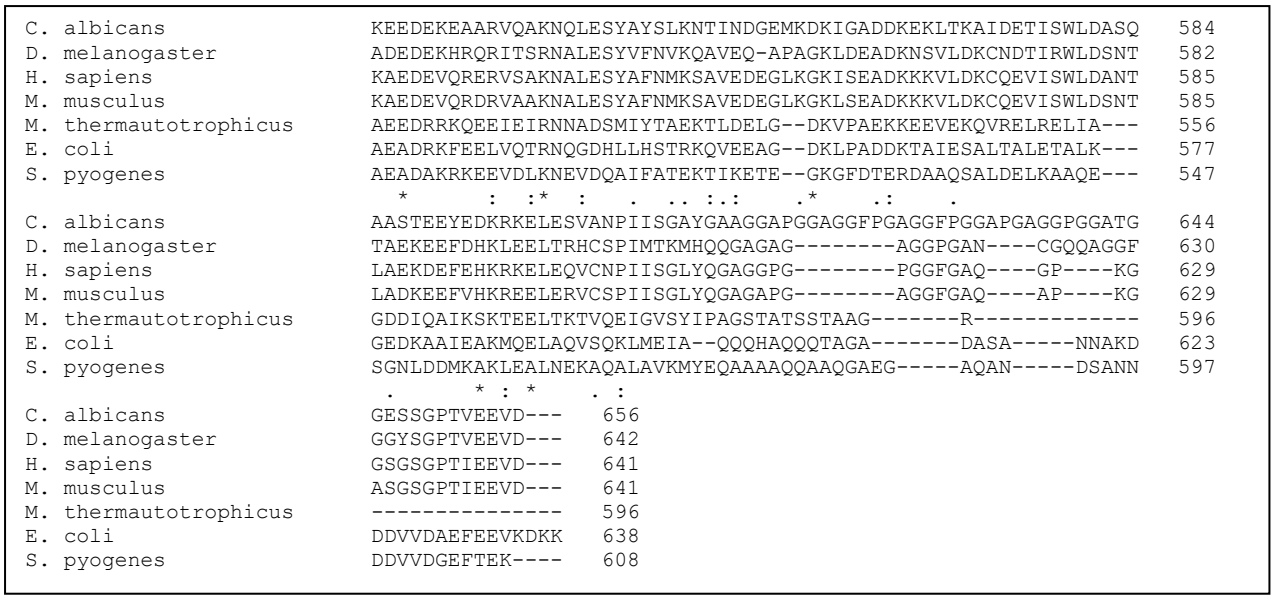


Figure 1.7 - Sequence alignment of DnaK/Hsp70 across species. *DnaK/Hsp70 sequences from Candida albicans, Drosophila melanogaster, Homo sapiens, Mus musculus, Methanothermobacter thermautotrophicus, Escherichia coli and Streptococcus pyogenes were aligned to show the conserved nature of DnaK across organisms, including bacteria, archaea and eukaryotes.. Fully conserved residues are indicated with *, partially conserved are indicated with: and similar residues are indicated with.. Protein sequences obtained from Uniprot.com, sequence alignment performed using Clustal Omega.*

1.5 Project Aims

The aims of this project were to further study the molecular interaction between Cas1 and DnaK in CRISPR systems. The approach to this was two-fold, firstly in *E. coli*, where the interaction between the two proteins is already established, specific amino acids of Cas1 were identified as potential interacting sites with Cas1. These sites were mutagenized and studied using gel-based assays along with fluorescence anisotropy *in vitro*, as well as *in vivo* assays to analyse physical interactions between proteins and functionality of these proteins. It was hypothesized that the variant Cas1 proteins produced in this study would interact with DnaK less strongly than the wild type Cas1 protein, and therefore bind DNA more strongly in the presence of DnaK than the wild type.

In addition to this, this project attempted to determine if the interaction between Cas1 and DnaK is a conserved method for controlling adaptation in bacteria and archaea. This was done by studying Cas1 and DnaK from *Streptococcus pyogenes*, a bacterium with a Type II-A CRISPR system, and *Methanothermobacter thermautotrophicus*, an archaeon with a Type III-C CRISPR system and comparing them to the interaction observed in *E.*

coli, which contains a Type I-E CRISPR system. If the interaction was conserved amongst these 3 proteins, who are distantly related to each other evolutionarily, it would be strong evidence that the interaction was universal amongst CRISPR systems. It was therefore hypothesised that the Cas1 and DnaK proteins from *S. pyogenes* and *M. thermautotrophicus* would behave similarly to *E. coli* Cas1 and DnaK in assays, and that a similar inhibitory interaction would be observed.

Chapter 2 - Materials and Methods

2.1 Strains

Table 2.1 – E. coli strains used in this study and their uses

***E. coli* Strain Name**	**Features**	**Use**
DH5α (Invitrogen)	*recA1* and *endA1* mutations to enhance transformation efficiency	Cloning of plasmids
BL21AI (Invitrogen)	*lon* and *OmpT* deletions to prevent excess degradation of heterologous proteins. Inducible T7 polymerase expression system. Tet^R	Protein expression
EB377	MG1655 strain with the genotype *araB::T7RNAP-tetA* to allow for inducible protein expression	Protein expression for naïve acquisition
BL21 Codon Plus (Aligent)	BL21 derivatives with additional copies of genes encoding codons rare in *E. coli* to assist expression of heterologous proteins. Cam^R	Protein expression
Rosetta 2 (Novagen)	BL21 derivatives with additional copies of genes encoding rare codons rare in *E. coli* to assist expression of heterologous proteins. Cam^R	Protein expression

2.2 Plasmids

Table 2.2 - Commercially available vector plasmids used in this study

Plasmid Name	**Size (bp)**	**Antibiotic Resistance**
pET14b (Novagen)	4671	Amp^R
pBad-HisA (Invitrogen)	4100	Amp^R
pACYC Duet (Novagen)	4008	$Chlm^R$
pET Duet (Novagen)	5420	Amp^R

Table 2.3 – Protein expression plasmids used in this study

Plasmid Name	**Notes**	**Creator**	**Purpose**
pEB505	pET14b with N-terminal 6x His tagged *E. coli cas1* cloned in using NdeI and EcoRI restriction sites	Edward Bolt	Protein overexpression
pEB628	pBAD-HisA with N-terminal 6x His tagged *E. coli cas1-cas2* construct cloned in using NcoI and XhoI restriction sites	Edward Bolt	Naïve acquisition assays
pTK37	Strep tagged *cas1-cas2* expression vector in pBadHisA	Tom Killelea	Pull-down Assays
pTK82	*E. coli* ^{His}*dnaK* expression vector in pACYC Duet	Tom Killelea	Protein overexpression
pALS01	Strep tagged *cas1 V76L-cas2* in pBadHisA. Made from pTK37.	Alison Stewart	Pull-down assays
pALS02	Strep tagged *cas1 R41G-cas2* in pBadHisA. Made from pTK37.	Alison Stewart	Pull-down assays

pALS03	^{His}*cas1 R41G* in pET14b. Made from pEB505.	Alison Stewart	Protein overexpression
pALS04	$^{His\text{-}}$*cas1 V76L* in pET14b. Made from pEB505.	Alison Stewart	Protein overexpression
pALS07	^{His}*cas1-cas2 V76L* in pBadHisA. Made from pEB628.	Alison Stewart	Naïve acquisition assays
pALS08	^{His}*cas1-cas2 R41G* in pBadHisA. Made from pEB628.	Alison Stewart	Naïve acquisition assays
pALS09	^{His}*cas1* from *M. thermautotrophicus* in pACYC Duet, cloned using BamHI and SalI restriction sites	Alison Stewart	Protein overexpression
pSpyCas1	^{His}*cas1* from *S. pyogenes* cloned into pET151. Amp^{R}	GeneArt	Protein overexpression
pSpyDnaK	^{His}*dnaK* from *S. pyogenes* cloned into pET151. Amp^{R}	GeneArt	Protein overexpression
pMthCas1	^{His}*cas1* from *M. thermautotrophicus* cloned into a pET100 D-TOPO vector. Amp$^{R.}$	GeneArt	Protein overexpression
pMthDnaK	^{His}*dnaK* from *M. thermautotrophicus* cloned into pET151. Amp^{R}	GeneArt	Protein overexpression

2.3 Antibiotics

Antibiotic	**Working Concentration**
Ampicillin (Sigma Aldrich)	100µg/ml
Chloramphenicol (Sigma Aldrich)	34µg/ml

Table 2.4 – Antibiotics used in this study and the concentrations they were used at.

2.4 Commercial Enzymes

All commercial enzymes used in this study were purchased from New England Biolabs.

Table 2.5 - A list of commercial enzymes used in this study.

Enzyme	**Features**
Q5 Polymerase	PCR polymerase
Vent Polymerase	PCR polymerase
Taq Polymerase	PCR polymerase
BamHI-HF	Restriction endonuclease with the cut site 5'-G*GATCC-3'
SalI-HF	Restriction endonuclease with the cut site 5'-G*TCGAC-3'
T4 DNA Ligase	Catalyses the formation of phosphodiester bonds between DNA nucleotides
T4 Polynucleotide Kinase (PNK)	Phosphorylates the 5' end of DNA for ligation
DpnI	Restriction endonuclease responsible for cutting methylated DNA with the cut site CH_3 | 5'-GA*TC-3'
Quick CIP	Dephosphorylates the ends of DNA strands to prevent relegation of linearised plasmids
Proteinase K	Protease that degrades peptide bonds

2.5 DNA Oligonucleotides and Substrates

All DNA oligonucleotides used in this study were supplied dried by Sigma-Aldrich, resuspended in sterile H_2O to 100μM and stored at -20°C.

Table 2.6 - Primers used for PCR amplification in this study

Primer Name	**Sequence (5'-3')**	**Notes**	**Calculated Tm for PCR**
oTK209 F	AACATTGTTGCTGTGGGTGGGGG	Used for V76L site directed mutagenesis of *E. coli* Cas1	62°C when used with Q5 polymerase
oTK209 R	CCAACTTGCGCAGC		
oALS01 F	GACAGGGATCACTGGCCATATTC	Used for R41G site directed mutagenesis of *E. coli* Cas1	59°C when used with Q5 polymerase
oALS01 R	TTGTCGATAAGTACAAACG		
oALS02 F	CGATGGATCCGAACTCTGCTGG	Used for amplifying Cas1 from *M. thermautotrophicus* genomic DNA, cloning in BamHI - SalI restriction sites	69°C when used with Q5 polymerase
oALS02 R	GCATGTCGACTTACCACCACATC		
SW1	AGCTGATCTTTAATAATAAGGAAAT	Used for amplifying the CRISPR locus of MG1655 derived strains of *E. coli*	44 °C when used with Taq Polymerase
SW2	CGGATTTATAAAGCTGACGGTT		

Table 2.7 – List of oligonucleotides used for DNA substrates in this study, their sequences and uses of each substrate

Substrate	**Oligo Name**	**Sequence (5'-3')**	**Assay the substrate was used for**
Fork 2a	MW12	GTCGGATCCTCTAGACAGCTCCATGATCA CTGGCACTGGTAGAATTCGGC	EMSAs
	Cy5-MW14	Cy5-CAACGTCATAGACGATTACATTGCTA CATGGAGCTGTCTAGAGGATCCGA	
Partial Fork 2a	MW12	GTCGGATCCTCTAGACAGCTCCATGATCA CTGGCACTGGTAGAATTCGGC	Disintegration assays
	MW14-Cy5	CAACGTCATAGACGATTACATTGCTACATG GAGCTGTCTAGAAGGATCCGA-Cy5	
	PM16	TGCCGAATTCTACCAGTGCCAGTGAT	
FAM - DNA 12	FAM-IIB1	FAM-AATCAAAGTGGACCCAACTCGAAAT CAACCGTAACAAGCAACAAGCAGGC	Fluorescence anisotropy
	IIB12	GCCTGCTTGTTG	
Cy5 - DNA12	Cy5-IIB1	Cy5-AATCAAAGTGGACCCAACTCGAAAT CAACCGTAACAAGCAACAAGCAGGC	EMSAs
	IIB12	GCCTGCTTGTTG	

2.6 Laboratory Recipes

All reagents used for this study were supplied by Sigma-Aldrich unless stated otherwise.

Table 2.8 – Gel types used in this study and their compositions

Gel Type	**Percentage**	**Composition**
Agarose	1%	1g agarose powder 100ml 1X TBE
SDS PAGE separating gel	10%	5.43ml dH_2O 3.75ml 30% acrylamide/bis-acrylamide

		1.4ml 3M tris-HCl pH 8.8 112µl 10% SDS 84µl 10% APS 9.5µl TEMED
	12%	4.48ml dH_2O 4.7ml 30% acrylamide/bis-acrylamide 1.4ml 3M tris-HCl pH 8.8 112µl 10% SDS 84µl 10% APS 9.5µl TEMED
SDS PAGE stacking gel	5%	1.75ml dH_2O 0.5ml 30% acrylamide/bis-acrylamide 0.75ml 0.5M tris-HCl pH 6.8 30µl 10% SDS 30µl 10% APS 3µl TEMED
1X TBE native polyacrylamide gel	5%	6.7ml 30% acrylamide/bis-acrylamide 4ml 10x TBE 29.3ml dH_2O 200µl 10% APS 50µl TEMED
	10%	13.3ml 30% acrylamide/bis-acrylamide 4ml 10X TBE 22.7mL dH_2O 200 µL 10% APS 50 µL TEMED
	12%	16ml 30% acrylamide/bis-acrylamide 4ml 10x TBE 20ml dH_2O 200µl 10% APS 50µl TEMED

1X LIS native polyacrylamide gel	5%	6.7ml 30% acrylamide/is-acrylamide 4ml 10x LIS Buffer 29.3ml dH_2O 200µl 10% APS 50µl TEMED

Table 2.9 - Commonly used buffer recipes in this study and their compositions

Buffer	**Composition**
10X tris, boric acid, EDTA (TBE)	1M tris 1M boric acid (Fisher Bioreagents) 20mM EDTA (Fisher Bioreagents)
10X SDS running buffer	250mM tris 1.92M glycine (Fisher Bioreagents) 1% (w/v) SDS
10X low ionic salt (LIS) buffer	67mM tris-HCl pH 8.0 33mM sodium acetate (Fisher Bioreagents)
5X helicase buffer	100mM Tris-HCl pH 8.0 0.5mg/ml BSA 35% (v/v) glycerol (Honeywell)
4X SDS loading buffer	0.2M tris-HCl pH 6.5 0.4M DTT 8% (w/v) SDS 6mM Bromophenol Blue 32% (v/v) glycerol
5X DNA loading dye	80% (v/v) glycerol Orange G dye to colour preference
1X cleavage buffer	20mM tris-HCl pH 7.5 10mM NaCl 1mM DTT

	5mM $MgCl_2$ (Acros Organics)
1X stop buffer	2 mg/mL Proteinase K (New England Biolabs) 2.5% (w/v) SDS 200 mM EDTA
1X dilution buffer	20mM tris-HCl pH 8.0 1mM DTT 100mM NaCl (Fisher Bioreagents)
10X annealing buffer	100mM tris-HCl pH 7.5 500mM NaCl
Elution buffer	4mM tris-HCl pH 7.5 10mM NaCl
1X resuspension buffer	20 mM MOPS pH7 (Calbiochem) 200 mM NaCl 0.1µM PMSF
1X wash buffer	20 mM MOPS pH7 100 mM NaCl 50 mM imidazole
Nickel charge buffer	0.1M $NiCl_2$ (Alfa Aesar)

Table 2.10 - Commonly used reagents in this study and their compositions

Reagent	**Composition**
Coomassie blue stain	0.05% (w/v) Brilliant Blue R-250 (Fisher Bioreagents) 10% (v/v) acetic acid (Fisher Bioreagents) 40% (v/v) methanol (VWR Chemicals)
Coomassie destain solution	10% (v/v) acetic acid 20% (v/v) methanol
Mu broth	10g/L tryptone (VWR Chemicals) 10g/L NaCl 5g/L yeast extract (Thermofisher Scientific) 2mM NaOH (Fisher Bioreagents)
Mu agar	3g agar powder (VWR Chemicals) 200ml Mu broth Antibiotic to working concentration where appropriate

2.7 Gel Electrophoresis

2.7.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used for the visualisation of unlabelled DNA strands, such as PCR products or plasmids. 6X Purple Loading Dye (New England Biolabs) was added to DNA samples, and these were loaded alongside 3µl 1kb ladder or 100bp ladder (New England Biolabs) on a 1% agarose gel containing 0.2µg/ml ethidium bromide to stain DNA. Gels were electrophoresed at 120V for 60 minutes in 1X TBE before being visualised with the Syngene™ U: genius 3 using UV light.

2.7.2 SDS PAGE Gel Electrophoresis

SDS protein agarose gel electrophoresis (SDS PAGE) was used for the visualisation of proteins following overexpression or purification. SDS PAGE gels consist of a low pH, low percentage acrylamide stacking gel poured on top of a higher pH, higher percentage

acrylamide separating gel to separate proteins according to their molecular weight. Samples to be analysed were added to 1X SDS loading buffer and heated to 95°C for 10 minutes to denature proteins. Samples were loaded alongside 3µl Coloured Protein Standard (CPS) (New England Biolabs) before being electrophoresed at 120V for 60 minutes. Gels were then stained using Coomassie blue stain for 15 minutes, before being rinsed in water, following which Coomassie destain solution was added. For protein purification, a 10% acrylamide separating gel was used and for pull-down assays, a 12% acrylamide separating gel was used.

2.7.3 Native TBE Gel Electrophoresis

Native TBE gel electrophoresis was used for the separation of DNA molecules from *in-vitro* assays. 12µl of samples was loaded onto a native TBE gel, and gels were electrophoresed at 120V for 90 minutes unless otherwise stated.

2.8 Microbiology and Cloning

2.8.1 Polymerase Chain Reactions

PCR reactions were used for the amplification of DNA for site-directed mutagenesis, cloning of plasmids and analysis of results of naïve acquisition assays. Forward and reverse primers were used to amplify a gene of interest. All PCR reactions were performed using an Applied Biosystems™ Veriti™ Thermal Cycler, 96-well.

2.8.1.1 Q5 Polymerase PCR

PCR mixtures using Q5 Polymerase were assembled according to the protocol available at https://www.neb.com/en-gb/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491 and thermocycled using the following conditions:

Table 2.11 - Thermocycling conditions for a PCR Reaction using Q5 Polymerase

Step	Temperature (°C)	Length of Time (s)	Number of Repeats
Initial Denaturation	98	30	1
Denaturation	98	5-10	25-35
Primer Annealing	50-72 depending on annealing temperature of primes used	10-30	
Extension	72	20-30 per kb	
Final Extension	72	120	1

2.8.1.2 Vent Polymerase PCR

PCR mixtures using Vent Polymerase were assembled according to the protocol available at https://www.neb.com/en-gb/protocols/2012/09/06/protocol-for-a-routine-vent-pcr-reaction and thermocycled using the following conditions

Table 2.12 – Thermocycling conditions for a PCR Reaction using Vent Polymerase

Step	Temperature (°C)	Length of Time (s)	Number of Repeats
Initial Denaturation	95	120-600	1
Denaturation	95	15-30	20-30
Primer Annealing	55-65 depending on annealing temperature of primes used	15-30	
Extension	72	60 per kb	
Final Extension	72	300	1

2.8.1.3 Taq Polymerase PCR

PCR mixtures using Taq Polymerase were assembled according to the protocol available at https://www.neb.com/en-gb/protocols/2012/09/11/protocol-for-onetaq-quick-load-2x-master-mix-with-standard-buffer-m0486.

Table 2.13 - Thermocycling conditions used for a PCR reaction using Taq Polymerase

Step	**Temperature (°C)**	**Length of Time (s)**	**Number of Repeats**
Initial Denaturation	94	30	1
Denaturation	94	15-30	30
Primer Annealing	45-68 depending on annealing temperature of primes used	15-60	
Extension	68	60 per kb	
Final Extension	68	300	1

2.8.2 Site Directed Mutagenesis

Mutagenic primers were designed using NEBase Changer (available at https://nebasechanger.neb.com/) and used to amplify a mutagenized version of the relevant plasmid. PCR amplification was confirmed via analysis on a 1% agarose gel, following which 2µl mutagenic PCR product was incubated with 1µl of T4 PNK, T4 DNA Ligase and DpnI) in 1x T4 DNA Ligase buffer (New England Biolabs) in a 20µl reaction. Reaction mix was left to incubate at room temperature for 1 hour before being transformed into DH5α cells. A single colony was then used to set up an overnight culture, from which plasmid DNA was extracted. Resulting pure plasmid was then sequenced using Sanger Sequencing via Genewiz from Azenta Life Sciences.

2.8.3 Restriction Digests

Restriction digests were performed on plasmid vector DNA and insert DNA to create complementary ends suitable for ligating together. Restriction digests were performed using BamHI-HF and SalI-HF according to the protocol available at

https://nebcloner.neb.com/#!/protocol/re/double/BamHI-HF,SalI-HF (last accessed 29/08/24). For restriction digestions of plasmid vector, 1µl nuclease free water was omitted and 1µl Quick CIP was added to dephosphorylate the ends of the DNA, preventing ligation of plasmid vector to itself.

2.8.4 Ligations

Ligations were performed to join complementary ends of DNA strands together to clone a gene of interest into a plasmid vector. Ligations were carried out using T4 DNA Ligase according to the protocol available at https://nebcloner.neb.com/#!/protocol/Ligation/M0202. Ligation mix was then transformed into DH5α cells using standard transformation protocol. A single colony was then used to set up an overnight culture, from which plasmid DNA was extracted. Successful ligation was confirmed using double restriction digest with BamHI and SalI, according to the protocol available at https://nebcloner.neb.com/#!/protocol/re/double/BamHI-HF,SalI-HF following which the reaction was run on a 1% agarose gel.

2.8.5 Extraction of DNA from Agarose Gels and PCR Clean Up

DNA was extracted and purified using the Wizard® SV Gel and PCR Clean-Up System (ProMega), protocol available at https://www.promega.co.uk/resources/protocols/technical-bulletins/101/wizard-sv-gel-and-pcr-cleanup-system-protocol/

2.8.6 Production of Competent Cells

Experimental procedure was carried out under sterile conditions. Glycerol stocks of competent cells were streaked out onto Mu agar with no antibiotic and incubated overnight at 37°C. A single colony was used to inoculate 50ml Mu broth in an overnight culture and left to grow for 16-18 hours. 10ml overnight culture was used to inoculate 1L Mu broth and cells were left to grow at 37°C with shaking until OD600 reached 0.4-0.6. Cells were then pelleted at 1878g at 4°C for 5 minutes. Cells were resuspended in 10ml ice cold sterile 0.1M $CaCl_2$ and left to incubate on ice for 1-6 hours. Cells were pelleted at 1878g at 4°C for 5 minutes, before being resuspended in 31.25ml ice cold sterile 0.1M

$CaCl_2$ and 18.75ml 80% glycerol. Cells were aliquoted in 100-400µl aliquots before being flash frozen with dry ice and stored at -80°C.

2.8.7 Transformation of Competent Cells

Experimental procedure was carried out under sterile conditions. 1µl plasmid was left to incubate on ice with 100µl competent cells for 30 minutes, before being heat shocked at 42°C for 90 seconds. Mixture was left on ice for 5 minutes, before 900µl Mu broth was added, and cells were left to incubate at 37°C with shaking for 1 hour. Cells were pelleted via centrifugation at 20238g for 1 minute, and 800µl Mu broth was removed. Cells were resuspended in remaining 200µl Mu broth and plated onto Mu agar containing relevant antibiotic(s). Cells were left to grow for 16-18 hours at 37°C.

2.8.8 Overnight Cultures

5ml Mu broth with relevant antibiotic were inoculated with a single bacterial colony and left to grow shaking at 114rpm for 16-18 hours at 37°C.

2.8.9 Extraction of Plasmid DNA from Cells

Plasmid DNA was extracted from cells using the Wizard® Plus SV Minipreps DNA Purification Systems (ProMega), protocol available at https://www.promega.co.uk/resources/protocols/technical-bulletins/0/wizard-plus-sv-minipreps-dna-purification-system-protocol/

2.9 Protein Production and Overexpression

2.9.1 Protein Overexpression

All proteins were first subjected to a pilot protein overexpression, during which 50ml Mu broth was inoculated with 500µl overnight culture and allowed to grow at 37°C with shaking until OD600 reached 0.6. At this point, protein expression was induced with the addition of L-arabinose to 0.2% final concentration, and IPTG (VWR Chemicals) to 1mM final concentration. Cells were then left to incubate at 37°C for a further 3 hours with shaking, after which 1ml of culture was taken and pelleted at 15871g. Pellet was resuspended in 150µl sterile H_2O and 1x SDS loading buffer and heated to 95°C for 10 minutes before being stored at -20°C.

Once protein overexpression was confirmed via SDS PAGE analysis, a larger scale overexpression was performed to obtain sufficient protein for purification. All proteins were expressed using the same conditions as in the pilot protein overexpression, except *S. pyogenes* Cas1, which was left to incubate at 18°C for 16-18 hours following the addition of L-arabinose and IPTG.

Following overexpression, cells were pelleted at 6000g for 15 minutes, before being resuspended in 5ml/L culture of resuspension buffer, before being stored at -20°C.

Table 2.14 - Different Resuspension Buffers used and their compositions

Resuspension Buffer	**Composition**
E. coli proteins	20mM tris-HCl pH 7.5 150mM NaCl 10% (v/v) glycerol 0.5mM PMSF 20mM imidazole
S. pyogenes proteins	300 mM NaCl 10% (v/v) glycerol 5 mM β-mercaptoethanol 20 mM tris-HCl pH 8.0 30mM imidazole
M. thermautotrophicus proteins	500mM NaCl 10% (v/v) glycerol 20mM Tris pH 8.0 20mM imidazole 0.5mM PMSF

2.9.2 Protein Purification

Stored biomass was thawed at room temperature and sonicated at 40-60% amplitude in 5 ml aliquots in 10 second intervals for 3 minutes using a Jencons Scientific ltd Vibracell™ sonicator. Cell debris was pelleted at 30966g for 30 minutes. All proteins were purified with an N-terminal hexahistidine tag using the ÄKTA™ start chromatography system

(Cytiva). All final protein concentrations were determined using either the Bradford Assay or the Bicinchoninic (BCA) Assay, and proteins were flash frozen using dry ice before being stored at -80°C.

Table 2.15 - Buffers used to purify proteins used in this study

Buffer	**Composition**
A	20mM tris-HCl pH 7.5, 500mM NaCl, 20mM imidazole, 10% glycerol
B	20mM tris-HCl pH7.5, 100mM NaCl, 1mM DTT, 10% glycerol
C	20mM tris-HCl pH7.5, 150mM NaCl, 10% glycerol
D	20 mM tris pH 7.5, 150 mM NaCl, 20mM imidazole, 10% glycerol
E	20 mM tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 10 mM $MgCl_2$, 5 mM ATP, 10% glycerol
F	50 mM tris pH 7.5, 100 mM NaCl, 1 mM DTT, 10% glycerol
G	300 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 20 mM imidazole, 20 mM tris-HCl pH 8.0
H	20mM tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 10% w/v glycerol
I	300 mM NaCl, 10% glycerol, 20 mM imidazole, 20 mM tris-HCl pH 8.0
J	300mM NaCl, 10% w/v glycerol, 20mM tris-HCl pH 8.0, 20mM imidazole, 10mM $MgCl_2$, 10mM ATP
K	500mM NaCl, 20mM tris pH 8.5, 20mM imidazole, 10% glycerol

L	20mM tris-HCl pH 8.5, 250mM NaCl, 1mM DTT, 10% glycerol
M	20mM tris pH 8.5, 10% glycerol, 1mM DTT, 100mM NaCl
N	50 mM tris pH 7.5, 100 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, 5 mM ATP, 10% glycerol

Table 2.16 – Columns used for protein purification in this study

Column	**Usage**
HiTrap chelating HP (5ml) (Cytiva)	Metal chelating column charged with Ni^{2+} before use with nickel charge buffer. Used for purification of hexa-His tagged proteins
HiTrap heparin HP (1ml) (Cytiva)	Purification of DNA binding proteins
HiTrap Q HP anion exchange (1mL) (Cytiva)	Q-Sepharose matrix used for anion exchange chromatography

2.9.2.1 *E. coli* Cas1, V76L and R41G Mutants

Cleared cell lysate was loaded onto a 5ml HiTrap chelating column pre-equilibrated with Buffer A. Protein was then eluted with 20-500mM imidazole over 30 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and fractions containing desired protein were dialysed overnight at 4°C in 2L Buffer B. Dialysed sample was then loaded on a 1ml HiTrap heparin HP column pre-equilibrated with Buffer C. Protein was eluted with 150-1000mM NaCl over 20 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and fractions containing desired protein were dialysed overnight at 4°C in storage buffer (20mM tris-HCl pH7.5, 100mM NaCl, 1mM DTT, 20% glycerol). Cas1 R41G was concentrated using a 4ml Spin Concentrator (5kDa) (Aligent Technologies) with centrifugation at 5000g before being flash frozen.

