DNA Repair in Antarctic Bacteria

Characterisation of Two Nuclease Proteins from Antarctic Dry Valley Metagenomes

A thesis

submitted in fulfilment

of the requirements for the degree

of

Masters (Research) [School of Science]

at

The University of Waikato

by

RONJA STELZER

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

ABSTRACT

DNA repair processes are crucial for the growth and survival of all organisms. This is especially true for organisms, that inhabit extreme habitats, like the Antarctic Dry Valleys. The Dry Valleys are one of the coldest and driest desert environments on Earth. The conditions in the Dry Valleys, including high levels of UV radiation, low temperatures, multiple daily freeze-thaw cycles, desiccation, and low nutrient levels, are highly damaging to the genomic DNA of the organisms inhabiting them. The Dry Valleys are dominated by microorganisms, which have a range of adaptations to survive in this extreme environment. Previous studies focus largely on the metabolic adaptation of Dry Valley inhabitants. In this thesis, the DNA repair machinery of bacteria in the Dry Valleys was examined to determine how organisms survive the DNA-damaging conditions.

Our understanding of DNA repair and replication in bacteria pathways is largely based on studies of isolated organisms and pathogens, which have been extensively studied. Methods such as metagenomic sequencing allow us to find novel proteins from currently unculturable organisms. In a search for novel DNA repair proteins in bacterial metagenomes from the Dry Valleys, two unique nuclease proteins were chosen for structural and biochemical characterisation. One of these nuclease proteins belongs to the currently uncharacterised protein domain UPF0102. This protein is distantly related to the archaeal Holliday junction resolvases and is a Type II restriction endonuclease type protein. The activity of DV-Hjc and the UPF0102 domain was characterised *in vitro* using biochemical assays and *in vivo* in *E. coli* knockout cells. Biochemical characterisation of the protein showed that the protein has strong binding affinity with double-stranded DNA substrates and Holliday-junction mimicking substrates. The structure of the protein was determined via X-ray crystallography at a resolution of 0.9 Å and structural features were related to its activity, revealing an extensive DNA binding surface and an active site similar to that of archaeal Holliday junction resolvases. The second nuclease characterised in this thesis is part of a unique nuclease-ligase fusion protein and belongs to the nuclease group of MBL-β-CASP proteins. The biochemical characterisation of the protein showed its ability to bind and cleave DNA, specific activity was observed with abasic site substrates.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr Adele Williamson. I would like to thank you for your continued support throughout my project. Your knowledge and mentorship have guided me from the start of this project and provided invaluable support and insight. Thank you for every opportunity you have given me during this project, including attending three conferences, multiple workshops, and the Australian synchrotron. These experiences have significantly expanded my knowledge and contributed to my development. It has been truly inspiring to be mentored by a woman in STEM, who not only conducts high-level research but also balances this demanding career with her family life. You have been a powerful role model throughout my journey and will continue to be one as I continue my career.

I want to express my deepest gratitude to my co-supervisor, Professor Craig Cary. He is one of the most passionate people I have worked with during my journey on this project and although he is no longer with us, I will continue to be inspired by his work and the impact he made on my future research. His way of thinking outside the box and seeing what others view as limitations as opportunities has left a lasting impression on me and continues to influence my research approach. I am grateful that I had the opportunity to work with an exceptional individual like Craig Cary and hope to one day inspire others as he inspired me.

I would also like to thank my co-supervisor, Dr Elizabeth Rzoska-Smith, with whom I began working on this project in 2021. You taught and guided me during my first lab research experience three years ago, and I have learned so much from you since then. If it wasn't for the knowledge, I have gained from working with you, I would not have finished this project to the extent that I have. Tackling the stubborn and challenging proteins from the Antarctic Dry Valleys wouldn't have been as fun without someone as passionate about them as I was. I would also like to thank you for your support during the writing process of this thesis and other reports on this project. Thank you for always my name first when opportunities came up and believing in my ability to fulfil them. Thank you for making working on this project not feel like work.

Next, I would like to thank Dr Judith Burrows. The positivity and team spirit you bring make this lab what it is. You have a talent for making everyone feel seen and appreciated in our group. Your support during both difficult and happy times has been invaluable, and it has made my experience in the lab much more rewarding and fulfilling. Thank you also for your help with using the equipment; I greatly appreciate your expertise. And most of all, I want to thank you for my pink safety glasses—thank you for making me look fashionable while working!

I would like to thank the members of our research group – Jolyn, Caitlin, Aakash, Chelsea, Avi, Liz and Ally – who have contributed to the great environment of our group. I appreciated the company of all of you, the collaborative atmosphere and mutual support. I would like to specifically thank Jolyn, for your expertise with DNA cloning, which has helped me during multiple steps of this project. To Ally, I would like to thank you for being the friend I needed during the past three years. Thank you for making me laugh when things weren't going as planned and helping me find a solution afterwards. I appreciate your support with the structural characterisation of my proteins. Furthermore, I would like to thank everyone in the C2 lab, who has made my time in the lab so enjoyable.

I would like to thank the University of Waikato for the financial support I have received through the Masters (Research) scholarship.

I would also like to thank Antarctica New Zealand and the New Zealand Post for financial support from the New Zealand Post Antarctic Scholarship. Furthermore, I would like to thank Antarctica New Zealand for the opportunity to attend and present at the New Zealand – Australia Antarctic Science Conference 2023, which allowed me to connect with fellow early career scientists from the Antarctic science research community and meet experts contributing to the field of Antarctic research.

To conclude, I want to thank my family and friends for their support during my master's research. A special thanks to my mum for setting me up in the best way possible and for your unwavering support from the other side of the world. Your encouragement and belief in me have been a great help.

Lastly, I would like to thank Rudi for reading all my research and looking at all my figures pretending to be interested. Your enthusiasm for everything is do means a lot to me and has helped me throughout my BSc and MSc. Thank you for forcing me to have a work-life balance when I needed it and for supporting me in every possible way during the busy times of my degree. Thank you for always believing in me. I could not have done this without you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 CHAPTER ONE: INTRODUCTION

1.1 DNA

Genomic DNA (deoxyribonucleic acid) carries the genetic or hereditary information of living organisms from all domains of life. Therefore, its integrity is essential to the survival and proliferation of all organisms.

DNA molecules consist of two polynucleotide chains that run antiparallel to each other forming a double helix (Figure 1, C) (Watson & Crick, 1953). The polynucleotide chains consist of four types of nucleotides: adenine, cytosine, guanine, and thymine (Figure 1, A and B). The four nucleotides consist of a nitrogen-containing base, a phosphate group, and a five-carbon sugar (deoxyribose) (Figure 1, A). In the case of RNA (ribonucleic acid), the five-carbon sugar is ribose. The nitrogenous bases differ between adenine, cytosine, guanine, and thymine. In RNA thymine is replaced by uracil. Cytosine, thymine and uracil are pyrimidine bases, which have a single pyrimidine ring with two nitrogen atoms. Adenine and guanine are purine bases, which consist of a pyrimidine and imidazole ring with four nitrogen atoms. Polynucleotide chains are formed by the nucleotides bound to each other through covalent bonds between the 5'hydroxyl group of the phosphate group of a nucleotide and the 3' hydroxyl group of the sugar of the adjacent nucleotide. Hydrogen bonds between the nitrogenous bases of the nucleotides in the polynucleotide chains hold them together forming the DNA double helix (Figure 1, B and C). Based on canonical Watson–Crick base pairing adenine pairs with thymine and cytosine pairs with guanine in the DNA double helix (Watson & Crick, 1953). The two nucleotide chains run antiparallel to each other so there is a 3'hydroxyl from one strand and a 5'phosphate group from the other strand on each end of a DNA double helix. Because of the hydrogen bonds between the nucleotide bases, the nitrogenous bases face the inside of the double helix, and the phosphate backbone is located on the outside of the helix. The phosphorous groups in the phosphate backbone are negatively charged. This makes the overall charge of DNA double helices, and single DNA strands negative. The two grooves that run along the DNA double helix have different widths, the major grove is 12 Å wide and the minor grove is 6 Å wide. Both, the shape of the DNA helix with the major and minor groove and its charge are essential for the interactions of proteins with the DNA.

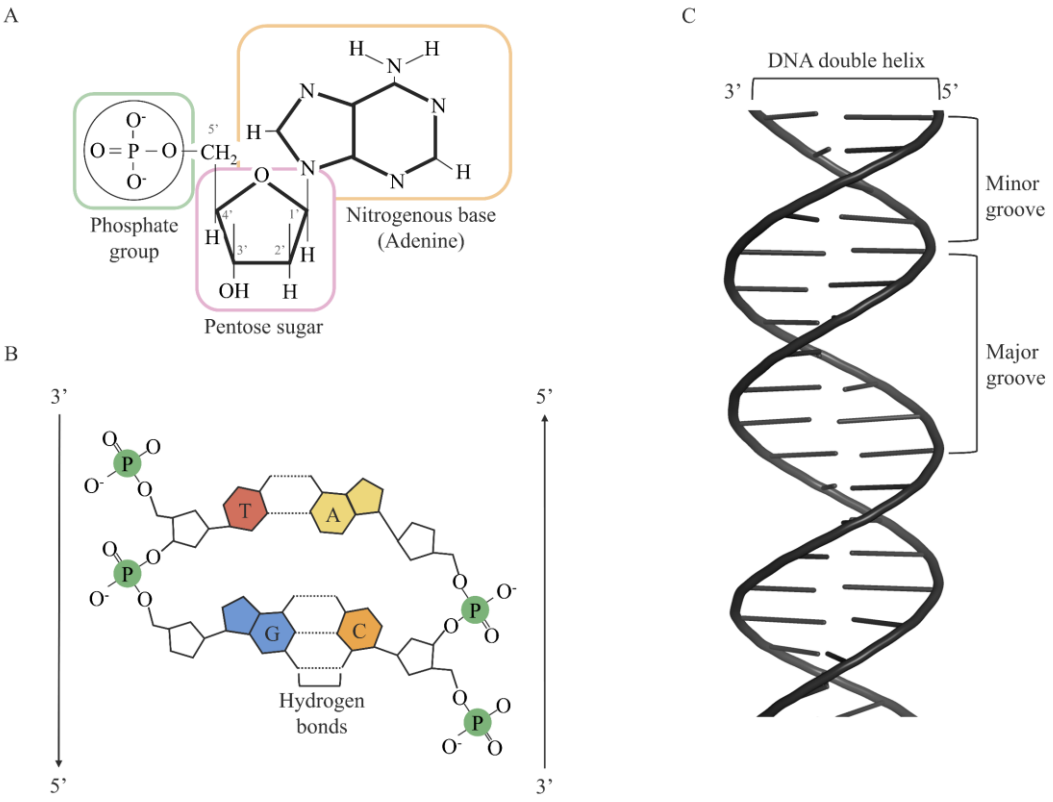


Figure 1: The basic structure of DNA.
***A:** A single nucleotide is shown on the example of adenine with the phosphate group in the green rectangle, the deoxyribose sugar in the pink rectangle, and the nitrogenous bases (adenine) in the orange rectangle. The 1' to 5' carbons are indicated in grey.*
***B:** Four nucleotides forming two base pairs. Thymine (T) and adenine (A) form a base pair, and guanine (G) and cytosine (C) form the second pair. Pairing is facilitated by hydrogen bonds as indicated by the dotted lines between the nucleotide bases. Nucleotides are connected by phosphate bonds between the phosphate groups and sugars of adjacent bases (here: G and T, and A and C), which form the backbone of the DNA.*
***C:** DNA double helix generated in PyMol (Schrödinger), showing the major and minor grove and the 3' and 5' end of the DNA double helix.*

The integrity of the DNA can be compromised by damage to the DNA, which can arise from DNA damaging agents, replication errors and spontaneously. If damages are not repaired and replication occurs over the damage it can lead to mutations, which can cause heritable disease. Irreparable damages can also directly lead to cell death due to an inability to replicate DNA. Most damages, however, are repaired through different DNA repair mechanisms in the cell so that the genetic integrity is restored (Figure 2).

The effects of mutations in the genome vastly range from rare beneficial ones which give a fitness advantage, to mutations with no effects, and finally to mutations causing disorders or disease-causing in multicellular organisms, or that are lethal in an individual cell. The effects of a mutation depend on its location in the genome. Mutations located within a gene, in the regulatory region of a gene or regions coding for non-translated RNAs have the greatest effects on the organism. Mutations in the regulatory

region of a gene can affect the gene's expression level (Krokan et al., 1997). This can be observed in bacterial strains under stress from antibiotic treatment, where mutations in the regulatory region of a gene coding for antibiotic-resistance genes cause upregulation of its expression leading to increased antibiotic resistance (McCallum et al., 2010). Mutations located within the gene, i.e. the protein-coding sequence, will change the gene's transcript and more importantly it may change the protein's amino acid sequence. Changes in the protein sequence are rare, but they are the underlying mechanism in the evolution of proteins. The change of a single amino acid in the sequence of a protein may not affect the protein, depending on the position of the amino acid in the protein structure or the amino acid change. For *Escherichia coli* a rate of beneficial mutations in the order of 10^{-9} per cell and a rate of deleterious mutations in the order of 10^{-4} per cell (Imhof & Schlötterer, 2001; Kibota & Lynch, 1996). While these mutation rates do not reflect the mutation rates of cells in their natural habitat, they give an idea of the ratio of deleterious to beneficial mutations, highlighting the importance of effective DNA replication and repair machinery to reduce the number of mutations in the genome.

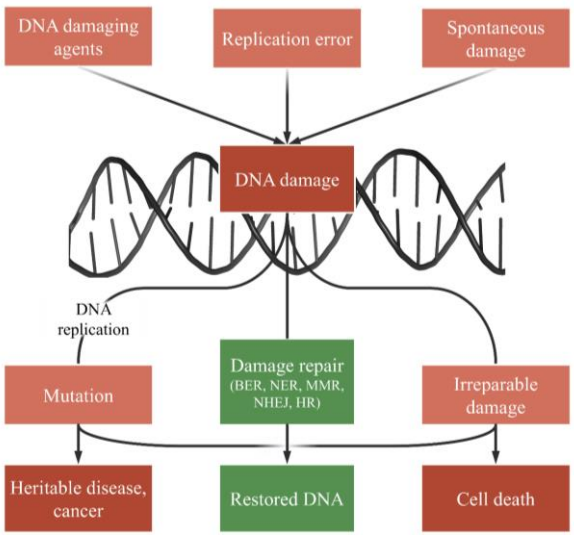


Figure 2: Overview of DNA damage, its repair and possible outcomes.
DNA damage may occur due to DNA damaging errors, replication errors, or spontaneously. If unrepaired and replicated DNA damages can lead to mutations which may cause heritable mutations. Irreparable damage can cause cell death or may lead to mutations. If the DNA damage is repaired through a DNA repair pathway (Base Excision Repair (BER), Nucleotide Excision Repair (NER), MisMatch Repair (MMR), Non-Homologous End Joining (NHEJ) and Homologous Recombination Repair (HR) are named as a few examples) the DNA may be restored.
The figure is inspired by that from Krokan et al. (1997).

1.1.1 DNA Lesions

As any other molecule, DNA can undergo chemical modifications leading to DNA damage. As per the definition of Chakarov et al. (2014) DNA damages are: "Any modifications in the physical and/or chemical structure of DNA resulting in an altered DNA molecule which is different from the original

DNA molecule concerning its physical, chemical and/or structural properties". DNA modifying agents or spontaneous reactions can cause DNA damage. They can change the information encoded in the DNA or even modify it to a point where replication and translation are inhibited. Lesions may be limited to one of the DNA strands of a double-helix or include both strands. The most harmful type of DNA lesion is a double-stranded DNA break, which can be caused by breaking of the phosphodiester backbone on both strands of the DNA double helix (Figure 3, A). If unrepaired, double-strand breaks can lead to translocations, loss of chromosomal material, and cell death (Chapman et al., 2012; Kaina, 2003; Lieber, 1998).

Interstrand cross-links can form as a result of a DNA modifying agent, that produces covalent adducts on the nucleotide bases on both strands of the DNA (Figure 3, D). These bases then form a covalent bond and cross-link, resulting in the inability of the strands to be separated (Noll et al., 2006). Cross-links may also form between bases on the same strand (Figure 3, F). Up to 90% of such lesions occur between guanine residues or adenine and guanine residues (Noll et al., 2006). If unrepaired inter- and intra-strand cross-links, interfere with or block genome replication and translation; thus, they can lead to mutations and cell death (Dronkert & Kanaar, 2001; Noll et al., 2006).

Single-stranded DNA breaks are much less harmful than double-stranded lesions since a complete template strand remains intact. They are a common intermediate during DNA repair processes, but can lead to double-stranded breaks if they are not repaired (Figure 3, B) (Abbotts & Wilson, 2017). Abasic (apurinic or apyrimidinic) sites are the most commonly occurring single-stranded lesions in cells (Figure 3, E). The coding nucleotide base part of the DNA is missing in abasic sites, which can interfere with DNA replication and translation by blocking DNA polymerase and RNA polymerase, respectively (Abbotts & Wilson, 2017). If polymerisation is carried out over an abasic site by a DNA polymerase the wrong base may be incorporated since the template is missing for the base at the abasic site, which is a source of DNA mismatches. DNA mismatches can generally occur through the misincorporation of bases during DNA replication (Figure 3, G, i) or deamination of cytosine bases, causing the formation of uracil (Figure 3, G, ii). If not repaired, mismatches can change the information encoded and lead to genome mutations. Nucleotide modifications, which are often caused by alkylating and oxidating agents, can also lead to the misincorporation of bases during replication causing a mismatch (Figure 3, H). These agents can also modify the phosphodiester backbone of the DNA. The causes of DNA modifications will be discussed in more detail below.

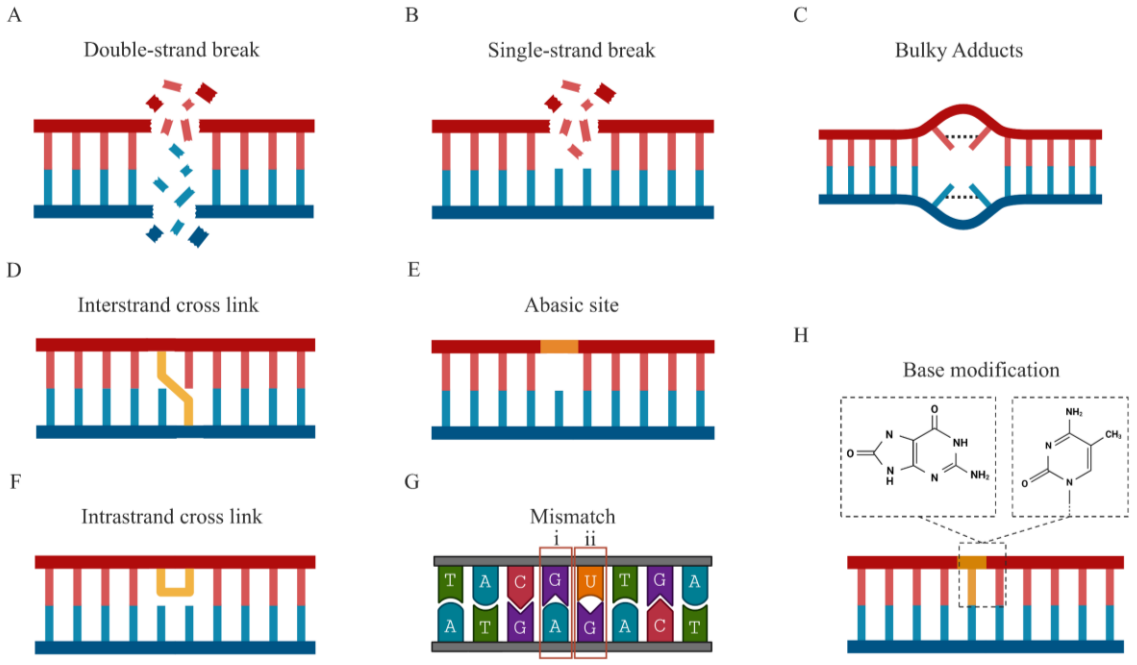


Figure 3: Illustration of common DNA lesions.
***A:** Double-stranded break. **B:** Single-stranded break. **C:** Bulky adducts. **D:** Interstrand cross link. **E:** Abasic site. **F:** Interstrand cross link. **G:** Base mismatch; **i:** guanine-adenine mismatch; **ii:** Uracil Mismatch. **H:** Base modifications shown on the example of 8-oxo guanine (left) and methylated cytosine (right). (Created with BioRender.com)*

1.1.2 DNA Damaging Agents

DNA modifications can be caused by exogenous and endogenous factors; exogenous factors originate from outside the organism and endogenous factors originate from within the organism due to cellular functions. While their origin may differ, the agents resulting in DNA damage may be the same for exogenous and endogenous factors. Furthermore, DNA-damaging or genotoxic agents can be divided into physical and chemical agents. Common physical agents are UV light and other ionising radiation, and common chemical agents include oxygen radicals, spontaneous reactions, polycyclic aromatic hydrocarbons, and replication errors (Almeida et al., 2021). These agents can be more or less harmful by causing different DNA damages, as will be discussed in more detail in the following.

Reactive Oxygen Species and Oxidising Agents

The oxygen paradox describes the ambiguity between the absolute requirement of aerobic organisms for oxygen and the threat that oxygen-induced damages pose to those organisms (Davies, 1995). Reactive oxygen species, like the superoxide anion radical ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2) and hydroxyl radical ($^{\cdot}OH$), are a byproduct of aerobic cellular metabolism (Fasnacht & Polacek, 2021). Oxidative stress can also be caused by external stressors, like exposure to ionising radiation, which can deprotonate water to produce hydroxyl radicals (de Jager et al., 2017). Reactive oxygen species are a

major cause of DNA damage and hydroxyl radicals are especially reactive with DNA and pose the largest threat to its integrity (Collins, 2009).

Reactive oxygen species can cause oxidation of the DNA molecule by adding oxygen atoms or removing hydrogen atoms from a DNA component. Oxidation may occur at the nucleotide base or on the DNA backbone. Modified nucleotide bases, abasic sites, single- or double-stranded DNA breaks and intra- and interstrand crosslinks may form as a result of DNA oxidation (Cadet et al., 2017; Cadet & Wagner, 2013; Devasagayam et al., 1991). Over 100 modifications to the nitrogenous bases and the deoxyribose sugars of nucleotides have been observed to date caused by oxidation (Cadet et al., 2012). The best-studied oxidative base damage is 7,8-dihydro-8-oxoguanine (8-oxo-G), which is often used as a biological marker for oxidative stress in cells (Figure 4, A) (Shih et al., 2019). Due to its low oxidation potential guanine is especially vulnerable to oxidative damage (Neeley & Essigmann, 2006). In the case of 8-oxo-G, the oxidative damage results from the addition of an oxygen-containing group at C8 and a proton at N7 of the nucleotide base (van Loon et al., 2010). which if unrepaired can lead to G·C→T·A transversion mutations (Figure 4, A) (Grollman & Moriya, 1993). This is caused by the erroneous incorporation of adenine opposite of 8-oxo-G by DNA polymerases during the DNA replication, as illustrated in Figure 4 (B).

Repair machinery that directly reverses oxidised nucleotides exists in most cells, for example, in humans, the oxidised purine nucleoside triphosphatase, MTH1, hydrolyses a range of oxidised deoxynucleotide triphosphates to prevent their incorporation into a newly synthesised DNA strand (Nakabeppu et al., 2006). The mammalian 8-oxo-dG DNA glycosylase (OGG1) recognises an 8-oxo-dG in DNA strands opposite of cytosine and it cleaves the bond between the base and sugar of the 8-oxo-dG to produce an abasic site that is consequently repaired via the base excision repair pathway (Krokan & Bjørås, 2013; Rosenquist et al., 1997).

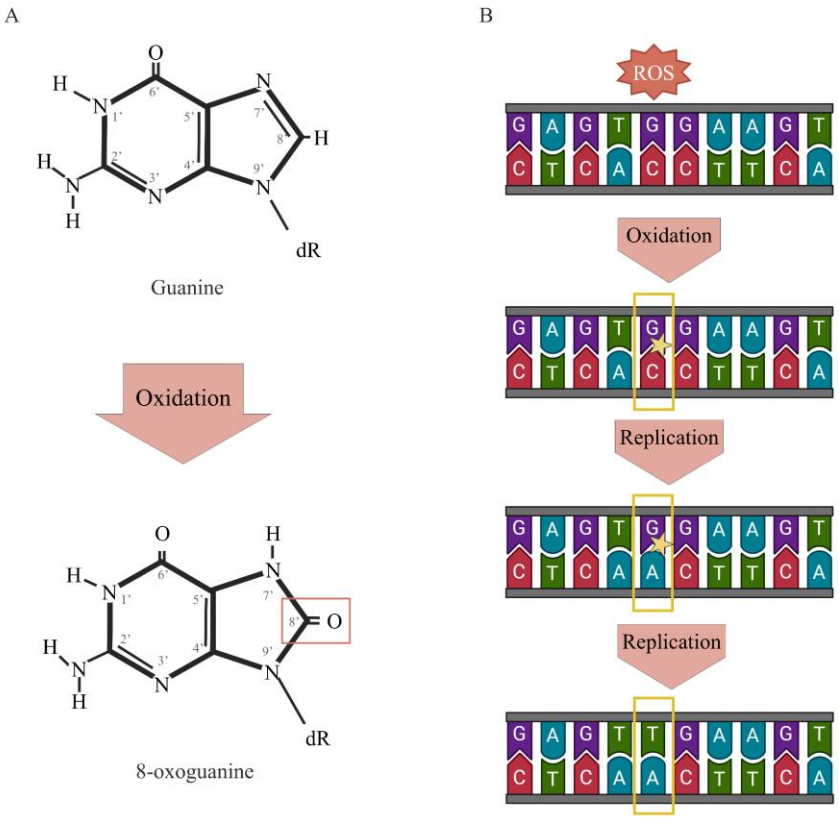


Figure 4: Illustration of possible effects of reactive oxygen species on DNA.
A: *Illustration of possible damage caused by reactive oxygen species on guanine. Guanine is oxidised to 8-oxo-guanine through oxidation of the 8' carbon as highlighted by the orange rectangle.*
B: *Possible outcomes of the oxidation of guanine to 8-oxoguanine. Reactive oxygen species (ROS) caused guanine oxidation to 8-oxo-guanine (indicated by the yellow star). Replication of 8-oxo-guanine leads to misincorporation of adenine opposite the 8-oxo-guanine. In the following round of replication, thymine will be incorporated in the place of the 8-oxoguanine leading to a G·C to T·A transversion mutation. (Created with BioRender.com)*

Alkylating/Methylating Agents

Alkylating agents are electrophilic compounds that can introduce methyl or ethyl groups to the nitrogenous bases or the phosphodiester backbone through an attack of their nitrogen and oxygen atoms (Couvé et al., 2013; Mielecki et al., 2015). Methyl and ethyl groups are covalently bound to the DNA resulting from an attack by alkylating agents. Alkylation of the bases or phosphodiester backbone can have mutagenic effects if it causes incorporation of the wrong base opposite the alkylated base or cytotoxic effects if it disrupts DNA replication. There are exogenous and endogenous sources for alkylating agents (Mielecki et al., 2015).

The most common nucleotide base lesions caused by alkylation are 7-methylguanine, 3-methyladenine, and O^6-methylguanine (Figure 5) (Couvé et al., 2013). If an unrepaired O^6-methylguanine provides the template for the human pol β thymine is preferably incorporated into the newly synthesised strand (Singh et al., 1996). Therefore O^6-methylguanine is likely to cause a G → A transition mutations (Singh et al., 1996). The same can be observed in prokaryotic cells (Wang & Wang, 2018). The DNA damage

caused by oxygen-alkylation can also lead to apoptosis and cell death, for example, due to replication stalling at an O^6-methylguanine (Groth et al., 2010; Meikrantz et al., 1998).

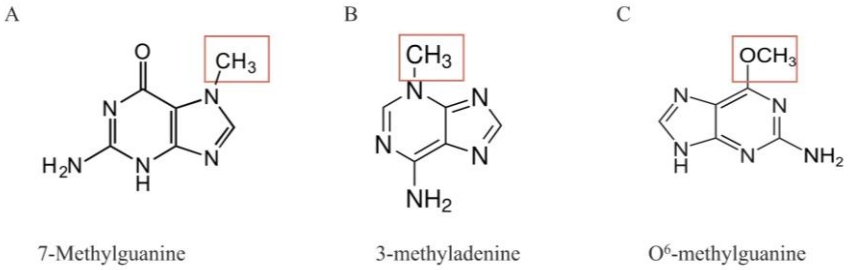


Figure 5: Methylated bases.
***A:** 7'nitrogen methylated guanine. **B:** 3'nitrogen methylated adenine. **C:** 6'oxygen methylated guanine.*

UV Radiation

DNA is the main cellular target of UV light. The longer wavelength UV radiation UV-A (315-400 nm) and UV-B (280-315 nm) are more damaging to DNA than the shorter wavelength UV-C radiation (<280 nm). While shorter wavelength UV is more damaging, less UV-C reaches DNA due to it being absorbed by oxygen and the ozone layer in the atmosphere, making UV-B radiation the main UV radiation causing DNA damage.

UV light is absorbed by DNA and can lead to the binding of carbon atoms of two pyrimidine bases, as shown in Figure 6. The most common UV photoproducts are cyclobutene-pyrimidine dimers followed by 6-4 photoproducts (Figure 6) (Goosen & Moolenaar, 2008; Kciuk et al., 2020). 6-4 photoproducts in double-stranded DNA can interact with the opposite strand and cause single-stranded breaks (Banaś et al., 2020). Both, cyclobutene-pyrimidine dimers and 6-4 photoproducts can block replication of the DNA strand by DNA polymerases (Chan et al., 1985). Therefore, they can lead to double-stranded break formation and cell death if replication is blocked. Additionally, UV radiation can cause DNA damage indirectly through the production of reactive oxygen species.

Organisms have evolved a range of enzymes and repair pathways to repair damages caused by UV radiation. Enzymes specialised for the repair of UV-induced damages include photolyases, which directly reverse the covalent link between two pyrimidines of pyrimidine dimers, pyrimidine dimer-DNA glycosylases, which remove the damaged nucleotide base from the deoxyribose sugar followed by its removal through base excision repair, and UV-damage endonucleases, which recognise both cyclobutene-pyrimidine dimers and 6-4 photoproducts and cuts 5' to the damage (reviewed in Goosen and Moolenaar (2008)).

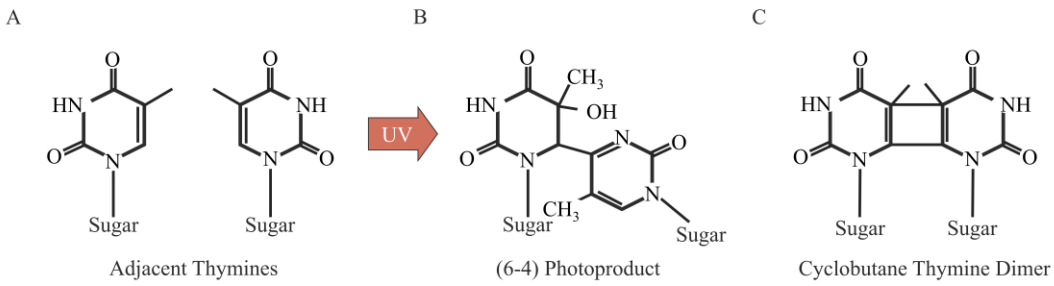


Figure 6: The effect of UV radiation on adjacent thymine bases.
*(**A**) leading to the formation of (6-4) photoproducts (**B**) and cyclobutane thymine dimers (**C**).*

Ionising Radiation

Ionising radiation is a form of radiation, that carries enough energy to free an electron from the orbits of an atom. This can cause breaking of the bonds in a molecule, including the disruption of covalent bonds (Borrego-Soto et al., 2015). Short wavelength photons, including ultraviolet rays, X-rays, and gamma rays, carry enough energy to disrupt chemical bonds. Gamma-ray photons have the shortest wavelength and therefore, the highest amount of energy, making them the most damaging type of radiation. Other common types of ionising radiation are alpha, and beta, which mainly originate from decaying radioactive elements. Living organisms are exposed to low levels of background ionising radiation from the environment, that originates from natural radioactivity in rocks and soils.

Ionising radiation, especially in high doses, can damage the DNA. 80% of this is caused by the production of reactive oxygen species, like hydroxyl radicals (OH), which can be produced by the ionisation of water molecules (Confalonieri & Sommer, 2011). These reactive oxygen species can lead to a variety of DNA lesions as described above. Approximately 20% of DNA lesions caused by ionising radiation are caused by the radiation itself. The transfer of high-energy particles or photons onto the DNA molecule breaks bonds in the DNA. This can generate single- and double-stranded DNA breaks if the ionising radiation breaks bonds in the phosphodiester backbone of the DNA (Borrego-Soto et al., 2015; Cannan & Pederson, 2016; Enderle et al., 2019). Ionising radiation can also induce interstrand cross-link formation. This occurs specifically in mismatched regions of the DNA (Dextraze et al., 2010). By producing reactive oxygen species and causing DNA damage, ionising radiation can cause cell death and is even used for sterilisation purposes (Yusof, 2018).

The bacterium *Deinococcus radiodurans* is known for its resistance to high levels of ionising radiation. It can survive an ionising radiation dose of up to 5 kGy, which is over 1000 times the lethal dose for humans (Byrne et al., 2014). *D. radiodurans* was discovered in a can of ground meat, which was treated with 4 kGy for sterilisation. The ionising radiation resistance of *D. radiodurans* is partially due to protection against the effects of reactive oxygen species and partially due to the DNA repair machinery

of the organism (Confalonieri & Sommer, 2011). This involves using adapted enzymes in conventional DNA repair pathways (Hassan & Gupta, 2018). How *D. radiodurans* survives these high ionisation levels and the associated DNA and protein damage is still not fully understood.

Spontaneous Hydrolysis

DNA is a biological macromolecule in an aqueous solution and as such, it can undergo spontaneous hydrolysis. This can lead to either the loss of an amine group from the nucleotide bases called deamination or the formation of an abasic (apurinic and apyrimidinic) site through spontaneous hydrolysis of the N-glycosylic bond. Abasic sites are one of the most common DNA lesions with around 10,000 to 20,000 per human cell per day (Thompson & Cortez, 2020). Additionally to spontaneous hydrolysis, abasic sites are also formed as intermediates during multiple repair processes.

Three of the four bases in DNA – cytosine, adenine, and guanine – have an amino group ($-NH_2$) and they can lose this amino group, which is called deamination. Deamination can occur spontaneously but is enhanced by the interaction of DNA with DNA-damaging agents, leading to the formation of uracil, hypoxanthine, and xanthine from cytosine, adenine, and guanine, respectively (Figure 7). These deaminated bases are highly mutagenic. In the case of adenine, deamination and formation of hypoxanthine often leads to A·T → G·C transition mutations since hypoxanthine preferably binds to cytosine (Hill-Perkins et al., 1986). Deamination of adenine can also lead to other mispairing since hypoxanthine can hydrogen bond with adenine, cytosine and thymine in DNA (Hill-Perkins et al., 1986). Deamination of guanine causes the formation of xanthine which preferably binds thymine, causing G·C → A·T transition mutations.

Deamination of cytosine to uracil occurs more frequently compared to the deamination of adenine or guanine. Cytosine deamination forms uracil, which leads to G·C → A·T transition mutations (Kow, 2002). Uracil nucleotides present in the DNA are detected by general glycosylases or the uracil-specific uracil-DNA glycosylase, which remove the uracil base to create an apurinic/apyrimidinic site and initiate the base-excision pathway (Yonekura et al., 2009). The main repair pathways for deaminated bases are base excision repair, initiation by the recognition of the lesion by glycosylases, and alternative excision repair, initiated through recognition of the lesion by endonuclease V (Kow, 2002).

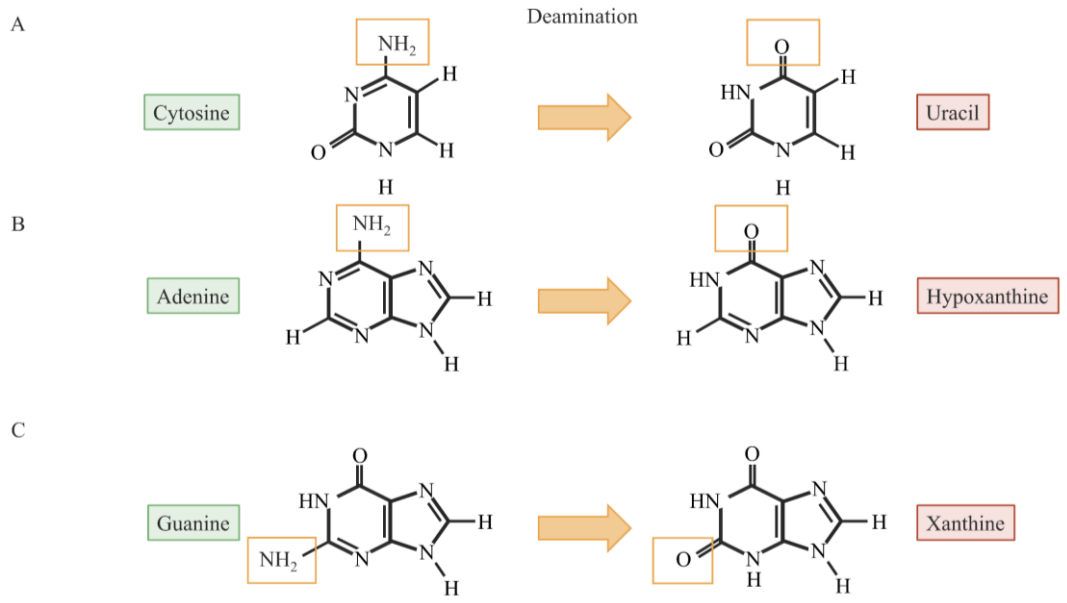


Figure 7: Deamination of three nitrogenous bases.
***A:** Cytosine to uracil, **B:** Adenine to hypoxanthine, and **C:** Guanine to xanthine. The place of deamination is indicated by the orange rectangle.*

Replication Errors

Errors during DNA replication may occur due to base modifications caused by DNA modifying agents as discussed above, but they may also occur spontaneously due to the incorporation of the wrong nucleotide by DNA polymerases. The role of DNA polymerases is to synthesise new DNA strands from a template strand by incorporating the correct nucleotide as per the Watson and Crick base pairing rules. Most replicative polymerases have very high fidelity and only incorporate the wrong base once per every 10^8 to 10^{10} nucleotides polymerised. This high fidelity is partially due to their proofreading ability (Fijalkowska et al., 2012).

Replication errors may also lead to the formation of double-stranded breaks. When a polymerase encounters a DNA lesion during DNA replication, like bulky lesions, oxidative lesions, abasic sites, inter- or intra-strand cross-links and single-stranded breaks, it can lead to stalling of the replication causing collapse of the replication fork (Cannan & Pederson, 2016; Pfeiffer et al., 2000). When this occurs, the leading strand may anneal to the lagging strand forming a Holliday-junction-like structure, called a ‘chicken foot’. The ‘chicken foot’ can be cleaved by Holliday junction resolving enzymes creating double-stranded breaks (Cannan & Pederson, 2016). Alternatively, double-stranded breaks can be caused by defects in the maturation of Okazaki fragments (Pfeiffer et al., 2000).

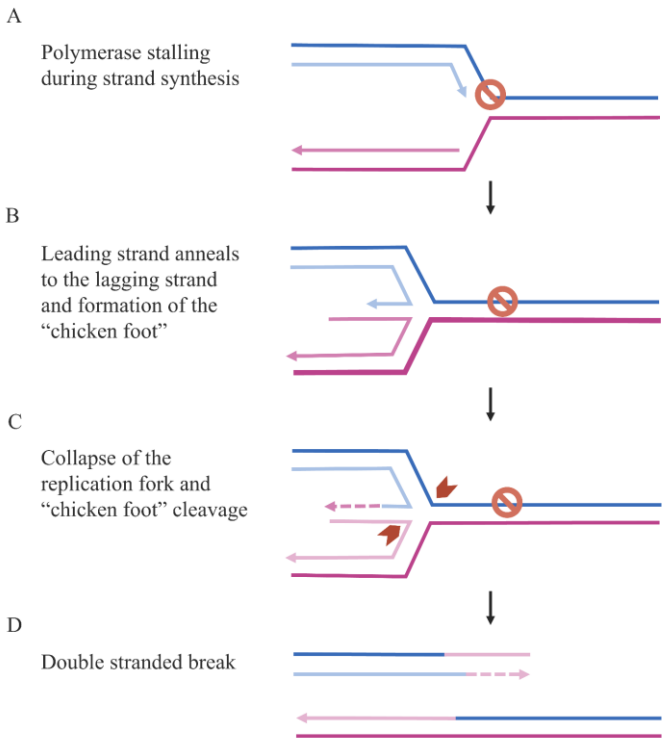


Figure 8: Creation of double-stranded DNA breaks by stalling of the replication fork.
***A:** The replication fork stalls due to a polymerase encountering a damaged DNA base. Stalling is indicated by the stop sign.*
***B:** Annealing of the leading and lagging strand forming a Holliday-junction-like DNA conformation called a “chicken foot”.*
***C:** DNA replication may continue from the annealed leading and lagging strand until the end is reached. The “chicken food” is cleaved by a Holliday junction resolving enzyme (indicated by the red arrows).*
***D:** The two double strands dissociate, and a double-stranded break is present on the upper double-strand. This figure was adapted from Cannan and Pederson (2016)*

1.1.3 DNA Lesion Repair

The DNA lesions described above threaten the integrity of an organism’s genome. Therefore, organisms have developed a range of DNA repair pathways to repair those lesions and protect the information contained in the genome. Some of the best-studied repair pathways are base excision repair, nucleotide excision repair, mismatch repair, non-homologous end joining and homologous recombination, which will be described below. The repair pathway chosen for repair depends mainly on the type of damage.

Direct Reversal

Certain types of DNA damage, specifically DNA lesions resulting from alkylation, UV and cross-linking agents, may be directly reversed through direct reversal repair. This is the simplest DNA repair pathway, as it does not involve incision of the DNA backbone and synthesis of a new DNA strand, making it an error-free repair pathway. A single enzyme usually carries out the direct reversal of DNA adducts. Photolyases, O^6-alkylguanine-DNA alkyltransferases and AlkB family dioxygenases are the major groups of enzymes known to be involved in direct reversal (Figure 9).

For example, photolyases can directly remove photoproducts like cyclobutane pyrimidine dimers and pyrimidine pyrimidones (6–4) photoproducts, as shown in Figure 9 (A). Cyclobutane pyrimidine dimer (CPD) photolyases use light and $FADH^-$ to repair cyclobutane pyrimidine dimers, as shown in Figure 9 (A) (Brettel & Byrdin, 2010).

The alkylation of nucleotide bases can be directly reversed through the activity of alkyltransferases. In bacteria DNA damage by alkylating agents triggers the Ada response. The Ada response causes increased expression of four proteins, Ada (methyltransferase), alkB (dioxygenase), alkA (DNA glycosylase) and aidB (Nieminuszczy & Grzesiuk, 2007). Ada is an O^6-methylguanine-DNA-methyltransferase and is part of the O^6-alkylguanine-DNA alkyltransferase group. Ada removes the methyl group from the methylated bases O^6-methylguanine and O^4-methylthymine (Figure 9, B) (Nieminuszczy & Grzesiuk, 2007; Sedgwick & Lindahl, 2002). Ada removes the methyl group from the guanine by transferring it from the O^6 of the base to a conserved cysteine residue in its active site (Tano et al., 1990; Yu et al., 2019). O^6-methylguanine preferably pairs with thymine during DNA replication, therefore removal of the methyl group from O^6-methylguanine by the methyltransferase prevents a possible G·C to A·T transversion mutation (Gutierrez & O'Connor, 2021). While DNA lesions caused by alkylation can be removed via mismatch and nucleotide excision repair, it has been shown that the removal of O^6-methylguanine via O^6-alkylguanine DNA alkyltransferase enzymes is more important (Taira et al., 2013).

Another example is the *E. coli* AlkB dioxygenase (or 1meA/3meC-DNA dioxygenase), which catalyses the removal of methyl groups from nitrogen atoms from 1-methyladenine and 3-methylcytosine (Figure 9, C) (Sedgwick et al., 2007). Both of these have a methyl group on a nitrogen atom in the ring structure of the nucleotide base. These lesions and their repair by AlkB only occur in single-stranded DNA due to the hydrogen bonds, which the nitrogen atoms involved form with the bases in the complementary DNA strand (Bodell & Singer, 1979; Dinglay et al., 2000).

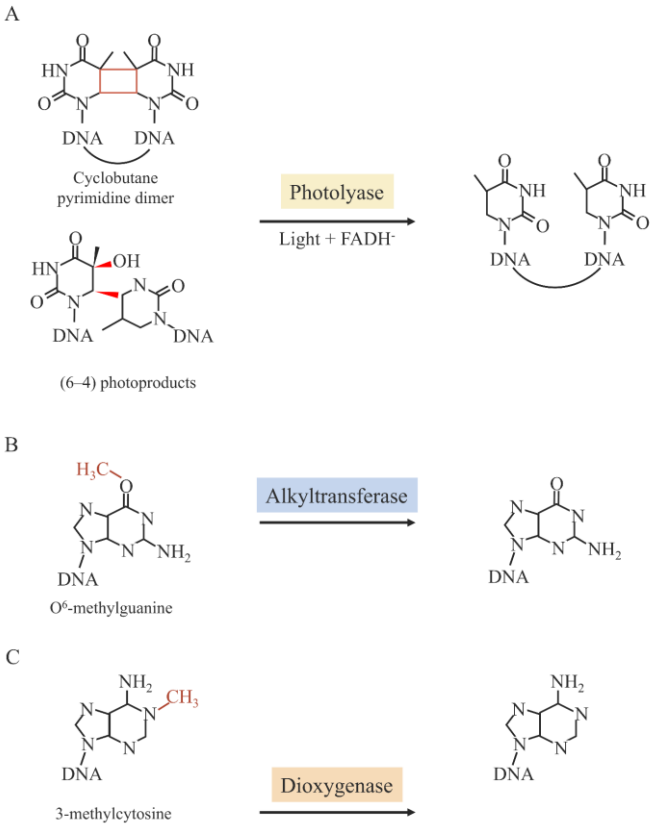


Figure 9: General mechanisms of direct reversal of modified DNA bases.
***A:** Reversal of the cyclobutane pyrimidine dimer or (6-4) photoproducts to by photolyase in a light- and $FADH^-$- dependent process.*
***B:** Reversal of O^6-methylguanin to guanine by an alkyltransferase protein.*
***C:** Reversal of 3-methylcytosine to cytosine by a dioxygenase enzyme (Yi & He, 2013)*

Polymerase Proofreading

If the wrong nucleotide base is incorporated by a polymerase during DNA replication, it can be repaired directly through the proofreading ability of some polymerases. DNA polymerases can have 3’ to 5’ and/or 5’ to 3’ exonuclease activity. Multiple of these polymerases have been identified. The 5’ to 3’ exonuclease in polymerase proteins removes RNA primers, while the 3’ to 5’ exonuclease served in the proofreading function. This proofreading allows for high-fidelity DNA replication.

DNA polymerase I (PolI) from *E. coli* was the first DNA polymerase to be discovered (Lehman, 2003). PolI is a monomeric protein with three enzymatic activities: polymerase activity, 3’ to 5’ exonuclease activity, and 5’ to 3’ exonuclease activity (Brutlag et al., 1969; Klenow & Henningsen, 1970; Ollis et al., 1985). The main function of PolI is the processing of Okazaki fragments and DNA fragments during DNA repair (Nagata et al., 2002; Okazaki et al., 1971). DNA PolI 3’ to 5’ exonuclease activity is dependent on two metal ions and four acidic amino acid residues are involved in the binding of those (Beese & Steitz, 1991; Derbyshire et al., 1991). DNA Polymerase III has been identified as the main polymerase for DNA replication in prokaryotes. This protein has a 3’ to 5’ proofreading exonuclease

subunit. This subunit allows for the removal of non-complementary nucleotides that may be incorporated by the polymerase during DNA replication. Mutations in the exonuclease subunit of PolIII lead to a loss of the proofreading function of PolIII, and a significant increase in the error rate during DNA replication (Fijalkowska & Schaaper, 1996; Lehtinen & Perrino, 2004).

Base Excision Repair

Base excision repair is an important DNA repair process for the prevention of mutations being introduced into the genome through different small DNA lesions, like methylated, alkylated, oxidised, abasic sites, mismatched nucleotides and DNA crosslinks (Krokan & Bjørås, 2013; Wozniak & Simmons, 2022). This pathway is highly conserved and exists in both eukaryotes and prokaryotes and generally follows similar steps in all organisms, which are shown in Figure 10 (Wallace, 2014).

Short patch base excision repair starts with the recognition and excision of the damaged base by a damage-specific glycosylase. The discovery of the *E. coli* uracil-DNA glycosylase was the first step in the characterisation of the base excision pathway (Lindahl, 1974). Uracil-DNA glycosylase removes uracil bases in DNA strands that are produced through cytosine deamination. To remove the modified base the glycosylase cleaves the *N*-glycosidic bond between the base and the deoxyribose group of the nucleotide. This creates an abasic site in the DNA strand, which is recognised by an AP endonuclease, that creates a cut next to the AP site (Pearl, 2000). Some glycosylases also exhibit AP endonuclease function, so the same enzyme carries out both steps. The AP-endonuclease creates a free 3'OH and a 5'deoxyribose-phosphate group in the strand. RecJ exonuclease removes the 5' terminal residue to create a substrate for re-synthesis of the DNA strand by a DNA polymerase. Finally, a DNA ligase fills the nick resulting in a repaired DNA strand (Dianov & Lindahl, 1994).