2.9.2.2 *E. coli* DnaK

E. coli DnaK was purified according to the method set out in (Killelea *et al.*, 2023). Cleared cell lysate was loaded onto a 5ml HiTrap chelating column (GE Healthcare) pre-equilibrated with Buffer D before being washed with Buffer E. Protein was eluted with 20-500mM imidazole over 30 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and fractions containing DnaK were dialysed overnight at 4°C in Buffer F. Dialysed sample was loaded onto a 1 ml HiTrap Q HP column and eluted with 100-1000mM NaCl over 20 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and DnaK containing fractions were dialysed into storage buffer (50 mM tris pH 7.5, 50 mM NaCl, 1 mM DTT, 20% glycerol) overnight at 4°C.

2.9.2.3 *S. pyogenes* Cas1

S. pyogenes Cas1 was purified based on the method set out in (Ka *et al.*, 2016). Cleared cell lysate was loaded onto a 5ml HiTrap chelating column pre-equilibrated in Buffer G. Elution was carried out using 20-500mM imidazole over 25 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and *S. pyogenes* Cas1 containing fractions were dialysed overnight in Buffer H. Dialysed sample was then loaded onto a 1ml HiTrap Heparin HP Column (GE Healthcare) pre-equilibrated in Buffer H. Elution was carried out using 100-1000mM NaCl over 20 1ml fractions. Fractions were analysed on a 10% SDS Page Gel and *S. pyogenes* Cas1 containing fractions were dialysed overnight into storage buffer (300mM NaCl, 20% glycerol, 1mM DTT, 20mM tris-HCl pH 8.0).

2.9.2.4 *S. pyogenes* DnaK

Cleared cell lysate was loaded onto a 5ml HiTrap chelating column pre-equilibrated in Buffer G and washed with Buffer J. Elution was carried out using 20-500mM imidazole over 25 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and *S. pyogenes* DnaK containing fractions were dialysed overnight in Buffer H. Dialysed sample was loaded onto a 1 ml HiTrap Q HP column (Cytivia™) pre-equilibrated in Buffer H. Elution was carried out using 100-1000mM NaCl over 20 1ml fractions. Protein containing fractions were analysed on a 10% SDS PAGE gel, and *S .pyogenes* DnaK containing fractions were dialysed overnight into storage buffer (20mM tris-HCl pH 8.0, 50mM NaCl, 1mM DTT, 20% w/v glycerol).

2.9.2.5 *M. thermautotrophicus* Cas1

Cleared cell lysate was loaded onto a 5ml HiTrap chelating column pre-equilibrated in Buffer K. Elution was done with 20-500mM imidazole over 30 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and fractions containing *M. thermautotrophicus* Cas1 were dialysed overnight at 4°C into Buffer L. Dialysed sample was loaded onto a 1ml HiTrap heparin HP column pre-equilibrated with Buffer M. Protein was eluted using 100-1000mM NaCl over 20 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and *M. thermautotrophicus* Cas1 containing fractions were dialysed overnight at 4°C into storage buffer (20mM tris pH 8.5, 500mM NaCl, 1mM DTT, 20% glycerol).

2.9.2.6 *M. thermautotrophicus* DnaK

Cleared cell lysate was loaded onto a 5ml HiTrap Chelating column pre-equilibrated in Buffer D. Elution was done with 20-500mM imidazole over 30 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and fractions containing *M. thermautotrophicus* DnaK were dialysed overnight at 4°C into Buffer F. Dialysed sample was loaded on to a 1ml HiTrap Q HP column and washed with Buffer N. Elution was carried out with 100-1000mM NaCl over 20 1ml fractions. Fractions were analysed on a 10% SDS PAGE gel and *M. thermautotrophicus* DnaK containing fractions were dialysed into storage buffer (50 mM tris pH 7.5, 50 mM NaCl, 1 mM DTT, 20% glycerol) overnight at 4°C.

2.9.3 Bradford Assay

Bradford assay was performed using Bradford Reagent (Sigma Aldrich) following the microplate procedure available at https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/358/973/b6916bul-mk.pdf. Absorbance readings were taken using the FLUOstar Omega Microplate Reader (BMG Labtech). Readings were analysed using GraphPad Prism using a linear regression model, and unknown protein concentrations were interpolated from a standard curve.

2.9.4 BCA Assay

BCA assays were performed using the Pierce™ BCA Protein Assay Kit (Thermofisher Scientific) microplate procedure, protocol available at https://assets.thermofisher.com/TFSAssets/LSG/manuals/MAN0011430_Pierce_BCA_P

rotein_Asy_UG.pdf. Absorbance readings were taken using the FLUOstar Omega Microplate Reader (BMG Labtech). Readings were analysed using GraphPad Prism using a linear regression model, and unknown protein concentrations were interpolated from a standard curve.

2.10 In Vitro Biochemical Analysis

2.10.1 DNA Substrate Preparation

DNA oligos was added to desired concentration in a final volume of 100µl in 1X annealing buffer and heated to 95°C for 10 minutes. Samples were allowed to cool slowly to room temperature overnight. 5X DNA loading dye was added, and samples were loaded onto a 10% 1X TBE native gel. Gel was run at 120V for 180 minutes to separate unannealed single stranded DNA from double stranded DNA. Gel containing double stranded DNA substrate was excised, and DNA was left to elute for up to 72 hours in elution buffer (20mM Tris-HCl pH 7.5, 50mM NaCl). DNA substrate was then stored at -20°C until use.

2.10.2 Electrophoretic Mobility Shift Assays (EMSAs)

Proteins were diluted to 10x final concentrations in 1X dilution buffer and allowed to incubate on ice. 2µl protein dilution was added to 18µl reaction mix (20nM Cy5 labelled DNA substrate, 0.5X helicase buffer). Reaction was allowed to incubate for 15 minutes at 37°C, before the addition of 5µl 5X DNA loading dye before being loaded on a 5% native TBE gel. Gel was electrophoresed at 120V for 90 minutes before being visualised using Amersham™ Typhoon™ Biomolecular Imager (Cytiva).

2.10.3 Anisotropy

Proteins were diluted to 10x final concentrations in 1X dilution buffer and left to incubate on ice. Cas1 dilutions were left to incubate for 5 minutes at 37°C with an equal volume of dilution buffer or DnaK. 15µl protein mix was added to 60µl reaction mix (40nM FAM-DNA12, 0.5X helicase buffer) and maintained at 37°C while measurements were taken. Polarisation readings were taken using the FLUOstar Omega Microplate Reader (BMG Labtech) at 0-, 5- and 10-minute intervals following the addition of the protein mix. Readings were analysed using GraphPad Prism using the one-site binding model.

2.10.4 Disintegration Assays

Proteins were diluted to 10x working conditions in 1X cleavage buffer, before 2µl diluted Cas1 was added to 16µl reaction mix containing 20nM partial fork 2a in 1X cleavage buffer, along with 2µl cleavage buffer or DnaK. Samples were left to incubate at 37°C for 30 minutes, before 5µl 1X stop buffer was added. Reaction mix was left for 15 minutes at 37°C, following which 6.5µl 5x DNA loading buffer was added. Samples were loaded on a 12% native TBE gel and electrophoresed at 120V for 90 minutes and visualised on the Amersham™ Typhoon™ Biomolecular Imager (Cytiva).

2.11 In Vivo Experimental Procedures

2.11.1 Pull-Down Assays

50ml of Mu Broth containing relevant antibiotics was inoculated with 0.5ml overnight culture of cells containing relevant plasmids. Cells were left to grow with shaking at 37°C until OD600 reached 0.6, at which point protein expression was induced with the addition of L-arabinose to 0.02% and IPTG to 1mM. Cells were left for a further 3 hours at 37°C with shaking, before cells were harvested by centrifugation at 1878rcf for 5 minutes. Pelleted cells were stored at -20°C. Cell pellets were thawed at room temperature and resuspended in 2ml 1X resuspension buffer, following which cells were sonicated for 1 minute at 40-60% amplitude in 5 second intervals to lyse cells. Cell debris was removed via centrifugation of 1ml aliquots of sonicated sample at 20817g for 30 minutes at 4°C. 100µl of resuspended iminodiacetic acid sepharose (Sigma Aldrich) was charged with 1 ml of 0.1 M $NiCl_2$, before being washed with 1ml 0.25M imidazole. Iminodiacetic acid sepharose was then washed 3 times with 1ml resuspension buffer before the addition of 1ml cleared cell lysate. Iminodiacetic acid sepharose and cleared cell lysate were then left to incubate for 60 minutes at 4°C while rotating. Iminodiacetic acid sepharose was then washed 3 times with 1X wash buffer with gentle agitation at room temperature for 5 minutes between each wash. 100µl 1x SDS loading buffer was added and samples were heated to 95°C for 10 minutes, before analysis on a 12.5% SDS PAGE gel.

2.11.2 Naïve Adaptation Assays

EB377 cells were transformed with relevant plasmids and plated onto Mu agar plated with relevant antibiotic(s) and left to incubate at 37°C for 16-18 hours. A single colony

was used to inoculate 5ml Mu broth with 0.2% L-arabinose under sterile conditions and cultures were left to shake at 114rpm for 16-18 hours to generate P1 samples. 17µl P1 sample was used to inoculate 5ml fresh Mu broth with 0.2% L-arabinose under sterile conditions and cells were left to grow at 37°C with shaking at 114rpm until OD600 reached 0.4 to make P2 samples. 50µl of sample was added to 200µl sterile H_2O and samples were denatured at 95°C for 10 minutes. CRISPR locus was amplified via PCR with SW1 and SW2 primers and Taq polymerase and PCR products were run on a 1% agarose gel. Gel was visualised using Syngene™ U:genius 3.

2.12 Bioinformatic Analysis

2.12.1 Online Databases used to Predict Structures of Proteins

Structural prediction was carried out using Alphafold 3, accessed via Alphafold Server available at https://golgi.sandbox.google.com/ (last accessed 26/08/24). Protein sequences were obtained from UniProt, available at https://www.uniprot.org/ (last accessed 26/08/24). Predicted structures were opened in PyMol, available at https://pymol.org/2/ and structures were annotated.

2.12.2 Amino Acid Sequence Alignment

Multiple sequence alignment was carried out using Clustal Omega, available at https://www.ebi.ac.uk/jdispatcher/msa/clustalo?stype=protein. Protein sequences were obtained from UniProt, available at https://www.uniprot.org/.

2.12.3 Nucleotide Sequence Extraction

Nucleotide sequences for the production of protein expression plasmids were taken from Kegg Orthology, *S. pyogenes* Cas1 reference SPy_1047, *S. pyogenes* DnaK SPy_1760, *M. thermautotrophicus* Cas1 reference MTH_1084, and *M. thermautotrophicus* DnaK MTH_1290.

Chapter 3 - Production and Biochemical Analysis of *E. coli* Cas1 and Mutants

3.1 Introduction

Although the interaction between Cas1 and DnaK was established in *E. coli* by Killelea et al. (2023), little is understood about the mechanisms of the interaction, including mapping the site of interaction. Therefore, the goal of this area of research was to determine amino acids of Cas1 required for the interaction with DnaK by predicting interacting sites and mutagenizing them to see if the interaction was affected.

The first amino acid selected to mutate was V76, as this mutant has been shown to be hyperactive in acquisition by Yosef et al. (2023). It was thought that the reason for the hyperactive acquisition seen with this mutant might be due it having lost the interaction with DnaK. It was therefore hypothesised that this mutation would remove the interaction between Cas1 and DnaK, which would prevent DnaK from being able to restrain Cas1 leading to observed hyperactive acquisition *in vivo*.

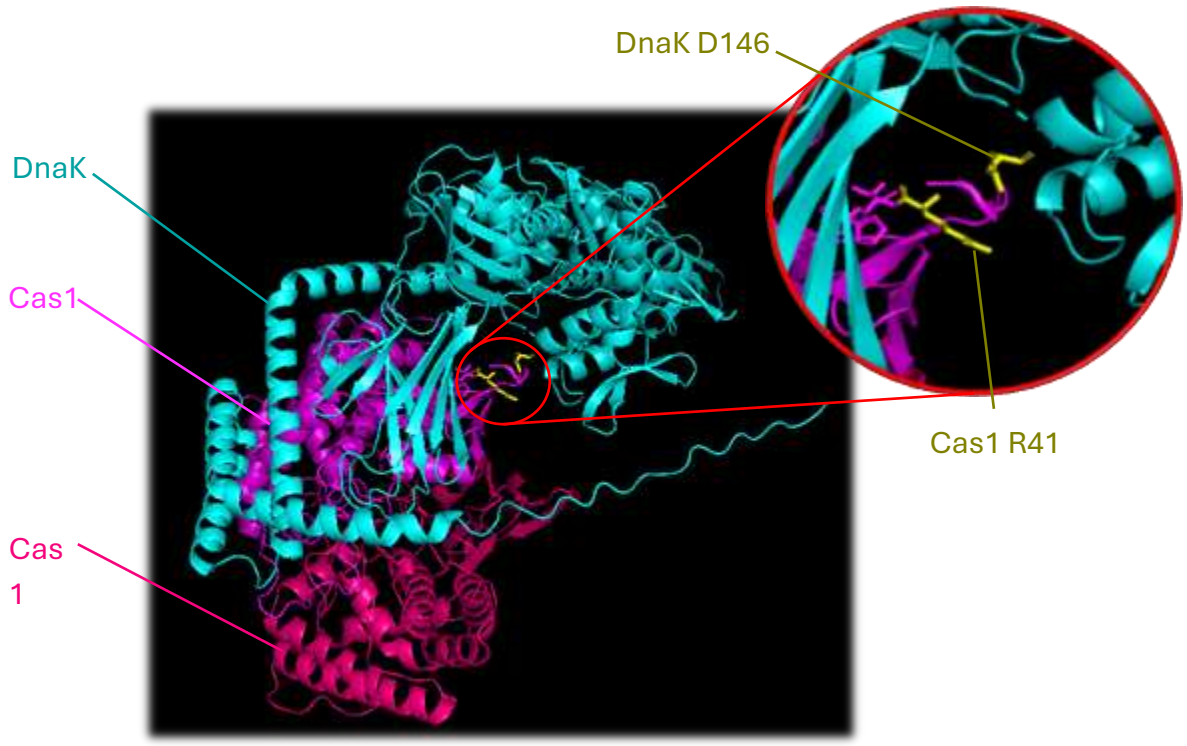


Figure 3.1 – Structural prediction of E. coli DnaK interacting with E. coli Cas1 homodimer*. Protein sequences taken from Uniprot, structure created using Alphafold 3, annotated In PyMol. Highlighted in yellow are potential interacting residues of Cas1 and DnaK which were the basis for R41G mutagenesis in this study.*

The second amino acid selected to mutate was R41, and this was mutated based on bioinformatic modelling of Cas1 and DnaK using Alphafold 3, which can be seen in Figure 3.1. Structural models indicate that the β-pleated sheet of Cas1 containing R41 interacts closely with DnaK, therefore this residue was selected for mutation to a glycine to test whether this would knock out this interaction.

3.2 Site Directed Mutagenesis of cas1 gene

Once amino acids of interest had been identified based on research by Yosef et al. (2023) and by structural modelling, mutagenic primers were designed to introduce a V76L mutation and an R41G mutation into the Cas1 protein so the impact of these mutations could be studied separately from each other. Mutagenic primers were designed using NEBase changer, and template DNA was amplified via PCR amplification with either Vent or Q5 polymerase using these mutagenic primers. The resulting PCR product then underwent a kinase, ligase, digestion (KLD) reaction using T4 PNK, T4 DNA Ligase and DpnI for 1 hour at room temperature. This allows phosphorylation and ligation of mutated DNA to form a circular plasmid and degradation of template DNA. The reaction mixture

was then transformed into DH5α cells and plasmid DNA was extracted, and successful mutagenesis was confirmed via Sanger sequencing, and the mutated DNA sequences can be seen in Figure 3.2.

pEB505, a 6xHis-Cas1 expression plasmid, pEB628, a 6xHis-Cas1-Cas2 plasmid and pTK37, a strep-Cas1-Cas2 expression plasmid, were all mutated with both R41G and V76L mutations for this study to allow for comprehensive analysis of the mutants using both in vitro and in vivo methodologies. pEB505 and its mutants were used to overexpress proteins for purification. pEB628 and its mutants were used to study naïve acquisition *in vivo*. pTK37 and its mutants were used to investigate a physical interaction between Cas1 and DnaK *in vivo* using pull down assays.

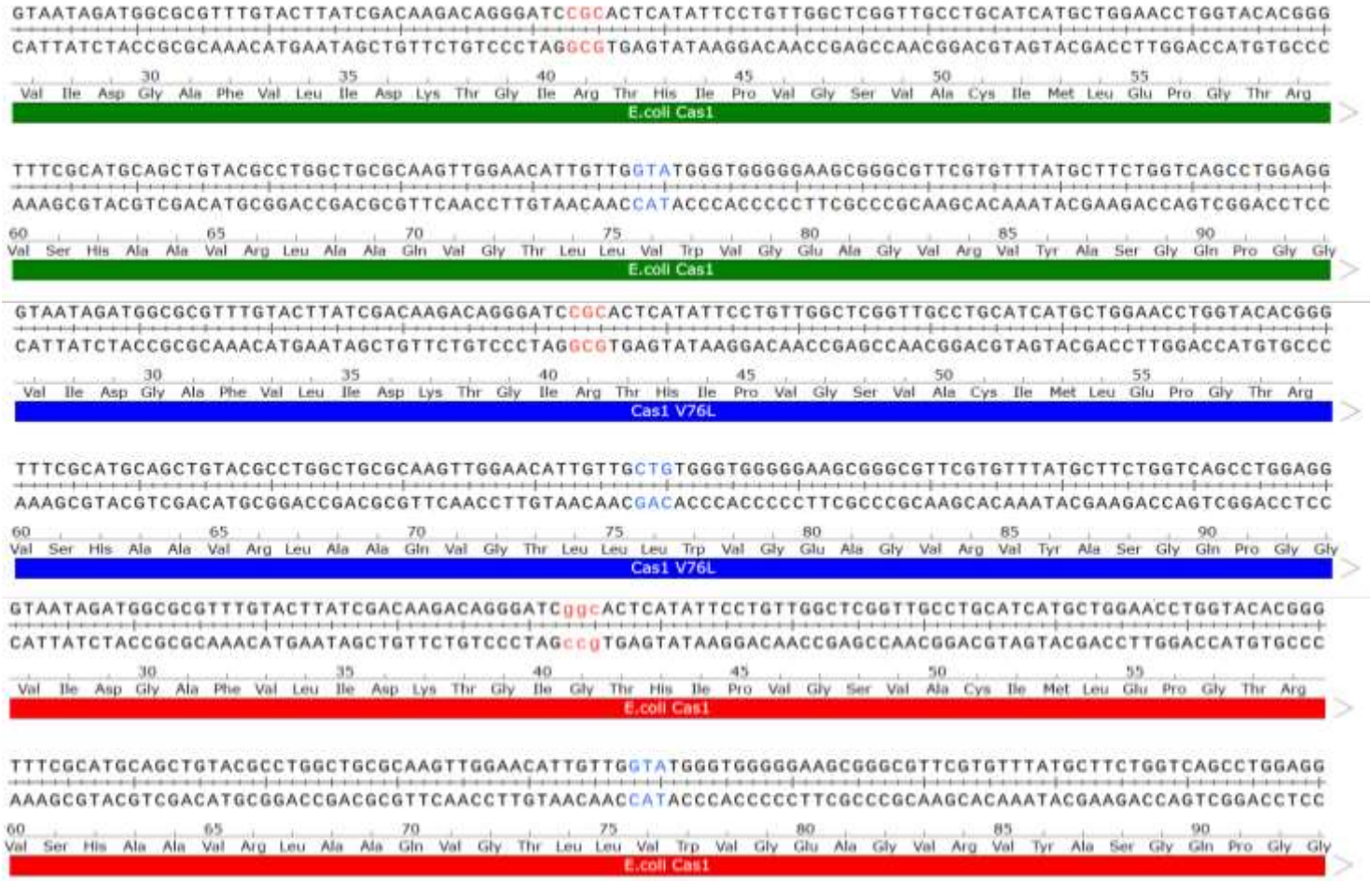


Figure 3.2 - Sequencing data from site directed mutagenesis of cas1 gene. *A section of the E. coli Cas1 protein sequence can be seen in green, Cas1 V76L can be seen in blue and Cas1 R41G can be seen in red. Positions R41 and V76 have been highlighted in red and green respectively to indicate the original vs mutated sequences.*

3.3 Protein Overexpression and Purification

3.3.1 Overexpression of Proteins

Once successful mutagenesis had been confirmed, plasmids were then subjected to a pilot protein overexpression to confirm that the mutant proteins expressed similarly to WT Cas1 in BL21AI cells. This required the transformation of relevant plasmids into BL21AI to allow for controlled overexpression of proteins, which was then subjected to pilot protein overexpression as described in Method 2.9.1 Protein expression was induced with L-arabinose and IPTG to appropriate concentrations. L-arabinose allows for controlled expression of proteins via the *ara*Bad promoter, which leads to expression of the T7 RNA polymerase. IPTG binds to the *lac* repressor on the T7 promotor of plasmids, removing it and allowing the T7 RNA polymerase to bind. Following protein overexpression, prepared SDS PAGE samples were ran on a 10% SDS PAGE gel alongside 3µl CPS before being stained with Coomassie blue stain and destained. SDS PAGE analysis, seen in Figure 3.3, showed similar levels of expression of WT Cas1 and both relevant mutants. This process was then scaled up to produce biomass following the procedure described in 2.9.1.

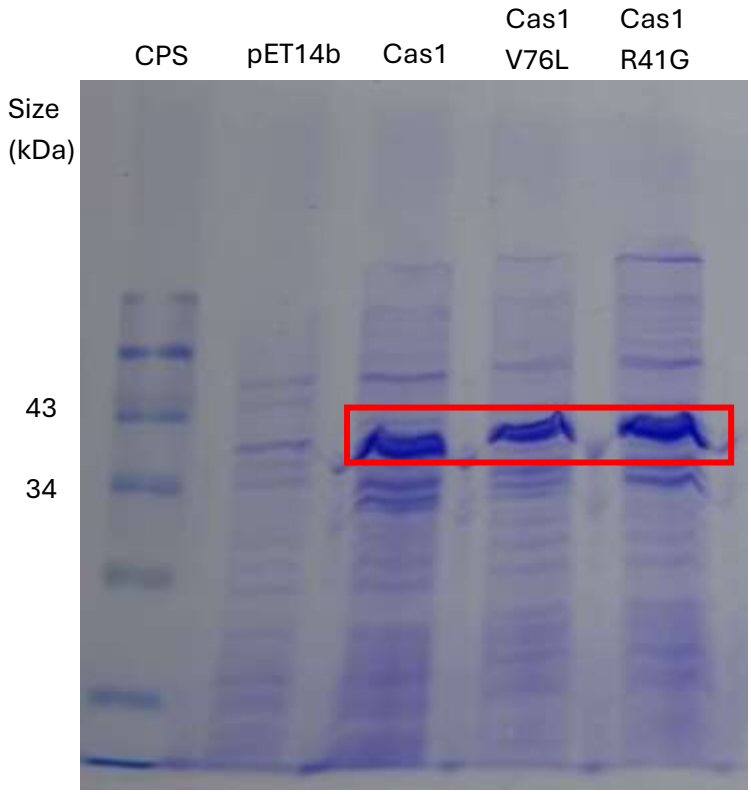


Figure 3.3 - Pilot protein overexpression of WT Cas1 and the mutants that were created during this study. *10% SDS Page gel showing expression of proteins in BL21AI cells. Proteins of interest are indicated with the red box.*

3.3.2 Protein Purification

Following production of cell biomass, proteins were purified using the method set out in 2.9.2. An overview of the purification steps used in this chapter can be seen in Figure 3.4. UV chromatograms produced by the AKTA Start machine used to purify proteins allowed identification of fractions containing proteins, which were then analysed using SDS Page analysis to identify fractions containing specific proteins of interest, as seen in Figure 3.5. Fractions from the Ni^{2+} NTA column were ran alongside cleared cell lysate to allow comparison to the overexpression, and the wash through/flow through containing proteins that did not bind or bound weakly to the column to show these had been removed. Fractions from the heparin column elution were run alongside the dialysed sample from the previous column, and the wash through/flow through to show the removal of any contaminants.