While the steps are generally the same, the mechanism and machinery used for base excision repair depend on the DNA glycosylase that carries out the first step and therefore, depends on the type of lesion (Krokan & Bjørås, 2013).

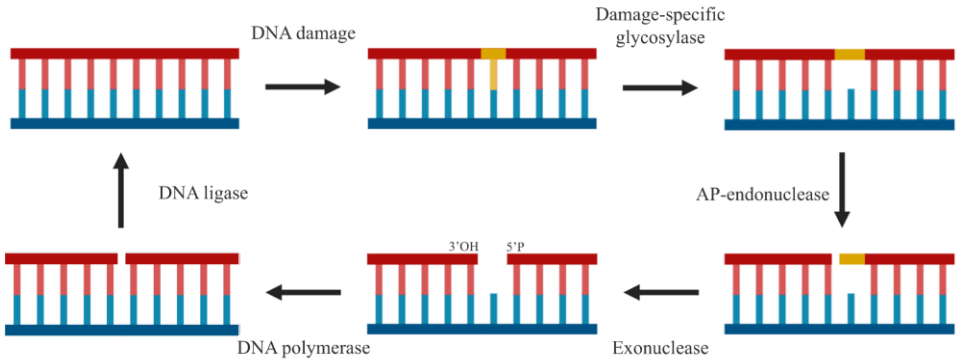


Figure 10: Illustration of the general base excision repair pathway steps.
Starting on the top left with an intact double-stranded DNA, that is damaged resulting in a damaged base (indicated by the yellow nucleotide). A glycosylase enzyme removes the damaged base creating an abasic site (top right). An AP endonuclease creates a nick near the abasic site (bottom right). RecJ exonuclease creates a substrate suitable for polymerisation (bottom left) and ligation to create intact double-stranded DNA (top left).
This illustration was adapted from that of David et al. (2007). (Created with BioRender.com)

Nucleotide Excision Repair

Nucleotide excision repair is a DNA repair mechanism conserved across all domains of life. Nucleotide excision repair starts with the detection of the damage, which in prokaryotes is carried out by UvrA, which moves along the DNA double-strand to search for DNA damage. Upon DNA lesion recognition, UvrB is recruited to UvrA in an ATP-dependent delivery process. UvrA and UvrB form a heterodimeric protein-DNA complex (Orren & Sancar, 1990; Verhoeven et al., 2002). After recognition of the damage by UvrB, UvrA dissociated from the UvrAB-DNA complex so a UvrB-DNA pre-incision complex can form (Orren & Sancar, 1990). The UvrC protein binds to this complex and creates a cut four nucleotides in the 3' direction and seven nucleotides in the 5' direction of the damage. UvrC is removed by the UvrD helicase along with the nucleotides in the incisions, creating a gap of twelve nucleotides in the DNA, where the damage was located (Caron et al., 1985). DNA polymerase I re-synthesises the excised 12 nucleotides allowing for the dissociation of UvrB (Caron et al., 1985; Husain et al., 1985). Finally, the nick created by the polymerase reaction is sealed by a DNA ligase.

Mismatch Repair

The goal of mismatch repair is to correct incorrect base pairs. As mentioned above, per Watson–Crick base pairing rules, in DNA adenine pairs with thymine and cytosine pairs with guanine. If base pairs do not follow this pattern, it is a DNA mismatch. Mismatches can be caused by errors during DNA replication, incorporation of RNAs into a DNA strand, spontaneous deamination or through the activity of DNA modifying agents (Rossetti et al., 2015). Unrepaired mismatches cause mutations in the cell's genome and can be passed on to the next generation. Mismatch repair is an important DNA repair mechanism and in its absence, mutation rates can increase up to 1000-fold (Iyer et al., 2006).

Mismatch repair follows the same steps in all prokaryotes and eukaryotes: mismatch recognition, incision, nucleotide removal, nucleotide re-synthesis, and ligation (Fukui, 2010). This process can be carried out by different pathways using different machinery. One well-studied pathway is the MutHLS mismatch repair pathway found in *E. coli*, shown in Figure 11.

E. coli methyl-directed mismatch using the MutHLS pathway was first described by Wagner and Meselson (1976) and has been reviewed many times since (Kunkel & Erie, 2005; Li, 2008; Modrich, 2016; Reyes et al., 2015). The process starts with mismatch recognition by the MutS homodimer, which forms a complex with the DNA (Su & Modrich, 1986). MutS recognises a range of mismatches, which allows for the repair of different wrongly paired bases. The MutL homodimer binds to the MutS-DNA complex in an ATP-dependent process (Figure 11, A) (Grilley et al., 1989). The DNA is then looped in a search for a hemimethylated d(GATC) restriction site near the 3' or 5' of the mismatch, this can be up to 1000 bases from the mismatch (Allen et al., 1997). The d(GATC) site of the newly synthesised strand is not methylated, which allows for the identification of the newly synthesised and therefore, error-containing DNA strand (Pukkila et al., 1983; Wagner & Meselson, 1976). When the restriction site is found MutH restriction endonuclease is recruited, which creates a cut at the d(GATC) site (Figure 11, B) (Au et al., 1992). UvrD helicase unwinds the DNA double-strand starting at the break produced by MutH towards the mismatch. A 3' to 5' or 5' to 3' exonuclease (RecJ, ExoI, ExoVI, ExoX), depending on the orientation of the restriction site to the damage, degrades the unwound strand from the cut and creates a gap until approximately 100 nt bases past the damaged base (Figure 11, C) (Cooper et al., 1993; Grilley et al., 1993). The generated single-stranded DNA is bound by a single-stranded binding protein (SSB) to prevent nuclease degradation. The gap is then filled by DNA polymerase III, which re-synthesizes the strand (Figure 11, D). The repair is completed by ligation of the nick between the old and newly synthesised strand to produce a repaired double-stranded DNA (Figure 11, E).

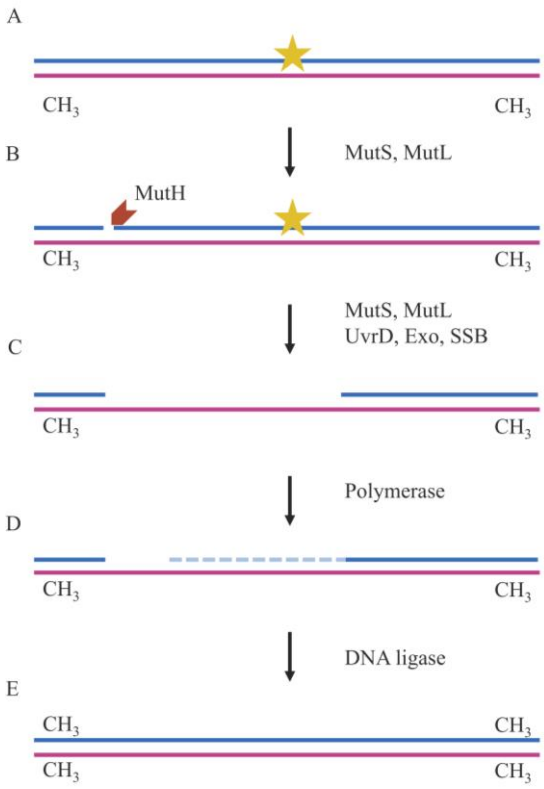


Figure 11: Overview of the E. coli mismatch repair pathway MutHLS.
***A:** Double-stranded DNA with a mismatch (indicated by the yellow star). The damage is recognised by MutS and bound by MutL. DNA strand methylation is indicated by CH_3.*
***B:** MutH is recruited and creates a cut at the GACT restriction site of the unmethylated strand.*
***C:** The double strand is unwound by UvrD. The damaged strand is removed by an exonuclease. The single strand is protected by single-strand binding proteins (SSB).*
***D:** The damaged strand is newly synthesised by a polymerase.*
***E:** The newly synthesised strand is joined to the original strand by a DNA ligase resulting in a repaired DNA double strand.*

Double-Stranded Break Repair

Double-stranded breaks are one of the most deleterious types of DNA damage. Therefore, organisms have a range of repair pathways, specifically for the repair of double-stranded breaks. These include homology-directed recombination-mediated repair, canonical non-homologous end joining, alternative end-joining (also referred to as microhomology-mediated end joining) and single-strand annealing. The repair pathway choice is affected by the cell cycle phase. Homologous recombination repair can only be carried out when a homologous template is present in the cell. It is therefore restricted to cell cycle phases where a second genome is present. The use of non-homologous end-joining, alternative end-joining versus single-strand annealing is thought to be determined by the length of the resection at both ends of the double-stranded break (Bhargava et al., 2016).

Non-Homologous End Joining

Non-homologous end joining is a pathway for the repair of double-stranded breaks, that does not require the presence of a homologous DNA template. It was thought that this pathway exists only in eukaryotes, but non-homologous end-joining specific genes have been identified in some archaea and bacteria (Aravind & Koonin, 2001; Bowater & Doherty, 2006; Doherty et al., 2001; Weller et al., 2002). In prokaryotes with non-homologous end-joining, the pathway is initiated by binding by the Ku protein homodimer to both ends of the double-stranded break (Figure 12, A). The Ku homodimer is a homologue of the eukaryotic Ku 70/80 heterodimer. The Ku homodimer brings the ends of the break together and facilitates hybridisation of the DNA overhangs, if possible. It recruits the LigD ligase protein, which is a multifunctional ligase protein with nuclease and polymerase function, to the break (Figure 12, B) (Öz et al., 2021; Weller et al., 2002). At the end of the breaks, LigD can create 3' overhangs through its nuclease activity (Amare et al., 2021). Through the LigD polymerase activity, single-stranded overhangs can be extended and are ligated through the LigD ligase activity at the break interface (Amare et al., 2021). In eukaryotic organisms, the function of LigD is carried out by multiple proteins, including a polymerase (polmu and lambda), nuclease (Artemis:DNA-PKcs), kinase/phosphatase (PNK), and a ligase (DNA ligase IV) (Lieber, 2010).

Non-homologous end joining is the only double-stranded break repair pathway that does not require an annealed intermediate before ligation of the two double-strands can occur. If the ends of the double-stranded break are processed by nuclease or polymerase activity, nucleic acids may be deleted from or added to the original sequence, as indicated in Figure 12 (C). This can lead to genome mutations. Because of this, non-homologous end-joining is often referred to as the more error-prone double-stranded repair pathway. However, Bétermier et al. (2014) argue that it is not necessarily an error-prone mechanism, but rather one that can adapt to double-stranded break ends that are not fully complementary and maximise the annealing process since an annealed intermediate is not required. This is unlike single-strand annealing and alternative non-homologous end-joining processes, which both require an annealed intermediate. If the ends of the double-stranded break are directly ligatable, non-homologous end-joining can lead to error-free repair (Chang et al., 2017). Nonetheless, non-homologous end-joining is more error-prone than the error-free homology-directed recombination-mediated repair.

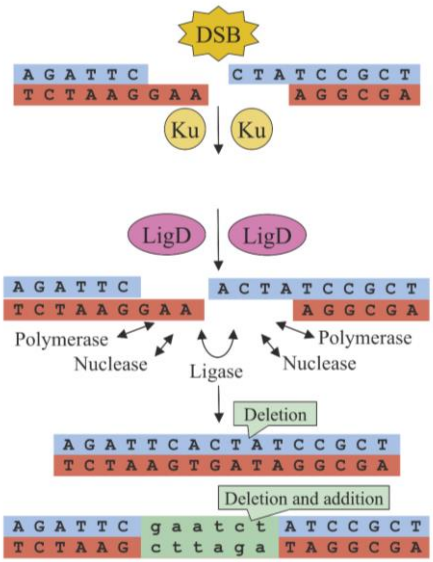


Figure 12: Illustration of the non-homologous end-joining repair pathway for double-stranded breaks.
***A:** A double-stranded break is created in a DNA strand and the Ku homodimer binds both ends of this break.*
***B:** LigD is recruited to the site of the damage and through its nuclease, polymerase and ligase activity the break is repaired.*
***C:** If nucleotides are removed from either side of the break this can lead to deletions (i). Nucleotides can be added to either side of the break by the polymerase function of LigD (ii) (Chang et al., 2017)*

Single-Strand Annealing and Alternative End-joining

Single-strand annealing and alternative end-joining, like non-homologous end joining do not require the presence of a homologous DNA template repairing double strands (Bhargava et al., 2016). Both require resection to generate a 3' ssDNA at the ends of the break, but the extent of the resection differs between the two. For single-strand annealing to be possible at least approximately 50 nucleotides on each side of the double-stranded break must be homologous. Alternative end-joining only requires homology between a few nucleotides of the double-stranded break ends. These homologous regions then anneal. The annealed intermediate is prepared for ligation by an endonuclease, which cleaves non-homologous single-stranded tails. A polymerase fills in gaps if any are present. The final step is the ligation of the intermediate to form an intact DNA double strand. The single-strand annealing process is inherently mutagenic and can lead to the loss of multiple thousands of bases. Alternative end-joining is also mutagenic and creates a deletion at or near the site of the break (Bhargava et al., 2016).

Homology-Directed Recombination-Mediated Repair

Homologous or template-directed recombination repair allows error-free repair of double-stranded breaks (Kowalczykowski, 2015; Yeeles & Dillingham, 2010). It requires the presence of a homologous DNA strand as a template and is therefore not possible in every stage of the cell cycle. Homologous recombination is the only repair pathway for a double-stranded break that allows for error-free repair of the break. Mechanisms for homologous recombination exist in Prokaryotes, Eukaryotes and Archaea (Li & Heyer, 2008).

In *E. coli* double-stranded break repair is mediated by the RecA protein and the RecCBD protein complex, as shown in Figure 13. Homologs for the RecA protein can be found in mammals, they serve the same function as RecA and are called Rad51 (Li & Heyer, 2008). The RecCBD protein has nuclease and helicase functions. It binds to double-stranded DNA ends, like those at a double-stranded break, and the RecB and RecD helicase subunits separate and unwind the DNA strands in both 3' to 5' and 5' to 3' direction (Dillingham et al., 2003; Phillips et al., 1997). The nuclease activity of the RecCBD complex can also create 3' overhangs if the ends of the break are blunt (Chen et al., 1997). In the next step, the broken strands bind to homologous strands of a homologous DNA duplex. This is enabled by the RecA protein in *E. coli* and results in the formation of what is called the D-loop where the homologous double-stranded DNA unwinds to allow invasion by the broken DNA strands (Kim, 2018). The invading strands provide a 5' end allowing for 5' to 3' synthesis by a polymerase to repair the break using the homologous strand as a template (Pavlov et al., 2006). After synthesis of the strands two complement and two coding DNA strands are present that are connected through two Holliday junctions. Holliday junctions are branched DNA structures. The Holliday junctions are resolved by Holliday junction resolving enzymes. In *E. coli* this is carried out by the RuvC, which creates symmetrical cuts at each side of the junction, as indicated in Figure 13 by the red arrows (Bennett et al., 1993; Matos & West, 2014). The resulting two nicked DNA duplexes are ligated resulting in two intact DNA duplexes (Matos & West, 2014).

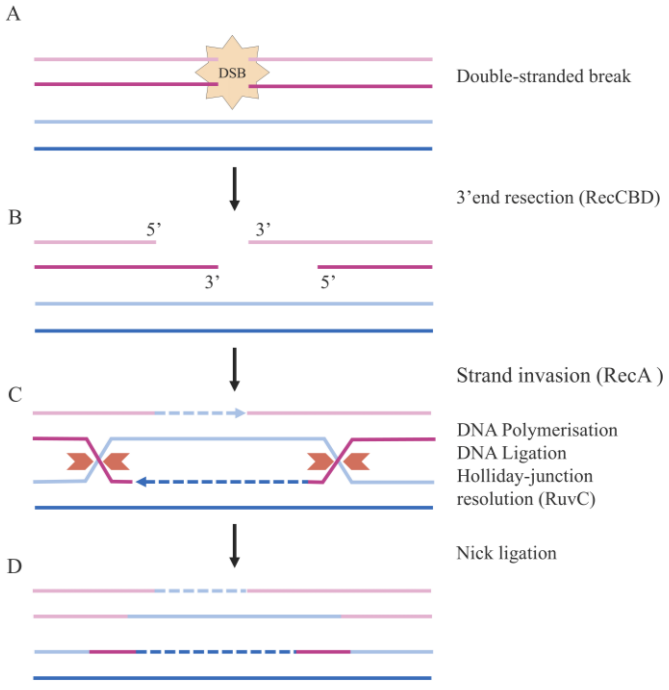


Figure 13: Illustration of the homologous recombination repair process in E. coli.
A: *DNA duplex with a double-stranded break (DSB) (pink) and a homologous DNA duplex (blue).*
B: *End resection by RecCBD creating 3' overhangs on the ends of the damaged DNA duplex.*

C: *Strand invasion is facilitated by RecA. The broken strand is re-synthesised by a DNA polymerase and nicks are ligated. Holliday junctions are resolved by the Holliday junction resolvase RuvC (indicated by the red arrows).*
D: *Nicks created by RuvC are ligated resulting in two intact DNA duplexes.*

1.2 Nucleases

As described above, DNA repair and replication pathways require many different enzymes. The key players in the replication and repair of DNA are polymerases, ligases, helicases and nucleases. Nucleases are phosphodiesterases that hydrolyse the bond between the sugar and phosphate groups of nucleic acids within oligonucleotides. They are a large group of enzymes and catalytic RNAs, that cleave the phosphodiester backbone of DNA or RNA polymers (Yang, 2011). They are indispensable for many processes, including DNA repair, replication and recombination, RNA processing, and even microbial defence mechanisms (Hsia et al., 2005; Yang, 2011). The focus here will be on their roles in DNA repair and maintaining the integrity of the genome.

Due to the high diversity of nucleases, they can be classified in multiple ways. The first differentiation is usually made based on their activity as endo- and/or exonucleases. Endonucleases cleave the phosphodiester bond within a polynucleotide, while exonucleases cleave from the end of a polynucleotide and require a 5' or 3' end to hydrolyse the phosphate bond (Figure 14). Exonucleases can be further divided by the direction they cleave nucleotides from the DNA strand, into 5' to 3' end processing and 3' to 5' end processing exonucleases (Figure 14, ii and iii, respectively). Some nuclease proteins have both endo- and exonuclease activity. For example, the human abasic endonuclease Ape1 has specific endonuclease cleavage activity 5' to abasic sites, but it also has 3' to 5' exonuclease activity (Chou & Cheng, 2003; Li & Wilson, 2014; Wilson, 2003).

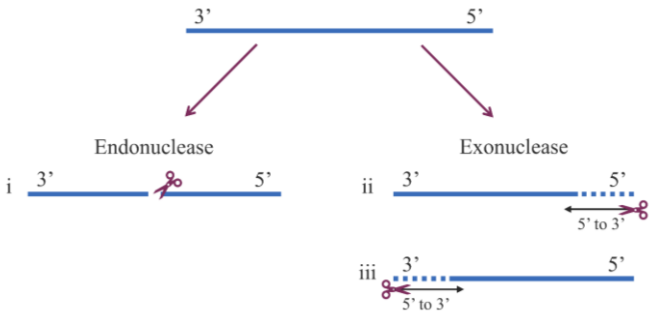


Figure 14: Illustration of the difference between endo- and exonucleases.
An intact single DNA strand on the top is cut by an endonuclease, which cuts within the DNA strand (i). Exonucleases cut nucleotides from the DNA strand in 5' to 3' direction (ii) or 3' to 5' direction (iii).

Furthermore, nucleases may be classified based on their preferred substrate. While many nucleases have a broad substrate specificity, some have specific substrate requirements. They may catalyse the cleavage of DNA or RNA and more specifically single- or double-stranded DNA/RNA substrates (Yang, 2011).

They can be specific for DNA in helical or base-stacked conformation (Yang, 2011). Additionally, specificity for a specific sequence or DNA lesion may be present.

For example, restriction enzymes are a group of sequence-specific endonucleases, which are common tools in molecular biology applications, specifically for molecular cloning, and because of this, they have been studied extensively (Di Felice et al., 2019). Restriction enzymes have been found mainly in bacteria. *In vivo*, restriction endonucleases protect bacteria against foreign DNA, specifically that from bacteriophages through the identification of the specific sequences by the enzymes (Alekseeva & Kuznetsov, 2023). Type II restriction enzymes are the best-studied restriction enzymes since they are the most useful restriction enzyme type for molecular biology applications (Loenen et al., 2014). Type II restriction enzymes cleave the DNA close to or within the recognition site. A well-studied Type II restriction enzyme is EcoRI, which was found in *E. coli*. (Pingoud et al., 2014) (Figure 15, A). Another example is the EcoRV restriction endonuclease, which was also discovered in *E. coli*. Unlike EcoRI, EcoRV produces blunt ends as shown in Figure 15 (B).

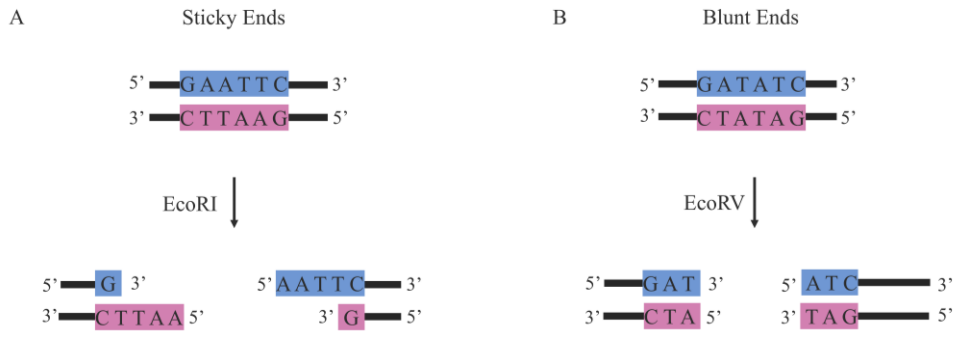


Figure 15: Creation of sticky and blunt ends by restriction endonucleases.
***A:** Creation of sticky ends by EcoRI. **B:** Creation of blunt ends by RecoRV.*

As mentioned above some nuclease enzymes create cuts at specific DNA damages. These enzymes are often involved in the repair processes. For example, AP endonucleases cut specifically at the abasic sites and are involved in the removal of abasic sites and the removal of other DNA lesions during base excision repair as described in Section 1.1.3. AP endonucleases can be found in bacteria, archaea, and eukaryotic organisms. The main AP endonuclease from *E. coli*, exonuclease III (AP endonuclease VI), for example, cleaves the DNA at the DNA 5' to the abasic site in both double- and single-stranded DNA (Shida et al., 1996; Takeshita et al., 1987).

Holliday junction resolvases are a group of nuclease enzymes that specifically cleave four-way DNA junctions. DNA junctions are common intermediates during DNA repair processes, like homologous recombination repair as described in Section 1.1.3. Holliday junction resolvases have been identified in organisms from all domains of life. Some of these enzymes exhibit sequence-specificity, including the *E. coli* RuvC, and Cce1 and Ydc2 from yeast mitochondria. Most Holliday junction resolvases

significantly distort the four-way junction to allow for substrate cleavage and they share a common structural feature of mixed α–β folds (Lilley, 2017).

1.2.1 DNA Nuclease Reaction Mechanism

There is not a high level of sequence conservation across nuclease proteins; however, the reaction mechanisms in which the cleavage of the phosphodiester backbone is carried out are fundamentally the same in all nucleases. In most cases, phosphodiester cleavage is carried out by general base-acid catalysis as described below.

Cleavage of the phosphodiester backbone can occur at two positions of the phosphodiester bond, which is either at the 3' oxygen or 5' oxygen of the P-O bond (Yang, 2011). First, a hydroxyl group is required, which in many cases is generated by the deprotonation of water by a general base in the active site (Figure 16, 1.). This hydroxyl group carries out a nucleophilic on the bond between the 3' or 5' oxygen and phosphate causing the formation of an unstable penta-covalent intermediate (Figure 16, 2.). This intermediate is stabilised by cations in the active site (Figure 16, 3.). Finally, the leaving group is protonated by a general acid in the active site (Figure 16, 4.).

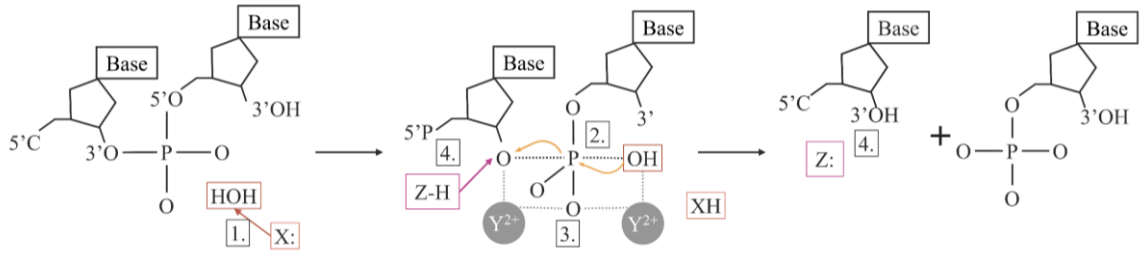


Figure 16: General mechanisms of the hydrolysis of the phosphodiester bond between two nucleotides.
1. *Deprotonation of water by a general base (X) to generate a hydroxide (OH).*
2. *Nucleophilic attack of the scissile bond by the hydroxide.*
3. *Stabilisation of the pentacovalent intermediate by two cations (Y^{2+}).*
4. *Protonation of the leaving group by a general acid (Z). This figure was inspired by that from (Nishino & Morikawa, 2002).*

The nucleophilic attack is often carried out by a hydroxyl group, which is produced via de-protonation by a basic side chain or a metal ion in the active site (Lovett, 2004; Yang, 2011). The hydroxyl group may also be provided by a histidine, serine or tyrosine side chain in the protein active site. Even the hydroxyl group of nucleotides or nucleic acids may serve for the hydrophilic attack (Yang, 2011).

Many nucleases use cations in the active site to stabilise the negatively charged pentacovalent intermediate during phosphodiester bond cleavage. Often these cations are in the form of metal ions, that are held and positioned in the active site by aspartic acid, glutamic acid and histidine residues (Nishino & Morikawa, 2002). Commonly used metal ions are Mg^{2+}, Ca^{2+}, Mn^{2+} or Zn^{2+}. Nucleases use between zero and three metal ions in their active sites (Hamdan et al., 2002; Ivanov et al., 2007; Mol et

al., 1995; Sasnauskas et al., 2007). Two metal ion-dependent nucleases, as shown in Figure 16, are the most commonly found ones (Yang, 2011). For example, the active site of the restriction enzyme EcoRI consists of two acidic residues (Asp91 and Glu111), which are involved in the binding of metal ions in the active site, and a basic residue (Lys113), which stabilises the pentacovalent phosphorus intermediate during the cleavage of the phosphate backbone (Selent et al., 1992).

1.2.2 DNA Binding

Nuclease proteins usually have a large positively charged surface formed by basic residues. This positive surface allows for the binding of the protein to the negatively charged phosphodiester backbone of the DNA. DNA recognition and binding are often facilitated by DNA recognition loops in the protein. These loops are positively charged with a crescent shape that complements the shape and negative charge of the DNA (Mol et al., 2000). Proteins with sequence or damage specificity require additional binding features to recognise these idiosyncratic features in the DNA.

For enzymes with sequence specificity, initial non-specific and loose binding to the DNA allows the enzyme to move along the DNA and scan for a specific sequence or damage (Nishino & Morikawa, 2002). The loose binding interactions between the DNA and proteins are mainly electrostatic and allow linear protein diffusion along the DNA (Pingoud & Jeltsch, 2001). Once the target site of the protein is encountered the binding changes from non-specific to specific binding. Specific binding is characterised by a change in the protein and DNA conformations. This conformational change allows for more extensive binding interactions between the protein and DNA. The restriction enzyme EcoRV causes a particularly extensive conformational change in the DNA. Upon binding by EcoRV, the DNA bends around 50° (Winkler et al., 1993). The closer association between the protein and DNA allows for the orientation of the phosphodiester backbone of the DNA in the active site of the protein. Bendability and kinking of the DNA are important features that allow restriction enzymes to recognise specific DNA sequences (Rohs et al., 2010).

Repair Nucleases

Less is known about the binding of damage-specific nuclease proteins to DNA and how these proteins recognise the damage. Insertion of aromatic side chains into a DNA duplex and base flipping are two important factors allowing DNA repair enzymes to recognise DNA lesions (Nishino & Morikawa, 2002).

Base flipping is an important process in the recognition of DNA lesions by DNA repair proteins, including nucleases, glycosylases, photolyases and methyltransferases (Fuxreiter et al., 2002). Base flipping has been studied extensively in glycosylase enzymes, but can also be found in nuclease

enzymes, like AP endonucleases (Brooks et al., 2013; Mol et al., 2000; Shi et al., 2021). During base flipping the nucleotide is turned out of the DNA duplex or strand into the binding pocket of the protein, as shown in Figure 17 (Nishino & Morikawa, 2002). The abasic site-specific endonuclease Q (EndoQ) from archaea and bacteria, recognises the abasic sites and deaminated bases by a combination of base flipping and interactions between specific residues of the protein and the backbone of the DNA (Shi et al., 2021). The protein can cut both double- and single-stranded DNA by interactions with the scissile strand of the DNA. Basic side chains were shown to be important for DNA binding by the EndoQ protein (Shi et al., 2021). Another example is Endonuclease IV, which cleaves the phosphodiester bond 5' of abasic sites in double-stranded DNA. The crystal structure of the protein showed, that when Endonuclease IV binds DNA with an abasic, the damaged strand is bent approximately 90° so that the damage is facing the protein's active site (Hosfield et al., 1999).

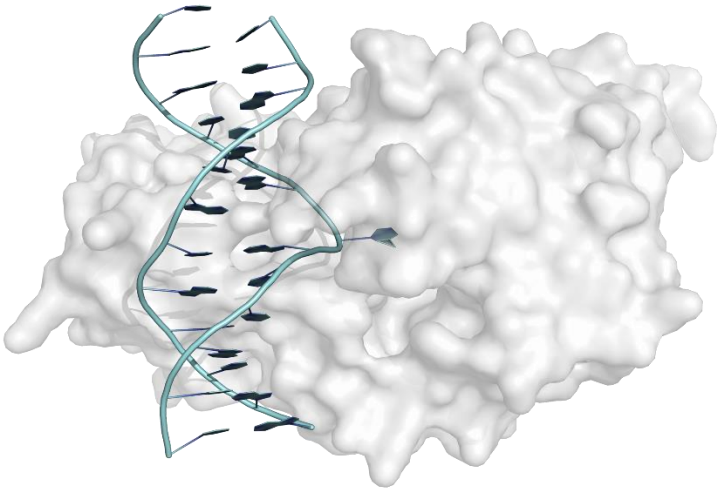


Figure 17: Base flipping, shown with the example of the methyltransferase from Hha I and double-stranded DNA. The DNA is shown in blue and Hha I is shown in grey. The image was taken from https://atdbio.com/nucleic-acids-book/Base-flipping.

1.2.3 Nuclease Structure

The amino acid sequences of nucleases are not conserved, apart from the catalytic site motifs, as described above. Structurally there is a lot more conservation across nuclease proteins, which allows for classification of nuclease proteins based on their structure. Nishino and Morikawa (2002) divided nuclease proteins into eight categories based on their structure: RNaseH-like, resolvase-like, restriction endonuclease, RecJ, metallophosphatase, DNasI, TIM β/α barrel, and His-Me finger nuclease-like fold (Table 1). Proteins belonging to the restriction endonuclease-like fold belong to the PD-(D/E)XK nuclease-like superfamily and share a common motif, which is part of the protein's active site (Nishino & Morikawa, 2002; Steczkiewicz et al., 2012). It is important to note that although this categorisation may be helpful in some cases, many nuclease proteins do not fit into any of those categories.

Table 1: Structural categorisation of common nuclease proteins.
This table was extended from that of Nishino and Morikawa (2002).

Category	Features	Examples
RNase-like fold	5-stranded β-sheet (32145) with flaking α-helixes	RuvC, RNaseH, ExoI
Resolvase-like fold	5-stranded β-sheet (21345) with flaking α-helixes	FEN1
Restriction endonuclease	5-stranded β-sheet (12345) with flaking α-helixes	MutH, Hjc, T7 endoI, restriction endonucleases
RecJ	5-stranded parallel β-sheet (21345), with 6 flanking α-helices	RecJ
Metallophosphatases	2 β-sheets between α-helices	Mre11
DNaseI	α/β sandwich	ExoIII, Ape1, DNaseI
TIM β/α barrel	α8/β8 barrel structure	EndoIV
His-Me finger nuclease-like fold	Core of β-hairpin flanked by two helices	T4 endoVII, some nuclease colicins, Serratia nuclease

1.3 Antarctic McMurdo Dry Valleys

The Antarctic continent is one of the most isolated environments in the world, located almost centrally around the South Pole. Due to its location, the Antarctic summer and winter seasons vary extremely from nearly six months of complete darkness in the winter months to six months of continuous sunlight in the summer months at the most southern point of Antarctica. Over 95% of the Antarctic continent is covered with snow. The McMurdo Dry Valleys (Dry Valleys from here on) are the largest permanently ice-free area of Antarctica (Bockheim, 1997). They are located between the Ross Sea and the Polar Plateau of Victoria Land Antarctica as shown in Figure 18. The Dry Valleys are one of the coldest and driest desert environments on Earth. They are characterised by large temperature fluctuations, and variability in light availability, low nutrient levels and water availability, and high levels of UV-B radiation (Cary et al., 2010). This makes the Dry Valleys one of Earth's harshest and most extreme environments. Due to the present conditions, the Dry Valleys have been described as an analogue to Mars on Earth (Heldmann et al., 2013).

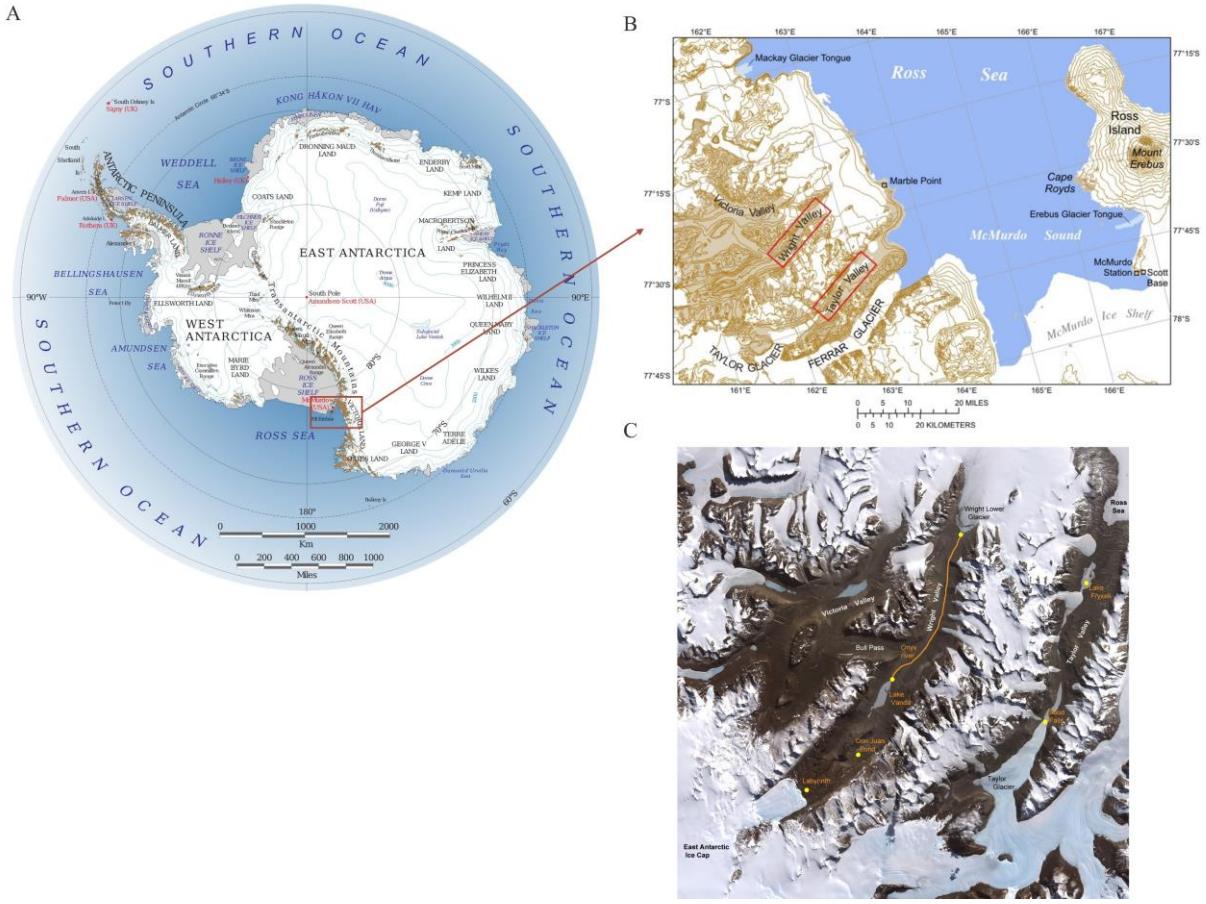


Figure 18: Location of the Antarctic McMurdo Dry Valleys.
***A:** Map of Antarctica. The rectangle shows the McMurdo Sound region.*
***B:** Map of the McMurdo Sound region including the McMurdo Dry Valleys on the left.*
***C:** Image of the Dry Valleys.*
Images are taken from www.coolantarctica.com.

The Dry Valleys largely consist of ice-free mineral soils with a few streams and lakes (Cary et al., 2010). They cover an area of approximately 5000 km^2. Due to the harsh conditions in the Dry Valleys, no animal or plant life is present, and the environment is dominated by microbes with a few mosses present in protected areas (Cary et al., 2010; Horowitz et al., 1972; Yung et al., 2014). Precipitation in the area is extremely low, ranging from 3 to 50 mm per year, mostly during the winter month and dominantly in the form of snow (Fountain et al., 2010). The temperatures in the Dry Valleys average at -20°C and the lowest temperatures are reached during the winter, where soil temperatures may be as low as -60°C. (Obryk et al., 2020). During the summer, the soil temperatures are above freezing for 25 to 75 days and may reach 15°C to 27°C (Horowitz et al., 1972). The temperatures fluctuate frequently and can range from -15° to 27°C within 3 hours. Compared to other cold desert environments the Dry Valleys have notably lower levels of precipitation, organic carbon in the soils, and microbial biomass, which may be an indication of the hostility of this environment (Cary et al., 2010).

High temperatures in the area lead to glacier melting, which feeds into streams that flow into basin lakes in the Dry Valleys (Figure 18, C). Those glacier melts are the largest source of water in the Dry Valleys

and most microbial life centres around those streams and lakes (Wlostowski et al., 2016). The lakes are permanently frozen and only partially melt around the edges of the lakes when soil temperatures increase above freezing. The surroundings of those lakes, where water availability is higher than in the rest of the Dry Valleys, are dominated by cyanobacteria. A layer of permafrost is present throughout the soils of the Dry Valleys. Between the permafrost layer and the top of the soil lays the active layer, which undergoes frequent freeze-thaw cycles (Bockheim et al., 2007). Due to the increased water availability in this layer microorganisms are most prevalent there (Cary et al., 2010).

Microorganisms in the Dry Valleys

Due to the extremely harsh conditions in Dry Valleys, it was thought that all life including microbial life was very limited. With more sophisticated methods for measuring microbial life in soil environments, including more sophisticated culturing and sequencing methods, it was clear that there are much higher levels of microbial biomass and diversity than expected (Cowan et al., 2002). The 16S and 18S rRNA sequencing revealed the presence of known and unique bacterial taxa in the Dry Valleys (Niederberger et al., 2008; Pointing et al., 2009; Smith et al., 2006). The Dry Valley soils are dominated by prokaryotic organisms with only a very low relative abundance of eukaryotic, mostly fungal, and archaeal taxa based on metagenomic data (Goordial et al., 2017; Le et al., 2016; Yung et al., 2014). The highly diverse bacterial communities are dominated by Actinobacteria, Proteobacteria, Cyanobacteria and Bacteroidetes (Le et al., 2016; Wei et al., 2016).

The development of more high-throughput and less expensive sequencing techniques has made metagenomic analysis more accessible and popular for studying soil microbial communities and their function. Metagenomics allows for cultivation-independent assessment of microbial communities and their metabolism and the discovery of novel biomolecules, including enzymes. Instead of culturing organisms for subsequent DNA sequencing, in metagenomic approaches, DNA is extracted directly from the environment and sequenced (Nam et al., 2023). Extraction of DNA especially from extreme environments like the Dry Valleys can be challenging due to the relatively low biomass present in the soil. However, it has been done for the Dry Valley soils as demonstrated by Chan et al. (2013), Niederberger et al. (2019), Tiao et al. (2012), and Lee et al. (2018).

1.3.1 Microbial Survival in Cold Environments

Psychrophily is the ability of an organism to survive and grow in cold conditions. There is no clear definition of the growth optima of psychrophilic and psychrotolerant organisms but they are often defined as follows (Casanueva et al., 2010; Cavicchioli, 2016). Psychrophilic organisms have optimal growth temperature at or below 15°C and a maximum growth temperature at or below 20°C. Psychrotolerant organisms can grow at temperatures below 15°C, however, they preferably grow at

temperatures between 20°C and 35°C. Permanently cold environments dominate the Earth's biosphere with over 80% and they are dominated by small organisms including bacteria, archaea, and small eukaryotes (Collins & Margesin, 2019; Rodrigues & Tiedje, 2008). The development of new tools, like metagenomics, allows for a better understanding of these communities of mostly unculturable organisms.

Low temperatures often come with a range of other stressors, including freezing, an increase in oxygen availability causing an increase in reactive oxygen species, decreased solubility of solutes and nutrients, reduced levels of diffusion, and increased osmotic stress (Collins & Margesin, 2019; Tribelli & López, 2018). The Dry Valleys and other cold environments have additional features, making them even more hostile. These include high levels of salinity, low levels of nutrients, freeze-thaw cycles, extremely low or high light exposure, high levels of UV radiation, and desiccation (Collins & Margesin, 2019). Therefore, organisms inhabiting these extreme environments must adapt to multiple stressors. The co-occurrence of these stressors makes it difficult to decipher, which of them an adaptation in an organism is responding to. The main problems that organisms face as a result of low temperatures are a loss of membrane fluidity, freezing, ice-crystal growth, loss of protein, DNA and RNA stability, reduced reaction rates and reduced protein activity.

A decrease in the cell membrane fluidity can interfere with the function of the membrane, which is important for transport and signalling, and transmembrane electrochemical gradient, which powers many cell functions (Strahl & Errington, 2017). Loss in membrane fluidity therefore leads to cell death and must be maintained at cold temperatures, which is generally achieved through modification of its fatty acid composition, by increasing the content of unsaturated fatty acids (Collins & Margesin, 2019). Adaptations in the fatty acid and phospholipid synthesis have been observed in multiple psychrophilic and psychrotolerant bacteria (Ayala-del-Río Héctor et al., 2010; Goordial et al., 2015; Koh et al., 2017; Methé et al., 2005; Raymond-Bouchard et al., 2018) and cold adaptive traits have been observed in proteins involved in fatty acid and phospholipid alterations and synthesis required to maintain cell membrane fluidity (Goordial et al., 2015).

Temperatures below freezing can cause the formation of ice crystals in the cellular fluid as well as the environment of the cell. This can cause cryoinjury, cell rupture and cell death due to physical damage (Collins & Margesin, 2019). Ice crystal formation inside the cells can occur when temperatures drop at a fast rate like they do in the Dry Valleys. Ice crystal formation in the cellular fluid is lethal to cells, therefore, compatible solutes, ice-binding proteins, extracellular polymeric substances (EPS) and/or biosurfactants can be produced by cells to weaken the effects of freezing and stop ice crystal formation. The most commonly used compatible solutes are glycine betaine, trehalose, glycerol, sucrose, sarcosine, mannitol and sorbitol (Collins & Margesin, 2019) and glycine betaine, choline and sarcosine have been

identified in species from the Dry Valleys and Arctic permafrost bacteria (Goordial et al., 2015). In addition to preventing ice crystal formation inside the cells, compatible solutes also help to increase enzyme activity and stability (Zhao, 2005). Some organisms secrete extracellular polymeric substances, including complex sugar polymers, proteins and small amounts of nucleic acids, lipids, phenols and humic substances (Deming & Young, 2017). These help create a biofilm and allow the cells to shape their environment to prevent ice crystal formation around the cells (Deming & Young, 2017). Ice-binding proteins also prevent ice crystal formation in cells by binding to the surface of ice to inhibit its growth (Bar Dolev et al., 2016). They have been found in metagenomic sequences from the Dry Valleys and other Antarctic environments (Gilbert et al., 2004; Munoz et al., 2017).

As a response to osmotic and oxidative stress, cold-adapted bacteria express a range of proteins specific to osmotic and oxidative stress response as well as general stress response proteins (Raymond-Bouchard et al., 2018). Additionally, it has been observed that oxidative metabolism is depressed at low temperatures to decrease the production of reactive oxygen species (Tribelli & López, 2018).

To help with DNA and RNA stability and protein synthesis bacteria may have cold shock proteins. Cold shock proteins serve as nucleic acid chaperones at low temperatures and increase the stability of mRNA molecules at low temperatures to allow for protein synthesis to proceed (Heinemann & Roske, 2021). They also regulate transcription and translation by binding to single- and double-stranded DNA and RNA (Horn et al., 2007). Some cold shock proteins even aid in DNA repair through direct interaction with DNA repair proteins (Heinemann & Roske, 2021). Apart from cold temperatures, cold shock proteins have been shown to contribute to the stress response to osmotic, oxidative, starvation, and pH stress (Keto-Timonen et al., 2016).

Since enzymes carry out all molecular cell functions it is important that they can catalyse reactions even at cold temperatures in psychrophilic and psychrotolerant organisms. The speed of all reactions is reduced at low temperatures, including those catalysed by enzymes (D'Amico et al., 2006). Therefore, many enzymes in psychrophilic organisms are adapted to the cold and can catalyse reactions with high specific activities at low temperatures compared to their mesophilic counterparts (D'Amico et al., 2006). It is thought that this is achieved mainly through increased protein flexibility. Specific chaperones assist with protein folding and stability in psychrophilic organisms at low temperatures (Feller, 2013).

1.3.2 Cold Adaption at the Protein Level

As mentioned above, the cold-adapted proteins allow the catalysis of reactions at high specific activity levels at low temperatures compared to their mesophilic counterparts (D'Amico et al., 2006). This allows for molecular processes required for cellular functions and growth to carry on at low temperatures. One adaptive trait of cold-adapted enzymes is increased protein flexibility, which has

been observed in some psychrophilic enzymes compared to their mesophilic counterparts (Feller, 2013). However, increased flexibility is not a requirement for psychrophilic enzymes and many psychrophilic enzymes have been identified, which have no increased flexibility compared to their mesophilic counterparts (Isaksen et al., 2014, 2016; Michetti et al., 2017). The increased flexibility may be limited to certain parts of the enzyme, or the entire protein may be more flexible. This is especially true for the protein active site or surface loop regions of the proteins, which are particularly flexible in some cold-adapted proteins (Collins et al., 2003; Georlette et al., 2003; Isaksen et al., 2014; Siddiqui et al., 2005; Sočan et al., 2019; Wallon et al., 1997). Additional features observed in the active site of cold-active proteins are a larger opening facilitated by smaller side chains near the active site or deletions in loop regions of the active site (Feller, 2013). A more accessible active site allows them to bind their substrate with a lower energy investment and also allows for easier release of the product (Feller, 2013). The open conformation of the active site also means that many cold-adapted enzymes have a broader substrate range (Feller, 2013).

Amino acid changes underlying the cold adaption of proteins can be difficult to observe at the protein level. However, genome or proteome-wide changes can be more obvious when comparing cold-adapted organisms with mesophilic ones. Genome-wide changes include lower amounts of proline residues, acidic and aliphatic side chains and a lowered arginine-to-lysine ratio (Aghajari et al., 1998; Ayala-del-Río Héctor et al., 2010; Goordial et al., 2015; Methé et al., 2005; Metpally & Reddy, 2009; Mykytczuk et al., 2013; Raymond-Bouchard et al., 2018). Metpally and Reddy (2009) propose that psychrophilic proteins contain less glutamic acid, phenylalanine, lysine, asparagine and tryptophan residues and more alanine, aspartic acid, glycine, serine, and threonine compared to mesophilic proteins, when comparisons are made at the proteome level. These amino acid changes can be observed at the protein level of some cold-adapted proteins but may be less obvious (Ayala-del-Río Héctor et al., 2010; Georlette et al., 2000; Goordial et al., 2015; Huston et al., 2004). A reduced proline content is common in cold-adapted proteins and is thought to increase the flexibility of the protein structure (Feller, 2013). Additionally, the negative effects of proline isomerisation on protein folding at low temperatures are reduced by a lower proline content (Feller, 2013).

In *Rhodococcus* sp. JG3 from the Dry Valley permafrost soils, proteins involved in fatty acid and phospholipid synthesis, cold, osmotic and oxidative stress response, general stress response, and transcription, translation and DNA repair were found to have high levels of the changes described above (Goordial et al., 2015). However, as shown in Figure 19, not all traits were present in all proteins and some proteins even showed features associated with hot-adapted proteins (Goordial et al., 2015). This highlights that while the features mentioned above are observed in a range of cold-adapted proteins, there are exceptions to these rules; and while some of these features may be present in cold-adapted

proteins and can help when trying to identify a cold-adapted protein, the only reliable method to date to determine the temperature dependence of a protein is to experimentally test its activity and stability.