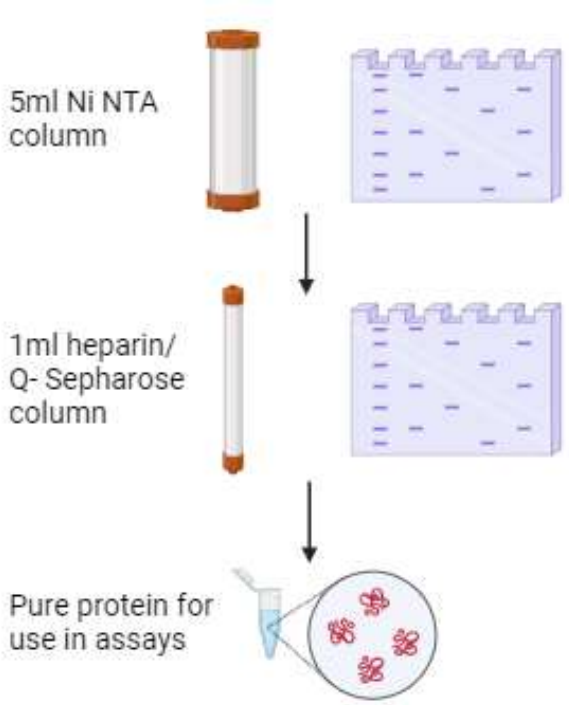


Figure 3.4 - Purification stages used for E. coli Cas1, Cas1 V76L, Cas1 R41G and DnaK. *Cleared cell lysate was loaded onto a 5ml Ni NTA column, and fractions from the elution were run on a 10% SDS PAGE gel to validate the fractions containing the relevant protein. Relevant fractions were run on a 1ml heparin column for Cas1 and its mutants, or a Q-Sepharose column for DnaK. Fractions were then run on a 10% SDS PAGE gel to validate the fractions containing the relevant protein. Final pure protein was then stored for use in assays.*

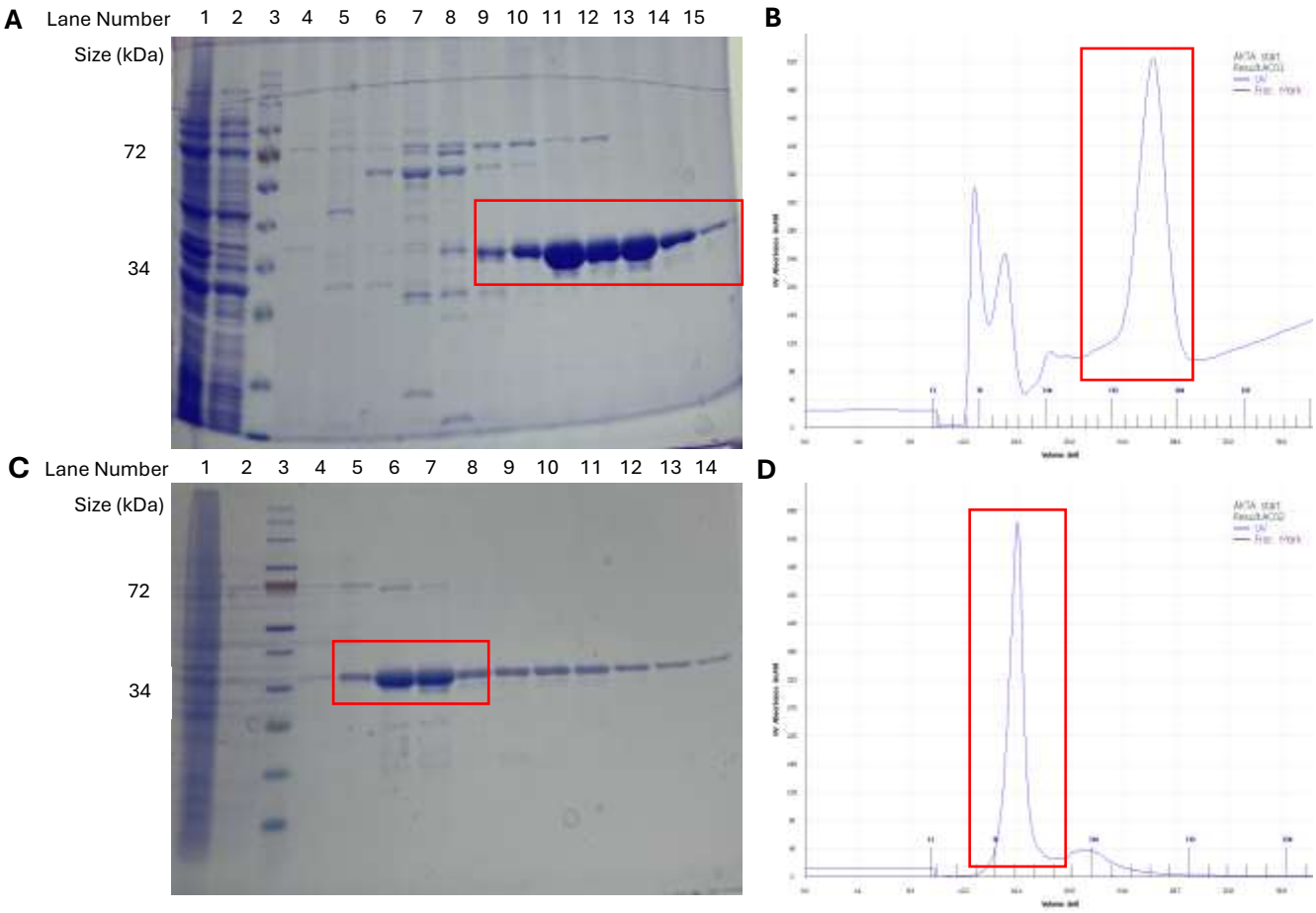


Figure 3.5 - E. coli Cas1 purification using a Ni^{2+} NTA column and a heparin column. *(A) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Ni^{2+} NTA column elution. Cas1 is indicated by the red box at ~34kDa. Lane 1 contains the cleared cell lysate, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to Cas1 V76L indicated by the red box. Cas1 containing fractions were pooled and ran on a heparin column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from heparin column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-14 contain purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being stored at -80°C.*

Once the purification method of WT Cas1 had been established, this same method was then used for the purification of the V76L and R41G mutants, as seen in Figures 3.6 and 3.7. This method involved first running the cleared cell lysate on a 5ml Ni^{2+} NTA column to separate proteins with a 6xHis tag from contaminants. Following an overnight dialysis at 4°C, the sample was run on a 1ml heparin column. Heparin functions as a DNA mimic and is therefore able to bind DNA binding proteins. Although this method left small amounts of contaminant proteins in the final sample, they were at a far lower concentration than Cas1, and the sample was considered pure enough for the assays

performed in this study. Cas1 containing fractions were therefore pooled and dialysed into storage buffer overnight at 4°C, before being stored at -80°C.

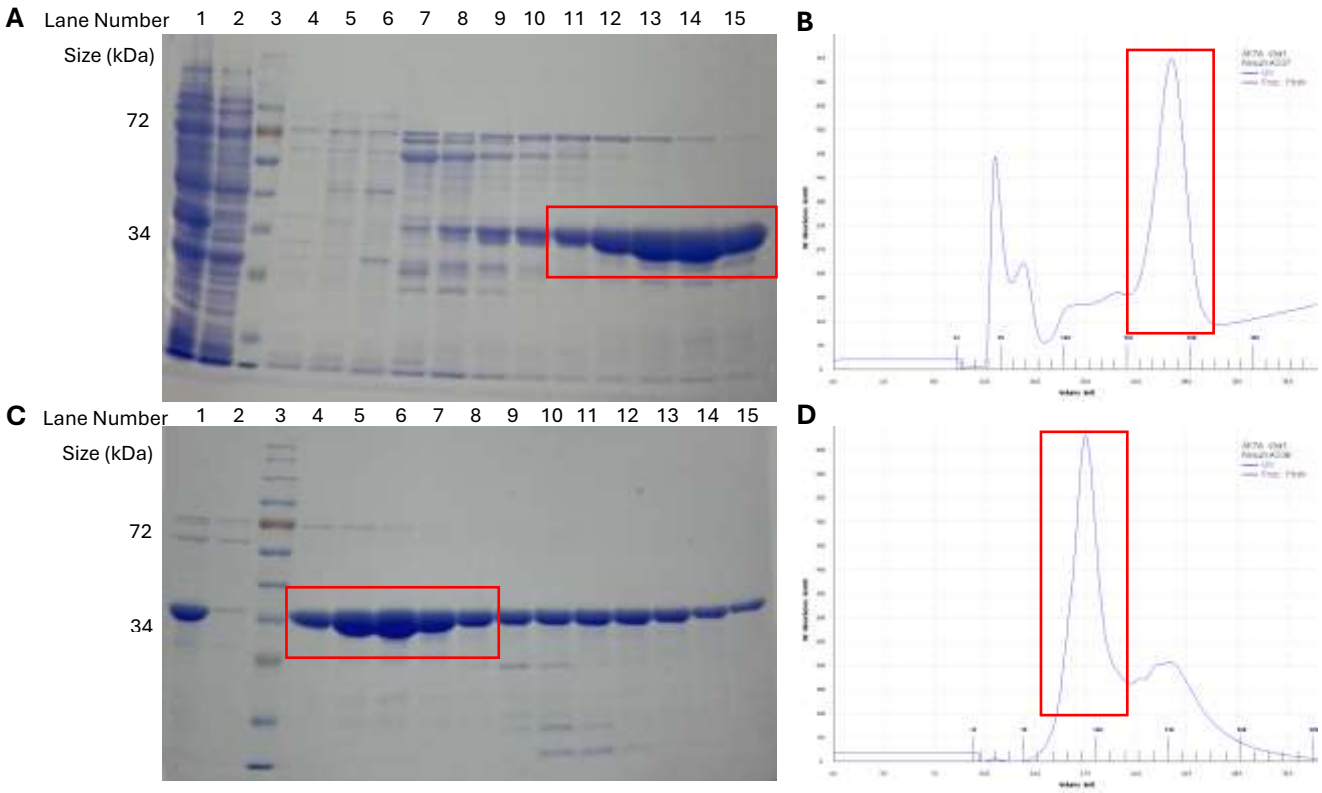


Figure 3.6 - E. coli Cas1 V76L purification using a Ni^{2+} NTA column and a heparin column. *(A) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Ni^{2+} NTA column elution. Cas1 V76L is indicated by the red box at ~34kDa. Lane 1 contains the cleared cell lysate, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to Cas1 V76L indicated by the red box. Cas1 containing fractions were pooled and ran on a heparin column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from heparin column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being stored at -80°C.*

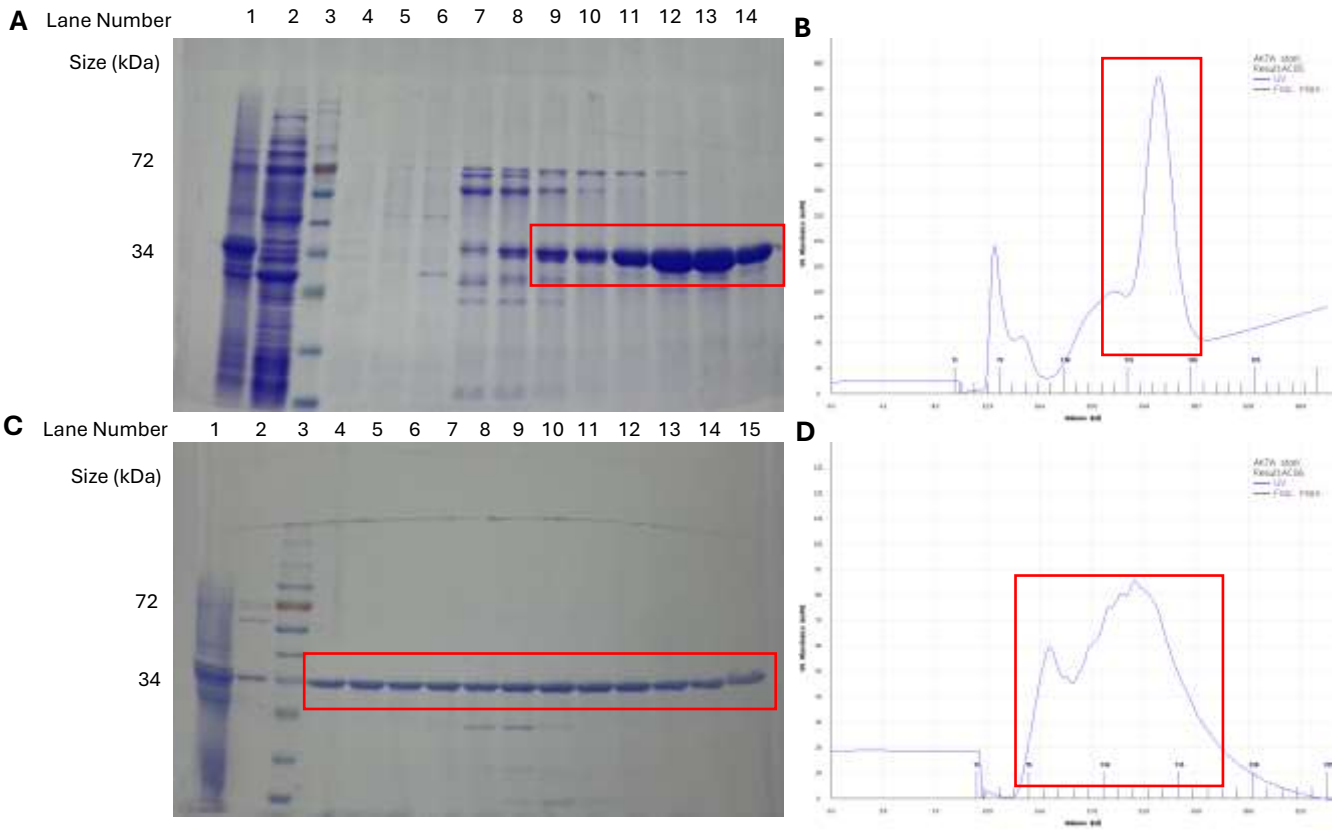


Figure 3.7 - E. coli Cas1 R41G purification using a Ni^{2+} NTA column and a heparin column. *(A) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Ni^{2+} NTA column elution. Cas1 R41G is indicated by the red box at ~34kDa. Lane 1 contains the cleared cell lysate, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-14 contain purification fractions. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to Cas1 R41G indicated by the red box. Cas1 containing fractions were pooled and ran on a heparin column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Heparin column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being stored at -80°C.*

Although Cas1 V76L behaved very similarly to WT Cas1 during purification, Cas1 R41G eluted from the Heparin across a much wider range of fractions. It was therefore decided to concentrate this protein using a spin concentrator before storage at -80°C, a step that was not required for the WT or V76L Cas1 proteins.

E. coli DnaK was also purified for this study, and it was purified according to the method set out by Killelea et al. (2023). Like Cas1 and its mutants, DnaK was purified with an N-terminal 6xHis tag, which allowed it to first be ran on a 5ml Ni^{2+} NTA column, following which protein containing fractions were analysed on a 10% SDS Page gel. The buffer used was supplemented with ATP and Mg^{2+} so that DnaK would release any proteins bound by its substrate binding domain, and therefore minimise the number of contaminant

proteins in the sample. Unlike Cas1, DnaK is not a DNA binding protein, so it was not able to be purified using a heparin column. Instead, a 1ml Q-Sepharose anion exchange column was used, and an overview of this purification can be seen in Figure 3.8. As with Cas1, the purification method used did not remove all contaminant proteins from the sample, but contaminants were deemed insignificant in proportion to DnaK, so the protein was considered pure enough to be used for the assays required in this study.

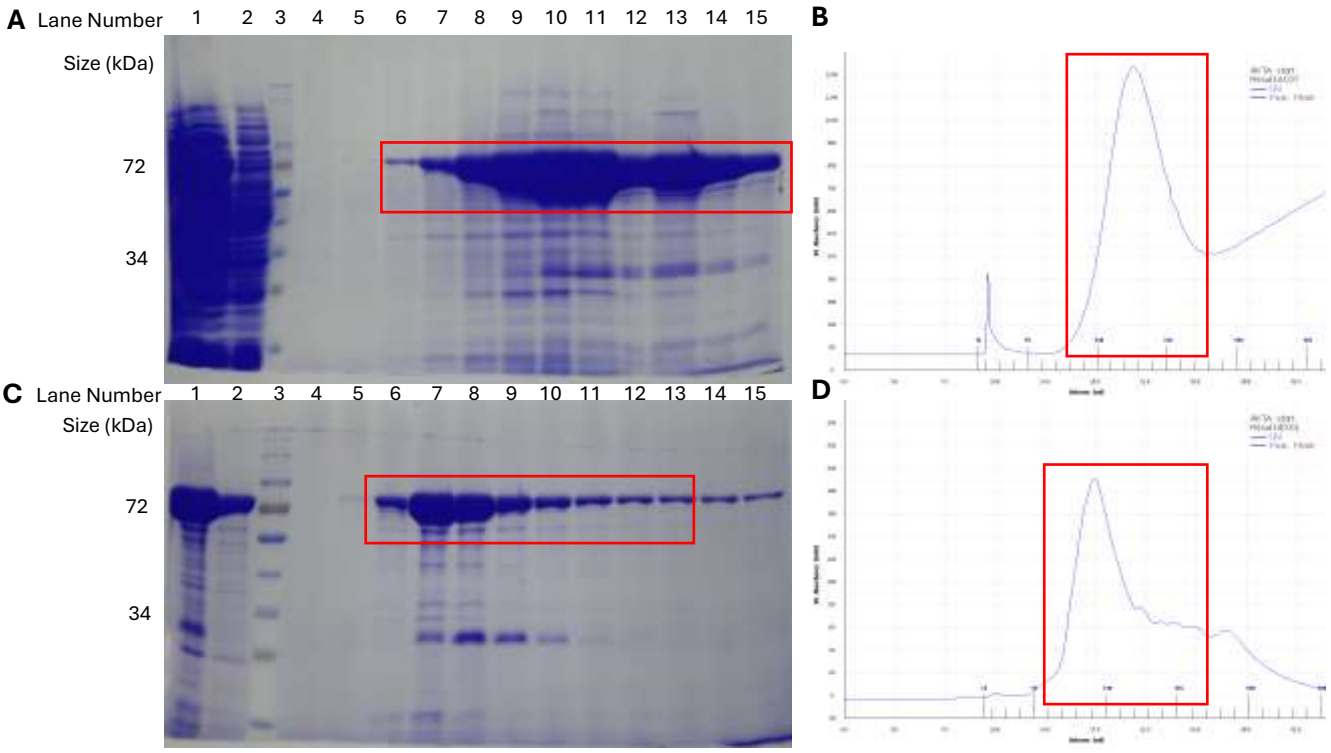


Figure 3.8 - E. coli DnaK purification using a Ni^{2+} NTA column and a Q-Sepharose column. *(A) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Ni^{2+} NTA column elution. DnaK is indicated by the red box at ~70kDa. Lane 1 contains the cleared cell lysate, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to DnaK indicated by the red box. DnaK containing fractions were pooled and ran on a Q- Sepharose column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from Q-Sepharose column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/flow through, lane 3 contains CPS and lanes 4-15 contain purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being concentrated and stored at -80°C.*

Following purification, protein concentration was determined using the Bradford assay and the BCA assay. A combination of the two assays was used, as proteins were purified at high concentrations which meant that they were outside the range that can be accurately calculated using these methods. Neat protein, ½ and ¼ dilutions were therefore used to determine protein concentration and compared across the two assays to ensure results were consistent with each other. Unknown protein concentrations were calculated in mg/ml compared to a standard curve of known concentrations of BSA, then mg/ml was converted to µM concentration based on the molecular weight of the protein purified as calculated by ProtParam.

A final 10% SDS PAGE gel containing 1,5 and 10µg of *E. coli* Cas1, Cas1 V76L, Cas1 R41G and DnaK was ran to judge the final purity of the proteins and ensure that the method of calculating protein concentration had been accurate. The SDS PAGE gel, as seen in Figure 3.9, showed consistent levels of protein across the different purifications, so the method of calculating protein concentration was accurate. Following the calculation of protein concentration, the pure proteins were used for *in vitro* assays.

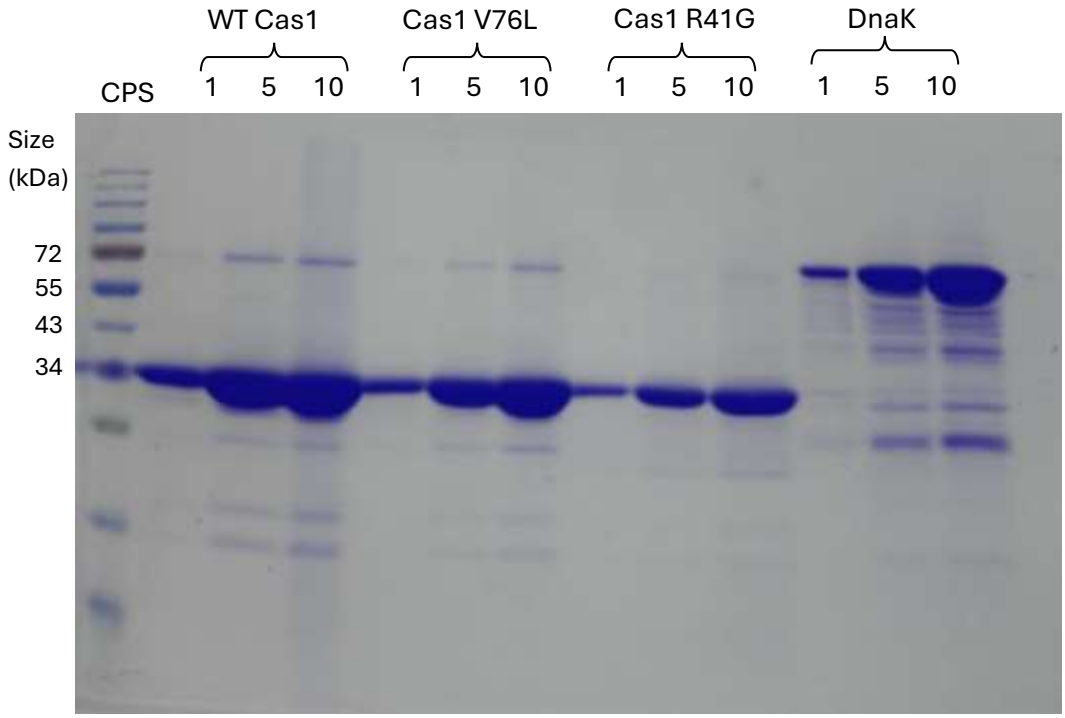


Figure 3.9 – Comparative 10% SDS PAGE gel showing 1, 5 and 10µg of E. coli Cas1, Cas1 V76L, Cas1 R41G and DnaK. *Gel was stained using Coomassie blue stain and destained before visualisation.*

3.4 *In Vitro* Biochemical Characterisation and Analysis of *E. coli* Cas1 and Mutants

3.4.1 Analysis of Cas1 and Mutant Binding of DNA

The ability of Cas1 and its mutants to bind DNA was studied in 2 ways; electrophoretic mobility shift assays (EMSAs) to detect the formation of a stable Cas1-DNA complex on gels, and fluorescence anisotropy to detect the formation of less stable complexes in solution.

EMSAs were used to detect stable Cas1-DNA complex formation in Figure 3.9, as they rely on the ability of a native TBE gel to separate complexes based on molecular weight. Complexes with a larger molecular weight are restrained by the gel and migrate a shorter distance, thus separating bound from unbound DNA. This allows for visualisation of the Cas1 concentration at which all DNA is bound. Reaction mixes were run on a 5% native TBE gel alongside a no protein (NP) control using a fluorescently labelled partial duplex as the substrate, as this mimics the structure of the DNA protospacer bound by Cas1 during naïve acquisition. EMSA analysis indicated that the V76L mutant has a similar ability to bind DNA as WT Cas1, but the R41G mutant lacks the ability to form the same stable complexes at equivalent concentrations. However, the shift seen at higher concentrations of Cas1 R41G indicates that it does interact with DNA, but it doesn't form the same stable protein-DNA complex as seen with WT Cas1 and Cas1 V76L.

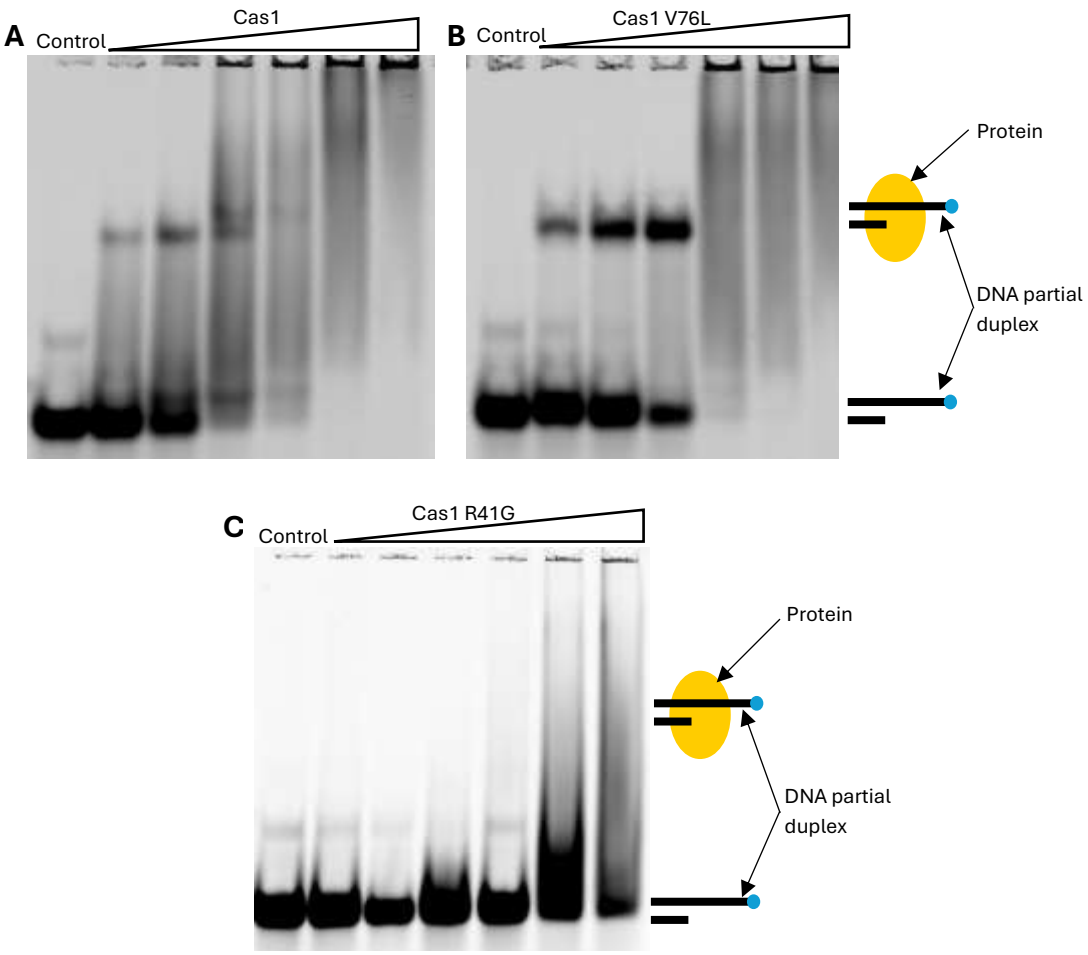


Figure 3.10 - Comparative EMSAs showing the ability of E. coli Cas1 (A), Cas1 V76L (B) and Cas1 R41G (C) mutants to bind DNA substrate. *All reactions were run on a 5% native TBE gel and contained 20nM Cy5-DNA12 partial duplex, with protein concentrations ranged from 62.5 - 2000nM. All reactions were also run alongside a NP control.*

In addition to the EMSAs seen in Figure 3.10, DNA binding was also assessed via fluorescence anisotropy. This technique requires the use of a fluorescently labelled DNA substrate which, when excited with energy, will emit light. The fluorescent label on DNA is excited in a certain axis of polarisation, and then the emission of light by the fluorescent label is detected in the same axis as that it was originally excited in. As all molecules spin in solution, fluorescent labels on unbound DNA emit very little light in the same axis of polarisation as that it was excited in. However, bound DNA spins more slowly than unbound DNA, and therefore when DNA is bound by a protein, more of the light emitted by the fluorescent label on the DNA is in the same axis of polarisation as that which it was initially excited by, and therefore a higher polarisation value is observed. Therefore, the higher polarisation value obtained, the more DNA binding is occurring. A diagram illustrating this principle can be seen in Figure 3.11.

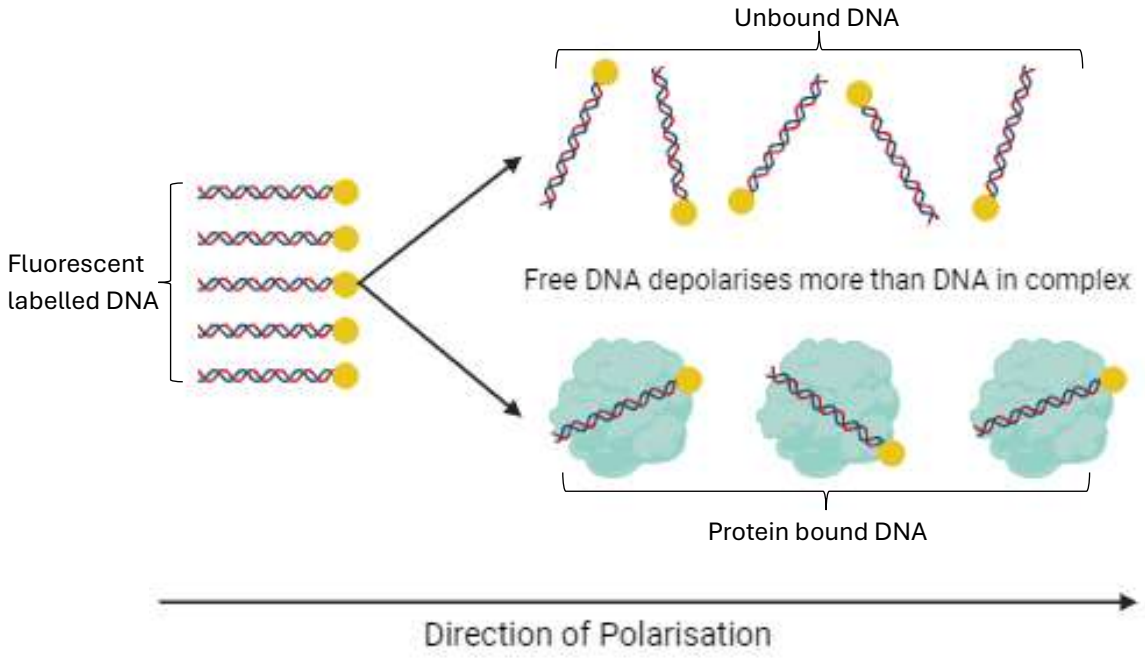


Figure 3.11 – A diagram illustrating the principles of fluorescence anisotropy. *A fluorescent label on DNA (illustrated by the yellow circle) is excited in a certain direction of polarisation, which causes the fluorescent label to emit light. The amount of light emitted in the same direction of polarisation as that it was excited in is then measured. As molecules spin in solution, DNA will rotate, and so little light will be emitted in the initial axis of polarisation. However, molecules in complex spin more slowly, and so DNA in complex with proteins will give a higher polarisation reading. Image made using Biorender.com*

Anomalous data was defined as any data point with a polarisation value of over greater than 100 units than equivalent data points. Data was analysed using GraphPad Prism software using the One Site-Total model of binding, using the equation Y=Bmax*X/(Kd+X) + NS*X + Background, where NS represents non-specific levels of DNA binding. The results of this experiment can be seen in Figure 3.12. Kd values (a measure of binding affinity) and R^2 values (a measure of goodness of fit) were calculated and averaged across the 3 time points.