Role	Gene ID	Gene name	A	B	C	D	E	F	G
Fatty acid and phospholipid alteration and biosynthesis	2529302193	Short-chain dehydrogenase							5
	2529298839	Glycerol-3-phosphate dehydrogenase							3
	2529299166	3-oxoacyl-[acyl-carrier-protein] synthase III							4
	2529302006	3-oxoacyl-[acyl-carrier-protein] synthase II							3
	2529299529	1-acylglycerol-3-phosphate O-acyltransferase							3
	2529298632	acetyl-coenzyme A synthetase							4
	2529302639	Acyl-CoA dehydrogenase							4
	2529302839	Acyl-CoA dehydrogenase							4
	2529299434	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligases II							4
	2529298474	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)							4
Cold, Osmotic and Oxidative Stress response	2529298663	Choline dehydrogenase and related flavoproteins							3
	2529298910	choline/carnitine/betaine transport							3
	2529299592	Predicted flavoprotein involved in K+ transport							4
	2529299729	Catalase							4
	2529299255	amino acid ABC transporter substrate-binding protein, PAAT family							6
	2529302676	ABC-type proline/glycine betaine transport systems, ATPase components							3
General stress response and regulation	2529299449	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain							3
	2529301347	SOS-response transcriptional repressor, LexA							4
transcription, translation and DNA repair	2529300888	ATP dependent helicase, Lhr family							4
	2529299669	Molecular chaperone							3
	2529299301	primary replicative DNA helicase							4
	2529302961	transcription antitermination protein nusG							5
	2529299297	single-strand binding protein							3
	2529303222	ATPases involved in chromosome partitioning							3
	2529299962	chaperone protein DnaJ							3
	2529301716	Formamidopyrimidine-DNA glycosylase							4
	2529301020	RNA polymerase, sigma subunit, ECF family							3
	2529301786	2'-5' RNA ligase superfamily							3
Carbon metabolism	2529301512	glucose-6-phosphate 1-dehydrogenase							5
	2529298836	glycogen debranching enzyme GlgX							4

A = aliphatic index, B = RK ratio, C = acidic residue, D = aromaticity, E = GRAVY, F = Proline %, G = Cold Adaptation score. Blue boxes indicate the protein was found to be cold adapted for that index, red is hot adapted and orange coloured boxes were not significantly different than mesophile homologs ($P < 0.05$). Cold adaptation scores were given to each protein based on the six indices. For each index, a point of 1 was given for each index which was determined to be cold adapted, and a point of −1 for each index found to be hot adapted. No points were given for indices which were not significantly different between *Rhodococcus* sp.JG3 and the 14 mesophillic strains.

Figure 19: Table of selected cold-adapted proteins from Rhodococcus sp. JG3.
This figure was taken from Goordial et al. (2015).

Adaptations in the Dry Valleys

Studies on the adaptations and stress responses of Dry Valley bacteria are limited. Advances in metagenomic sequencing and culturing techniques have allowed a better understanding of the microbial communities in the Dry Valleys and how they survive in such a harsh environment. Metagenomic data has been utilised to determine the presence of known stress response pathways in Dry Valley microorganisms and showed that pathways for the osmotic, radiation, desiccation, cold shock, heat shock, and low nutrient stress response are present in the Dry Valley metagenomes (Chan et al., 2013;

Goordial et al., 2017). These pathways are more abundant in the communities inhabiting the exposed surfaces of the Dry Valley soils compared to those living in the more sheltered soil habitats (Chan et al., 2013). *Halomonas* isolates with high ionisation and desiccation resistance have been found in the Dry Valley soils (Musilova et al., 2015). High ionisation resistance is thought to go hand in hand with resistance to desiccation, which may be due to the similar consequences of both in terms of DNA damage (Billi et al., 2000; Musilova et al., 2015).

A comprehensive study by Goordial et al. (2015) of a single isolated organism from the permafrost of the Dry Valley permafrost soils reveals insight into the cold adaptive traits at the level of a single organism. *Rhodococcus* sp. JG3 can grow at temperatures below freezing (up to -5°) (Goordial et al., 2015). A genome comparison of this isolate to mesophilic *Rhodococcus* species showed that it has higher copy numbers of genes involved in general stress response and for protection from UV radiation, cold shock, osmotic stress and oxidative stress (Goordial et al., 2015). Multiple photolyase and photorepair proteins were found in the Dry Valley isolate (Goordial et al., 2015). Proteome comparisons of the *Rhodococcus* sp. JG3 with mesophilic *Rhodococcus* species shows that both at the genome level and the protein level its proteins have adaptations in terms of their amino acid sequence to cold temperatures as described above (Goordial et al., 2015). These adaptations were present predominantly in proteins involved in cold adaptation of the cell, including those involved in fatty acid and phospholipid alterations and synthesis required to maintain the fluidity of the cell membrane, proteins involved in the stress response to cold, osmotic, oxidative and general stress. Furthermore, it was demonstrated, that proteins involved in transcription, translation, and DNA repair and replication in *Rhodococcus* sp. JG3 have cold adaption traits in terms of their amino acid sequence (Goordial et al., 2015). A similar trend was observed in the psychrotolerant *Psychrobacter arcticus* in genes belonging to the COG category of replication, recombination and repair (Ayala-del-Río Héctor et al., 2010). Comparisons of Dry Valley soil metagenomic data with that from other soil metagenomes and genomes from isolated pathogens showed that an increased proportion of proteins belonging to the 'Replication, Recombination and Repair' COG category are present in the Dry Valley metagenomes (Rzoska-Smith; et al., 2023).

Collectively, this highlights that maintaining membrane fluidity, general stress responses, protein synthesis, and DNA repair and replication appear to be prioritised by bacteria experiencing cold stress and other stressors. DNA repair appears to play a crucial role in the survival of Dry Valley microorganisms in this extreme environment. Therefore, it is hypothesised that organisms from the Dry Valleys possess novel DNA repair pathways and proteins, that are adapted to the extreme conditions in this environment.

1.4 DNA Repair in Extreme Environments

High levels of UV radiation, regular freeze-thaw cycles, osmotic stress and low nutrition in the Dry Valleys threaten the integrity of genomic DNA. Metagenomic sequences from the Dry Valleys show a higher proportion of genes belonging to the COG categories of replication, recombination and repair (Le et al., 2016; Rzoska-Smith; et al., 2023). Additionally, it was found that some proteins involved in DNA repair and replication have cold adaptations in an organism from the Dry Valley permafrost soil (Goordial et al., 2015). Those include an ATP-dependent helicase from the Lhr family, a primary replicative DNA helicase, single-stranded binding proteins, and formamidopyrimidine-DNA glycosylase (Goordial et al., 2015).

A study by Albarracín et al. (2012) conducted on *Acinetobacter* strains from the High-Altitude Andean Lakes of the South American Andes, where conditions are comparable to the Dry Valleys with high UVB levels, frequent strong changes in temperature and high salinity, showed that *Acinetobacter* isolates from the environment had high resistance to UV-B irradiation. UV-B irradiation caused an increased number of DNA lesions which were repaired partially through direct repair by photolyases and partially through excision repair pathways. Repair was most effective and led to greater survival upon treatment in the strains from the Andean Lakes compared to pathogenic clinical *Acinetobacter* isolates. The increased resistance was thought to be inferred through novel photolyase proteins, which were found in the resistant strains (Albarracín et al., 2012). Furthermore, *Acinetobacter* isolates from this environment have what is called a UV-resistome, which consists of genes that help the organism evade or repair damages caused by UV radiation (Kurth et al., 2015; Portero et al., 2019).

Deinococcus radiodurans is the model organism for the survival of extremely DNA-damaging agents including high levels of ionising radiation, desiccation, ionising radiation, reactive oxygen species (ROS), and toxic chemicals. Under extreme conditions, *D. radiodurans* cells stop growing but can recover and resume growth in suitable growth conditions (Timmins & Moe, 2016). D. radiodurans can resume growth even after exposure to the low Earth orbit for one year (Ott et al., 2020). *D. radiodurans* encodes for the same DNA repair machinery found in most studied organisms, like base excision repair, nucleotide excision repair, double-stranded break repair and mismatch repair (White et al., 1999). On top of the more common DNA repair machinery *D. radiodurans* also possesses some unique DNA repair proteins. These include the DdrA, DdrB, DdrC, DdrD, and PprA, which are highly expressed upon radiation exposure and function in DNA repair of radiation-induced damages (Banneville et al., 2022; Harris et al., 2004; Norais et al., 2009; Selvam et al., 2013; Szabla et al., 2024). Its highly effective DNA damage repair systems allow *D. radiodurans* to repair hundreds of double-stranded breaks per cell and remain viable (Cox & Battista, 2005).

D. radiodurans can repair its genome after it is fragmented into hundreds of DNA pieces by exposure to high radiation levels. This repair is achieved through a process called 'extended synthesis-dependent strand annealing' (ESDSA) (Zahradka et al., 2006). The starting point of ESDSA is a large number of DNA fragments resulting from a large number of double-stranded DNA breaks (Zahradka et al., 2006). The fragments are processed by an exonuclease protein to produce 3' overhangs. RecA binds the single-stranded overhangs in a RecFOR-dependent mechanism, which mediates the binding of the single-stranded regions to homologous parts of an overlapping fragment (Bentchikou et al., 2010). Similar to homologous recombination a D-loop is formed by the invading strand with the homologous region and the broken strand is re-synthesised by Polymerase I or III based on the template strand (Slade et al., 2009). The extensions run until the end of the template fragment and fragments now have longer single-stranded overhangs. These steps may be repeated multiple times until the fragments find a complementary strand they can anneal to. Alternatively, during single-strand annealing (SSA), the single-stranded overhangs of the initial fragments anneal to the complementary single-stranded overhangs of another fragment. In both pathways after overhang annealing, the fragments are connected. To restore the genome the long linear double-stranded fragments or linear intermediates join together through a crossover step similar to that in homologous recombination (Zahradka et al., 2006).

1.5 Biotechnological Applications of Extremophilic Enzymes

Metagenomic sequencing opens access to a mostly unexplored reservoir of novel enzymes from extreme environments, like hot springs, deserts, hypersaline environments, and cold environments like the Dry Valleys. These environments harbour a range of biocatalysts with activity adapted to these extreme conditions (Adrio & Demain, 2014; Ferrer et al., 2009; Handelsman, 2004; Steele et al., 2008). Exploration of those biocatalysts has largely been focused on those suitable for industrial applications, like amylases (Ariaeenejad et al., 2021), cellulases (Ilmberger et al., 2012), proteases (Pessoa et al., 2017) and lipases (Yan et al., 2017). The Dry Valleys have a high level of endemism and could harbour a range of unique and novel cold-adapted proteins.

Cold-adapted proteins are of particular interest for biotechnological applications due to their activity at low temperatures. Allowing catalysis at low temperatures, reduces the amount of heat input required, making them more cost-effective compared to mesophilic and thermophilic enzymes, which require high temperatures for their activity (Aliyu et al., 2017). This is particularly useful in the detergent, food, paper and textile industries (Aliyu et al., 2017). Cold-adapted proteins often have relatively low melting temperatures, which makes it easy to inactivate those enzymes through heating.

DNA repair proteins are commonly used biotechnological tools. They have applications in recombinant DNA technologies and molecular cloning, polymerase chain reaction and ligase chain reaction, genome editing, mutagenesis, DNA sequencing, whole genome amplification, synthetic biology, and molecular diagnostics among others. Cold-adapted DNA repair and replication proteins could be valuable for biotechnological applications and their ability to function at low temperatures makes them especially useful when working with heat-labile substrates.

1.6 Research Aim and Objectives

This thesis is part of a project, which aims to provide insight into mechanisms enabling the Dry Valley inhabiting microorganisms to survive the DNA-damaging conditions in this environment. Bacteria there survive and even grow under extremely DNA-damaging conditions. Therefore, the hypothesis is that highly effective and potentially novel DNA repair proteins and pathways are used by those bacteria. Previous to the commencement of this project the metagenomes of a subset of soil samples from the Dry Valleys, which were collected as part of the most comprehensive interdisciplinary survey of terrestrial Antarctic ecosystems (Lee et al., 2019) were sequenced and analysed as described in Rzoska-Smith et al. (2023). Analysis of these metagenomes showed that 5.5% of the annotated genes belong to the COG category 'Replication, Recombination and Repair', which is higher than the proportion of the same category found in forest soils, other soil communities, freshwater sediments and the genomes of human pathogen isolates (Rzoska-Smith; et al., 2023).

To further characterise the DNA repair in Dry Valley organisms, hmmsearch for distant homology to known DNA repair proteins and the predicted protein-coding genes from the metagenomes was done. This revealed the presence of multiple NucS/EndoMS endonucleases, Holliday junction resolvase-type endonucleases, UV-damage excision endonucleases, light-activated photolyase proteins, *Deinococcus* radiation resistance proteins DdrA and DdrB, as well as multi-domain families such as ATP-dependent DNA ligases, excision-repair glycosylases and mismatch-recognition proteins (Rzoska-Smith; et al., 2023). Biochemical characterisation of multiple LigB-type ligases and NucS-type nucleases were described in the thesis by Rzoska-Smith (2023) and the publication by Rzoska-Smith; et al. (2023). One of the ligase proteins described in the thesis by Rzoska-Smith (2023) is part of a nuclease-ligase fusion protein (NucDom and LigDom, respectively), which is named DV1-1. The characterisation of the full-length DV1-1 protein and the ligase domain of this protein was done by Rzoska-Smith (2023). The purification and characterisation of the nuclease domain of this protein were done by myself and my co-supervisor Dr Rzoska-Smith prior to the start of this project. For this purpose, the purification was optimised using different constructs of the protein and the protein activity was characterised using different DNA substrates, which gave insight into the activity of the proteins, as will be described in

more detail in section 4. This project has contributed to the characterisation of the NucDom by showing the DNA-binding ability of the protein and characterising the protein activity with different Zn metal concentrations, and over a time and temperature gradient. Specifically, the activity of the protein at extremely low temperatures was characterised. Furthermore, crystallisation of the NucDom was attempted.

Additionally, an archaeal Holliday junction resolvase-type protein homologue named DV-Hjc was identified as described in Rzoska-Smith (2023). DV-Hjc belongs to the currently uncharacterised UPF0102 protein family. Preliminary characterisation of this protein was undertaken by my supervisor Dr Williamson and was published in Rzoska-Smith; et al. (2023). This showed that this protein binds DNA, however no activity was observed with DNA. X-ray diffraction data of crystallised DV-Hjc protein was obtained by Dr Williamson. This project contributes to the characterisation of DV-Hjc by characterising and optimising the protein's activity with DNA substrates and by solving the structure of DV-Hjc using X-ray diffraction data. Furthermore, the growth of an *E. coli* knockout of the UPF0102 protein was characterised under UV and H_2O_2 treatment to determine the function of the UPF0102 protein *in vivo*.

2 CHAPTER TWO: METHODOLOGY

2.1 Recombinant Protein Expression and Purification

The optimised expression conditions, bacterial strains and plasmids are summarised in Supplementary Table 1, Supplementary Table 2 and Supplementary Table 3, respectively.

2.1.1 Cloning and Transformation of Expression Vectors

The genes of interest were selected as described in (Rzoska-Smith; et al., 2023). The clonal genes expressing the proteins of interest were codon optimised for recombinant protein expression in *E. coli* and ordered from Twist Bioscience in the pENTR plasmid and include a cleavage recognition site for the tobacco mosaic virus protease (TEV-protease), allowing for the removal of the MBP-tag and His-tag by TEV-protease. Clonal genes were sub-cloned from pENTR into the pHMGWA (GenBank #Eu680841), which includes an MBL-tag and 6xHis-tag at the N-terminus of the protein, using the Gateway™ LR Clonase™ Kit (Thermo Fisher). For the reaction, 2 µL of pENTR plasmid with the clonal gene and 2 µL of empty pHMGWA plasmid were combined with 1 µL of the Gateway™ LR Clonase™ (Thermo Fisher) and incubated at room temperature overnight. This was followed by treatment of the reaction with 1 µL of Proteinase K (Invitrogen) at 37°C for 10 min. The resulting expression plasmid was then transformed into chemically competent DH5α cells for storage and plasmid propagation following the protocol for the transformation of chemically competent cells described in Section 7.1.2.. Cloning success was determined via colony PCR. For colony PCR a single colony was selected and resuspended in 6 µL MQ H2O. PCR reactions were set up with 4 µL of 5x HOT FIREPol® Master Mix (Solis BioDyne), 0.6 µL T7 forward and reverse primers (Supplementary Table 13), and 2 µL of the resuspended cells. The PCR reaction was run on the setting shown in Table 2. The PCR results were analysed by agarose gel electrophoresis on 1 x SYBR Safe^{TM} DNA gel stain (Invitrogen, USA) strained 1% agarose gels in 1x TAE buffer (40 mM Tris-acetate, 20 mM EDTA). The gels were visualised on an iBright imaging system (Invitrogen) using the nucleic acids setting. The product size was determined by comparison to the 1kb-Plus DNA ladder (Invitrogen) as a molecular standard. Successful transformants were used to set up overnight cultures in 50 mL LB media (10 g/L bacto-peptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7) with 50 µg/mL ampicillin, which were grown at 37°C with 160 rpm shaking for 16 hours. For long-term storage, the overnight cultures were combined with glycerol at a 1:1 ratio and stored at -80°C.

The plasmid was extracted from the DH5α cells using the QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. The pHMGWA plasmid containing the gene of interest,

as determined via PCR, was then transformed into the optimal *E. coli* DE3 expression strains for recombinant protein expression as previously determined via small-scale expression trials, which are Origami™ (DE3) for the DV-Hjc protein and BL21(DE3) pLysS for the DV1-1 NucDom protein. Transformants were grown on LB Agar plates overnight and successful transformation was determined via colony PCR using the T7 forward and reverse primers as described above. Colony PCR results were analysed as described above. As described above, successful transformants were grown in 50 mL LB media cultures overnight. For long-term storage, the overnight cultures were combined with 50% LB:glycerol at a 1:1 ratio and stored at -80°C.

Table 2: Colony PCR conditions using T7 primer for pHMGWA plasmid

Step	**# of PCR cycles**	**Temperature (in °C)**	**Duration**
Initial denaturation	1	95	15 min
Denature	30	95	30 sec
Anneal		51	20 sec
Extend		72	30 sec
Final Extension	1	72	10 min

2.1.2 Large Scale Expression

Starter cultures for purification were grown from successful transformants of the clonal genes in *E. coli* strains for recombinant expression. Starter cultures were inoculated into 30 mL LB media (10 g/L bacto-peptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7) with 50 µg/mL ampicillin with the clonal gene transformants in the *E coli* expression strains. After overnight incubation, the starter cultures were used to inoculate 1 L cultures with TB media (24 g/L yeast extract, 12 g/L tryptone, 4 mL/L glycerol), phosphate salts (2.31 g/L kH_2PO_4, 12.54 g/L $k_2H_2PO_4$) and 50 µg/mL ampicillin. Cultures were grown at 37°C with 160 rpm shaking until they reached an OD_{600} of 0.4. The incubation temperature was reduced to 15°C for the DV-Hjc protein and to 30°C for the DV1-1 NucDom protein for 30 min before the cultures were induced to 0.5 mM with IPTG (Isopropyl-β-D-thiogalactopyranosid) to induce protein expression. After a 16-hour incubation period, the cells were pelleted through centrifugation at 4°C and were either used directly for protein purification or stored at -30°C for later use.

2.1.3 Protein Purification

Purifications were carried out on an Äkta Prime FPLC at 4°C or a BioRad NGC Chromatography system using the ice bath fraction collector to keep the protein fractions at a low temperature.

Cell Lysis

The pelleted cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 750 mM NaCl, 1 mM $MgCl_2$, 5% v/v glycerol) and lysed through sonication on ice for 3 min, with an amplitude of 12 with 1 sec on, and 1 sec off pulses. The cell lysate was pelleted through centrifugation at 14,000 xg for 45 min and the lysate was passed through a 0.2 µm Minisart® syringe filter (Sartorius).

Immobilised Metal Affinity Chromatography

Using the Äkta Prime FPLC or BioRad NGC Chromatography system the filtered lysate was applied to a high-performance immobilised metal affinity chromatography (IMAC) column (Cytiva), which was pre-equilibrated with 25 mL Buffer A (50 mM Tris pH 8.0, 750 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole). The his-tagged protein was eluted over 75 mL with a gradient of 0% to 100% of Buffer B (50 mM Tris pH 8.0, 750 mM NaCl, 5% (v/v) glycerol, 500 mM imidazole). For the DV1-1 NucDom, an additional wash step with 8% Buffer B was done before the Buffer B gradient to remove contaminants.

MBP- and His-Tag Removal using TEV Protease and Reverse IMAC

For His-tag and MBP-tag removal, fractions containing the protein of interest, as determined by analysis via SDS-PAGE gel electrophoresis, were pooled and exchanged into Buffer C (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, 5% (v/v) glycerol) using a HiPrep 26/10 Desalting column (Cytiva). The pooled protein was aliquoted into 1 mL batches, which were incubated with 0.025 mg/mL of TEV protease at 4°C overnight on a tilting shaker. For the DV-Hjc protein only the pooled protein was diluted to below 0.8 mg/mL before incubation with TEV protease. The cleaved protein was pooled and applied to an IMAC column (Cytiva), which was preequilibrated with Buffer C. The protein was eluted over 100 mL with a 0% to 100% Buffer B gradient. For the DV-Hjc protein, the now untagged protein of interest was collected from the flow-through of the IMAC (i.e. reverse IMAC). For the DV1-1 NucDom, an additional wash step with Buffer A, was done to elute the protein before the Buffer B gradient step. The protein-containing fractions were determined through electrophoresis on a 15% or 12% acrylamide SDS-PAGE gel for the DV-Hjc and the DV1-1 NucDom, respectively.

Gel Filtration Chromatography

The protein-containing fractions were pooled and up-concentrated to a volume below 5 mL while maintaining a concentration below 0.2 mg/mL for the DV-Hjc protein. The protein was then applied to a Superdex75 16/600 (S75) or Superdex200 16/600 (S200) column for DV-Hjc or DV1-1 NucDom, respectively. The fractions containing the protein of interest, as determined by electrophoresis on a 15% or 12% acrylamide SDS-PAGE gel for DV-Hjc and DV1-1 NucDom, respectively. For the DV-Hjc

protein only, the pooled protein was exchanged into a storage buffer (50 mM Bis-Tris pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 1 mM DTT) using a HiPrep 26/10 Desalting column (Cytiva). The pooled proteins were concentrated and stored at -80°C.

2.2 *In vitro* Assays of Nuclease Activity

2.2.1 Differential Scanning Fluorimetry

Differential scanning fluorimetry was used to determine the thermal stability of DV-Hjc and DV1-1 NucDom over a range of pH levels as described by Ericsson et al. (2006). A pH range of 5.0 to 9.5 was achieved by diluting DV-Hjc protein into Britton-Robinson universal buffers to give a final concentration of 0.2 mg/mL DV-Hjc protein. SYPRO Orange (Life Technologies) dye was added to a concentration of 1.2x. Using a RotoGene Q thermocycler (Qiagen) the unfolding of the protein was measured continuously through fluorescence measurements (excitation 470, emission 550 nm) over a temperature range of 25°C to 98°C.

2.2.2 Short Oligonucleotide Assay

A range of DNA substrates was used to determine the function of DV-Hjc and DV1-1. These include short and long fluorescently labelled DNA substrates and plasmid DNA.

Short Oligonucleotide Substrate Synthesis

Gel-based activity assays with modified synthetic oligonucleotides were used to determine the activity of both DV-Hjc and DV1-1 NucDom. The substrates were designed and assembled as described in Stelzer et al. (2024). Oligonucleotides listed in Supplementary Table 9 were ordered from IDT and used to generate the substrates listed in Supplementary Table 10. Oligonucleotides were annealed to give a final concentration of 80 nM for the 5'FAM labelled oligonucleotide, and 112 nM for the complement oligonucleotide.

Nuclease Activity Assay Reaction Setup

Reactions were set up in PCR tubes with a final volume of 25 µL. The DV1-1 NucDom protein assays were done in the 'standard assay setup' with 50 mM Tris pH 8, 50 mM NaCl, and 10 mM DTT (Supplementary Table 11). The DV-Hjc protein assays were carried out in 'DV-Hjc-specific' assay conditions with 50 mM HEPES (pH 6.5), 50 mM NaCl, 0.1 mg/mL BSA (bovine serum albumin), 0.5 mM DTT and 6% glycerol (Supplementary Table 12). Where added, the concentration of $MgCl_2$, $MnCl_2$ and $ZnCl_2$ as well as any other changes to this standard setup are indicated with the corresponding

results. To initiate the reaction 2.5 µL protein was added at varying final concentrations to the assay reactions, which were incubated at temperatures and times as indicated with the corresponding results. The reactions were terminated by the addition of quench buffer to a final concentration of 25% formamide, 0.1% SDS, 20 mM EDTA and 0.01% bromophenol blue and heating to 95°C for 5 min.

Assay Product Analysis via TBE-Urea PAGE

The assay products were analysed via TBE-Urea PAGE denaturing electrophoresis on 10% to 20% denaturing TBE urea gels (10-20% acrylamide, 7 M Urea, 1x TBE), which were run with 47°C heating at 10 mA per gel. For the 3'tail and 5'tail substrates loading dye to a final concentration of 5% glycerol, 20 mM EDTA and 0.01% bromophenol blue were added to the reaction. The reactions were electrophoresed with cooling at 60 V on 10% TBE gels (10% acrylamide, 1x TBE), where indicated a 7% TBE stacker (7% acrylamide, 1x TBE) was used. Gels were visualised on an iBright imager (Invitrogen) using a filter set that gives excitation at 495/519 nm for the FAM-labelled substrates.

2.2.3 pUC19 Plasmid-Based Assay

Nuclease assays with the pUC19 plasmid were used to assess the activity of the DV-Hjc protein on supercoiled plasmid DNA. The plasmid was stored in *E. coli* DH5α cells and extracted from those using the QIAprep® Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. DV-Hjc protein was incubated for varying times and temperatures with 150 ng/uL pUC19 plasmid, 50 mM HEPES (pH 6.5), 50 mM NaCl, 0.1 mg/mL BSA, 0.5 mM DTT, 6% glycerol, and $MgCl_2$, $MnCl_2$ or $ZnCl_2$. The details of the reaction conditions are indicated with the corresponding results. The reaction was terminated by incubation with 0.1% SDS and 10 mM EDTA for 10 min at 37°C. For reaction analysis, 5 µL of DNA loading dye (30% glycerol (v/v), 0.25% bromophenol blue dye (w/v), and 0.25% xylene cyanol FF dye (w/v)) was added per 20 µL reaction and the reactions were electrophoresed on thiazole orange stained 1% agarose/ 1x TAE (40 mM tris acetate, 10 mM EDTA) gels. Agarose gels were visualised on the iBright imager (Invitrogen) on the nucleic acid gels setting. The ImageJ program was used to quantify the band intensity of the products and substrates on the gels and the percent product was calculated with the formula below. The results were visualised using GraphPad Prism 10.

$$\%\ \text{Product} = \frac{P}{(P+S)} \times 100$$

2.2.4 DNA Binding Assay using Electrophoretic Mobility Shift Assay

The binding affinity of DV-Hjc and DV1-1 NucDom were tested using electrophoretic mobility shift assay (EMSA). Binding assays were carried out using the substrates in Supplementary Table 10, which

were assembled as described by Stelzer et al. (2024) with the DNA oligonucleotides listed in Supplementary Table 9.

Binding Assay Reaction Setup

DV-Hjc protein was incubated with the DNA substrates in 50 mM HEPES (pH 6.5), 50 mM NaCl, 0.1 mg/mL BSA, 0.5 mM DTT, 6% glycerol and 5 mM EDTA for 30 min on ice. DV1-1 NucDom protein was incubated with the DNA substrates in 50 mM Tris (pH 8.0), 50 mM NaCl, 10 mM DTT and 5 mM EDTA for 30 min at 10°C. For the DV1-1 NucDom protein only, an additional cross-linking step was done to achieve protein binding. After the 30-min incubation period, 1% formaldehyde was added to the reaction followed by a 10 min incubation at 10°C. 15 mM Tris was added to the reaction.

For all binding assays, loading dye to 100 mM EDTA, 0.25% (v/v) glycerol and 0.25% bromophenol blue was added to the reaction, before electrophoresis of the reactions on 10% TBE gels (10% acrylamide, 1x TBE). Any changes to this setup are specified with the results. The gels were visualised as described for UREA-TBE gels in Section 2.2.2..

2.3 Assay of Nuclease Activity *in vivo*

To determine the function of the DV-Hjc protein *in vivo*, *E. coli* K-12 mutants from the Keio collection (listed in Supplementary Table 2) (Baba et al., 2006) were subjected to high levels of UV radiation and H_2O_2. The DV-Hjc homologue knockout from the Keio collection, which is the *yraN* gene knockout (JW3117) was used for this purpose. Additionally, based on the literature, a knockout highly sensitive to the relevant stressors UV or H_2O_2 tested, and the wild-type *E. coli* K-12 was used as a negative control.

2.3.1 UV Treatment

The UV treatment method was adapted from Sargentini et al. (2016) to fit the purpose of this research. Based on the results from Sargentini et al. (2016) the *ybaB* gene knockout (JW0460) was used as a positive control for the UV treatment. Each strain was grown overnight in LB medium (at 37°C with 25 mg/mL kanamycin to select for the knockout strains, except for the wild type, which was grown without any antibiotics. The overnight cultures were diluted 1:100 in LB and grown at 37°C with 160 rpm shaking until they reached an OD_{600} reading of 0.4. 100 µL of each culture was placed in UV transparent 96 well plates and treated with 10 Jm^{-2} of UV radiation using a UV cross-linker (Model BLX 254 UV Crosslinker, Life Technologies). Serial dilutions of the cultures were done in 0.25% sterile saline solution, which were incubated on LB agar plates for 16 hours to determine the CFU/mL. Data

for the continuous growth of UV-treated cells was obtained by treating cells with UV over 30 min, but otherwise using the same protocol as described. The experiment was done in triplicates and untreated samples were included for each, which were treated as described apart from the UV treatment.

2.3.2 H_2O_2 Treatment

The method for the H_2O_2 treatment was based on that of Dhawale et al. (2021), which was adapted to fit the purpose of this research. Based on the same article, the *ruvC* gene knockout strain (JW1852) was used as a positive control for the H_2O_2 treatment. Overnight cultures of the WT, *yraN* (JW3117) or *ruvC* (JW1852) gene knockouts were grown for 16 hours at 37°C with 160 rpm shaking in LB media. The overnight cultures were diluted 1:100 in 100 mL LB media and grown at 37°C with 160 rpm shaking until an OD_{600} of 0.4 was reached. Each culture was then split into two 50 mL cultures, one of which was treated with 3 mM H_2O_2 and the other of which was left untreated. The OD600 was continuously measured every 15 min for 4.5 hours. After 90 min 2 mL of each culture was transferred into Eppendorf tubes, pelleted and resuspended to 2 mL in LB media twice to remove the H_2O_2. The cells were diluted in 0.25% sterile saline solution and 100 µL of each culture was plated on LB agar plates, which were grown for 16 hours to determine the CFU/mL.

For both treatments, the survival and the fold difference in survival were determined as follows

$$\%\, Survival \frac{CFU/ml\,(Treated)}{CFU/ml\,(Untreated)}$$

$$Fold - difference = \frac{\%\, Survival\,(WT)}{\%\, Survival\,(knockout)}$$

2.3.3 DV-Hjc Transformation into *yraN* Gene Knockout *E. coli* Cells

To test the effect of introducing the DV-Hjc protein into the *yraN* gene knockout strains to replace the YraN protein, the DV-Hjc protein-coding gene was cloned into the pProEx plasmid via restriction cloning.

Introduction of BamHI Restriction Sites to DV-Hjc Gene Construct

Primers (p.20240423.1 and p.20240423.2), listed in Supplementary Table 13, which include restriction sites for BamHI restriction enzyme (RE) and anneal to both ends of the DV-Hjc gene, were designed in Geneious Prime.

PCR was used to amplify the DV-Hjc coding gene in pHMGWA plasmid and introduce the BamHI restriction sites at both ends of the gene. PCR was set up with 1x HOT FIREPol® Master Mix (Solis

BioDyne), 0.3 µM forward primer, 0.3 µM reverse primer (p.20240423.1 and p.20240423.2, listed in Supplementary Table 13) and 100 ng template DNA and run at the conditions listed in Table 3. The PCR product was analysed on a 1 x SYBR SafeTM DNA gel stain (Invitrogen, USA) stained 1% agarose gel in 1x TE buffer. The product was retrieved from the gel and purified using the QIAquick® Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The cleaned-up product was re-amplified using the same PCR conditions as before and the PCR product was again cleaned up using the QIAquick® PCR Purification Kit (Qiagen) as per the manufacturer's instruction.

Table 3: PCR conditions for p.20240423.1 and p.20240423 primers with DV-Hjc template DNA

Step	**# of PCR cycles**	**Temperature (in °C)**	**Duration**
Initial denaturation	1	95	15 min
Denature	30	95	30 sec
Anneal		68	20 sec
Extend		72	30 sec
Final Extension	1	72	10 min

Restriction Digest of DV-Hjc and pProEx

DV-Hjc with BamHI restriction sites and pProEx plasmid were digested separately with 2 µL of BamH1 restriction enzyme, 1x rCutSmart™ Buffer (New England Biolabs), and 1 µg DNA via incubation at 37°C for 30 min. The restriction digest products were cleaned up using the QIAquick® PCR Purification Kit (Qiagen) and ligated using T4 DNA ligase. The T4 DNA ligase reaction was set up with 1x T4 DNA Ligase Buffer, 50 ng pProEx DNA, 37.5 ng DV-Hjc DNA, and 1 µL per 20 µL reaction and incubated at 16°C overnight. The T4 ligase was heat-inactivated at 65°C for 10 min and transformed in *E. coli* DH5α cells (as described in Section 7.1.2) to amplify the plasmid. The DV-Hjc protein-coding gene in pProEx was then transformed (as described in Section 7.1.2) into chemically competent *yraN* knockouts (JW3117), which were prepared as described below. The presence of the DV-Hjc gene in the transformants was confirmed via colony PCR. A single colony was selected for colony PCR and was resuspended in 6 µL MQ H_2O. PCR reactions were set up with 4 µL of 5x HOT FIREPol® Master Mix (Solis BioDyne), 0.6 µL forward and reverse primers (p.20240423.1 and p.20240423.2, listed in Supplementary Table 13), and 2 µL of the resuspended cells. The PCR was carried out according to the conditions outlined in Table 3, with a modification to the extension step, extending the time to 60 sec. The PCR reaction was analysed on an agarose gel as described previously.

Production of Chemically Competent *E. coli YraN* Gene Knockout Cells

Chemically competent *yraN* gene knockout *E. coli* (JW3117) cells were prepared by growing overnight cultures of the cells in LB medium (10 g/L bacto-peptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7) with 180 rpm shaking at 37°C. Overnight cultures were used to inoculate 500 mL of LB medium, which was incubated at 37°C with 25 rpm shaking until the culture reached an OD_{600} of 0.4. The cells were pelleted by centrifugation of 3400 xg for 10 min. The pelleted cells were resuspended in 150 mL ice-cold 100 mM $CaCl_2$/10% glycerol solution. The resuspended cells were pelleted again via centrifugation at 3400 xg for 10 min and the pellets were resuspended in ice-cold 100 mM CaCl2/10% glycerol solution again. The resuspended cells were incubated on ice for 25 min, divided into 200 µL aliquots and snap-frozen in liquid N_2. Competent cells were stored at -80°C until use.

2.4 Structural Characterisation

2.4.1 Crystallisation and Diffraction Data Analysis of DV-Hjc

Conditions from the crystal screens that showed favourable conditions for crystal formation of the DV-Hjc protein were used and further optimised in fine screens to generate crystals suitable for diffraction data collection via X-ray.

Purified DV-Hjc apoprotein (288 µM) was used for crystallisation via hanging drop diffusion in 34% PEG 4,000, 100 mM ammonium citrate pH 5.6 and 200 mM ammonium acetate. Crystals were grown at 18 °C and were mounted in cryoloops and flash-frozen in liquid N_2. Diffraction data was collected at the Australian Synchrotron MX2 beamline (Aragao et al., 2018). Diffraction data was integrated, scaled and merged using XDS and Aimless (Kabsch, 2010; Winn et al., 2011). The quality of the collected data sets was assessed using *Xtriage* in *Phenix* and analysis was continued with the highest quality dataset (Adams et al., 2010). A model of the DV-Hjc protein was generated using AlphaFold Colab, which was used as a search model in molecular replacement using *phenix.phaser* in *Phenix* (Jumper et al., 2021; McCoy et al., 2007). The starting model was refined in *phenix.refine* and by using the graphical interface of *Coot* to rebuild and refine the model manually (Afonine et al., 2012; Emsley et al., 2010). Solvent molecules were manually added to the model in *Coot*. Collection data and statistics are detailed in Table 4.

Table 4: Data collection and refinement statistics of the DV-Hjc protein used for structural characterisation of the protein. Values for the highest-resolution shell are shown in parentheses.

Wavelength	0. 953732 Å
Resolution range	44.94 – 1.739 (1.76 – 1.74)

Space group	P 21 21 2
Unit cell	57.739 71.56 29.961 90 90 90
Total reflections	156406 (8523)
Unique reflections	22518 (1218)
Multiplicity	6.2 (4.0)
Completeness (%)	99.81 (97.36)
Mean I/sigma(I)	5.03 (1.54)
Wilson B-factor	17.33
R-merge	0.4823 (0.6659)
R-meas	0.5209 (0.7697)
R-pim	0.1955 (0.386)
CC1/2	0.887 (0.589)
CC*	0.97 (0.861)
Reflections used in refinement	13323 (443)
Reflections used for R-free	0 (0)
R-work	0.1869 (0.2358)
R-free	0.2225 (0.2816)
Number of non-hydrogen atoms	1126
macromolecules	965
ligands	13
solvent	148
Protein residues	117
RMS(bonds)	0.009
RMS(angles)	0.89
Ramachandran favoured (%)	97.35
Ramachandran allowed (%)	1.77
Ramachandran outliers (%)	0.88
Rotamer outliers (%)	2.20
Clash score	5.58
Average B-factor	24.15
macromolecules	22.96
ligand	18.65
Solvent	32.35

2.4.2 Crystallisation Screening

To determine the optimal conditions for protein crystallisation of DV-Hjc and DV1-1 NucDom apoprotein and for co-crystallisation of the proteins with DNA substrates, crystallisation conditions were screened using the PEG Rx (HR2-086) and Natrix 2 (HR2-116) Hampton Research Crystallisation Screens. To set up the screens 100 µL of the crystallisation reagents were transferred into a 96-well Intelli-Plate (Hampton Research). Using a Mosquito® crystallisation robot (TTP LabTech Ltd, UK), 100 nL of protein and 100 nL of the crystallisation reagent were transferred into the crystallisation well of the plates. The plates were sealed with ClearsealTM film (Hampton Research, USA) and kept at 18°C.

Protein-DNA Co-crystallisation

To determine the structure of the DNA-protein complexes protein-DNA co-crystallisation was attempted. The DNA oligonucleotides for this were purchased from IDT (Integrated DNA Technologies) and were resuspended in annealing buffer (50 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) to 9 mM. To form DNA duplexes oligonucleotides were combined at a ratio of 1:1 and heated in a 1 L boiling water bath until cooled to room temperature, to anneal the oligonucleotides and form a 4.5 mM DNA duplex. Purified DV-Hjc (198 nM) or DV1-1 NucDom (181.52 µM) was incubated on ice with the DNA duplex or single-stranded DNA to give a ratio of 1:1.2 (protein:DNA) and 5 mM EDTA for 30 min to form a DNA-protein complex.

DV-Hjc-DNA Co-crystallisation

DNA substrates for co-crystallisation were produced and incubated with the protein as described above. Based on the results of the high-throughput crystal screens, the protein-DNA complex was used for crystallisation in the hanging drop diffusion method with the protein-DNA complex and the crystallisation reagents at a ratio of 1:1 (1 µL to 1 µL) or 2:1 (2 µL to 1 µL). The crystallisation reagents listed in Table 5 were tested for this purpose. These fine screens were left at 18°C. Where crystals formed, they were mounted in crystalloops and soaked in cryoprotectant solution (crystallisation reagent and 12% ethylene glycol or 12% glycerol) for 15 sec and then flash frozen in liquid N_2, or the crystals were flash frozen in liquid N_2 directly.

Table 5: DV-Hjc crystallisation conditions for co-crystallisation with DNA.

Well	Composition
1A	1M Succinic acid pH 7.0; 0.1 M BICINE pH8.5; 26% PEG 550
1B	1M Succinic acid pH 7.0; 0.1 M BICINE pH8.5; 28% PEG 550

1C	1M Succinic acid pH 7.0; 0.1 M BICINE pH8.5; 30% PEG 550
1D	1M Succinic acid pH 7.0; 0.1 M BICINE pH8.5; 32% PEG 550
2A	0.2 M L-Proline; 0.1 M HEPES pH 7.5; 20% PEG 1,500
2B	0.2 M L-Proline; 0.1 M HEPES pH 7.5; 22% PEG 1,500
2C	0.2 M L-Proline; 0.1 M HEPES pH 7.5; 24% PEG 1,500
2D	0.2 M L-Proline; 0.1 M HEPES pH 7.5; 26% PEG 1,500
3A	80 mM Potassium chloride; 40 mM Sodium cacodylate trihydrate pH 7.0; 56% (+/-)-2-Methyl-2,4-pentanediol; 10 mM Spermine tetrahydrochloride
3B	80 mM Potassium chloride; 40 mM Sodium cacodylate trihydrate pH 7.0; 58% (+/-)-2-Methyl-2,4-pentanediol; 10 mM Spermine tetrahydrochloride
3C	80 mM Potassium chloride; 40 mM Sodium cacodylate trihydrate pH 7.0; 60% (+/-)-2-Methyl-2,4-pentanediol; 10 mM Spermine tetrahydrochloride
3D	80 mM Potassium chloride; 40 mM Sodium cacodylate trihydrate pH 7.0; 62% (+/-)-2-Methyl-2,4-pentanediol; 10 mM Spermine tetrahydrochloride

2.4.3 Crystallisation of DV1-1 NucDom

Based on the results of the crystal screening of the DV1-1 NucDom, fine screens of DV1-1 NucDom apoprotein and protein co-crystallised with DNA substrates were used. DNA substrates for co-crystallisation were produced and incubated with 182 µM of DV1-1 NucDom as described in Section 2.4.1. Based on the screening results, apoprotein and DNA-protein complex were mixed with the crystallisation reagents listed in Table 6 at a ratio of 1:1 (1 µL protein to 1 µL crystallisation reagent) for crystallisation in the hanging drop diffusion method at 18°C.

Table 6: DV1-1 NucDom crystallisation conditions for co-crystallisation with DNA.

Well	**Composition**
1A	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
1B	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
1C	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride

1D	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 60% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
2A	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
2B	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
2C	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
2D	80 mM KCl_2; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 60% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
3A	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
3B	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
3C	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
3D	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 60% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
4A	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
4B	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
4C	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
4D	80 mM KCl_2; 20 mM Magnesium chloride hexahydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 60% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
5A	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
5B	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride

5C	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 6.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
5D	0.2 M Ammonium acetate; 10 mM Magnesium acetate tetrahydrate; 50 mM Sodium cacodylate trihydrate; 30% PEG 8000
6A	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 35% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
6B	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 45% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
6C	80 mM $NaCl_2$; 20 mM Barium chloride dihydrate; 40 mM Sodium cacodylate trihydrate pH 7.0; 55% (+/-)-2-Methyl-2,4-pentanediol; 12 mM Spermine tetrahydrochloride
6D	0.2 M Ammonium chloride; 10 mM Magnesium chloride hexahydrate; 50 mM HEPES Sodium pH 7.0; 30% PEG 8000

2.4.4 Protein Structure Predictions using AlphaFold3

Where no crystal structure of the DV1-1 and DV-Hjc proteins was available, the structures were predicted using the web-based AlphaFold3 developed by Google DeepMind and Isomorphic Labs (Abramson et al., 2024). The structures were visualised in PyMol Version 2.5.4. (Schrödinger). The overall structure confidence was judged based on the ipTM and pTM values provided by AlphaFold3 for each prediction. The pTM (predicted template modelling) score indicates the accuracy of the predicted fold and a score above 0.5 was interpreted as a fold that is likely to resemble the true protein structure (Abramson et al., 2024). The ipTM (interface predicted template modelling) score was used to assess the accuracy of the positions of the subunits of the prediction within the whole predicted complex. An ipTM score above 0.8 was interpreted as a confident prediction and values below 0.6 were interpreted as a likely failed prediction (Abramson et al., 2024)

3 CHAPTER THREE: DV-HJC

3.1 DV-Hjc Protein Background

3.1.1 Identification of a Holliday Junction Resolvase-like Nuclease from the Dry Valley metagenomes

Metagenomes from the Dry Valleys were searched for proteins with distant homology to known DNA repair proteins. This included a search for archaeal Holliday junction resolvase family enzymes (IPR002732). Sequence similarity networks were used to identify groups of homologous to this protein family, that are unique to the Dry Valley metagenomes as described in Rzoska-Smith; et al. (2023). The DV-Hjc protein was found in a unique Dry Valley metagenome cluster of Holliday junction resolving enzymes, at an edge threshold of 30%, which suggests that the sequence of this protein could have specific adaptations. Almost 3,000 proteins with an Hjc domain were found in the Dry Valley metagenomes by Hmm search. A sequence similarity network of those sequences with Hjc UniRef sequences at a 26% identity threshold resulted in six clusters containing only sequences from the Dry Valley metagenomes (Figure 20, A). Two clusters contained proteins from the UPF0102 domain (Figure 20, A, i and ii). In a sequence similarity network with Dry Valley sequences from clusters i and ii (Figure 20, A, i and ii) and UPF0102 protein sequences, the Dry Valley sequences separated into a cluster at a 64% identity threshold (Figure 20, B) (Rzoska-Smith; et al., 2023). A Hjc protein homologue (Gene ID: Ga0136632_1000003432) (DV-Hjc from hereafter) was selected from the Dry Valley sequences for further characterisation.

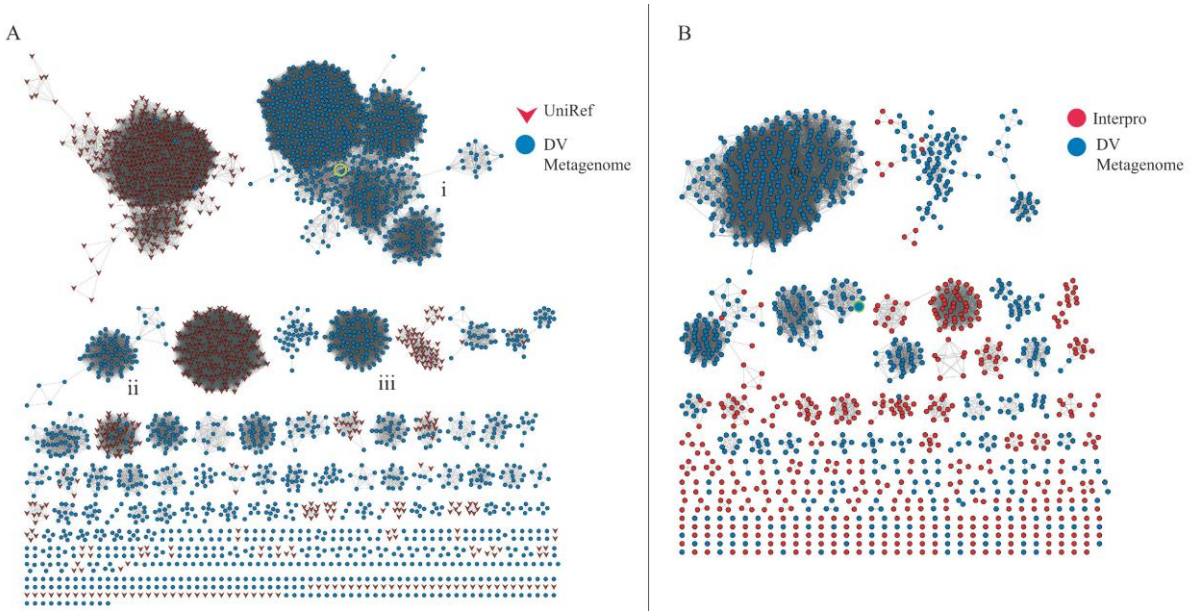


Figure 20: Sequences similarity networks for Hjc-type and UPF0102 domain proteins.
***A:** Sequence similarity network with Hjc-type protein UniRef (red arrows) and Dry Valley metagenome sequences (blue dots). i, ii, and iii indicate the cluster discussed in the text. The green circle indicates the node corresponding to DV-Hjc.*
***B:** Sequence similarity network of Dry Valley metagenome sequence hits to UPF0102 domain and Pfam members of the same domain. Pfam sequences are shown as red circles and Dry Valley metagenome sequences as blue circles.*

IMG predicts that the DV-Hjc protein-coding gene is part of the genome of *Cellulomonas timonensis* SN7, which belongs to the Actinomycetota phylum. The DV-Hjc protein was assigned to the Clusters of Orthologous Groups (COG) of 'replication, recombination and repair', and is a predicted endonuclease distantly related to archaeal Holliday junction resolvase. Annotation of DV-Hjc by sequence search against the Pfam database places DV-Hjc into the uncharacterised protein family UPF0102 (pfam02021). Several proteins with a UPF0102 domain are annotated as YraN transcription factors; however, DV-Hjc does not align well with YraN proteins. Furthermore, InterPro places the DV-Hjc protein in the tRNA endonuclease-like domain superfamily and the restriction endonuclease type II superfamily (Jones et al., 2014).