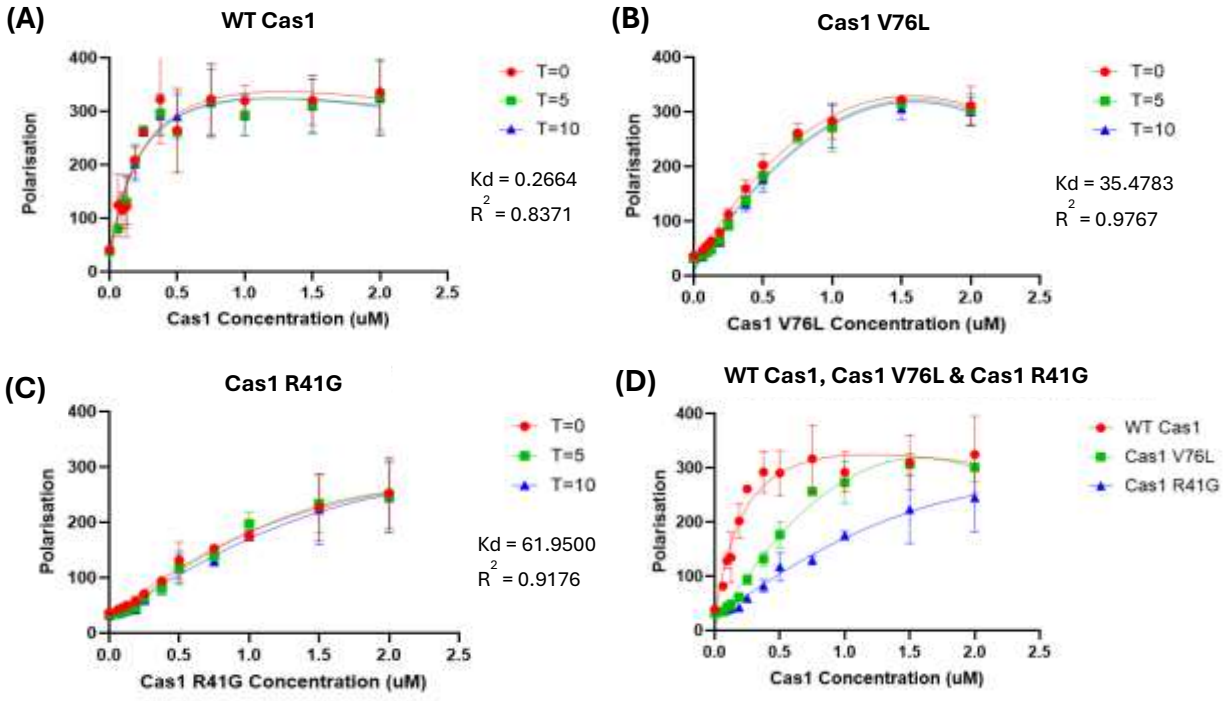


Figure 3.12 – DNA binding ability of E. coli Cas1 and mutants as calculated via Fluorescence Anisotropy. *Reaction mix contained 40nM FAM-DNA12 partial duplex, and 0.0625-2μM protein and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd values (a measure of binding affinity) and R^2 values (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point.* ***(A)*** *WT Cas1 data* ***(B)*** *Cas1 V76L data (****C****) Cas1 R41G data* ***(D)*** *T=10 data points for the 3 proteins plotted alongside each other on the same graph.*

These results also indicate that the R41G mutation greatly decreased the affinity of DNA binding by Cas1, as indicated by the Kd values increasing by 61.6836 units. The data do also indicate that the V76L mutation decreases the affinity of DNA binding, but to a much lower degree than the R41G mutation. All models display a similar goodness of fit, as indicated by the high R^2 values, and there is very little change in binding affinity across the 3 different time points, indicating that the interaction formed between Cas1 and DNA is stable.

3.4.2 Analysis of the Impact of DnaK on DNA Binding by Cas1 and Mutants

Once the binding of Cas1 to DNA had been confirmed, it was then necessary to investigate the impact that DnaK had on this binding. First a control EMSA had to be done to show that DnaK itself did not bind DNA, which can be seen in Figure 3.13. The results of this assay proved that the DnaK purified for this study did not bind DNA, as there was no change in the movement of DNA as DnaK was added in.

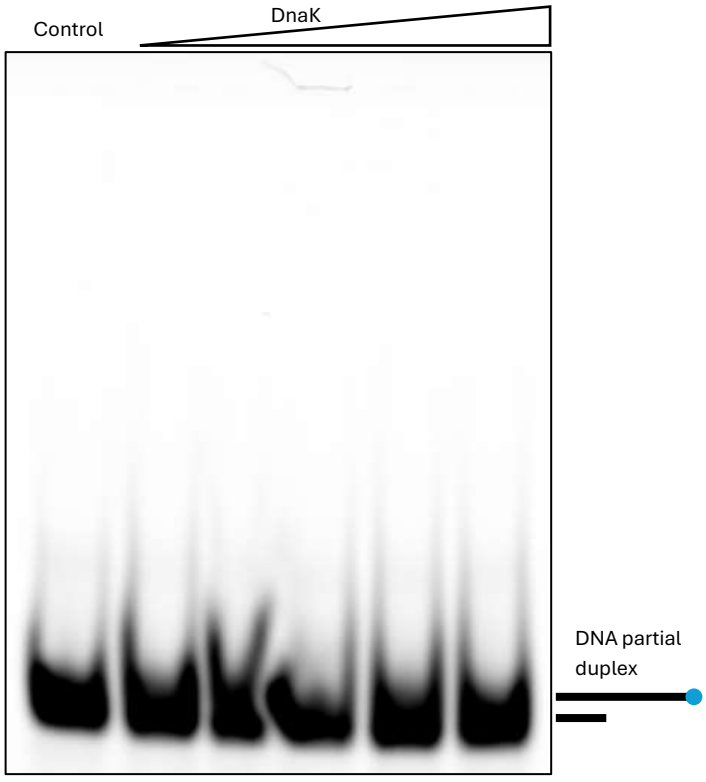


Figure 3.13 - Control EMSA showing DnaK does not bind DNA. *Reactions was run on a 5% Native TBE gel and contained 20nM Cy5-DNA12 partial duplex, with protein concentrations ranged from 125 - 2000nM. Samples were run alongside a NP control.*

After it had been confirmed that DnaK did not bind DNA, it was able to be added into the reaction mix alongside Cas1. Firstly, this was attempted by pre-incubating Cas1 and DnaK together for 5 minutes at 37 °C before adding them to the reaction mix containing the DNA. The results of this experiment can be seen in Figure 3.14.

Unfortunately, the pre-incubation of Cas1 before adding it to DNA seemed to cause protein aggregation, as a single stable complex was not visible on the final gel. However, as DnaK was titrated into the reaction mix, less aggregation was seen with WT Cas1 and as1 V76L, indicating that there was an interaction between Cas1 and DnaK in these samples and that the presence of DnaK was preventing aggregation. Unsurprisingly, no clear difference in binding was seen when Cas1 R41G was present, so it was not possible to observe a difference when DnaK was added to the reaction mix. There was however no increase in the levels of unbound DNA as DnaK was added in, so it could not be concluded from this assay alone that DnaK prevented Cas1 from binding to DNA.

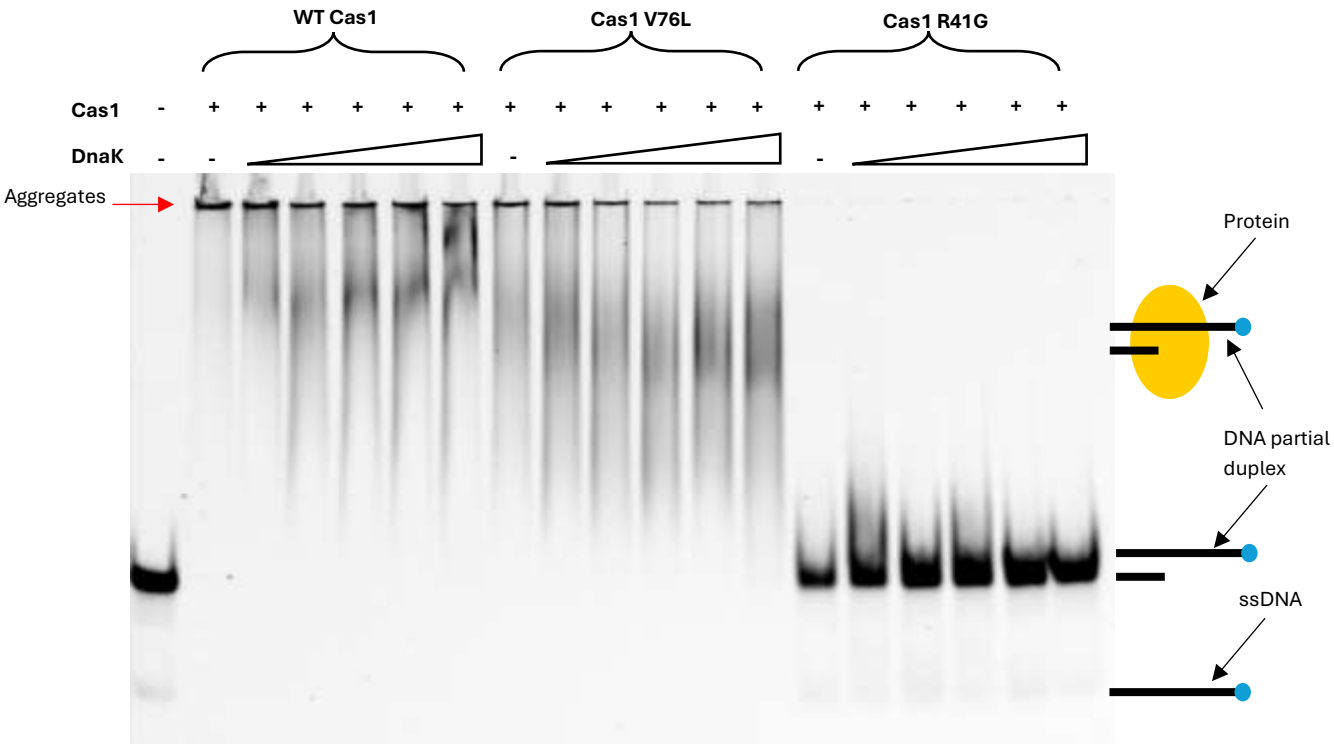


Figure 3.14 - Comparative EMSAs showing the impact of DnaK on Cas1, Cas1 V76L and Cas1 R41G binding DNA. *All reactions were run on a 5% Native TBE gel and contained 20nM Cy5-DNA12 partial duplex. Constant concentration of 250nM Cas1 was used, with DnaK concentration ranging from 125-2000nM. Proteins were pre-incubated together at 37°C for 5 minutes before being added to the reaction mix. All reactions were also run alongside a NP control.*

Due to the issues with aggregation seen when the two proteins were incubated together, the assay was attempted again, with Cas1 first being added to the reaction mix, followed by DnaK, and this experiment can be seen in Figure 3.15.

The results of this EMSA were clearer as it led to far less protein aggregation. For WT Cas1, it was not possible to see an increase in free DNA as DnaK concentration increased in the same way as observed by Killelea et al. (2023). This could be for a variety of reasons, such as the equipment not being optimised, or experimental conditions not being ideal for the protein. Interestingly, for Cas1 V76L mutant low levels of DnaK appeared to aid Cas1 binding, as the amount of bound DNA increased. However, as the levels of DnaK increased, so did the amount of unbound DNA, which suggests that the presence of DnaK is inhibitory to DNA binding as with WT Cas1. For Cas1 R41G, there was again no stable Cas1-DNA complex formed, so the presence of DnaK had no impact on this reaction.

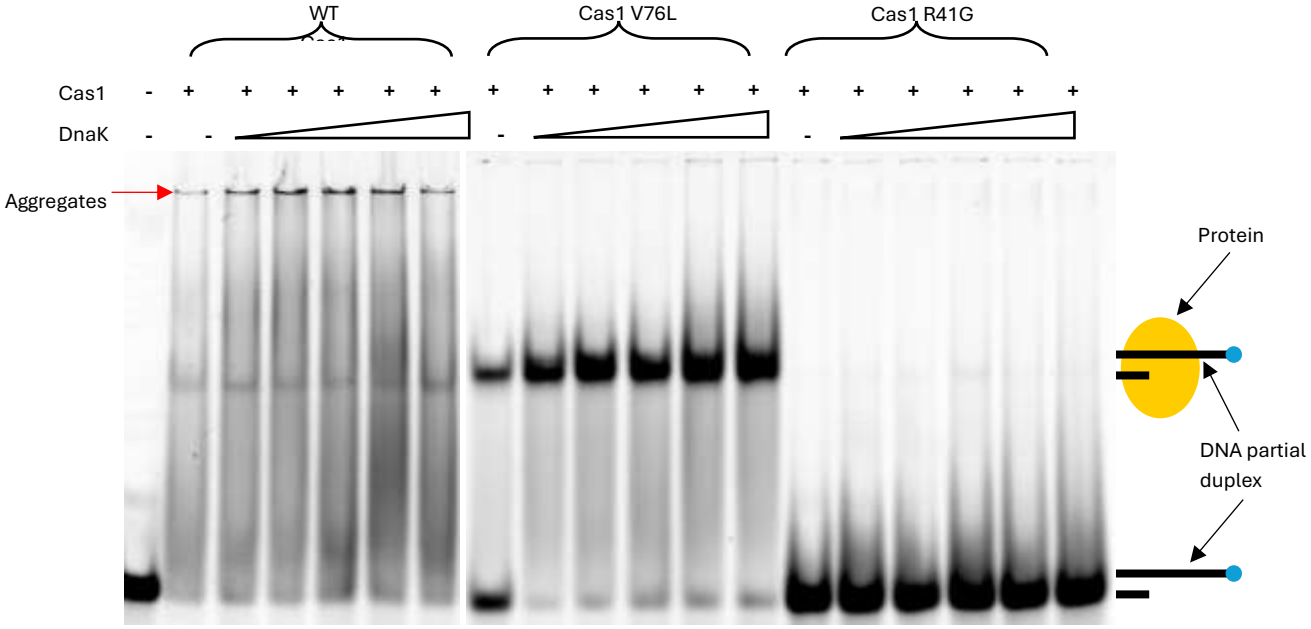


Figure 3.15 - Comparative EMSAs showing the impact of DnaK on Cas1, Cas1 V76L and Cas1 R41G binding DNA. *All reactions were run on a 5% Native TBE gel and contained 20nM Cy5-DNA12 partial duplex. Constant concentration of 250nM Cas1 was used, with DnaK concentration ranging from 125-2000nM. Cas1 was first added to the reaction mix, followed by DnaK at appropriate concentrations. All reactions were also run alongside a NP control.*

The effect of the interaction between Cas1 and DnaK was also studied with fluorescence anisotropy. As with the EMSAs, before the interaction between Cas1 and DnaK could be studied, first a control assay using only DnaK had to be performed to ensure that DnaK itself did not impact DNA binding, and this is seen in Figure 3.16.

This experiment also showed that DnaK in isolation does not bind DNA, as there was almost no change in polarisation value when DnaK was added in. Once this was confirmed, it was possible to titrate DnaK into the reaction alongside Cas1 to determine the impact had on DNA binding.

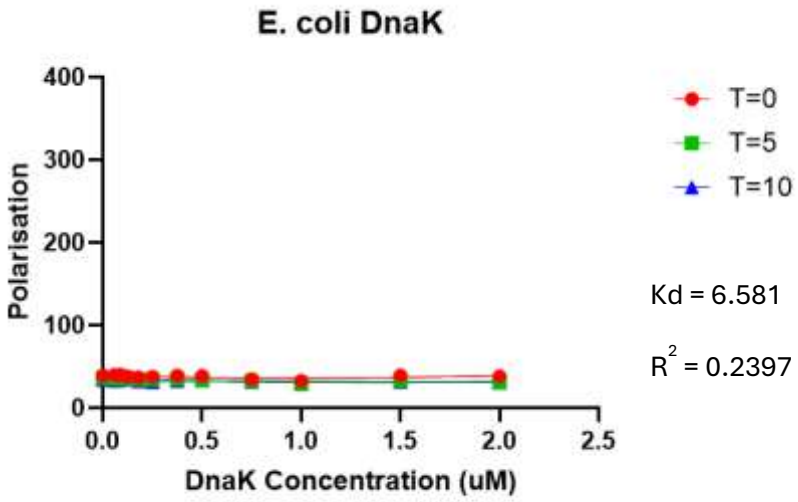


Figure 3.16 - DNA binding ability of E. coli DnaK as calculated via fluorescence anisotropy. *Reaction mix contained 40nM FAM-DNA12 partial duplex, and 0.0625-2μM protein and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd value (a measure of binding affinity) and R^2 value (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point. As there was very little deviation in the data for this experiment, the error bars are shorter than the size of the symbol used, and therefore hidden.*

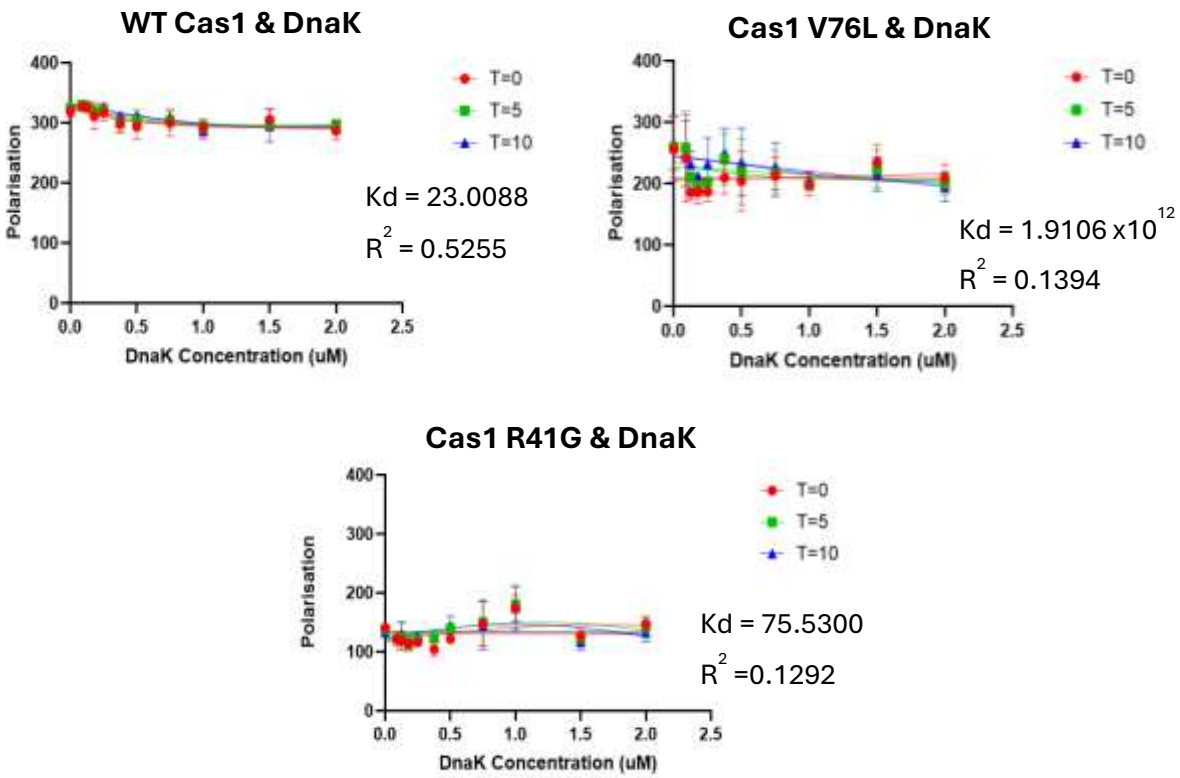


Figure 3.17 - DNA binding ability of E. coli Cas1 and mutants in the presence of DnaK as calculated via fluorescence anisotropy*. Reaction mix contained 40nM FAM-DNA12 partial duplex, 0.5μM Cas1, Cas1 V76L or Cas1 R41G and 0.09375-2μM DnaK and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd values (a measure of binding affinity) and R^2 values (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point.*

Figure 3.17 shows the impact that DnaK had on the binding of Cas1, Cas1 R41G and Cas1 V76L to DNA. The polarisation values calculated from this experiment do not change much upon the addition of DnaK, indicating that the conditions used for this experiment may not be optimal. However, WT Cas1 and Cas1 V76L both show a general negative trend in polarisation as DnaK concentration increases, whereas the polarisation values seen with Cas1 R41G remain fairly constant. The Kd values calculated for each experiment was higher than that calculated for Cas1, Cas1 V76L and Cas1 R41G respectively, which indicated that the presence of DnaK was interfering with DNA binding. For WT Cas1 and Cas1 V76L, the levels of polarisation decreased by similar amounts, indicating that DnaK can interact similarly well with these two proteins. Cas1 R41G however does not seem to show the same trend, with the polarisation levels remaining constant or increasing as DnaK concentration increases. This potentially supports the

hypothesis that R41 interacts with DnaK as predicted by bioinformatics models. The R^2 values for each of the data points are lower than for Cas1, Cas1 V76L and Cas1 R41G on their own, which indicates that the data is less reliable as the data fits the trend less well.

3.4.3 Investigating the Catalytic Activity of Cas1 and Mutants

Catalytic activity was investigated using disintegration assays. These require the presence of a partial fork DNA substrate which can be cleaved by Cas1. The DNA substrate was incubated in cleavage buffer with the relevant protein, after which protein was degraded via addition of stop buffer. Reaction mix was then run on a 12% native TBE gel to allow the separation of cleaved from un-cleaved DNA and the gel was imaged using an Amersham Typhoon.

Both the WT Cas1 and the V76L mutant showed catalytic activity in Figure 3.18, with the V76L mutant being slightly less able to cleave the DNA substrate. The R41G mutant however showed no catalytic activity, which is unsurprising as the EMSAs performed in this study indicate that this mutant is unable to form a stable complex in gels with DNA.

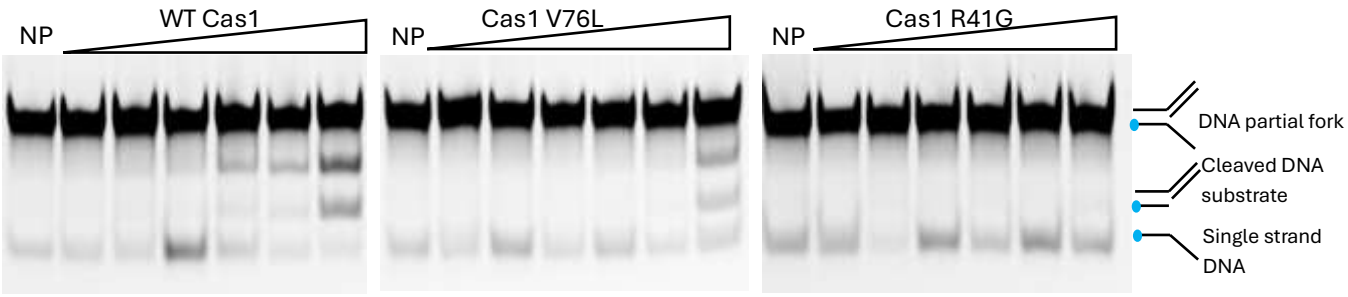


Figure 3.18 – Disintegration assay showing catalytic activity of E. coli Cas1 and Mutants. *20nM Cy5-labelled partial fork was incubated with protein concentrations from 12.5-100nM for 30 minutes at 37°C to allow for cleavage of DNA substrate, following which protein was degraded using Stop buffer. Samples were electrophoresed for 90 minutes on a 12% native TBE gel.*

Once catalytic activity had been confirmed, DnaK was titrated into the reaction mix to determine if the presence of DnaK was able to inhibit the ability of Cas1 to cleave DNA. Before the interaction was investigated, a control assay was performed to confirm that DnaK on its own is not able to cleave DNA, and that there were no contaminating nucleases. No degradation of DNA was seen in Figure 3.19, indicating that the DnaK purified was suitable for this assay.

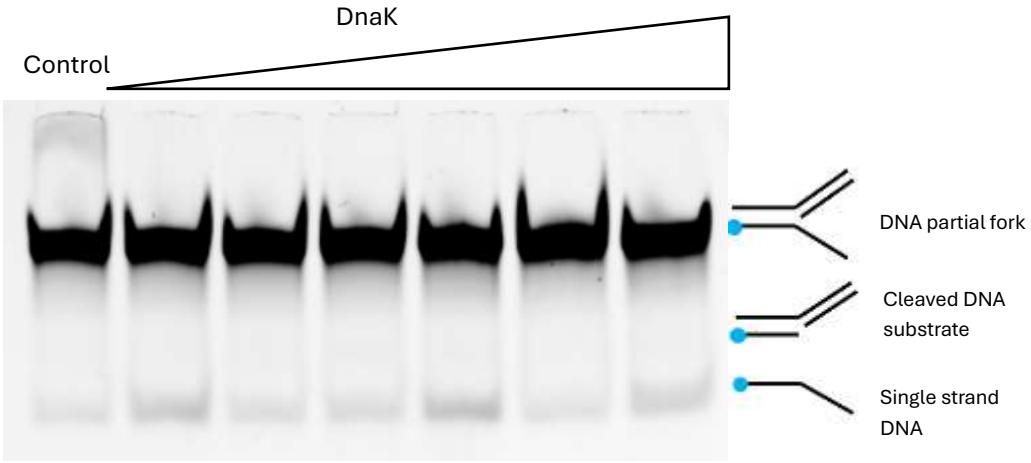


Figure 3.19 - DnaK Control Disintegration Assay. *20nM Cy5-labelled partial fork was incubated with DnaK from 12.5-100nM for 30 minutes at 37°C to allow for cleavage of DNA substrate, alongside a NP control, following which protein was degraded using Stop buffer. Samples were electrophoresed for 90 minutes on a 12% Native TBE Gel.*

Once it was confirmed that DnaK on its own would not cleave DNA, DnaK was titrated into disintegration assays with a constant concentration of Cas1 to determine if there was any inhibition of Cas1, the results of which are seen in Figure 3.20. No inhibition was detected with WT Cas1 or Cas 1V76L, and Cas1 R41G showed catalytic activity which had not been visible when only Cas1 R41G was present. This indicates that there was a potential mistake when assembling the reaction mix with Cas1 R41G. However, although DnaK can inhibit DNA binding by Cas1, it does not appear to inhibit catalytic activity directly *in vitro*. In fact, it appears to stimulate the catalytic activity of Cas1 and Cas1 V76L, as the amount of cleaved DNA increases as DnaK concentration increases. It would have been beneficial in this assay to have included a control with just 200nM DnaK to determine if the reaction conditions used in this exact assay had shown any cleavage from just the DnaK protein on its own.