The UPF0102 protein domain is present in most bacterial phyla, with a notable absence from the *Firmicutes* (Figure 21, A). The UPF0102 protein was found in only 14% of the *Firmicute* genomes, mainly in the *Paenibacillales*. The UPF0102 domain can also be found in some archaeal species, but there appears to be no pattern to their distribution in archaea (Figure 21, B). The presence of the archaeal Holliday junction resolvase homologue protein (UPF0102) in a large proportion of bacterial species is curious since bacteria have the RuvC Holliday junction resolvase protein for resolving Holliday junctions during homologous recombination and DNA repair (Wyatt & West, 2014).

The UPF0102 family protein structure from *Rhodopseudomonas palustris* was solved by X-ray crystallography in 2009 and the coordinates were deposited in the PDB; however, no functional characterisation was conducted and there is no journal publication accompanying the structure (Osipiuk

et al., 2009). This is the only structure solved by X-ray crystallography of this protein family so far and there are no publications on the function of a protein with UPF0102 protein domain family.

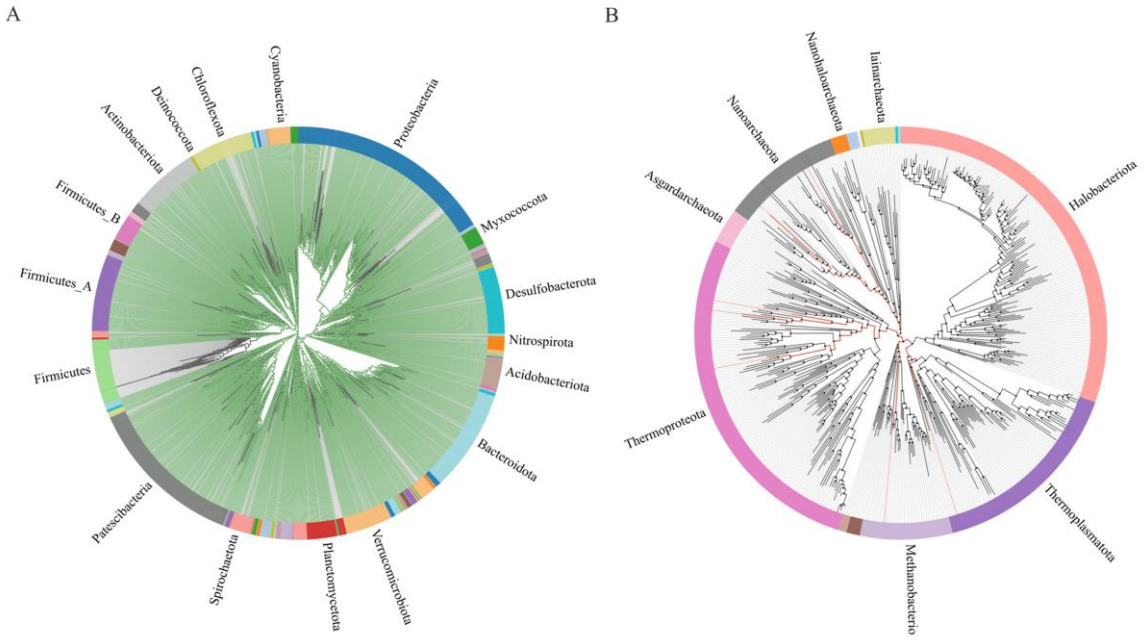


Figure 21: Distribution of the UPF0102 protein domain in bacteria and archaea.
The trees were generated using AnnoTree (Mendler et al., 2019).
***A:** For bacteria, the domain search was done at the family level with annotations at the phylum level, showing only the larger phyla.*
***B:** For archaea the domain search was done at the genus level with annotations at the phylum level, only showing the larger phyla.*

3.1.2 Holliday Junction Resolvases

DV-Hjc is a homologue of the archaeal Holliday junction resolvases (Hjc). Holliday junctions are an intermediate during DNA recombination and double-stranded DNA break repair by homologous recombination (Lilley, 2017). Holliday junctions are formed during the invasion of a homologous strand into a double helix, which results in four DNA helices, which meet at a single branchpoint, as shown in Figure 22. They form a junction between two DNA duplexes. The existence of Holliday junctions was first proposed by Robert Holliday, whose mechanism for gene conversion involved the formation of a Holliday junction (Holliday, 1964). Holliday junctions can exist in a stacked and open confirmation, as shown in Figure 22 (B and C, respectively). In solution, the conformation is mainly determined by metal ion availability. In the absence of cations, the junction is in an open conformation and in the presence of cations it is in a stacked conformation (Lilley & Clegg, 1993). In the stacked conformation, Holliday junctions exist in an antiparallel configuration not in a parallel conformation, as shown in Figure 22 (A and B) (Cooper & Hagerman, 1989; Lu et al., 1991; Murchie et al., 1989). The antiparallel configuration can differ depending on the antiparallel stacking partner. This configurational change can change the orientation of junction-resolving enzymes.

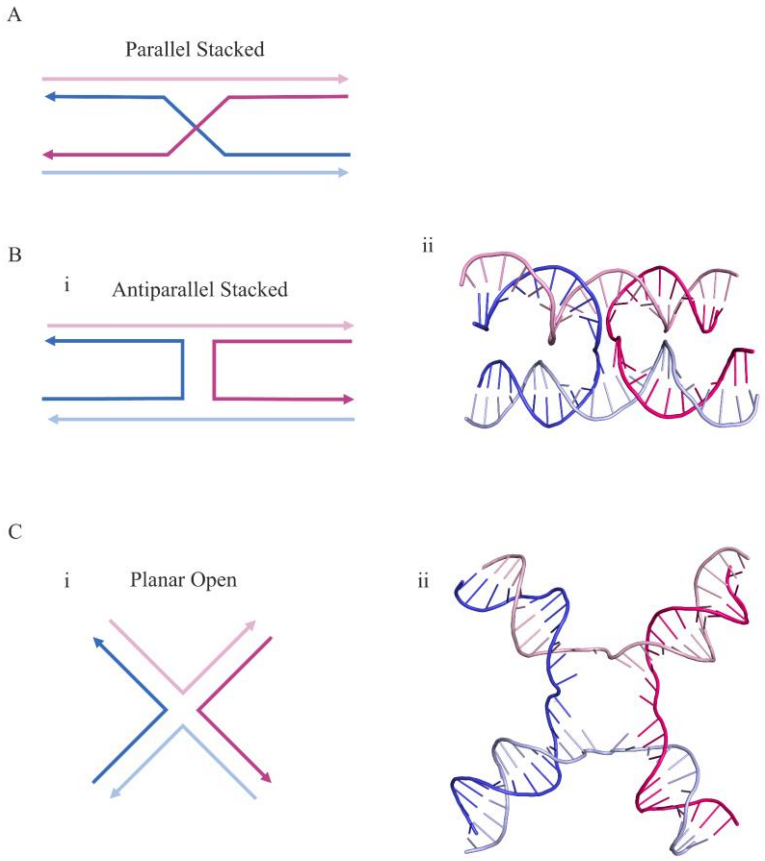


Figure 22: Conformation of Holliday junctions.
***A:** Illustration of the Holliday junction in a parallel stacked conformation.*
***B:** Holliday junction in an antiparallel stacked conformation shown illustrated (i) and with the 3D structure (2WJ0) from Carolis et al. (2009) (ii).*
***C:** Holliday junction in an open planar conformation shown illustrated (i) and with the 3D structure (7X5A) from Zhang et al. (2023).*

To produce intact and replicable DNA after a Holliday junction intermediate is formed during a DNA replication or repair, the junction must be resolved into two DNA duplexes. This resolution is carried out by Holliday junction resolving proteins, which are a group of structure-specific DNA endonucleases found in organisms from all domains of life. The most well-known Holliday junction resolving enzymes are, the bacterial RuvC, archaeal Holliday junction resolvases, T7 and T4 endonuclease, the human Gen1 protein and eukaryotic Cce1 proteins, which are summarised in Table 7 (Gorecka et al., 2013; Rass et al., 2010; Wyatt & West, 2014). The sequences and structures across these Holliday junction-resolving enzymes are not conserved. However, they share some common features. These features include a large amount of positively charged amino acids, which form large positively charged surfaces for DNA binding (Wyatt & West, 2014). Additionally, their active sites usually contain three to four acidic amino acids, required for nuclease activity and binding of a single metal ion (Wyatt & West, 2014).

Table 7: Summary of well-known Holliday junction resolving enzymes.

Organism	Enzyme Name	Oligomeric State	Junction Conformation	Sequence Specificity	Reference
Prokaryotes	RuvC	Homodimer	Open planar	5-A/TTT↓G/C-3′	(Górecka et al., 2019)
Archaea	Hjc Hje	Homodimer Homodimer	Open planar Stacked	No No	(Kvaratskhelia & White, 2000)
Eukaryotes	Gen1 (human) Cce1	Monomer Homodimer	Not determined Open planar	Weak preference for 2 guanine in a thymine-rich region 5'-CT ↓ -3’	(Shah Punatar et al., 2017) (White & Lilley, 1997)
Phage	T4 endo VII T7 endo I	Homodimer Heterodimer	Open planar Stacked	5’-AA/TTA↓G/CTC-3 No	(Hardie & Murray, 2018) (Hadden et al., 2007)

YraN Family Proteins

The DV-Hjc protein is homologous to YraN proteins, which are currently not well characterised but are annotated as a transcription factor. A preprint by Hardy et al. (2024) was released while the presented research was being undertaken, which suggests that the protein is involved in a homologous recombination process as part of bacterial natural transformation. During natural transformation, cells take up DNA from the environment and integrate it into their genome. External DNA integration into the host genome occurs in a process similar to homologous recombination (Hamilton & Dillard, 2006; Straume et al., 2015). The external DNA invades the host genome at the homologous region forming a D-loop. Hardy et al. (2024) suggest that the YraN protein resolves the Holliday junction formed during this invasion. The protein was tested *in vivo* in *Legionella pneumophila* and *Synechococcus elongatus*, where the efficiency of natural transformation was reduced by knocking out the yraN protein or mutations in the yraN protein by 40-fold in *L. pneumophila* and 120-fold in *S. elongatus*. Their results suggest that the yraN protein is involved in the process of chromosomal DNA integration rather than the uptake of the DNA from the environment (Hardy et al., 2024).

DV-Hjc Genomic Context

The DV-Hjc sits in a gene cluster with four proteins predicted to be involved in replication, recombination and repair, and five proteins predicted to be involved in translation, ribosomal structure and biogenesis, based on the COG classification (Figure 23). DV-Hjc is not in the same operon as those genes, however its proximity to them further suggests its DNA-modifying role. Based on a neighbourhood search for the DV-Hjc protein, homologues are often located next to a Mg-chelatase (77%) (PF13541, PF01078, PF13335), DNA recombination-mediator protein A (60%) (PF02481), a signal peptidase (60%) (PF10502), and RNase HII (59%) (PF01351). A similar synteny can also be seen in neighbourhoods of DV-Hjc homologues in the Dry Valley metagenomes. In some cases, the

RNase HII and UPF0102 domain are expressed as a multifunctional fusion protein (315 proteins in the UniProt database). In the case of DV-Hjc, an RNase HII and a Mg-Chelatase are positioned on either side of the protein (Figure 23).

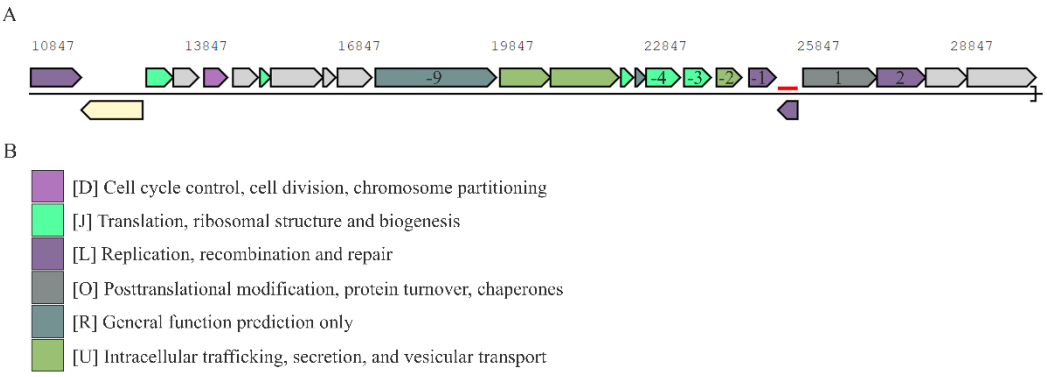


Figure 23: Gene contig of DV-Hjc.
***A:** Gene Contig of DV-Hjc. DV-Hjc is highlighted by the red line. The colours indicate the COG categories (**B**) the proteins belong to. Grey colour indicates the COG categories not listed in **B**. Numbered proteins refer to the Pfam predicted: -9 SMC_N, -4 tRNA (Guanine-1)-methyltransferase, -3 Ribosomal protein L19, -2 Peptidase_S26 -1 RNase_HII, 1 Magnesium chelatase, subunit ChlI, 2 DNA recombination-mediator protein A.*
***B:** Colour codes of COG categories of the protein shown in **A**.*

3.2 DV-Hjc Protein Purification

The DV-Hjc protein was recombinantly expressed in *E. coli* BL21 plysS cells and the protein was purified through multiple-step purification initially on the Äkta Prime FPLC which was stored in and operated at 4°C to keep a constant low temperature during protein purification. Due to the failure of the Äkta Prime, purification was carried out on a BioRad NGC purifier, which operated at room temperature. This was suboptimal for DV-Hjc purification and yielded lower protein concentrations due to decreased protein stability. To increase the DV-Hjc protein stability after purification the protein was transferred into a stabilising agent containing buffer (50 mM Bis-Tris pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). This led to an increase in the protein's shelf life of up to a week when stored at 4°C. However, it did not enable up-concentration of the protein to a concentration above 16 µM. Ideally, the Hjc is purified on a machine operating at 4°C and kept at this temperature throughout the purification process to enable up-concentration of the protein to higher concentrations. Additionally, changing DV-Hjc into the DV-Hjc-specific storage buffer increases the protein's stability for biochemical characterisation.

DV-Hjc Wild Type Purification

Crude *E. coli* cell lysate containing the recombinantly expressed MPB- and 6xHis-tagged DV-Hjc protein was loaded onto an IMAC column. The eluate was collected in 4 mL fractions and analysed via

SDS-PAGE gel electrophoresis. The protein eluted at about 0.4 M imidazole. DV-Hjc containing fractions contained several contaminating proteins (Figure 24, A).

The protein-containing fractions were pooled and changed into a suitable buffer for TEV-protease (Buffer C, Supplementary Table 7). The pooled fractions were diluted as described in Section 2.1.3, before incubation with TEV protease. The dilution of the protein ensured that the concentration of DV-Hjc after cleavage of the MBP-tag by TEV protease was below 0.2 mg/mL. At a concentration above 0.2 mg/mL, DV-Hjc precipitated following the removal of the MBP-tag by TEV protease. Following the overnight TEV incubation, a reverse IMAC elution was used to separate the DV-Hjc protein from the 6xHis-tagged MBP-tag and TEV protease. The eluate was collected in 2 mL fractions and analysed via SDS-PAGE gel electrophoresis. The reverse IMAC elution resulted in DV-Hjc protein with a small number of contaminating proteins (Figure 24, B).

To further purify the protein, protein-containing fractions from the previous step were applied to an S75 gel filtration column. The DV-Hjc protein eluted after approximately 60 mL and resulted in clean a protein product as determined by SDS-PAGE electrophoresis (Figure 24, C). The protein-containing fractions were pooled and changed into the storage buffer (Supplementary Table 8) to increase the protein stability.

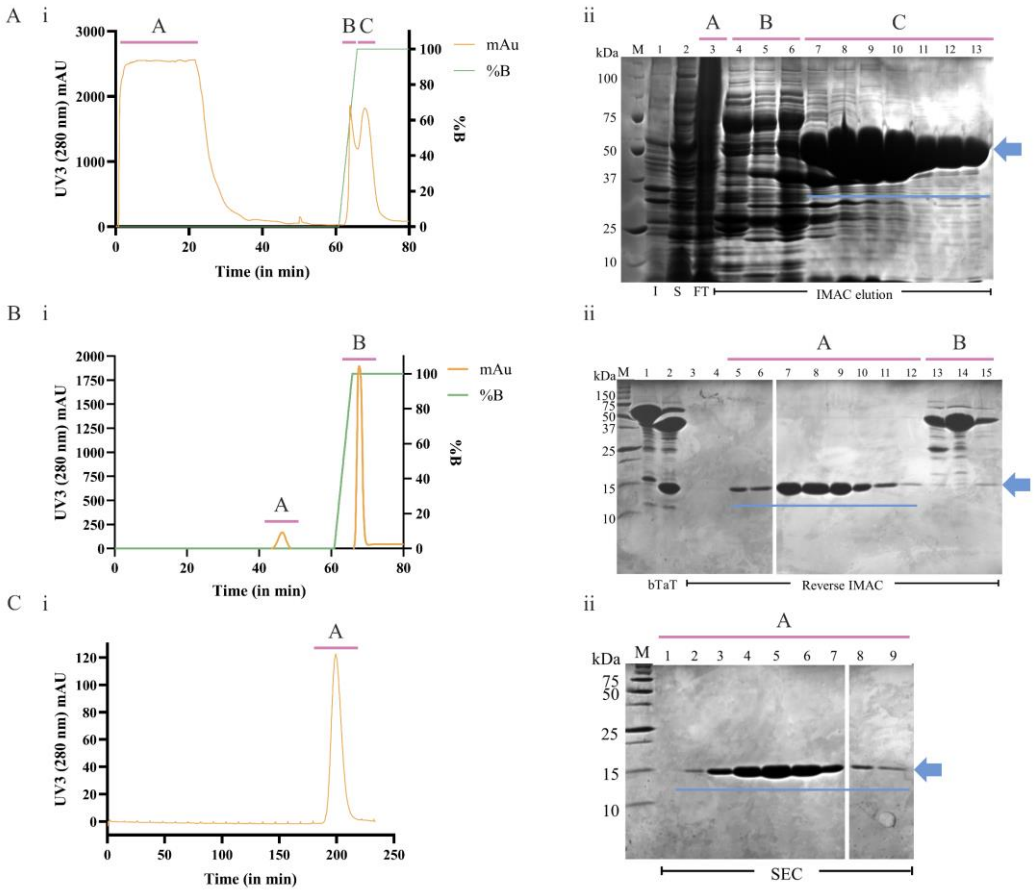


Figure 24: DV-Hjc protein purification chromatograms and SDS-PAGE gels.
All SDS-PAGE gels have the Precision Plus Protein™ standard as a molecular weight indicator (M). Blue arrows indicate the position of DV-Hjc on the gel. Blue lines indicate, which fractions were taken into the next step.
***A:** i: Chromatogram of the IMAC elution of MBP-tagged protein. ii: 12% SDS-PAGE gel, showing the insoluble proteins (I) in the first lane and crude soluble proteins (S) in the second lane. Lane 3 contains the flow through (FT). Lanes 4 to 13 are fractions from the IMAC elution. The pink underlined A, B, and C indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***B:** i: Chromatogram of the reverse IMAC elution of untagged DV-Hjc. ii: 15% SDS-PAGE gel showing the pooled protein from before and after TEV) incubation (BT and AT, respectively in lanes 1 and 2 (ii). Lanes 3 to 15 are fractions from the reverse IMAC elution. The pink underlined A and B indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***C:** i: Gel filtration chromatogram. ii:15% SDS-PAGE gel with fractions from the gel filtration elution. The pink underlined A indicates the position of the chromatogram peak on the SDS-PAGE gel.*

DV-Hjc^{D42A} Mutant Purification

The DV-Hjc^{D42A} mutant was purified as described in Section 2.1.3. DV-Hjc^{D42A} and behaved similarly to the DV-Hjc wild-type protein, described above, which indicates that DV-Hjc^{D42A} folds correctly. During IMAC elution of the MBP- and 6xHis-tagged protein DV-Hjc^{D42A}, the protein eluted at 0.2 M imidazole as shown in the SDS-PAGE gel in Figure 25 (A). The protein-containing fractions were changed into Buffer C and diluted as previously described before the protein was incubated with TEV protease overnight. After TEV incubation the protein was put through a reverse IMAC to separate DV-Hjc^{D42A} from the 6xHis-tagged MBP-tag and TEV protease protein. The eluate was collected in 2 mL fractions and analysed on an SDS-PAGE gel (Figure 25, B). After the reverse IMAC the protein was relatively free of contaminants, but to further purify the protein, protein-containing fractions were

pooled and applied to an S75 gel filtration column for size exclusion chromatography. This resulted in a single large peak and a few smaller peaks from contaminating proteins, which were analysed on an SDS-PAGE gel (Figure 25, C). The fractions containing the DV-Hjc^{D42A} protein were transferred into the storage buffer using a de-salt column to stabilise the protein.

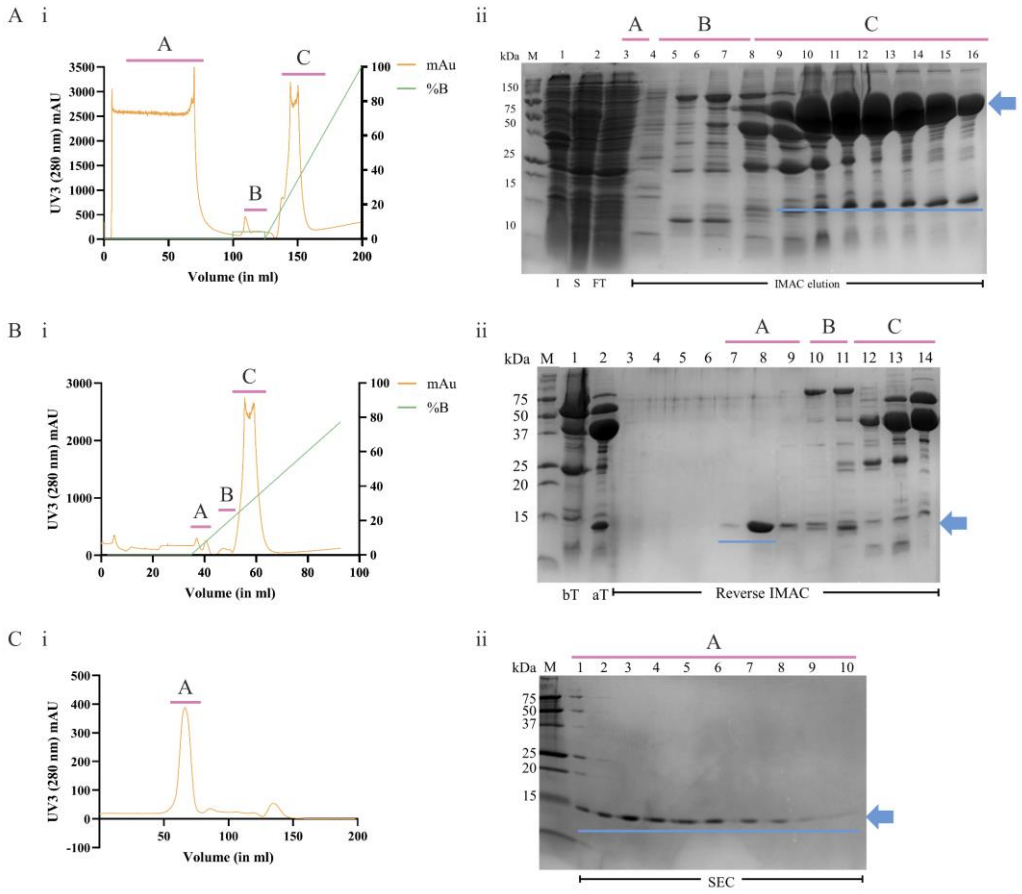


Figure 25: DV-Hjc^{D42A} protein purification chromatograms and SDS-PAGE gels.
All SDS-PAGE gels have the Precision Plus Protein™ standard as a molecular weight indicator (M). Blue arrows indicate the position of DV-Hjc on the gel. Blue lines indicate, which fractions were taken into the next step.
***A:** i: Chromatogram of the IMAC elution of MBP-tagged protein. ii: 12% SDS-PAGE gel, showing the insoluble proteins (I) in the first lane and crude soluble proteins (S) in the second lane. Lane 3 contains the flow through (FT). Lanes 4 to 16 are fractions from the IMAC elution. The pink underlined A, B, and C indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***B:** i: Chromatogram of the reverse IMAC elution of untagged DV-Hjc. ii: 15% SDS-PAGE gel showing the pooled protein from before and after TEV incubation in lanes 1 and 2, respectively (BT and AT, respectively) (ii). Lanes 3 to 14 are fractions from the reverse IMAC elution. The pink underlined A, B and C indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***C:** i: Gel filtration chromatogram. ii:15% SDS-PAGE gel with fractions from the gel filtration elution. The pink underlined A indicates the position of the chromatogram peak on the SDS-PAGE gel.*

3.3 Structural Characterisation of DV-Hjc

Crystals were grown in 35% PEG 4,000, 100 mM ammonium citrate (pH 5.6) and 200 mM ammonium acetate, and diffraction data was collected at the Australian Synchrotron by my supervisor, Dr

Williamson, before the commencement of my Masters Research. I processed this data and used it to determine the DV-Hjc structure. The structure of the DV-Hjc protein was determined at a resolution of 1.74 Å with an r-free of 0.2225. The data collection and refinement statistics of the solved DV-Hjc protein structure presented here are shown in Table 4. Residues Glu13, Glu18, Arg30, Arg77, and Lys81 show alternative side-chain formations and have been modelled as such. Residues 61 to 73 are absent in the final model due to poor electron density in this area. A single citric acid molecule is present near the α1 α-helix, which was a component of the crystallisation buffer. The DV-Hjc protein is an α/β protein with a five-stranded β-sheet sandwiched between two α-helixes (Figure 26). The β-sheets are in the order: 1-2-3-4-5 with sheets 2 and 5 running antiparallel to the other strand (Figure 26, B and C).