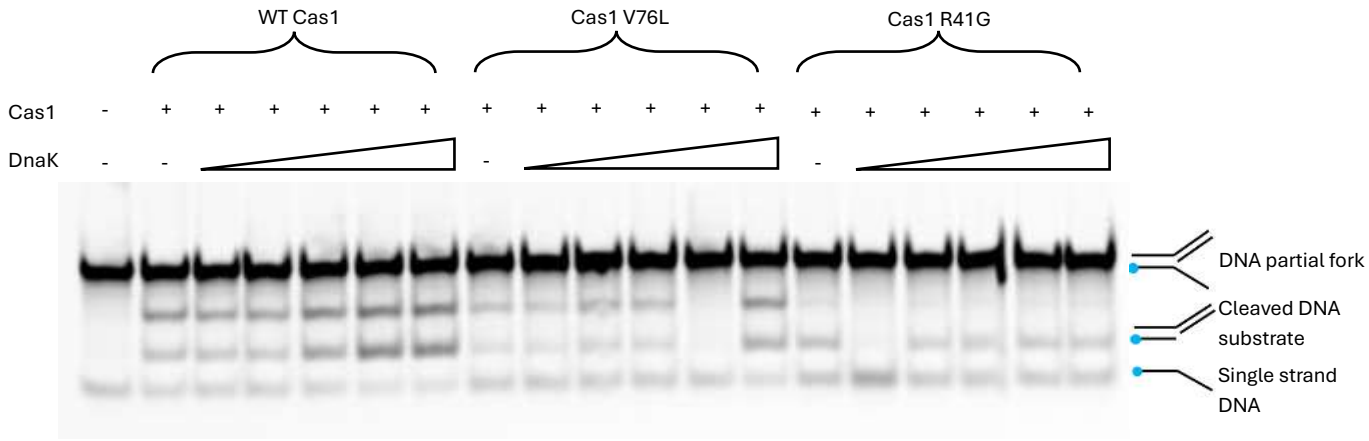


Figure 3.20 - Disintegration assay with Cas1 and DnaK to determine if DnaK inhibits the catalytic activity of Cas1. *20nM Cy5-labelled partial fork was incubated with 100nM Cas1 or mutant and DnaK from 12.5-200nM for 30 minutes at 37°C to allow for cleavage of DNA substrate, following which protein was degraded using stop buffer. Samples were electrophoresed for 90 minutes on a 12% native TBE gel.*

3.5 In Vivo Investigation of *E. coli* Cas1 and Mutants

3.5.1 Investigating the Physical Interaction between DnaK and Cas1 and Mutants

To determine whether there was a physical interaction between Cas1 and DnaK *in vivo*, pull down assays were used. They rely on the ability of Ni^{2+} charged iminodiacetic acid Sepharose resin to bind protein with a 6xHis tag, which will pull any proteins the 6x His tagged proteins are associated with out of the cleared lysate of cells. Samples then denatured in 1X SDS loading dye and ran on a 12% SDS PAGE gel to allow visualisation of proteins. This assay used 6xHis tagged DnaK to allow it to bind to the charged iminodiacetic acid Sepharose, however, Strep-tagged Cas1, Cas1 V76L and Cas1 R41G were used so these proteins did not have an affinity for the iminodiacetic acid Sepharose, and therefore any Cas1 present was due to it having bound to DnaK. The results of this assay can be seen in Figure 3.21, and it showed a strong interaction between Cas1 and its mutants with DnaK, as Strep-Cas1 can clearly be seen when it is overexpressed with His-DnaK. It does indicate that the association with Cas1 R41G is potentially weaker than with WT Cas1 or Cas1 V76L, as the band that corresponds to this protein is less intense.

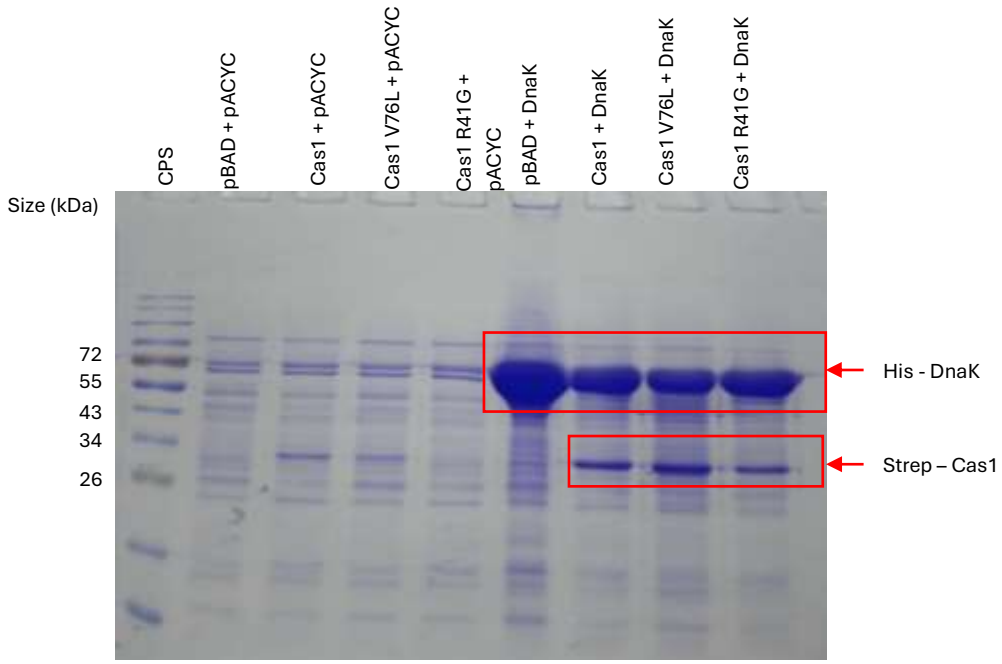


Figure 3.21 - 12% SDS PAGE gel showing denatured pull down assay samples. *Cleared cell lysate containing indicated proteins was mixed with Ni^{2+} charged iminodiacetic acid Sepharose, which has an affinity for 6xHis-tagged proteins such as His-DnaK. This allows for proteins that interact with DnaK to be identified as they do not have an affinity for the iminodiacetic acid Sepharose. Strep-tagged Cas1, Cas1 V76L and Cas1 R41G can all be seen as proteins that interact with DnaK, indicated by the red box.*

3.5.2 Investigating the Activity of E. coli Cas1 and Mutants in Naïve Acquisition

Naïve acquisition assays were performed to determine whether the Cas1 mutants used in this assay were functional *in vivo*. This required the overexpression of WT Cas1, Cas1 V76L and Cas1 R41G along with Cas2 in EB377 cells to allow for naïve acquisition to occur. EB377 cells are modified MG1655 strain *E. coli* with the addition of an arabinose inducible promoter into the genome to allow for controlled expression of proteins. Cells are subjected to multiple passages, whereby a small volume of the previous culture is used to inoculate fresh media and cells are allowed to grow again to allow for spacer acquisition to occur. The levels of acquisition can then be measured using PCR amplification of the CRISPR locus. Cells that have undergone spacer acquisition will have a larger CRISPR locus due to the additional spacer and repeat, therefore the amplified PCR product will travel less far through an agarose gel. The acquisition assays performed in this study can be seen in Figure 3.22.

The results of these naïve acquisition assays showed that WT Cas1 and Cas1 V76L are active in acquisition, but Cas1 R41G is not. This was not surprising given that Cas1 R41G had been shown to not be catalytically active in disintegration assays and binds DNA with lower affinity than WT Cas1 and Cas1 V76L. However, the hyperactive acquisition that has been previously described with Cas1 V76L was not observed in this assay.

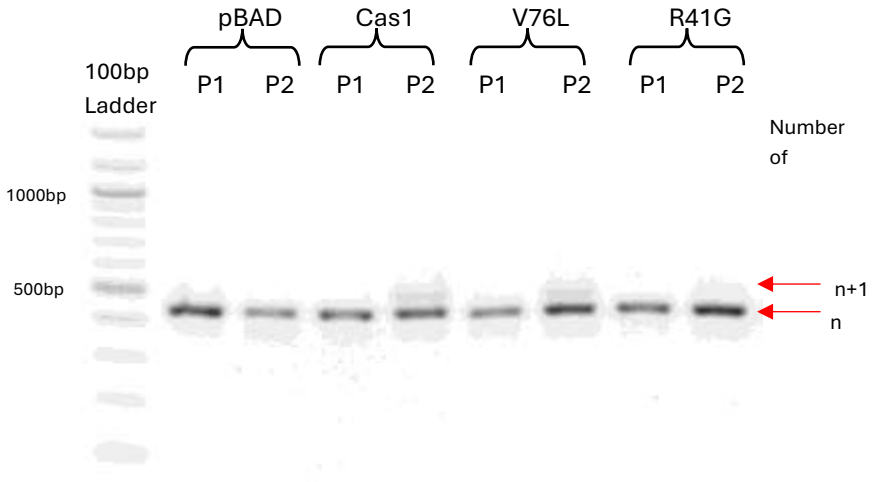


Figure 3.22 – DNA resulting from the PCR amplification of the CRISPR locus of E. coli following naïve acquisition assay, run on a 1% agarose gel. *WT Cas1-Cas2, Cas1 V76L-Cas2 and Cas1 R41G-Cas2 were overexpressed in EB377 cells to allow for naïve acquisition to occur, the levels of which were then measured using PCR amplification. P1 and P2 indicate the number of passages the cells had been subjected to. The appearance of a second, larger band (referred to as n+1) indicates expansion of the CRISPR locus and integration of a new spacer. This is seen with Cas1-Cas2 and Cas1 V76L-Cas2, but not Cas1 R41G-Cas2.*

3.6 Summary

Bioinformatics models predicts that DnaK interacts with Cas1 via a region containing R41, so this amino acid was mutated to attempt to knock out the association between the two proteins. V76 was also mutated as this mutation has previously been shown to cause hyperactive acquisition, so it was hypothesised that mutating this residue interferes with the Cas1-DnaK interaction.

The V76L mutation leads to similar behaviours both *in vivo* and *in vitro* to wild type Cas1. It shows similar levels of binding to DNA in both EMSAs and fluorescence anisotropy and is also catalytically active *in vitro* when used in disintegration assays. The effect of DnaK on DNA binding is similar in WT Cas1 and Cas1 V76L, with the presence of DnaK inhibiting DNA binding to similar degrees. It also binds to DnaK with similar affinity to WT Cas1 *in vitro* as similar levels of the protein are seen when pull down assays are performed. It has however been proven to be hyperactive in naïve acquisition assays by other lab groups, indicating that this residue is important *in vivo*, but that this residue does not interact with DnaK.

Mutating R41 has impacts on the ability Cas1 to bind DNA, as fluorescence anisotropy shows that it binds DNA, albeit with lower affinity than WT Cas1 or Cas1 V76L. EMSA analysis of Cas1 R41G also shows that it is unable to form a stable complex with DNA *in vitro*, and disintegration assays show that this mutation removes the catalytic activity of the protein. It is also unable to function catalytically *in vivo*, as it was not able to perform naïve acquisition. This is unsurprising, as Nuñez, Harrington, et al. (2015a) showed that this residue forms part of the arginine clamp, a series of key residues involved in DNA binding in the Cas1-Cas2 complex. The data gathered do however indicate that the residue may interact with DnaK as indicated by bioinformatic structural predictions, as the presence of DnaK does not seem to interfere with DNA binding by Cas1 R41G in fluorescence anisotropy. Pull down assays also indicate that Cas1 R41G may interact more weakly with DnaK than WT Cas1 and Cas1 V76L, which also supports the idea that this residue interacts with DnaK.

The R41 Mutation would be better studied using more methods to determine a physical interaction between proteins, such as bio immunofluorescence complementation (BiFC)

(discussed further in Chapter 5). It may also be that a single point mutation is not enough to entirely remove the interaction between the two proteins. It would therefore be ideal to study a protein which had 3-4 residues in the area containing R41 mutated to glycine residues and perform pull down assays and BiFC to determine if the interaction can be removed entirely.

A summary of the findings gathered in this chapter can be seen in Table 3.1, compiling DNA binding capabilities in EMSAs and fluorescence anisotropy, catalytic activity observed in disintegration assays and interaction with DnaK in pull down assays.

Table 3.1 - Summary table compiling the results gathered in Chapter 3.

Protein	**DNA binding in EMSAs**	**Kd calculated via fluorescence anisotropy**	**Kd calculated via fluorescence anisotropy with DnaK**	**Catalytic activity in disintegration assays**	**Physical interaction with DnaK in pull-down assays**
WT Cas1	Yes	0.2664	23.0088	Yes	Yes
Cas1 V76L	Yes	35.4783	1.9106 x10 12	Yes	Yes
Cas1 R41G	No	61.9500	75.5300	No	Yes, but weaker than that seen with WT Cas1 and Cas1 V76L

Chapter 4 - Investigating whether the Interaction between Cas1 and DnaK is Conserved across CRISPR Systems

4.1 Introduction

Although the interaction between Cas1 and DnaK is strongly supported in *E. coli* by recent data from Killelea et al. (2023), it has not been studied in other organisms. It is therefore unknown whether this method of controlling adaptation may be universal amongst CRISPR systems. In the study, Cas1 and DnaK from *S. pyogenes* and *M. thermautotrophicus* were cloned and purified to determine whether DnaK inhibits Cas1 binding to DNA.

These species were selected due to the differences of the structure of their CRISPR systems compared to *E. coli*. *S. pyogenes* for example is a bacterium which uses a Type II-A CRISPR-Cas system (Le Rhun et al., 2019), and *M. thermautotrophicus* is an archaeon which uses a Type I-B CRISPR-Cas system (Richter et al., 2017). These are both in contrast to *E. coli,* a bacterium which possesses a Type IE CRISPR-Cas system (Cass et al., 2015). Therefore, if an interaction could be proven as conserved amongst these 3 classifications of CRISPR systems, it could be hypothesised that it is likely universal amongst CRISPR systems.

4.2 Bioinformatic Investigation of Cas1 from *S. pyogenes* and *M. thermautotrophicus*

4.2.1 Structural Modelling of Mth and Spy Cas1 with DnaK

Structural modelling was performed to predict whether Cas1 and DnaK from *S. pyogenes* and *M. thermautotrophicus* would interact with each other. Protein sequences were extracted from Uniprot, and structural modelling was performed using Alphafold 3 via Alphafold Server and the resulting images were opened and annotated using PyMol.

Both *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 homodimers were predicted to interact with DnaK, as seen in Figures 4.1 and 4.2. It was therefore decided to proceed with the cloning and purification of these proteins so that the interaction could be tested for.

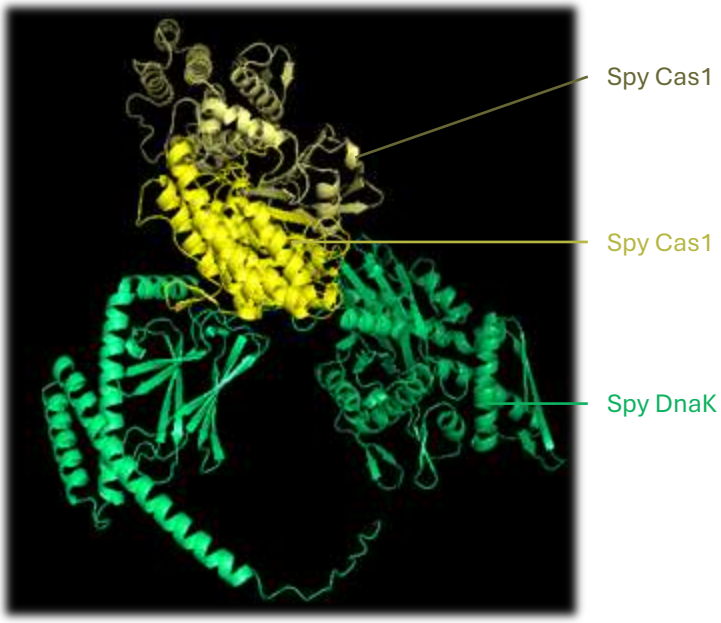


*Figure 4.1 - **Structural prediction of S. pyogenes DnaK (referred to as Spy DnaK) interacting with S. pyogenes Cas1 (referred to as Spy Cas1) homodimer**. Protein sequences taken from UniProt, Cas1 reference Q99ZW1, DnaK reference P0C0C6. Structure created using Alphafold 3, annotated In PyMol.*

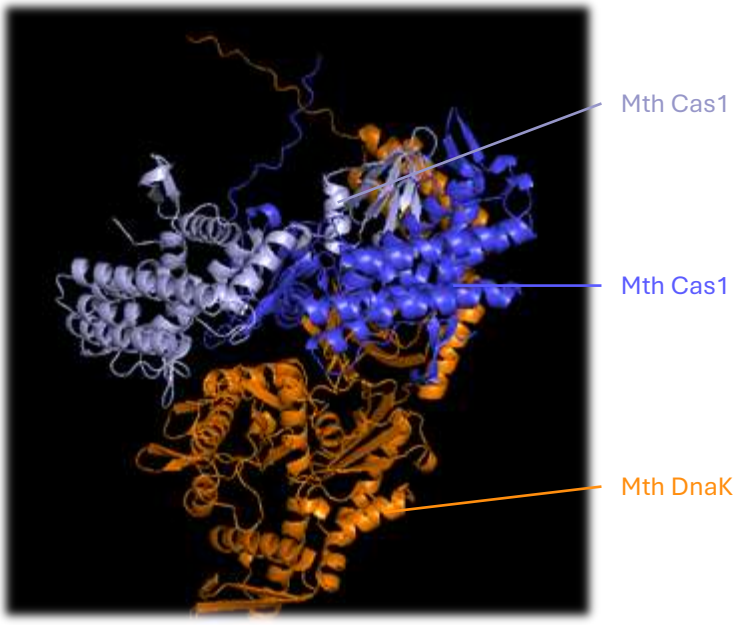


***Figure 4.2 - Structural prediction of M. thermautotrophicus DnaK (referred to as Mth DnaK) interacting with M. thermautotrophicus Cas1 (referred to as Mth Cas1) homodimer**. Protein sequences taken from UniProt, Cas1 reference O27156, DnaK refence O27351. Structure created using Alphafold 3, annotated In PyMol.*

4.2.2 Sequence Alignment

A sequence alignment of Cas1 from *E. coli, S. pyogenes* and *M. thermautotrophicus* was performed to identify conserved residues. Protein sequences were extracted from UniProt, and sequence alignment was performed using Clustal Omega. Sequence alignment was performed to see if the residues investigated in *E. coli* in Chapter 3 were conserved, which would hint at a universal mechanism for control of acquisition.

Several residues were identified as conserved amongst these 3 homologues of Cas1, as seen in Figure 4.3, however neither of the residues identified as potentially of interest in Chapter 3 were conserved. This may mean that the residues identified in Chapter 3 are not involved in the interaction between Cas1 and DnaK *in vivo*, or that other residues which are conserved are potentially also key. Alternatively, although the amino acid sequence does not show a conserved residue at this position, it may in fact be that the 3D structure at this position is conserved and binds to the canonical DnaK binding motif.

```
CLUSTAL O(1.2.4) multiple sequence alignment
E.      ----------------------------------MTWLPLNPIPLKDRVSMIFLQYGQI     25
S.      --MAGWR---------TVVVNTHSKLSYKNNHLIFKDAYKTELIHLSE-IDILLLETTDI     48
M.      MNSAGLEQCGMIMTRKNYYITTDGLLKRNENTLYFINKDLKRPIPINK-IYSI-YSYGAL     58
                                                . * :.. :  :  .   :
E.      DVIDGAFVLIDKTGIRTHIPVGSVACIMLEPGTRVSHAAVRLAAQVGTLLVWVGEAGVR-     84
S.      VLSTMLVKRLVDENVL---------------------------------VIFCDDKRLPT     75
M.      TISSQALNLLSKEGIP---------------------------------IHFFNRYGFYS     85
         :    .  : . .:                                   : : .   .
E.      -----VYASGQPGGARSDKLLYQAKLALDEDLRLKVVRKMFELR-------------FGE     126
S.      AMLMPFYGRHDSSLQLGKQM-SWSETVKS-----QVWTTIIAQKILNQSCYLGACSYFEK     129
M.      G---SFYPRET--LLSGDVIIKQAEHVIDHDKRIELARSFVRGAALNM---RRVLSYYGI     137
             .*         .. :   :: . .     ::  .:.              :
E.      PAP---------ARRSVEQLRGIEGSRVRATYALLAKQYGVTW-NGRRYDPKDWEKGDTI     176
S.      SQ----SIMDLYHGLENFDPSNREGHAARIYFNTL---------FGNDF---SRDLEHPI     173
M.      ENGISDTLMDLDSSNSVTDVMNVEGRIRSDYYNAIDSILPEGFRIGKRT---RRPPENMT     194
                      .  :  . **      :  :          *.         .
E.      NQCISAATSCLYGVTEAAILAAGYAPAIGFVHTGKPLSFVY--DIADIIKFDTVVPKAFE     234
S.      NAGLDYGYTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDIMEPF--RPLVD-KIV     230
M.      NAMISFGNSLLYSTVITELYNTQLNPTISYLHEPFERRYSLALDLSEIF--KPTLIDRMI     252
        *  :. . : * .     :  :     :.  *      :    *: : :     :   :
E.      IARRNPGEPDREVRLA----CRDIFRSSKTLAKLIPLIEDVLAAGEIQPPAPPEDAQPVA     290
S.      YENRNQPFPKIKRELFTL--FSDTFSYNGKEMYLTNIISDYT--------------KKV-     273
M.      ISLIKK--KAIKAEDFEHGMNHCLLNNSGKRKF----LAEYD--------------RRL-     291
            :      : .          :  . .       : :               : :
E.      IPLPVSLGDAGHRSS---------------------------- 305
S.      -VKALNNE--GKGVPEFRI------------------------- 289
M.      -GKTVKHRELGRKVSYRRLIRLEAYKLIKHLIGQKSYEPFVMWW 334
            :.    *:
```

Figure 4.3 – Multiple Sequence Alignment of Cas1 from E. coli, S. pyogenes and M. thermautotrophicus created using Clustal Omega. *Residues mutated in Chapter 3 are indicated in red and bold. Fully conserved residues are indicated with *, partially conserved are indicated with: and similar residues are indicated with .. Residues highlighted in red were studied in Chapter 3 as potential DnaK interaction sites.*

4.3 Cloning, Protein Expression and Purification of Proteins

4.3.1 Cloning of Protein Expression Plasmids

Nucleotide sequences were taken from Kegg Orthology. Codon optimised expression plasmids for *S. pyogenes* Cas1, *S. pyogenes* DnaK and *M. thermautotrophicus* DnaK were ordered from Addgene cloned into a pET151 Vector. *M. thermautotrophicus* Cas1 was cloned into pACYC Duet using BamHI and SalI restriction sites following PCR amplification from genomic DNA. Successful cloning was confirmed using a double digest with BamHI and SalI. This plasmid was found not to overexpress proteins *in vitro*, so *M. thermautotropicus* Cas1 was ordered codon optimised cloned into a pET-100 D-TOPO vector from Addgene.

4.3.2 Protein Overexpression

All proteins were first subjected to a pilot protein overexpression as described in Method 2.9.1 Samples were then analysed using SDS PAGE gels, which were stained with Coomassie blue stain, and destained accordingly as seen in Figures 4.4 and 4.5.

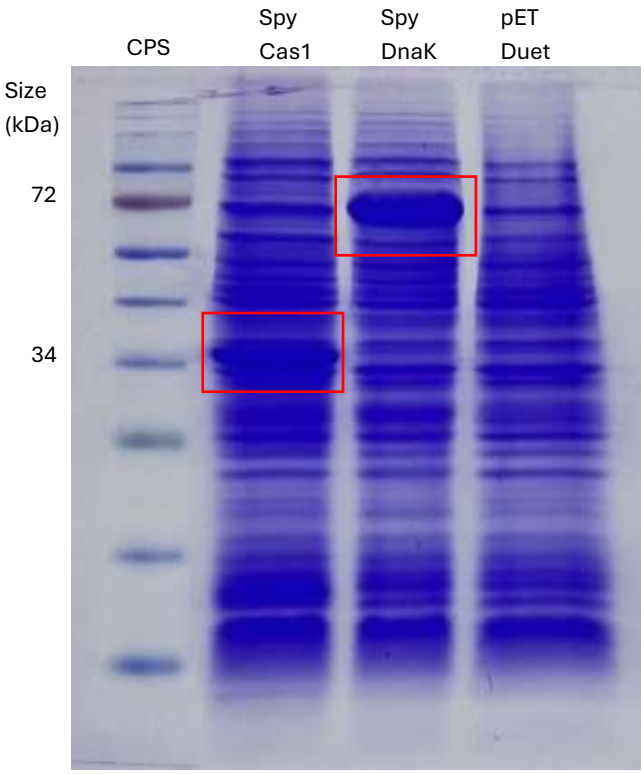


Figure 4.4 – 12% SDS PAGE gel showing pilot protein overexpression of S. pyogenes Cas1 and DnaK. *Indicated proteins were expressed in BL21AI cells before being denatured and ran on a 12% SDS PAGE gel. Gel was stained with Coomassie blue stain and destained. Proteins of interest are indicated with the red box.*

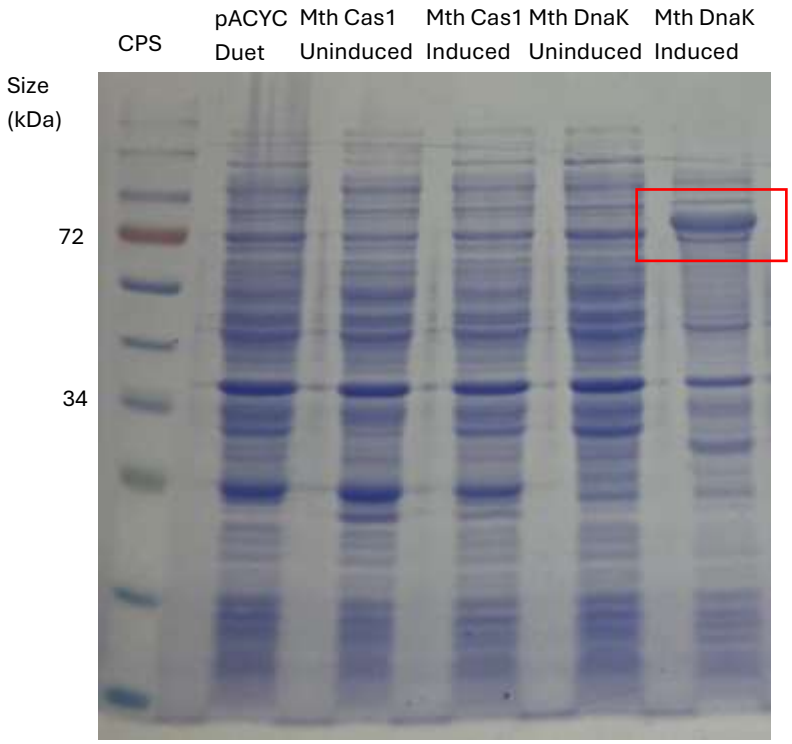


Figure 4.5 - 10% SDS PAGE gel showing pilot protein overexpression of cloned M. thermautotrophicus Cas1 and DnaK in BL21AI Cells. *Indicated proteins were expressed in BL21AI cells before being denatured and ran on a 10% SDS PAGE gel. Gel was stained with Coomassie blue stain and destained. Proteins of interest are indicated with the red box*

S. pyogenes Cas1, *S. pyogenes* DnaK and *M. thermautotrophicus* DnaK all overexpressed well in BL21AI cells, however the cloned plasmid of pACYC Duet containing *M. thermautotrophicus* Cas1 showed no visible overexpression. It was theorised that this was due to usage of rare codons in the *M. thermautotrophicus* Cas1 gene sequence, as it was cloned directly from genomic DNA. To overcome this, the plasmid was transformed into BL21 Codon Plus and Rosetta 2 cells. These are both BL21 derived *E.* coli strains that are engineered to contain additional copies of tRNAs with rare codons to allow optimal expression of heterologous proteins.

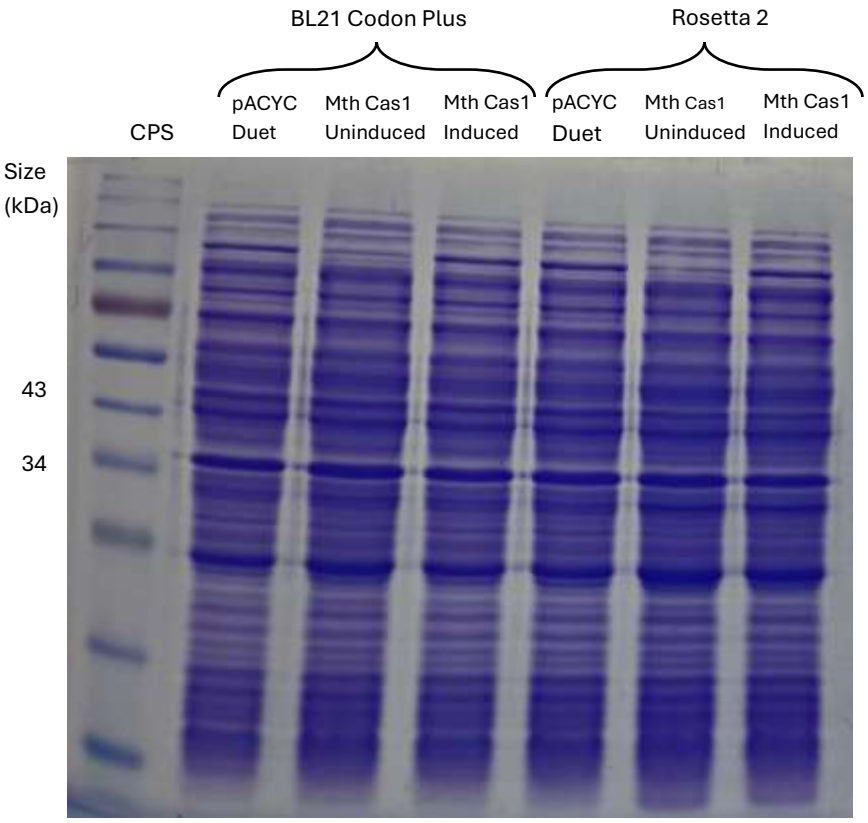


Figure 4.6 -10% SDS PAGE gel showing pilot protein overexpression of cloned M. thermautotrophicus Cas1 in BL21 Codon Plus and Rosetta 2 cells. *M. thermautotrophicus Cas1 was expressed in indicated cell strains before being denatured and ran on a 10% SDS PAGE gel. Gel was stained with Coomassie blue stain and destained.*

As both BL21 Codon Plus and Rosetta 2 cells are both BL21 derived, pilot protein overexpression was carried out using the same method. Unfortunately, as seen in Figure 4.6, no visible expression of *M. thermautotrophicus* Cas1 could be seen when BL21 Codon Plus or Rosetta 2 cells were used for protein expression. This indicated that there was a potential problem with the plasmid cloned, so a new plasmid was designed. This contained a codon-optimised gene sequence for *M. thermautotrophicus* Cas1 cloned into a pET-100 D-TOPO vector and was created by Addgene.

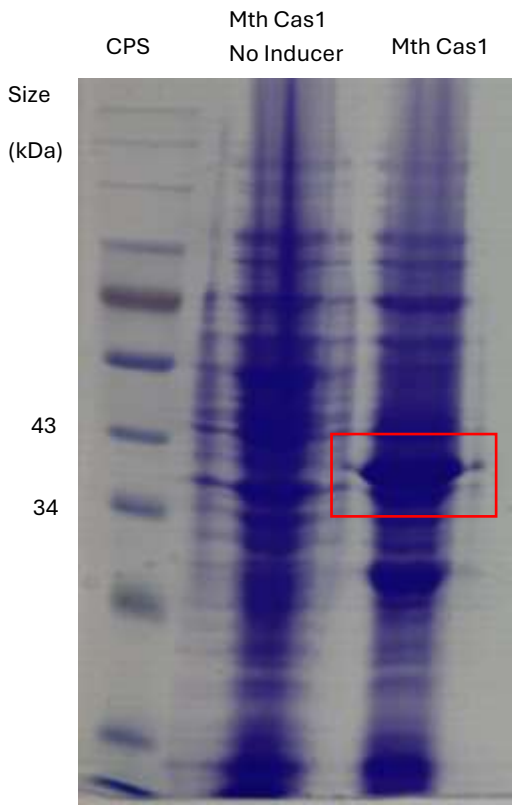


Figure 4.7 - 10% SDS PAGE gel showing pilot protein overexpression of bought M. thermautotrophicus Cas1 in BL21AI Cells. *M. thermautotrophicus Cas1 proteins were expressed in BL21AI cells before being denatured and ran on a 10% SDS PAGE gel. Gel was stained with Coomassie blue stain and destained. M. thermautotrophicus Cas1 is indicated with the red box.*

The new *M. thermautotrophicus* Cas1 plasmid was then transformed into BL21AI cells and underwent pilot protein overexpression. The final SDS PAGE gel of this plasmid (shown in Figure 4.7) showed strong expression of *M. thermautotrophicus* Cas1, confirming that the plasmid overexpressed the correct protein.

Once protein overexpression had been confirmed with SDS PAGE analysis, the reactions were scaled up to produce larger volumes of protein. This followed the same procedure as pilot protein overexpression, except for *S. pyogenes* Cas1 production. When grown at 37°C following induction, *S. pyogenes* Cas1 seems to form protein aggregates, so a very low yield of pure protein is acquired. Instead, *S. pyogenes* Cas1 was grown at 18°C for 16-18 hours with shaking following the addition of L-Arabinose and IPTG to allow optimal protein expression. Following the scaled-up production of cell biomass containing relevant proteins, it was possible to purify said proteins.

4.3.3 Protein Purification

Following production of cell biomass, proteins were purified according to the method set out in Chapter 2.9.2. All proteins used in this study were purified using an N-terminal 6x-His tag, and they were purified using AKTA Start machinery. The production of a UV chromatogram allowed for identification of protein containing fractions of interest, which were then analysed further using SDS PAGE to determine the exact fractions containing relevant proteins.

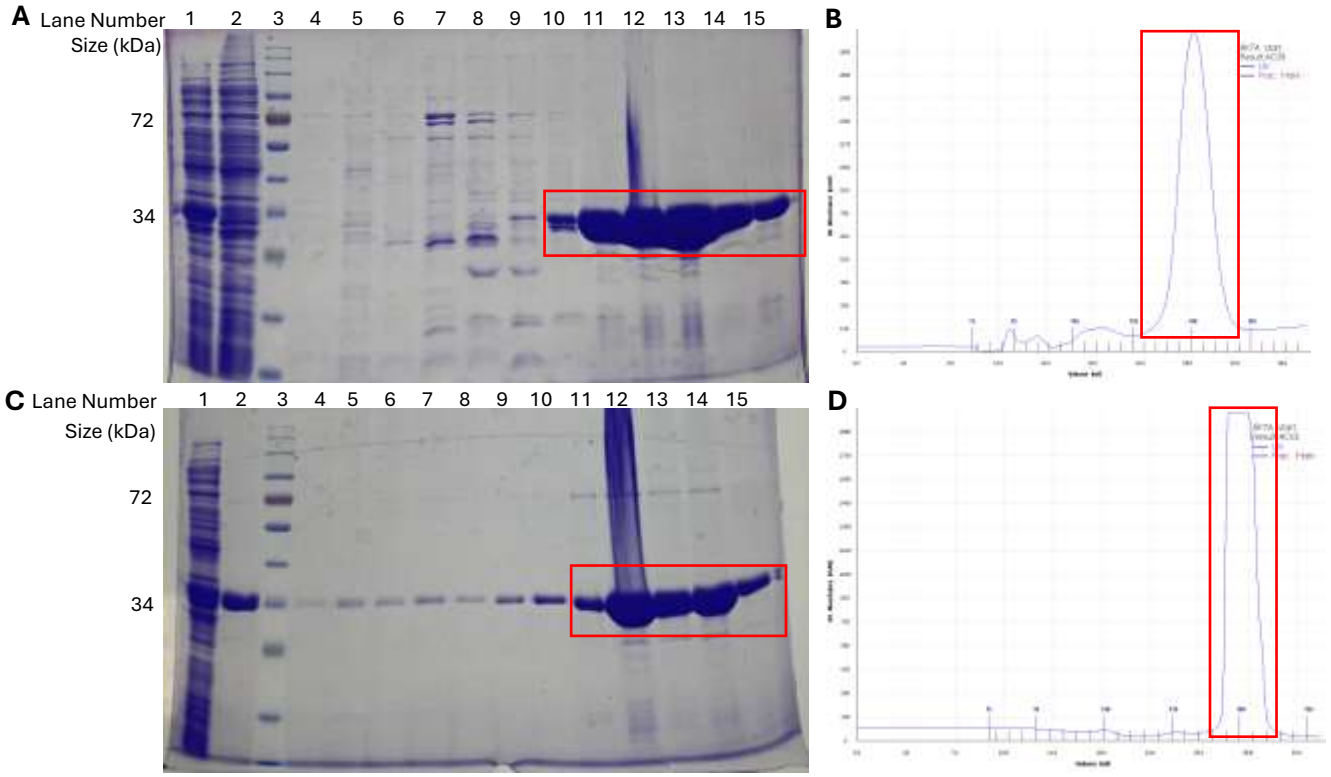


Figure 4.8 - S. pyogenes Cas1 purification using a Ni^{2+} NTA column and a heparin column. *(A) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Ni^{2+} NTA column elution. S. pyogenes Cas1 is indicated by the red box at ~34kDa. Lane 1 contains cleared cell lysate, lane 2 contains the wash through/ flow through, lane 3 contains CPS, and lanes 4-15 contain protein purification samples. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to S. pyogenes Cas1 indicated by the red box. S. pyogenes Cas1 containing fractions were pooled and ran on a heparin column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from heparin column elution. Lane 1 contains cleared cell lysate, lane 2 contains the wash through/ flow through, lane 3 contains CPS, and lanes 4-15 contain protein purification samples. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being stored at -80°C.*

Purification of *S. pyogenes* Cas1 was based on the method set out by Ka et al. (2016), and can be seen in Figure 4.8. As *S. pyogenes* Cas1 was purified with an N-Terminal 6x-His tag, it was first able to be purified using a 5ml Ni^{2+} NTA column, following which protein containing fractions were identified using SDS PAGE analysis. *S. pyogenes* Cas1 containing fractions were pooled and dialysed overnight at 4°C, before being ran on a 1ml heparin column. A heparin column was used as heparin is a DNA mimic and can therefore be used to bind to DNA binding proteins. Following purification using a heparin column, the sample was deemed as pure enough to be used for assays, so unlike the experimental procedures set out by Ka et al. (2016), the sample was not subjected to size exclusion chromatography. Instead, protein containing fractions were pooled and dialysed at 4°C overnight before being stored at -80°C.

Once successful purification of *S. pyogenes* Cas1 had been confirmed, *M. thermautotrophicus* Cas1 was able to be purified using the same method, shown in Figure 4.9. Both final protein solutions had small amounts of contaminant proteins in the mixture, but the concentration of these was deemed as insignificant compared to the protein of interest, so an additional purification stage was not needed.

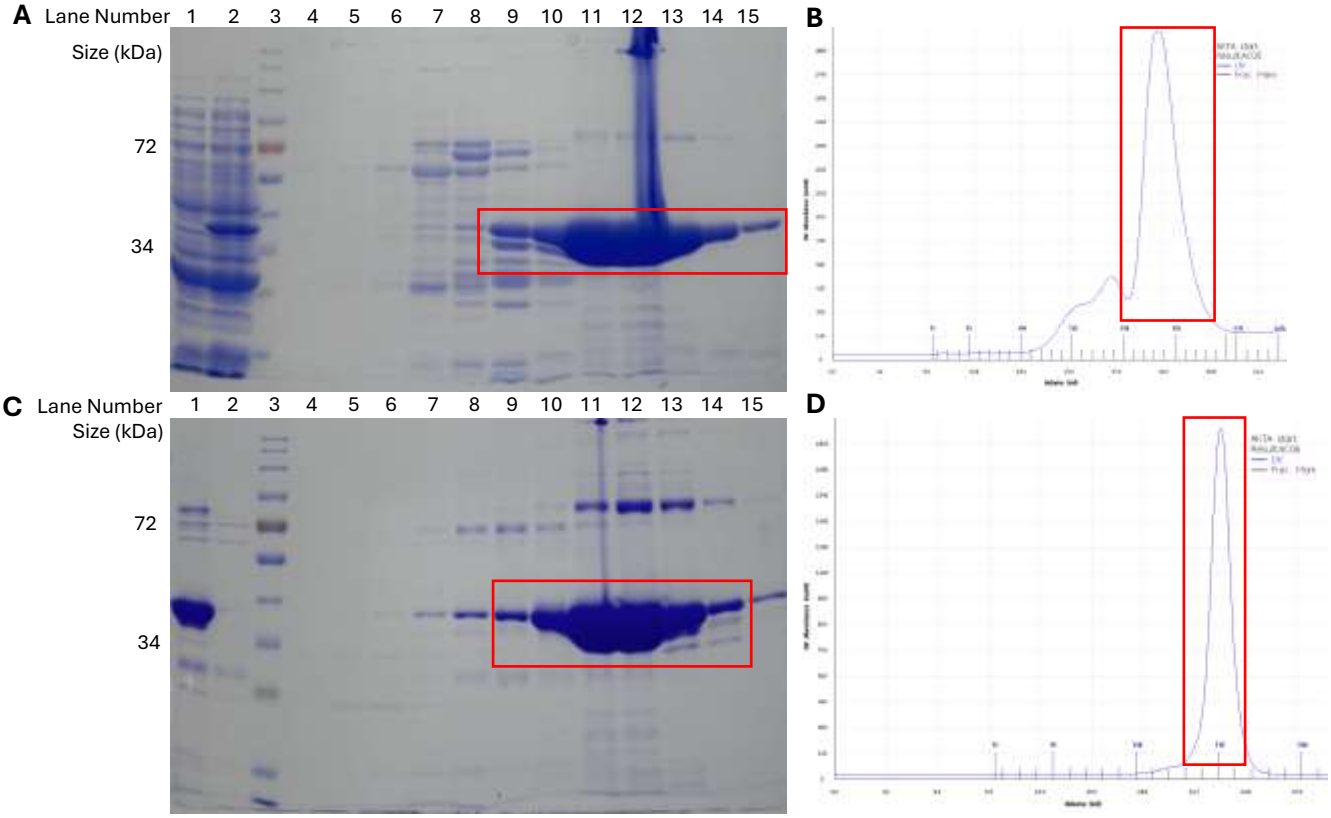


Figure 4.9 – M. thermautotrophicus Cas1 purification using a Ni^{2+} NTA column and a heparin column. *(A) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Ni^{2+} NTA column elution. M. thermautotrophicus Cas1 is indicated by the red box at ~34kDa. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to Mth Cas1 indicated by the red box. Lane 1 contains the wash through/ flow through, lane 2 contains the cleared cell lysate, lane 3 contains CPS, lanes 4-15 contain purification fractions. M. thermautotrophicus Cas1 containing fractions were pooled and ran on a heparin column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue showing fractions from heparin column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/ flow through, lane 3 contains CPS, lanes 4-15 contain protein purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Indicated fractions were pooled and dialysed before being stored at -80°C.*

For this study, DnaK from *S. M. thermautotrophicus* and *S. pyogenes* were also purified, shown in Figure 4.10 and 4.11 respectively. The method for this was based on the method of E. coli DnaK purification set out by Killelea et al. (2023). As both proteins were N-terminally 6x-His tagged, the first column used was a 5ml Ni^{2+} NTA column. Following this, as DnaK is not a DNA binding protein, it was purified using anion exchange chromatography a 1ml Q-Sepharose column.

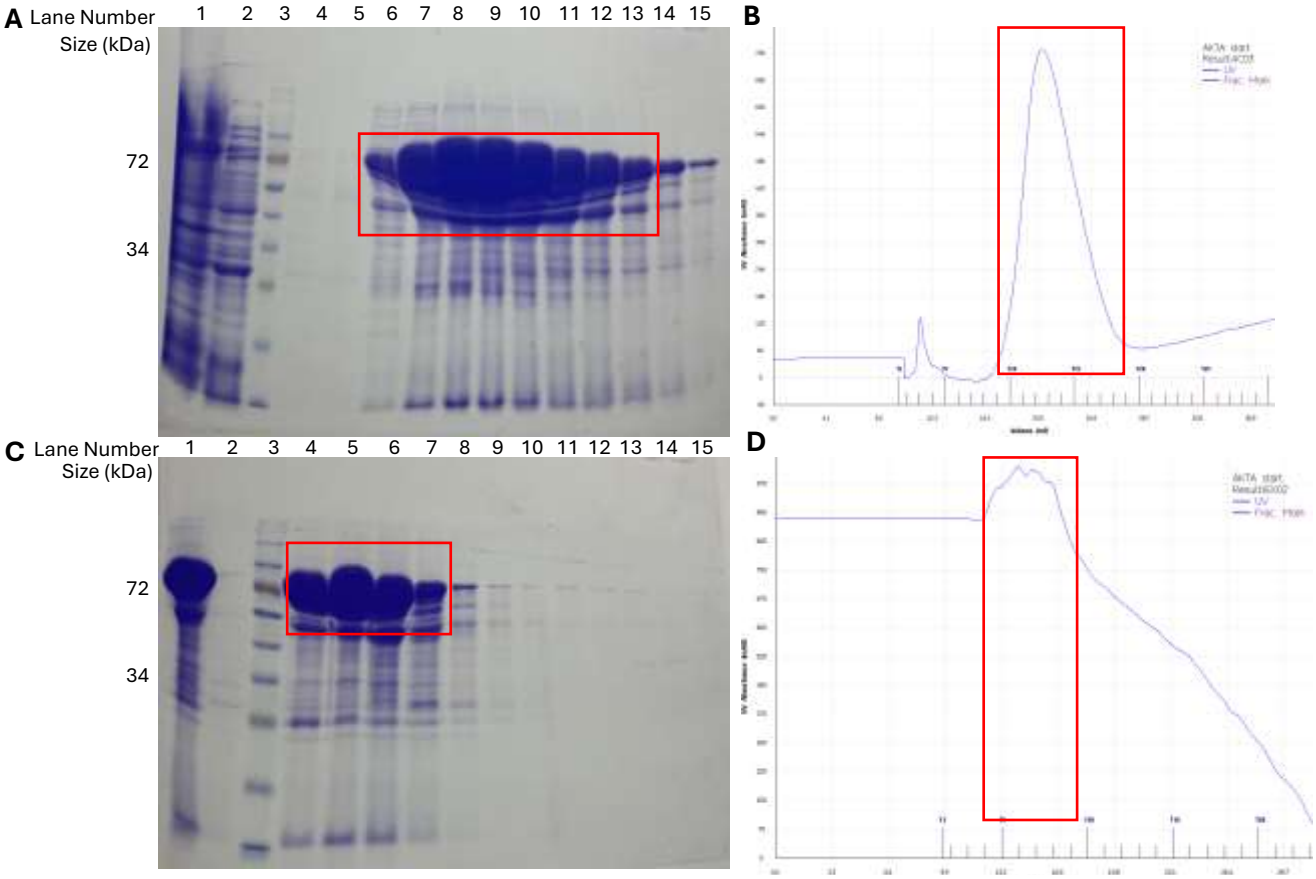


Figure 4.10 – M. thermautotrophicus DnaK purification using a Ni^{2+} NTA column and a Q-Sepharose column. *(A) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Ni^{2+} NTA column elution. M. thermautotrophicus DnaK is indicated by the red box at ~70kDa. Lane 1 contains the cleared cell lysate, lane 2 contains the wash through/ flow through, lane 3 contains CPS, lanes 4-15 contain protein purification fractions. (B) UV trace from AKTA start showing the UV absorbance throughout the purification, with the corresponding peak to M. thermautotrophicus DnaK indicated by the red box. M. thermautotrophicus DnaK containing fractions were pooled and ran on a Q- Sepharose column, the elution from which is shown in (C) and (D). (C) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Q-Sepharose column elution. Lane 1 contains the dialysed sample from the Ni^{2+} NTA column, lane 2 contains the wash through/ flow through, lane 3 contains CPS, lanes 4-15 contain protein purification fractions. (D) UV trace from AKTA start showing the UV absorbance throughout the purification. Note, the UV absorbance in (D) starts higher than in (C) due to ATP being present in the buffer used, which also absorbs UV light. Indicated fractions were pooled and dialysed before being concentrated and stored at -80°C.*

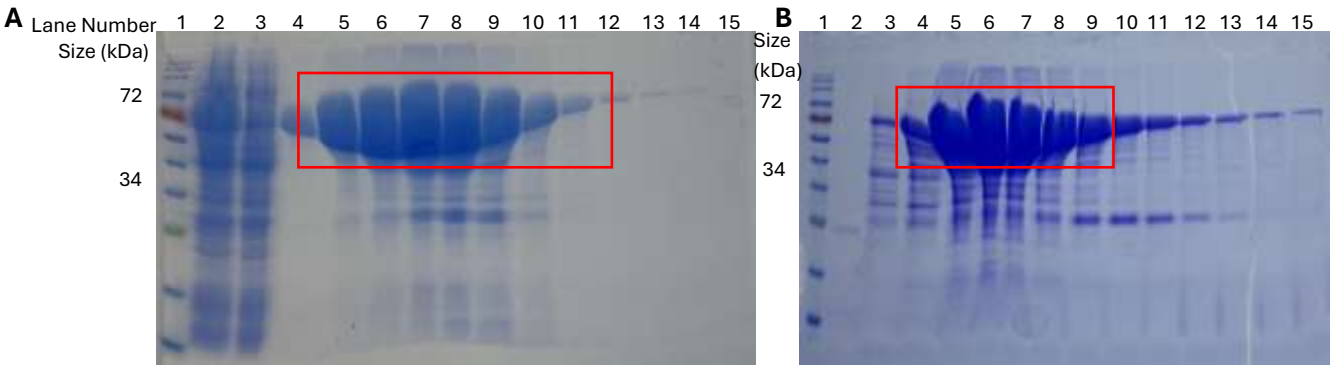


Figure 4.11 - S. pyogenes DnaK purification using a Ni^{2+} NTA column and a Q- Sepharose column. *(A) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Ni^{2+} NTA column elution. S. pyogenes DnaK is indicated by the red box at ~72kDa. Lane 1 contains CPS, lane 2 contains the cleared cell lysate, lane 3 contains the wash through/flow through, lanes 4-15 contain protein purification fractions. S. pyogenes DnaK containing fractions were pooled and ran on a Q-Sepharose column, the elution from which is shown in (B). (B) 10% SDS PAGE gel stained with Coomassie blue stain showing fractions from Q-Sepharose column elution. Lane 1 contains CPS, lane 2 contains the wash through/ flow through, lanes 3-15 contain protein purification fractions. Indicated fractions were pooled and dialysed, before being stored at -80°C.*

Both *S. pyogenes* DnaK and *M. thermautotrophicus* DnaK were supplemented with ATP and Mg^{2+} during the purification, *S. pyogenes* DnaK during the Ni^{2+} NTA column and *M. thermautotrophicus* DnaK during the Q Sepharose column. This is due to the fact that DnaK is a protein binding protein, so any proteins bound by DnaK were removed by supplementing the buffer with ATP and Mg^{2+}, as these cause the substrate binding domain of DnaK to release its substrate.

Once all the proteins needed for this study had been purified, their concentration was determined using the Bradford Assay, and the BCA assay. As in Chapter 3, the proteins purified in this study eluted at high concentrations, which were outside of the range of concentrations that can accurately be calculated using the Bradford or the BCA assay. Therefore, neat protein, ½ dilutions and ¼ dilutions were used in both assays to ensure that consistent results were gathered. Protein concentrations were calculated in mg/ml based on a standard curve of known concentrations of BSA, before being converted to µM concentrations using the molecular weight as calculated by ProtParam.

To ensure that protein concentrations calculated were accurate, a summary SDS PAGE gel was ran containing 1, 5 and 10µg of protein, shown in Figure 4.12. This ensured that the levels of protein across purifications was comparable. As this confirmed that protein concentrations calculated were accurate, they were able to be used in assays. It should be noted that the bands for *S. pyogenes* Cas1 appear to be more intense than for the other proteins, so the concentration calculted in this study may be lower than the true concentration of the protein.

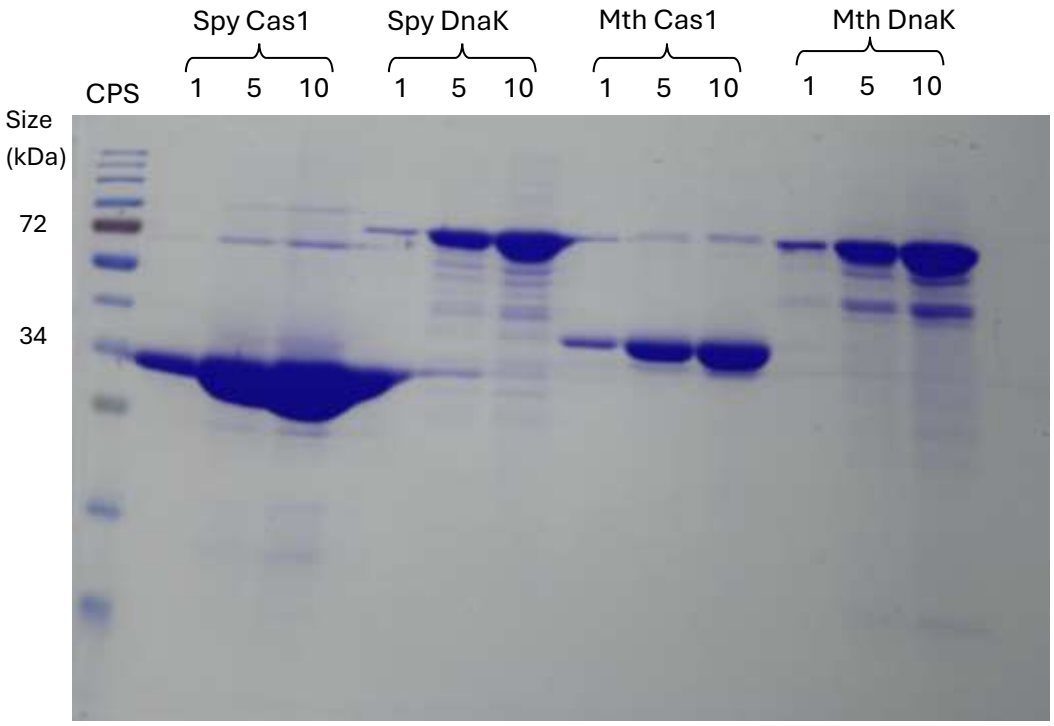


Figure 4.12 - Comparative 10% SDS PAGE gel showing 1, 5 and 10µg of S. pyogenes Cas1, S. pyogenes DnaK, M. thermautotrophicus Cas1 and M. thermautotrophicus DnaK. *Gel was stained using Coomassie blue stain and destained before visualisation*

4.4 Biochemical Analysis of *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 with DnaK

4.4.1 Analysis of Cas1 binding to DNA across species

The ability of the *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 purified in this study to bind to DNA was studied using the same methods as seen in Chapter 3.4.1. EMSAs were used to detect the formation of a stable protein-DNA complex on a gel. Fluorescence anisotropy was also used to detect the formation of more transient protein-DNA complexes in solution.

M. thermautotrophicus Cas1 was found to bind DNA well in EMSAs as seen in Figure 4.13, forming a clear stable complex in the gel at lower concentrations of Cas1. However, at high concentrations protein-DNA aggregates did form, which are complexes larger than the acrylamide matrix found in the gel. Due to the size of these, they are unable to travel into the gel and they instead get trapped in the wells of the DNA.

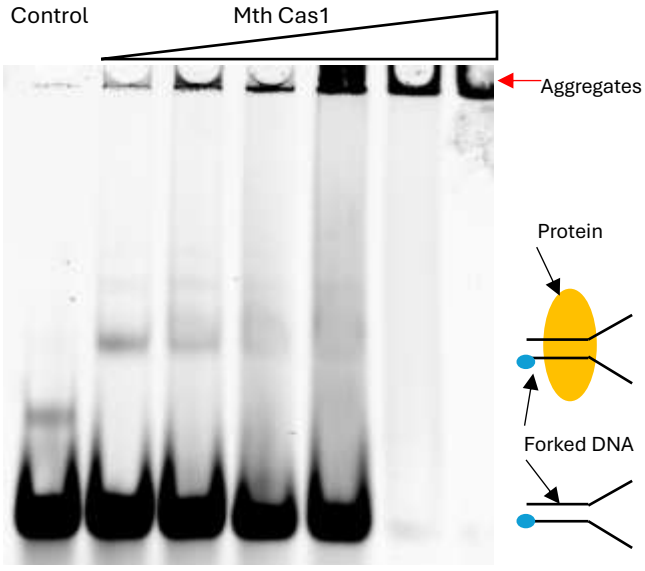


Figure 4.13 - EMSA showing the ability of M. thermautotrophicus Cas1 to bind forked DNA substrate. *Reaction was run on a 5% native TBE gel and contained 20nM Cy5-Fork 2A, with protein concentrations ranged from 62.5 - 2000nM. Reaction was run alongside a NP control*

S. pyogenes Cas1 however did not form clear, stable protein-DNA complexes when using the standard EMSA protocol used for other proteins. Immediately, large protein-DNA aggregates formed, as shown in Figure 4.14, so distinct DNA binding was not able to be identified. As a result of this, a variety of modifications were made to the protocol in attempt to break down the protein-DNA aggregates and isolate a stable protein-DNA complex, and these are seen in Figure 4.15.

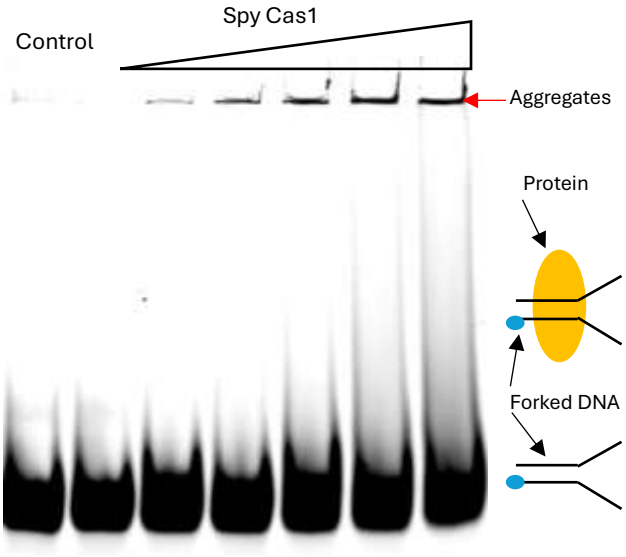


Figure 4.14 -EMSA showing the ability of S. pyogenes Cas1 to bind forked DNA substrate. *Reaction was run on a 5% native TBE gel and contained 20nM Cy5-Fork 2A, with protein concentrations ranged from 62.5 - 2000nM. Reaction was run alongside a NP control*

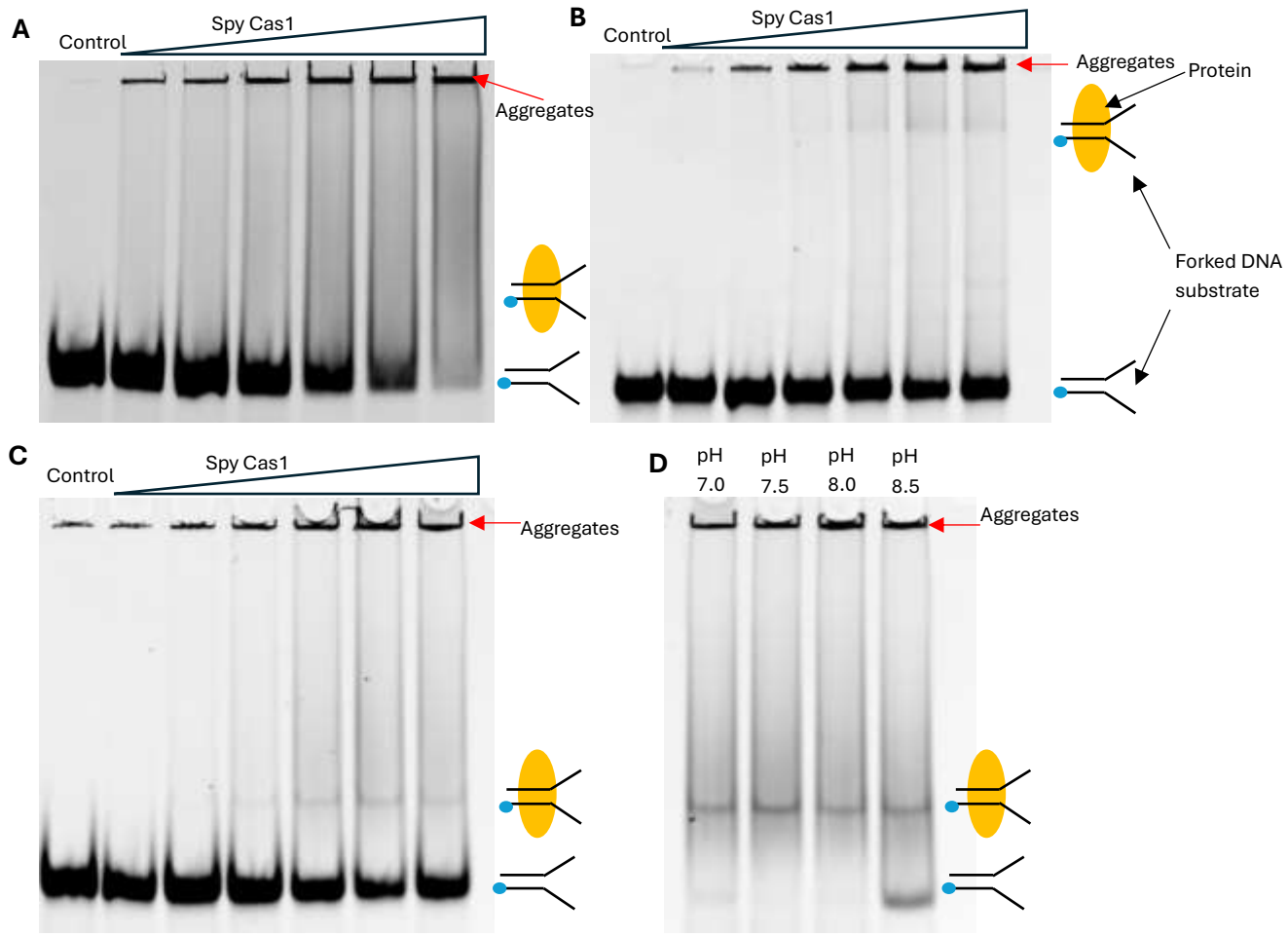


Figure 4.15 – Comparative EMSAs showing levels of S. pyogenes Cas1-DNA aggregation under different conditions. *All reactions contained 20nM Cy5-Fork 2A, with protein concentrations ranged from 62.5 - 2000nM. (A-C) reactions were run alongside a NP control. (A) S. pyogenes Cas1 binding DNA with 0.1% Tween 20 in the reaction mix, run on a 5% native TBE gel using TBE buffer. (B) EMSA run on a 5% LIS gel using LIS buffer rather than TBE. (C) S. pyogenes Cas1 binding DNA with 5mM EDTA in the reaction mix, run on a 5% native TBE gel using TBE buffer. (D) Comparative EMSA containing 1μM S. pyogenes Cas1 at varying pH to determine the impact had on levels of protein-DNA binding and aggregation. Run on a 5% native TBE gel using TBE buffer.*

Modifications made included adding 0.1% Tween 20 in the reaction mix. Tween 20 is a detergent and an emulsifier, so it was theorised that the presence of Tween 20 in the reaction mix would break down some of the protein aggregates. Unfortunately, no individual protein complex was able to be seen. Following this, the EMSA was run on a 5% LIS gel rather than a TBE gel, alongside being ran in 1X LIS buffer. Sodium acetate is more similar to the salt found inside cells than sodium chloride, so by trying to replicate cellular conditions it was predicted that the protein would be more stable and therefore fewer protein aggregates would form. This technique did have some benefit, as a small

amount of protein-DNA complex was able to be seen in the gel. However, the majority of the protein still aggregated in the wells, so the problem was not eliminated.

The reaction mix was also set up using 5mM EDTA in the reaction mix. EDTA is a metal chelating agent, so it was theorised that it would react with any Mg^{2+} ions present in the reaction mix, removing them and preventing them from causing protein aggregation. As with the LIS gel, adding EDTA to the reaction mix did break down some of the protein aggregates,but the majority of the protein was still getting trapped in the wells of the gels. As a result of this, it was theorised that the reaction conditions were not optimal for protein structure. The reaction was therefore repeated with 5mM EDTA in the reaction mix using a constant concentration of *S. pyogenes* Cas1, but varying the pH to test the impact that this had on DNA binding. Up to this point, the protein had been being diluted in pH 8.0 dilution buffer, so the pH was varied from pH 7.0 to pH 8.5. DNA binding was found to be optimal at pH 7.5 and pH 8.0, so it was concluded that the pH of the reaction had already been optimal for the protein.

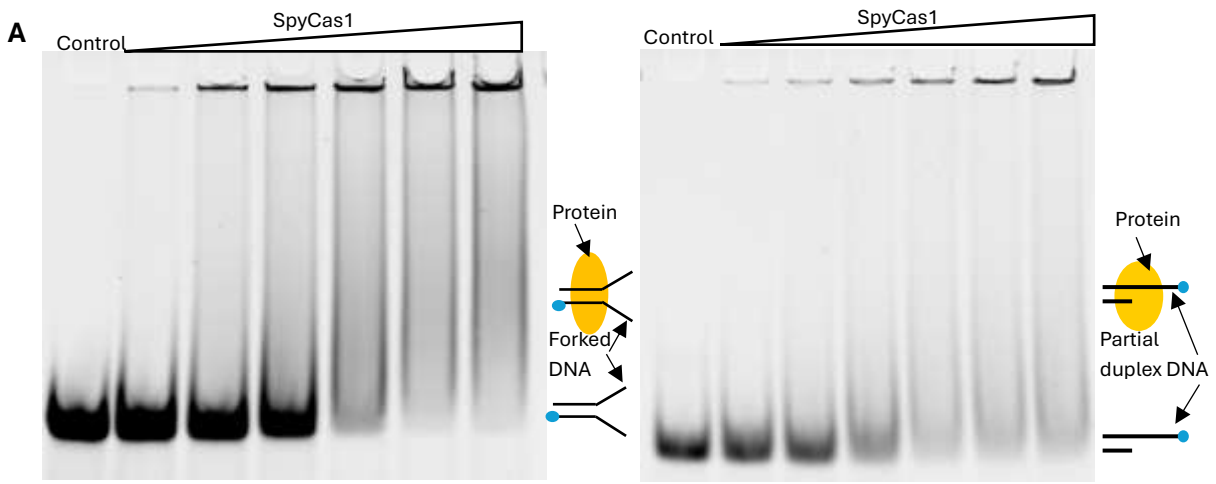


Figure 4.16 - Comparative EMSAs showing levels of S. pyogenes Cas1-DNA aggregation using different DNA substrates. *All reactions contained 20nM DNA Substrate with protein concentrations ranged from 62.5 - 2000nM. Reactions were run alongside a NP control, and all reactions were supplemented with 5mM EDTA in the reaction mix. (A) S. pyogenes Cas1 binding forked DNA substrate. (B) S. pyogenes Cas1 binding partial duplex DNA substrate.*

Although adding 5mM EDTA to the reaction mix was found to prevent some protein-DNA aggregation, the majority of protein was still aggregating in the wells of the gel, and the affinity of binding of DNA by *S. pyogenes* Cas1 was reduced. An EMSA was therefore also attempted using a partial duplex DNA substrate rather than a forked DNA substrate. The theory behind this was that a smaller, lighter substrate may allow for more individual protein-DNA complexes to be formed, rather than the larger aggregates seen up to this point. Unfortunately, no difference was seen in the levels of aggregation with the two different substrates, as seen in Figure 4.16.

Although the time constraints for this study meant that EMSA conditions were not able to be optimised for *S. pyogenes* Cas1-DNA binding, they were able to clearly show that *S. pyogenes* Cas1 did bind to DNA, albeit in higher molecular weight protein aggregates.

As in Chapter 3, DNA binding was also assessed using fluorescence anisotropy, and this is shown in Figure 4.17. Polarisation values were plotted and analysed using GraphPad Prism following the One Site Binding – Total model using the equation Y=Bmax*X/(Kd+X) + NS*X + Background, where NS represents non-specific levels of DNA binding. Kd and R^2 values were calculated for each time point and averaged to determine the final values shown on the graph. Kd and R^2 values indicate the binding affinity and goodness of fit of the graph respectively. Anomalous data points were defined as having a polarisation value greater than 100 polarisation units difference from equivalent data values and were excluded from the analysis. The results of this fluorescence anisotropy indicate that both *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 are capable of DNA binding, as seen in the EMSAs performed with the proteins. *S. pyogenes* Cas1 was found to bind with higher affinity than *M. thermautotrophicus* Cas1, as indicated by the higher Kd values calculated. It should be noted that *M. thermautotrophicus* Cas1 has an optimal temperature of between 55-65°C, and the experiments performed in this study were carried out at 37°C. This may therefore explain the high Kd value calculated for *M. thermautotrophicus* Cas1. This was found to be a better measure of DNA binding by *S. pyogenes* Cas1 than EMSA analysis, as the data are less affected by protein aggregation. Both models of data analysis have high R^2 values, indicating a high goodness of fit. Therefore, the model of analysis was considered appropriate to measure DNA binding by these proteins.

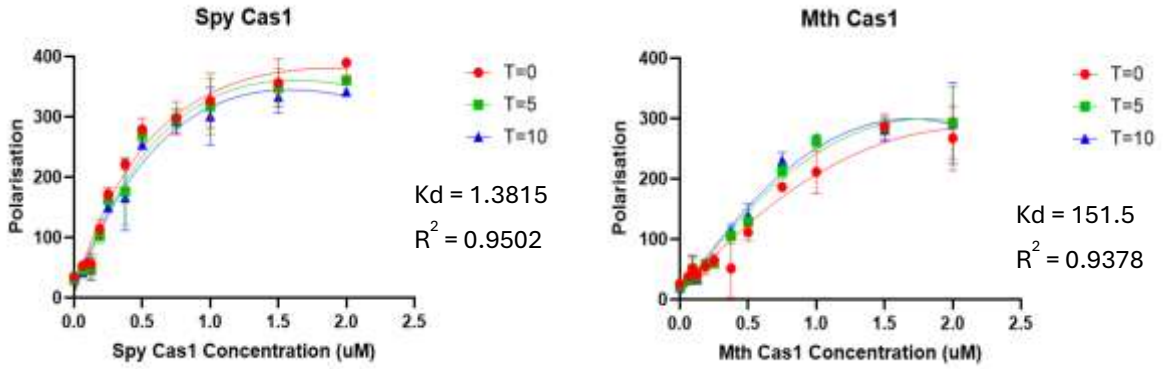


Figure 4.17 - DNA binding ability of S. pyogenes Cas1 and M. thermautotrophicus Cas1 as calculated via fluorescence anisotropy. *Reaction mix contained 40nM FAM-DNA12 partial duplex, and 0.0625-2μM protein and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd values (a measure of binding affinity) and R^2 values (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point.*

4.4.2 Investigating the Impact of DnaK on Cas1 binding DNA across species

Once DNA binding had been confirmed using EMSAs and fluorescence anisotropy, DnaK was titrated into the reaction to determine whether there was an inhibitory effect. Before this could be done, relevant controls had to be performed to ensure that DnaK on its own did not bind to DNA. Neither *S. pyogenes* DnaK nor *M. thermautotrophicus* DnaK showed any binding of DNA substrate, as seen in Figure 4.18, so it was deemed that the proteins purified in this study were sufficiently pure to test for an interaction.

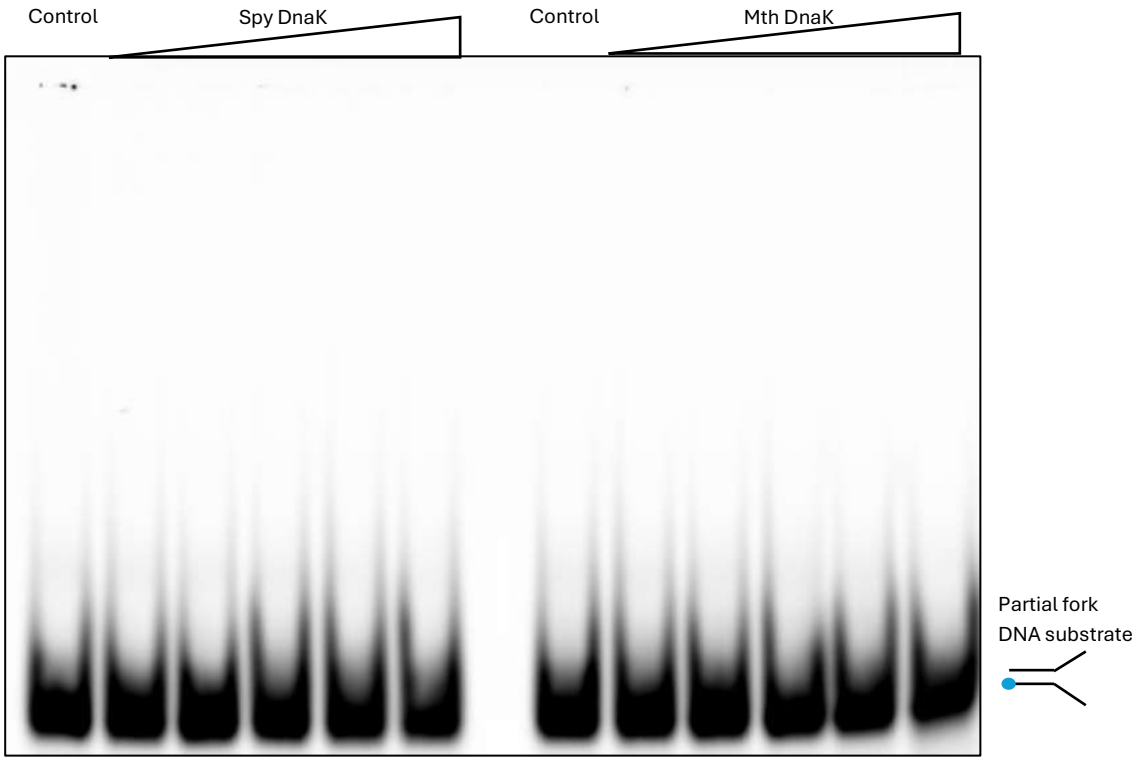


Figure 4.18 - Control EMSA showing that S. pyogenes DnaK and M. thermautotrophicus DnaK do not bind DNA. *Reactions were run on a 5% native TBE gel and contained 20nM Cy5-Fork 2a partial fork, with protein concentrations ranged from 125 - 2000nM. Samples were run alongside a NP control.*

Following this, EMSAs were set up to assess whether DnaK had an inhibitory effect on DNA binding by Cas1 in *S. pyogenes* or *M. thermautotrophicus*. To test for the interaction between *M. thermautotrophicus* Cas1 and *M. thermautotrophicus* DnaK, the two proteins were incubated together before being added to a reaction mix containing 20mM forked DNA substrate. A constant concentration of *M. thermautotrophicus* Cas1 was used with increasing concentrations of *M. thermautotrophicus* DnaK, and this can be seen in Figure 4.19. No clear trend was visible as *M. thermautotrophicus* DnaK concentration increased due to protein-DNA aggregates forming. It did however appear that higher concentrations of DnaK increased the amount of unbound DNA, indicating at a potential interaction between the two proteins.

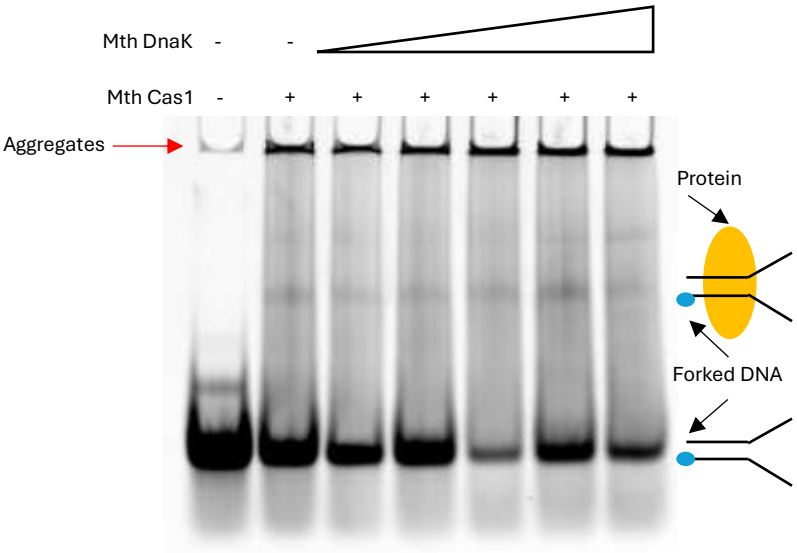


Figure 4.19 - Comparative EMSA showing the impact of M. thermautotrophicus DnaK on M. thermautotrophicus Cas1 binding DNA. *All reactions were run on a 5% native TBE gel and contained 20nM Cy5-Fork 2a DNA substrate. Constant concentration of 1000nM Mth Cas1 was used, with DnaK concentration ranging from 125-2000nM. M. thermautotrophicus Cas1 was first added to the reaction mix, followed by M. thermautotrophicus DnaK at indicated concentrations. All reactions were also run alongside a NP control.*

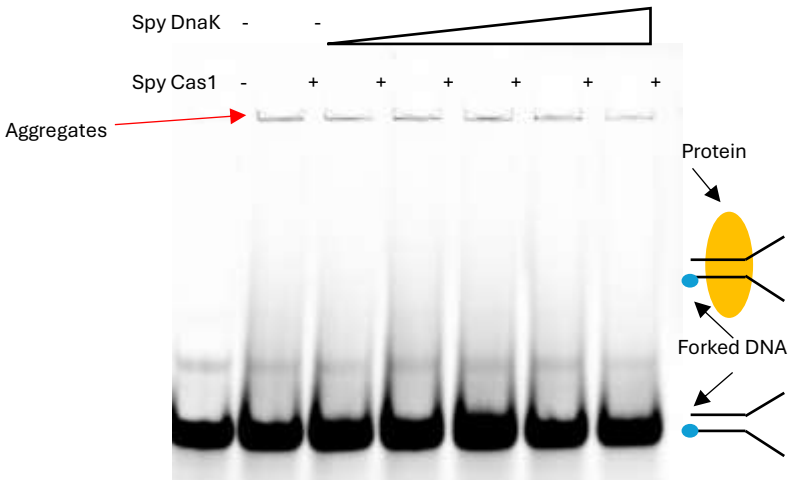


Figure 4.20 - Comparative EMSA showing the impact of Spy DnaK on S. pyogenes Cas1 binding DNA. *All reactions were run on a 5% native TBE gel and contained 20nM Cy5-Fork 2a DNA substrate. Constant concentration of 1000nM S. pyogenes Cas1 was used, with DnaK concentration ranging from 125-2000nM. S. pyogenes Cas1 was first added to the reaction mix, followed by S. pyogenes DnaK at indicated concentrations. All reactions were also run alongside a NP control.*

An equivalent EMSA was set up using *S. pyogenes* Cas1 and *S. pyogenes* DnaK to test for the presence of an interaction between the two proteins, shown in Figure 4.20. The method used was identical to that used with *M. thermautotrophicus* Cas1 and *M. thermautotrophicus* DnaK, except the reaction mixture was supplemented with 5mM EDTA, as this condition showed the least protein-DNA aggregation in previous EMSAs performed. A constant concentration of *S. pyogenes* Cas1 was used, and variable concentrations of *S. pyogenes* DnaK were added to the reaction mix. Unfortunately, the reaction showed very little binding of *S. pyogenes* Cas1 to DNA, and all of it was in the form of protein-DNA aggregates, so it was difficult to see any trends. It did appear however that there was potentially less aggregation as DnaK concentration increased, hinting at a potential inhibitory interaction between *S. pyogenes* Cas1 and *S. pyogenes* DnaK.

The interaction between the Cas1 and DnaK in *S. pyogenes* and *M. thermautotrophicus* was also assessed with fluorescence anisotropy. This first required performing a control reaction using only *S. pyogenes* DnaK and *M. thermautotrophicus* DnaK to ensure that these proteins alone did not bind DNA, which is shown in Figure 4.21. Fortunately, neither protein exhibited DNA binding in fluorescence anisotropy, so they were able to be used in assays with *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 respectively.

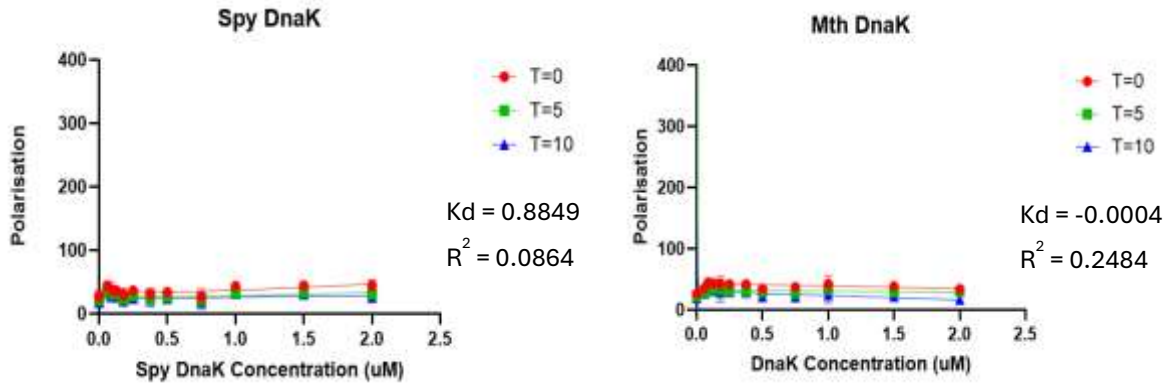


Figure 4.21 - DNA binding ability of S. pyogenes DnaK and M. thermautotrophicus DnaK as calculated via fluorescence anisotropy. *Reaction mix contained 40nM FAM-DNA12 partial duplex, and 0.0625-2μM protein and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd values (a measure of binding affinity) and R^2 values (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point.*

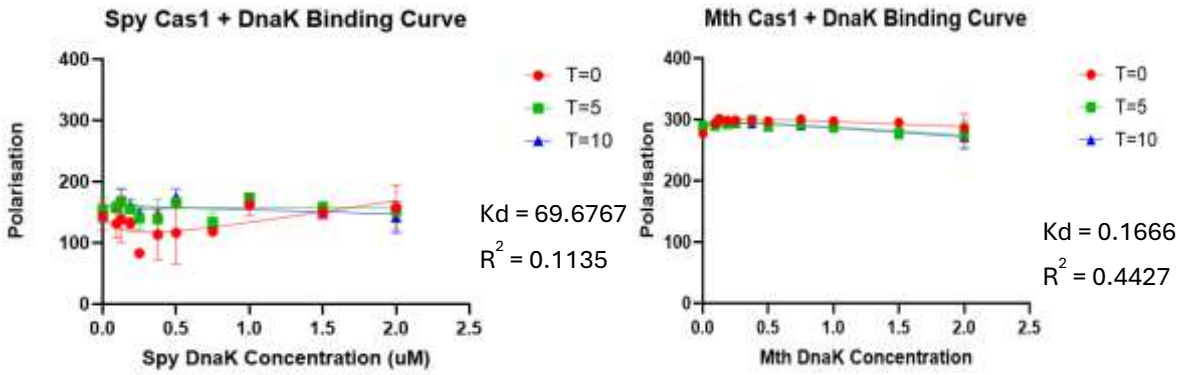


Figure 4.22 - DNA binding ability of S. pyogenes Cas1 and M. thermautotrophicus Cas1 in the presence of S. pyogenes DnaK and M. thermautotrophicus DnaK respectively, as calculated via fluorescence anisotropy*. Reaction mix contained 40nM FAM-DNA12 partial duplex, 0.5μM Cas1 and 0.09375-2μM DnaK and measurements were taken at 0, 5 and 10 minutes following the addition of relevant protein. Higher polarisation values indicate higher levels of DNA binding. Kd values (a measure of binding affinity) and R2 values (a measure of goodness of fit) calculated for each of the time points were averaged and presented alongside the graph. Data plotted is a mean value across triplicate experiments, and error bars indicate the standard deviation for each data point.*

For the fluorescence anisotropy experiments involving both Cas1 and DnaK, the two proteins were first incubated together for 5 minutes at 37°C, before being added to the reaction mixture, and the results of this can be seen in Figure 4.22. Anomalous data was defined as any data point with a difference of greater than 100 polarisation units from equivalent values.

For *S. pyogenes* Cas1 and DnaK, there was an increase in the Kd value from *S. pyogenes* Cas1 alone, which is indicative of lower binding affinity. This may mean that the presence of DnaK does inhibit Cas1 binding to DnaK in *S. pyogenes*. The trend in the change in polarisation values however was unclear, as initial binding showed an increase in polarisation as *S. pyogenes* DnaK concentration increased, however for binding after 5 and 10 minutes, there was an overall negative trend in the polarisation level, indicating that there was less DNA binding. The R^2 value for *S. pyogenes* Cas1 and DnaK was also low, indicating that the data had a low goodness of fit and may therefore not be particularly reliable.

For *M. thermautotrophicus* Cas1 and DnaK however, there is a decreased Kd value from *M. thermautotrophicus* Cas1 on its own, indicating that the binding affinity is increased in the presence of *M. thermautotrophicus* DnaK. There is a slight negative trend in the polarisation values, hinting that there may be a mild inhibitory effect from *M.*

thermautotrophicus DnaK on *M. thermautotrophicus* Cas1 binding DNA, but no substantial changes were observed. The R^2 value calculated for this data is much higher than that seen for *S. pyogenes* Cas1 and DnaK, indicating that there is a better fit of the data and that it is more reliable.

4.4.3 Investigating the Catalytic Activity of Cas1 across species

Catalytic activity of *S. pyogenes* and *M. thermautotrophicus* Cas1 was assessed using disintegration assays. These rely on the ability of Cas1 to cleave a partial fork DNA substrate which replicated DNA found *in vivo* during CRISPR acquisition. Proteins were diluted in 1X cleavage buffer, containing 100mM NaCl rather than 10mM NaCl as used for *E. coli* proteins. This is because both *S. pyogenes* and *M. thermautotrophicus* Cas1 required higher salt in their storage buffer to prevent precipitation of protein, so higher salt was also used in assays. Following this, experiments were carried out according to Method 2.10.4.

The results of these disintegration assay, shown in Figure 4.23, showed that neither of two Cas1 proteins purified for this area of research were as active in disintegration assays as the *E. coli* Cas1 purified for this study. *S. pyogenes* Cas1 showed only trace amounts of catalytic activity at 100nM protein concentration. *M. thermautotrophicus* Cas1 showed slightly higher levels of activity, with small amounts of catalytic activity being seen from 50nM protein, however this was also less active than the *E. coli* proteins used in this study. Interestingly, the amount of single stranded DNA seen decreased as the concentration of Cas1 increased. This indicated that the proteins used in this study had

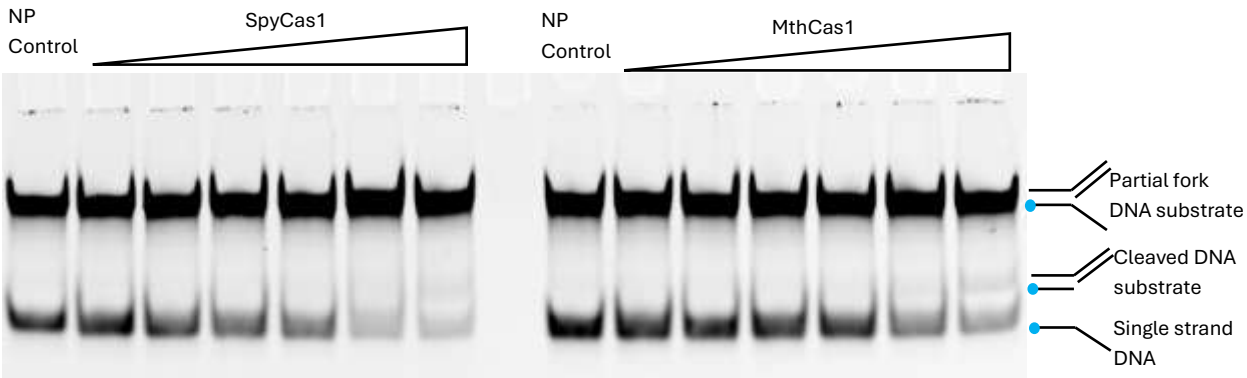


Figure 4.23 – Disintegration assay showing catalytic activity of S. pyogenes Cas1 and M. thermautotrophicus Cas1. *20nM Cy5-labelled partial fork was incubated with protein concentrations from 12.5-100nM for 30 minutes at 37°C to allow for cleavage of DNA substrate, following which protein was degraded using stop buffer. Samples were electrophoresed for 90 minutes on a 12% native TBE gel.*

exonuclease activity and were able to degrade the single stranded DNA seen. The gel type used in this assay does not allow for the resolution of such small DNA fragments, but a denaturing urea gel could have been used to visualise the products of this degradation were it not for the time constraints of this study.

4.5 Summary

Bioinformatic structural modelling indicates that there would be an interaction between *S. pyogenes* Cas1 with DnaK, and *M. thermautotrophicus* Cas1 DnaK in a similar manner to that which can be seen in *E. coli* Cas1 and DnaK. A sequence alignment was performed to see if any of the residues on Cas1 implicated in DnaK binding in *E. coli* were conserved across the Cas1 proteins studied in this research, but the residues were not conserved. Plasmids were ordered from Addgene to allow for the overexpression of *S. pyogenes* Cas1, *S. pyogenes* DnaK, *M. thermautotrophicus* Cas1 and *M. thermautotrophicus* DnaK in BL21AI cells to allow for protein overexpression, and these proteins were then purified to be used in *in vitro* assays.

The ability of *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 to bind DNA *in vitro* was studied using EMSAs and fluorescence anisotropy. *Spy* Cas1 showed high affinity binding to DNA in both EMSA and fluorescence anisotropy, but unfortunately when used in EMSAs large protein aggregates formed so a single protein-DNA complex was unable to be resolved on a gel. *M. thermautotrophicus* Cas1 showed lower affinity binding for DNA than *S. pyogenes* Cas1 in both EMSAs and fluorescence anisotropy, but fewer protein-DNA aggregates formed so it was possible to isolate a stable complex on a gel. As both proteins were found to bind to DNA, it was possible to titrate DnaK from each respective species into the reaction mix to determine if there was an inhibitory effect on Cas1 binding to DNA. Not enough conclusive proof could be gathered to indicate that the interaction was conserved across these species, as although EMSAs and fluorescence anisotropy both hinted at an interaction between the two proteins, no clear trend was able to be observed with either form of assay.

S. pyogenes Cas1 and *M. thermautotrophicus* Cas1 were both also tested for their catalytic activity using disintegration assays to determine if they were able to cleave a partial fork DNA substrate. Both showed low levels of catalytic activity at high

concentrations and were less active than the *E. coli* proteins purified for Chapter 3, indicating that the proteins may not have been purified in their optimal form, or that the reaction conditions were not optimised for protein functionality. As the plasmids used for this area of research were bought from Addgene, they were synthesized with a 6xHis tag and a linker region before the start of the Cas1 gene. This leads to amino acids being added at the N-terminal portion of the protein, and this may have interfered with the functionality of the Cas1 proteins purified *in vitro*.

Chapter 5 - Discussion and Future Work

The aim of this study was to further characterise the interaction between Cas1 and DnaK in CRISPR Systems that was set out in *E. coli* by Killelea et al. (2023). In E. coli, where the interaction between Cas1 and DnaK inhibits DNA binding by Cas1, the specific interaction between amino acids in Cas1 was studied to try to map the interaction. It was also attempted to establish whether this is a conserved mechanism for control of CRISPR Adaptation by investigating the interaction between the two proteins across bacteria and archaea that have different configurations of CRISPR systems, but also contain Cas1 and DnaK.

5.1 Molecular Mechanism for the Interaction of Cas1 and DnaK in *E. coli*

The data gathered in Chapter 3 indicate that R41 may be a potential interaction site for Cas1 and DnaK in E. coli, as predicted by structural predictions using Alphafold 3. Pull down assays performed seem to indicate a weaker interaction between Cas1 R41G and DnaK, compared with Cas1 V76L and WT Cas1 with DnaK. This conclusion is also supported by the anisotropy data gathered in this study, as both WT Cas1 and Cas1 V76L showed a decrease in DNA binding when DnaK was included in the reaction mix, but this was not seen with Cas1 R41G. The R41G mutant however was difficult to gain any confident conclusions from, due to the fact that the mutation interrupts the arginine clamp of the Cas1-Cas2 acquisition complex, which rendered Cas1 R41G non-functional in acquisition and disintegration assays, and having a far lower DNA binding affinity than WT Cas1 and Cas1 V76L, making it challenging to study the interaction between the two proteins using anisotropy and EMSAs. However, as Cas1 forms homodimers, there is still the possibility of R41 interaction with both DNA and DnaK *in vivo*, especially given that DnaK restrains DNA binding. It may therefore be that DnaK binds to Cas1 via R41, until MGE DNA is present upon which time, Cas1 is released and R41 is free to bind MGE DNA.

The data indicate that although the R41G mutation in Cas1 may lead to a lower binding affinity to DnaK, it does not remove the interaction altogether. It would therefore be advisable in future work to mutate a larger portion of the Cas1 protein, such as mutating the entire interacting face predicted by structural modelling, rather than a single amino acid. This could then be used in assays that examine a physical interaction between 2

proteins, such as pull down assays and BiFC to test for a physical interaction between the two proteins, and if the Alphafold prediction is correct, the interaction between the two proteins would be removed. A caveat to this is that the lack of acquisition by the Cas1 R41G-Cas2 complex would not allow conclusive removal of Cas1-DnaK interaction from a functional CRISPR acquisition complex, which would make it difficult to study the impact of the interaction between the proteins *in vivo*. For example, it would not be possible to study whether overexpressing DnaK would inhibit the Cas1 R41G-Cas2 complex in acquisition, as no acquisition is seen.

BiFC is a method of investigating protein-protein interactions whereby a fragment of a fluorescent protein such as M-Venus is fused to each of the proteins that are predicted to interact. When expressed *in-vivo*, if the two proteins interact then the fluorescent protein fragments will come together and fluoresce, and the levels of fluorescence can then be measured to determine the strength of the interaction (Kerppola, 2008). This could be used for the comparison of the interaction between WT Cas1 and DnaK, along with Cas1 R41G and DnaK, as you could fuse Cas1 or Cas1 R41G to either C- or N- M-Venus, and DnaK to the other half of the protein. If the R41G mutation does interfere with the interaction between Cas1 and DnaK, you would expect to see lower levels of fluorescence. A larger mutation involving the whole interaction face predicted by Alphafold could also be introduced to Cas1, and if this were to remove the Cas1-DnaK interaction entirely, no fluorescence would be seen with BiFC.

As Cas1 V76L behaves similarly to WT Cas1 with DnaK in gel based assays, anisotropy and pull-down assays, the data gathered in this study do not indicate that the V76 residue of Cas1 interacts with DnaK. It is however reported to be hyperactive in adaptation by Yosef et al. (2023), which raises an interesting question of why that is because it is not hyperactive in DNA binding or disintegration assays performed in this study, where it binds DNA with slightly lower affinity. It was not catalytically hyperactive in disintegration assays but may be stimulated by the presence of DnaK which would be another explanation for the hyperactive adaptation the V76L protein exhibits. For future work, it would therefore be advisable to study Cas1 V76L and WT Cas1 with DnaK in integration assays as well as disintegration assays to determine if it is hyperactive in this manner. If Cas1 V76L was not hyperactive in integration or disintegration, this would prove that the

mutant is not catalytically hyperactive. This may imply that the residue is involved in an interaction with a currently unidentified protein for the control of adaptation.

One method of identifying said interacting proteins would be to utilise BioID2 labelling. BioID2 labelling requires the fusion of BioID2 with the protein of interest, such as Cas1. BioID2 is a highly active biotinylating protein, so proteins that interact with the protein of interest, fused to BioID2 will become biotinylated. Biotinylated proteins can then be harvested and analysed via mass spectrometry to identify the interacting proteins (Killelea et al., 2024). The proteins that interact with Cas1 could then potentially be compared to those that interact with Cas1 V76L, and this could give an insight into why Cas1 V76L is hyperactive in spacer acquisition.

A limitation of this study is that although Cas1 and its mutants were studied for their activity to determine whether they were catalytically active and therefore functional, no such experiments were performed to assess whether DnaK was functional *in vitro*. It is therefore possible that the DnaK purified for this study was not functional, and therefore may not have acted *in vitro* with Cas1 and relevant mutants as it would in cells. Therefore, ATPase assays should have been performed to determine whether DnaK was capable of hydrolysing ATP. If it were found to be capable of hydrolysing ATP, this would indicate that the protein is functional and undamaged, and therefore the results of this study can be thought of as representative of the interaction between the relevant proteins. However, if the DnaK purified in this study was not capable of hydrolysing ATP, this may indicate that the protein was damaged or not in its optimal state, and therefore may not interact with Cas1 and the relevant mutants in the manner which is seen *in vivo*. These assays were beyond the scope of what could be achieved during a 10-month MRes but should be completed in any future work.

5.2 Comparing the Interaction of Cas1 and DnaK Across Species

The data gathered in these experiments were inconclusive. EMSAs and anisotropy data imply that there is a difference in binding of purified Cas1 to DNA in the presence of purified DnaK in *S. pyogenes* and *M. thermautotrophicus*, but the trend seen is inconclusive when compared with equivalent proteins from E. coli. The fact that the *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 purified for this study were not

catalytically active to the same degree as *E. coli* Cas1 in disintegration assays indicates that the reaction conditions used in this study were not optimal for the proteins, or that the purification method used was problematic for the proteins. It may therefore be the case that the *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 were not in the same conformation as would be found *in vivo*, and therefore the interaction with *S. pyogenes* DnaK and *M. thermautotrophicus* Cas1 may not be optimal. Interestingly, the disintegration assays performed in this study indicate that the *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 purified for this study exhibit exonuclease activity due to the degradation in single strand DNA seen. This would need to be studied further by repeating the assay and running the samples on a denaturing gel, rather than a native gel to allow for the resolution of smaller DNA fragments. If exonuclease activity is seen, then there may be contaminating proteins in the purified Cas1 proteins used in this study, which may interfere with their activity. Alternatively, it may indicate a previously unidentified capability of these Cas1 proteins, which may give interesting insights into their functionality in future studies.

Other studies have also found it challenging to get *S. pyogenes* Cas1 to behave optimally in assays, for example in Ka et al. (2016), the purification method used in this research, they were unable to show *S. pyogenes* Cas1 nuclease activity in assays. They suggested that the conformation of purified *S. pyogenes* Cas1 purified may represent a catalytically inactive form, and there was potential for the requirement of an unidentified viral cofactor to induce a conformational change for catalytic activity. All the proteins purified in this study had an N-terminal 6x-His tag to allow for purification with a Ni-NTA column, but it is also possible that this tag may have interfered with protein structure by causing a conformational change, thus affecting the catalytic activity of the proteins *in vivo*. In future work, it would be advisable to cleave the 6x-His tag from the protein following purification if possible, along with altering reaction conditions such as salt concentration, pH, and temperature. It should also be noted that Mn^{2+} ions have been shown to allow better catalytic activity in *E. coli* Cas1 than the Mg^{2+} ions found in the cleavage buffer used in this study (Nuñez, Harrington, et al., 2015b). Based on this, it is therefore possible the purified *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 used

in this study would also exhibit better catalytic activity should Mn^{2+} ions be used in the cleavage buffer.

This portion of the study also has the same limitation as that seen with the *E. coli* proteins, in that neither *S. pyogenes* DnaK nor *M. thermautotrophicus* DnaK were tested for their ATPase activity to determine if the proteins were functional *in vitro*. It may therefore be that the DnaK proteins purified for this area of research were not active, and as a result of this they may not have interacted optimally with the *S. pyogenes* Cas1 and *M. thermautotrophicus* Cas1 used in this study. Future work would therefore require determination of whether the DnaK proteins used for this study were active to ensure that the results gathered were reliable, and if so to utilise Cas1-DnaK assays both with and without ATP.

Although the presence of DnaK did alter Cas1 binding to DNA with both *S. pyogenes* and *M. thermautotrophicus* proteins in this study, inhibition of DNA binding was not seen. It is difficult to know whether this is caused by DnaK acting as a chaperone protein as normal and removing protein aggregates, or whether it is caused by a true inhibitory interaction between Cas1 and DnaK. To establish the nature of the interaction further, it would also be necessary to test for a physical interaction between *S. pyogenes* and *M. thermautotrophicus* Cas1 and DnaK, using similar methods to those performed and described for *E. coli* proteins. Pull down assays would be essential to determine whether there is a physical reaction between the two proteins *in vivo*, and BiFC and BioID2 could also be used to determine whether the proteins interact with each other. This would give better insight into whether the structures predicted by Alphafold 3 were accurate, and therefore further insight into whether the interaction between Cas1 and DnaK is conserved across all CRISPR systems. To conclude, more detailed investigations on the interaction between these proteins across species needs to be performed in future work.

Chapter 6 – References

Anderson, D. G., & Kowalczykowski, S. C. (1997). The Translocating RecBCD Enzyme Stimulates Recombination by Directing RecA Protein onto ssDNA in a χ-Regulated Manner. *Cell*, *90*(1), 77–86. https://doi.org/10.1016/S0092-8674(00)80315-3

Barrangou, R., & Marraffini, L. A. (2014). CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell*, *54*(2), 234–244. https://doi.org/10.1016/j.molcel.2014.03.011

Bertelsen, E. B., Chang, L., Gestwicki, J. E., & Zuiderweg, E. R. P. (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8471–8476. https://doi.org/10.1073/pnas.0903503106

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V, & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Calloni, G., Chen, T., Schermann, S. M., Chang, H.-C., Genevaux, P., Agostini, F., Tartaglia, G. G., Hayer-Hartl, M., & Hartl, F. U. (2012). DnaK functions as a central hub in the E. coli chaperone network. *Cell Reports*, *1*(3), 251–264. https://doi.org/10.1016/j.celrep.2011.12.007

Cass, S. D. B., Haas, K. A., Stoll, B., Alkhnbashi, O. S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A., & Bolt, E. L. (2015). The role of Cas8 in type I CRISPR interference. *Bioscience Reports*, *35*(3). https://doi.org/10.1042/BSR20150043

Dillingham, M. S., & Kowalczykowski, S. C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and Molecular Biology Reviews : MMBR*, *72*(4), 642–671, Table of Contents. https://doi.org/10.1128/MMBR.00020-08

Dixon, D. A., & Kowalczykowski, S. C. (1993). The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell*, *73*(1), 87–96. https://doi.org/10.1016/0092-8674(93)90162-j

Garrett, S. C. (2021). Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array. *Frontiers in Microbiology*, *12*, 664299. https://doi.org/10.3389/fmicb.2021.664299

He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., Yang, Q., Wu, Y., Zhao, X., Chen, S., Jia, R., Zhang, S., Liu, Y., Yu, Y., & Zhang, L. (2018). Cas1 and Cas2 From the Type II-C CRISPR-Cas System of Riemerella anatipestifer Are Required for Spacer Acquisition. *Frontiers in Cellular and Infection Microbiology*, *8*, 195. https://doi.org/10.3389/fcimb.2018.00195

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1707). https://doi.org/10.1098/rstb.2015.0496

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & 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*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Ivančić-Baće, I., Cass, S. D., Wearne, S. J., & Bolt, E. L. (2015). Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Research*, *43*(22), 10821–10830. https://doi.org/10.1093/nar/gkv1213

Jansen, R., van Embden, J. DA, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575.

Ka, D., Lee, H., Jung, Y.-D., Kim, K., Seok, C., Suh, N., & Bae, E. (2016). Crystal Structure of Streptococcus pyogenes Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure (London, England : 1993)*, *24*(1), 70–79. https://doi.org/10.1016/j.str.2015.10.019

Kerppola, T. K. (2008). Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics*, *37*, 465–487. https://doi.org/10.1146/annurev.biophys.37.032807.125842

Killelea, T., & Bolt, E. L. (2017). CRISPR-Cas adaptive immunity and the three Rs. *Bioscience Reports*, *37*(4). https://doi.org/10.1042/BSR20160297

Killelea, T., Dimude, J. U., He, L., Stewart, A. L., Kemm, F. E., Radovčić, M., Ivančić-Baće, I., Rudolph, C. J., & Bolt, E. L. (2023). Cas1-Cas2 physically and functionally interacts with DnaK to modulate CRISPR Adaptation. *Nucleic Acids Research*, *51*(13), 6914–6926. https://doi.org/10.1093/nar/gkad473

Killelea, T., Kemm, F. E., He, L., Rudolph, C. J., & Bolt, E. L. (2024). Repurposing Proximity-Dependent Protein Labeling (BioID2) for Protein Interaction Mapping in E. coli. *Methods in Molecular Biology (Clifton, N.J.)*, *2828*, 87–106. https://doi.org/10.1007/978-1-0716-4023-4_9

Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Hartl, F. U. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry*, *82*, 323–355. https://doi.org/10.1146/annurev-biochem-060208-092442

Koonin, E. V, Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Lau, C. H., Reeves, R., & Bolt, E. L. (2019). Adaptation processes that build CRISPR immunity: creative destruction, updated. *Essays in Biochemistry*, *63*(2), 227–235. https://doi.org/10.1042/EBC20180073

Le Rhun, A., Escalera-Maurer, A., Bratovič, M., & Charpentier, E. (2019). CRISPR-Cas in Streptococcus pyogenes. *RNA Biology*, *16*(4), 380–389. https://doi.org/10.1080/15476286.2019.1582974

Lee, H., Dhingra, Y., & Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *ELife*, *8*. https://doi.org/10.7554/eLife.44248

Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Liao, C., Sharma, S., Svensson, S. L., Kibe, A., Weinberg, Z., Alkhnbashi, O. S., Bischler, T., Backofen, R., Caliskan, N., Sharma, C. M., & Beisel, C. L. (2022). Spacer prioritization in CRISPR-Cas9 immunity is enabled by the leader RNA. *Nature Microbiology*, *7*(4), 530–541. https://doi.org/10.1038/s41564-022-01074-3

Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods in Molecular Biology (Clifton, N.J.)*, *1311*, 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4

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., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews. Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *The CRISPR Journal*, *1*(5), 325–336. https://doi.org/10.1089/crispr.2018.0033

McGinn, J., & Marraffini, L. A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Molecular Cell*, *64*(3), 616–623. https://doi.org/10.1016/j.molcel.2016.08.038

McGinn, J., & Marraffini, L. A. (2019). Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nature Reviews. Microbiology*, *17*(1), 7–12. https://doi.org/10.1038/s41579-018-0071-7

Mitić, D., Bolt, E. L., & Ivančić-Baće, I. (2023). CRISPR-Cas adaptation in Escherichia coli. *Bioscience Reports*, *43*(3). https://doi.org/10.1042/BSR20221198

Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, *36*(1), 244–246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica, F. J., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*, *62*(6), 824–833. https://doi.org/10.1016/j.molcel.2016.04.027

Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*, *527*(7579), 535–538. https://doi.org/10.1038/nature15760

Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015b). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*, *527*(7579), 535–538. https://doi.org/10.1038/nature15760

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V, Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

Nuñez, J. K., Lee, A. S. Y., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, *519*(7542), 193–198. https://doi.org/10.1038/nature14237

Radovcic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E. L., & 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 Research*, *46*(19), 10173–10183. https://doi.org/10.1093/nar/gky799

Ramachandran, A., Summerville, L., Learn, B. A., DeBell, L., & Bailey, S. (2020). Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. *The Journal of Biological Chemistry*, *295*(11), 3403–3414. https://doi.org/10.1074/jbc.RA119.012196

Richter, H., Rompf, J., Wiegel, J., Rau, K., & Randau, L. (2017). Fragmentation of the CRISPR-Cas Type I-B signature protein Cas8b. *Biochimica et Biophysica Acta. General Subjects*, *1861*(11 Pt B), 2993–3000. https://doi.org/10.1016/j.bbagen.2017.02.026

Rollie, C., Graham, S., Rouillon, C., & White, M. F. (2018). Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Research*, *46*(3), 1007–1020. https://doi.org/10.1093/nar/gkx1232

Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *ELife*, *4*. https://doi.org/10.7554/eLife.08716

Rüdiger, S., Germeroth, L., Schneider-Mergener, J., & Bukau, B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO Journal*, *16*(7), 1501–1507. https://doi.org/10.1093/emboj/16.7.1501

Shah, S. A., Erdmann, S., Mojica, F. J. M., & Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biology*, *10*(5), 891–899. https://doi.org/10.4161/rna.23764

Shiimori, M., Garrett, S. C., Graveley, B. R., & Terns, M. P. (2018). Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Molecular Cell*, *70*(5), 814-824.e6. https://doi.org/10.1016/j.molcel.2018.05.002

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., Severinov, K., Zhang, F., & Koonin, E. V. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. *Nature Reviews. Microbiology*, *15*(3), 169–182. https://doi.org/10.1038/nrmicro.2016.184

Vabulas, R. M., Raychaudhuri, S., Hayer-Hartl, M., & Hartl, F. U. (2010). Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harbor Perspectives in Biology*, *2*(12), a004390. https://doi.org/10.1101/cshperspect.a004390

van Soolingen, D., de Haas, P. E., Hermans, P. W., Groenen, P. M., & van Embden, J. D. (1993). Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, *31*(8), 1987–1995. https://doi.org/10.1128/jcm.31.8.1987-1995.1993

Wang, J., Ma, J., Cheng, Z., Meng, X., You, L., Wang, M., Zhang, X., & Wang, Y. (2016). A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. *Cell Research*, *26*(10), 1165–1168. https://doi.org/10.1038/cr.2016.103

Wang, W., Liu, Q., Liu, Q., & Hendrickson, W. A. (2021). Conformational equilibria in allosteric control of Hsp70 chaperones. *Molecular Cell*, *81*(19), 3919-3933.e7. https://doi.org/10.1016/j.molcel.2021.07.039

Wimmer, F., & Beisel, C. L. (2020). CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.03078

Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*, *550*(7674), 137–141. https://doi.org/10.1038/nature24020

Yosef, I., Mahata, T., Goren, M. G., Degany, O. J., Ben-Shem, A., & Qimron, U. (2023). Highly active CRISPR-adaptation proteins revealed by a robust enrichment technology. *Nucleic Acids Research*, *51*(14), 7552–7562. https://doi.org/10.1093/nar/gkad510

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., & Hendrickson, W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science (New York, N.Y.)*, *272*(5268), 1606–1614. https://doi.org/10.1126/science.272.5268.1606

Chapter 7 – Supplementary Information

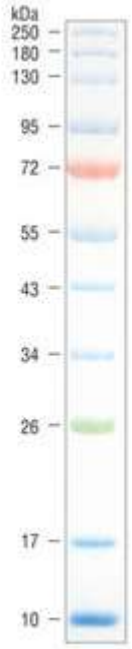


Figure 7.2 - Colour Protein Standard (NEB). *Molecular marker shows bands at corresponding sizes of protein when ran on an SDS PAGE gel.*

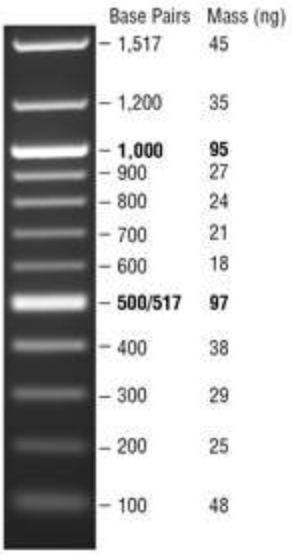


Figure 7.1 - 100bp DNA Ladder (NEB). *Molecular marker shows bands at corresponding sizes of DNA when ran on an agarose gel.*