The protein has a large positively charged surface, which is formed mainly by basic side chains on α1, α2 and in the loop regions (Figure 26, D). Embedded in this positively charged surface is a negatively charged pocket formed by acidic residues (Glu13, Asp22, Asp42, Asp47, Glu54) on α1, β3 and β2.

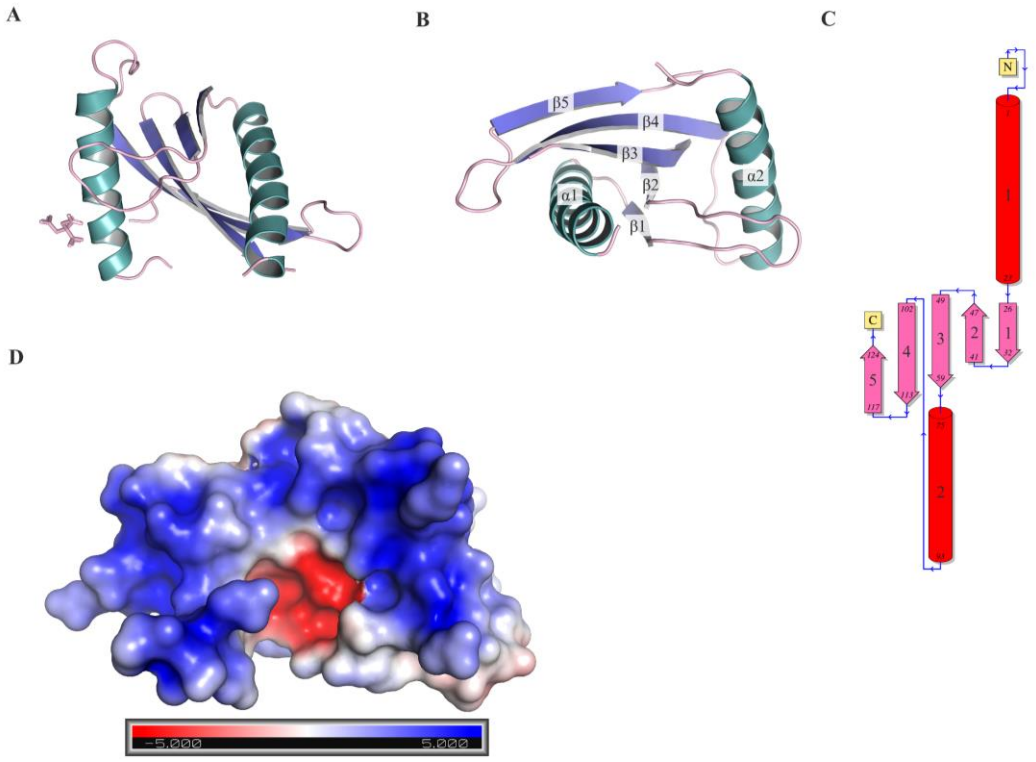


Figure 26: Structure of DV-Hjc.
A: *Cartoon representation of the structure, showing α-helices in green, β-sheets in blue and loops in pink. The citric acid molecule is shown in pink.*
B: *Structure of the protein with α-helices and β-sheets numbered ascending from the N-terminus to the C-terminus of the protein.*
C: *Topology diagram of DV-Hjc generated in PDBsum1 (Laskowski, 2022). α-helices are shown as red cylinders. β-sheets are shown as pink arrows. Helices and sheets are numbered like in B.*
D: *The electrostatic surface potential of DV-Hjc, red and blue coloured surfaces indicate negative and positive potential, respectively.*

The structure of *Rhodopseudomonas palustris* is the only structure representative of a UPF0102 domain protein so far (3FOV)(Osipiuk et al., 2009). The superimposition of the DV-Hjc with the *R. palustris* structure showed a high structural similarity between the proteins (Figure 27). Both proteins consist of a five-stranded β-sheet with two α-helices framing them. The structure of *R. palustris* UPF0102 has an unstructured region from 70 to 75, which lies on the same loop as the unstructured region in the DV-Hjc structure (residues 61 to 73). The length of the β-sheets and α-helices differ between the two proteins and the *R. palustris* protein has an additional short α-helix at the C-terminus of α2. The long loop regions of both proteins are located in very different positions, which may be due to their flexibility and function in DNA binding. A sequence alignment of DV-Hjc with other YraN and UPF0102 domain proteins shows a high level of conservation between the proteins (Figure 27).

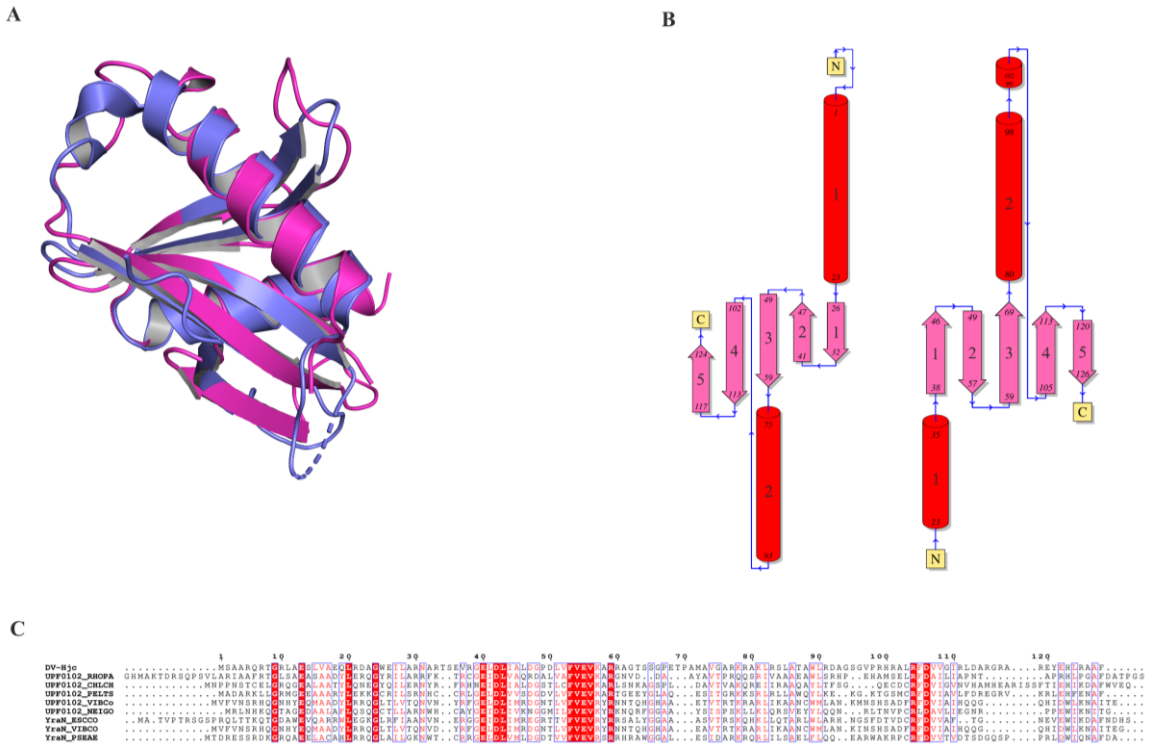


Figure 27: Comparison between DV-Hjc and the R. palustris UPF0102 protein structure.
***A:** Structural alignment of the two proteins in cartoon representation. DV-Hjc is shown in pink and the R. palustris UPF0102 (pdb ID: 3FOV) is shown in blue.*
***B:** Topology of DV-Hjc (left) and 3FOV (right). α-helices and β-sheets are numbered ascending from the N- to C-terminus. Topology diagrams were generated in PDBsum (Laskowski, 2022).*
***C:** Amino acid sequence alignment of DV-Hjc with UPF0102 domain of R. palustris (RHOPA), Chlorobium chlorochromatii (CHLCH), Pelotomaculum thermopropionicum (PELTH), Vibrio cholerae (VIBCO) and Neisseria gonorrhoeae (NEIGO) and YraN proteins from E. coli (ESCCO), Vibrio cholerae (VIBCO) and Pseudomonas aeruginosa (PSEAE).*
The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014)

Archaeal Hjc Protein Structures

Based on a Vector Alignment Search Tool (VAST) search with DV-Hjc the protein shares high structural similarity with the archaeal Holliday junction resolvase Hjc from *Pyrococcus furiosus* (1GEF) (Gibrat et al., 1996). Superimposition of these two structures shows a high similarity of the five-stranded

β-sheet core of the two proteins (Figure 28, A). The *P. furiosus* structure has an additional α-helix and two additional β-sheets compared to the DV-Hjc protein structure (Figure 28, A and B). The two additional β-sheets form a two-stranded β-sheet located opposite the β3, β4, and β5.

Other archaeal Hjc proteins share the same core of a five-stranded β-sheet and two α-helices. The order of these five-stranded β-sheet is 1-2-3-4-5 with strands two and five antiparallel to the others (1-2-3-4-5, ↑↓↑↑↓) in archaeal Hjc proteins and in DV-Hjc (Bond et al., 2001; Middleton et al., 2004; Nishino, Komori, Ishino, et al., 2001; Nishino, Komori, Tsuchiya, et al., 2001). As shown in the topology diagram of the archaeal Hjc proteins shown in Figure 28 (B), the archaeal Hjc proteins have additional structural elements to the core structure. These additional features consist mostly of short β-sheets and α-helices. For example, in the case of *P. furiosus* Hjc, an additional α-helix and two additional β-sheets are present, and in the case of the *Saccharolobus solfataricus* Hjc protein, two additional α-helices and four additional β-sheets are present, when compared to DV-Hjc. The function of these additional features has not yet been determined.

Sequence comparison of DV-Hjc with archaeal Hjc protein sequences showed overall little sequence similarity between DV-Hjc and archaeal Hjc proteins. The sequence alignments showed that glycine residues at positions 9 and 24, glutamic acid at positions 13 and 55, leucine at position 20, aspartic acid at position 42, alanine at position 45, and lysine at position 57 are conserved in all aligned proteins. These are mostly part of the catalytic site of the proteins as will be discussed in more detail below.

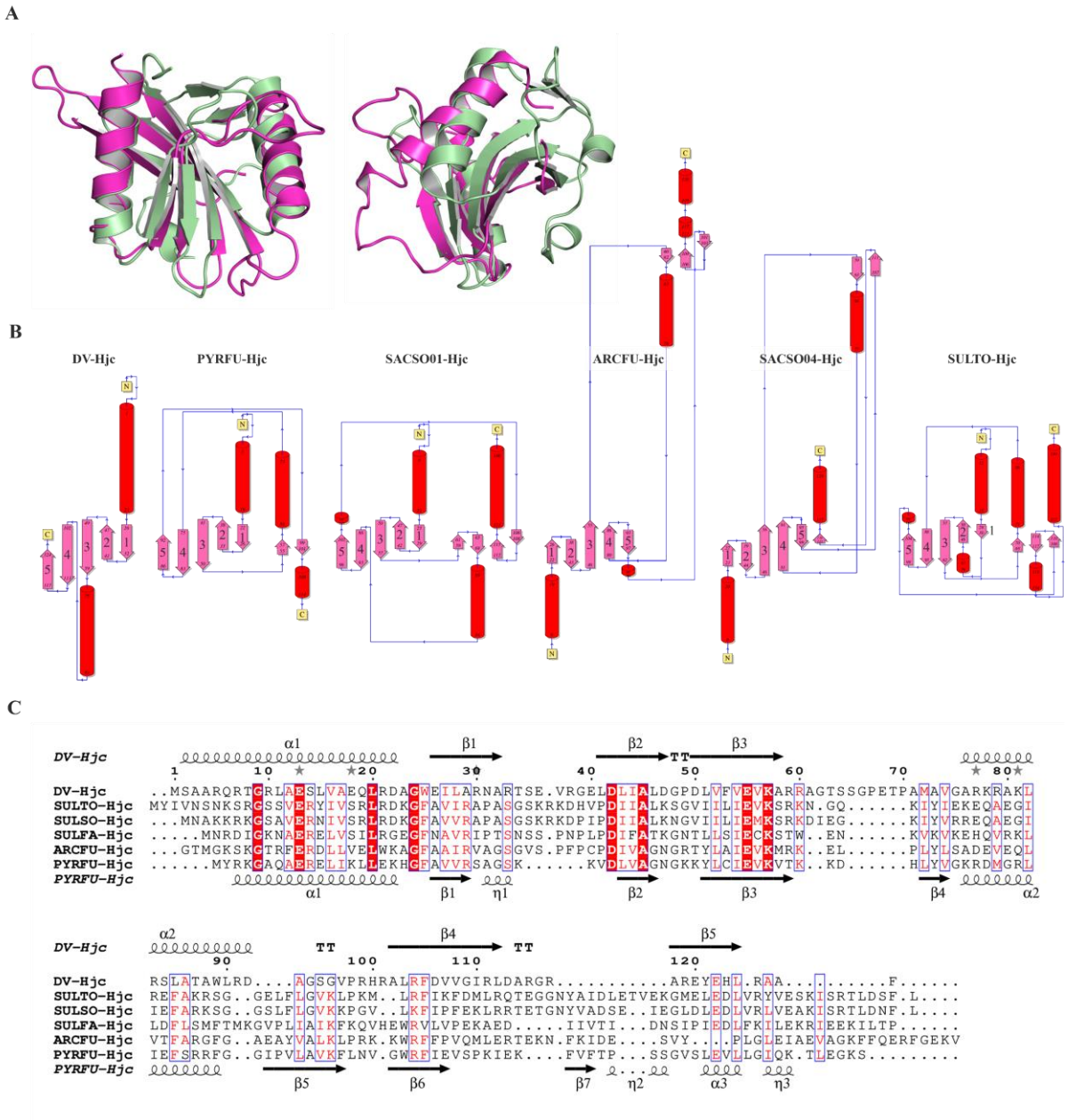


Figure 28: Structural comparison of DV-Hjc with archaeal Hjc proteins.
A: *Structural superimposition of DV-Hjc (in pink) with the archaeal P. furiosus Hjc protein (pdb ID: 1GEF) (in green).*
B: *Topological diagram of DV-Hjc and archaeal Hjc proteins. From left to right: DV-Hjc, P. furiosus Hjc (pdb ID: 1GEF), S. solfataricus Hjc (pdb ID: 1HH1), A. fulgidus (pdb ID: 2WCW), S. solfataricus Hje, S. tokodaii (pdb ID: 2EO0)(Bond et al., 2001; Carolis et al., 2009; Nishino, Komori, Tsuchiya, et al., 2001; Sarai et al., 2007).*
C: *Amino acid sequence alignment of the proteins named in B. The structure of DV-Hjc is shown on top and the structure of the P. furiosus Hjc on the bottom.*
The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014). Topology diagrams were generated in PDBsum (Laskowski, 2022).

Nuclease Proteins with a Five-Stranded β-Sheet

The structural motif of a five-stranded β-sheet flanked by several α-helices can also be found in other nuclease enzymes, as shown in the examples of the type II restriction endonucleases, like EcoRV and EcoRI, the RNase HI, RuvC, LHR exonuclease, RecJ, γ δ resolvase, and Colicin E5 riboendonuclease

in Figure 29 (Ariyoshi et al., 1994; Katayanagi et al., 1992; Lin et al., 2005; Pingoud et al., 2005; Rice & Steitz, 1994; Wakamatsu et al., 2010). The order of the β-sheets is the same in DV-Hjc, archaeal Hjc proteins and type II restriction endonucleases (1-2-3-4-5), but the direction of the β-sheets differs between the proteins. In RNaseHI and RuvC the order of the sheets is (5-4-1-2-3↑↑↑↓↑) (Ariyoshi et al., 1994; Breyer & Matthews, 2000; Katayanagi et al., 1992). The five-stranded β-sheet flanked by α-helices appears to be a common structural feature in nuclease. The proteins shown in Figure 29, have additional structural elements to the five-stranded β-sheet core, which are unique to those proteins. DV-Hjc does not have any additional features.

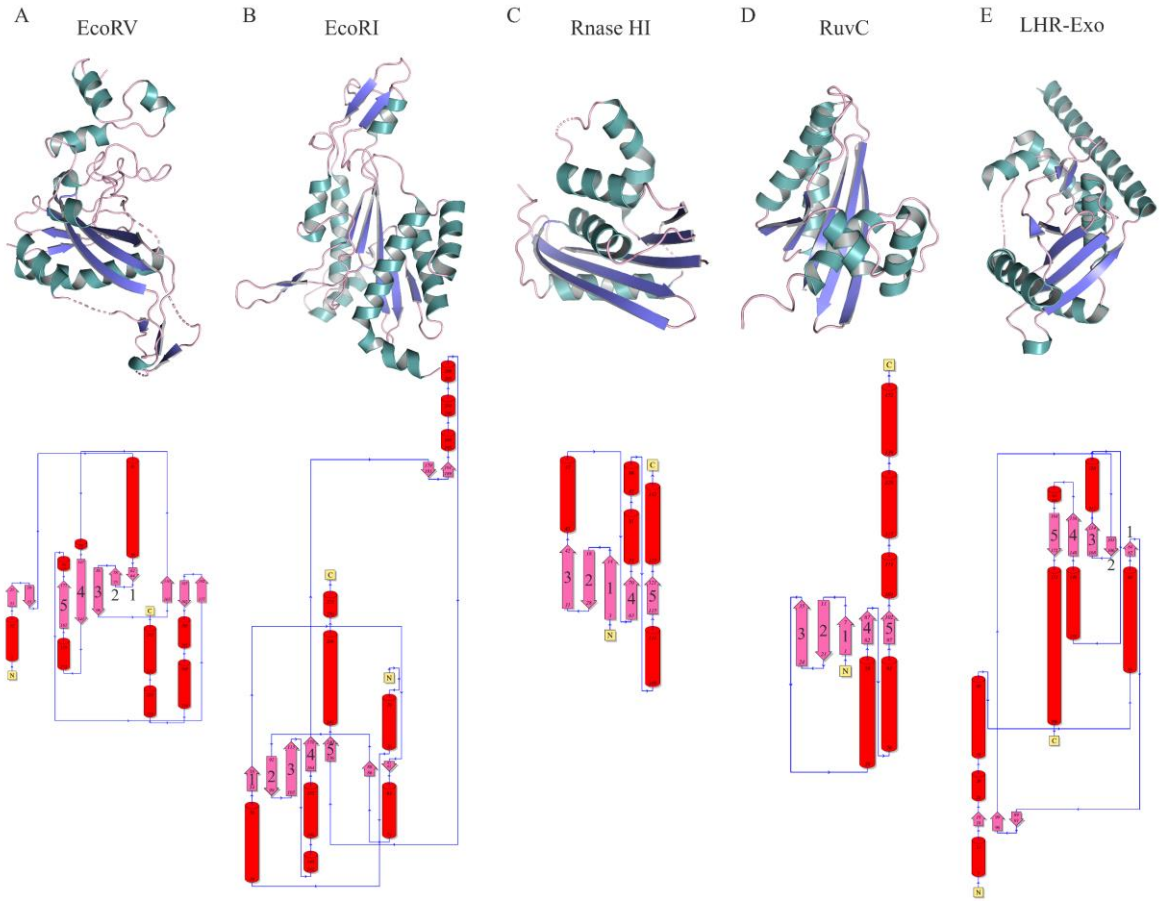


Figure 29: Structures and topology diagrams of the following enzymes:
***A:** Type II restriction endonuclease EcoRV (pdb ID: 1AZ0)(Perona & Martin, 1997).*
***B:** Type II restriction endonuclease EcoRI (pdb ID: 2OXV)(Sapienza et al., 2007).*
***C:** E. coli RNase HI (pdb ID: 1G15) (Goedken & Marqusee, 2001).*
***D:** E. coli RuvC (pdb ID: 1HJR)(Ariyoshi et al., 1994).*
***E:** Laribacter hongkongensis exonuclease (3SZ4)(Yang et al., 2011).*
Topology diagrams were generated in PDBsum (Laskowski, 2022).

3.3.1 Metal Binding Residues of DV-Hjc

Nucleolytic activity of nucleases may require one to three metal ions in the protein's active site (Yang, 2011). If a nuclease requires a metal for its activity, it is referred to as a metallonuclease. Metallonucleases commonly utilise Mn^{2+} and/or Mg^{2+} for their activity, which is abundant in all living

cells (Dupureur, 2008a; Wolf & Cittadini, 2003). They may also use Ca^{2+} or Zn^{2+} (Dupureur, 2008a). As shown in Section 3.4.3, DV-Hjc requires magnesium or manganese for nucleolytic activity. Metal ions are required for the activity of all metallonucleases because they coordinate the scissile phosphate and are predicted to activate the nucleophile, stabilise the unstable intermediate and protonate the leaving group in some nuclease mechanisms (Dupureur, 2008a, 2008b). Crystallisation of DV-Hjc with a metal ion was not attempted during this project; therefore, the DV-Hjc metal binding site was predicted with a Mg^{2+} and a Mn^{2+} ion using the AlphaFold3 (Abramson et al., 2024). The predicted DV-Hjc protein as a monomer with Mg^{2+} bound to it had the highest ipTM of 0.91, which indicates that the position of the Mg^{2+} ion in the complex with the DV-Hjc protein is likely to be accurate.

The prediction places the Mg^{2+} near the acidic residues Asp42 and Glu55 (Figure 30, A). Based on this prediction, Asp42 and Glu55 are hypothesised to be responsible for the orientation of a single metal ion in the protein's predicted active site. The alignment of the DV-Hjc protein with archaeal Hjc proteins and sequences for UPF0102 family proteins show that the glutamic acid residues and the aspartic acid residue are conserved across both archaeal Hjc and UPF0102 family proteins (Figure 30, B).

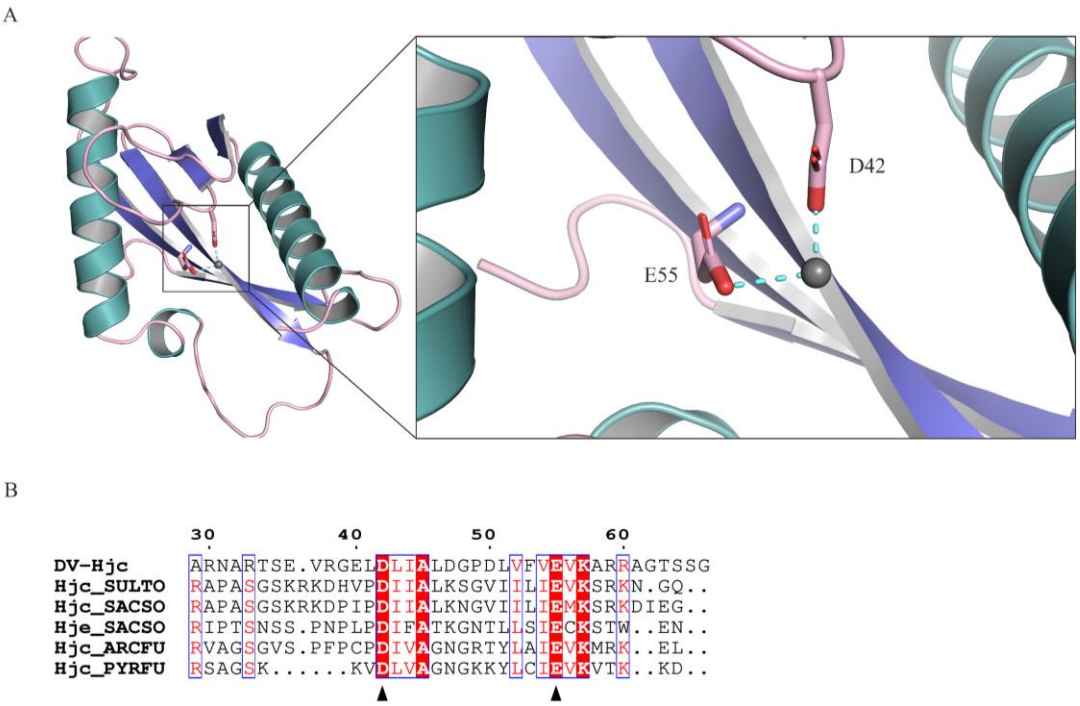


Figure 30: DV-Hjc metal binding site.
A: *Predicted metal binding site of DV-Hjc based on an AlphaFold3 (Abramson et al., 2024). The prediction was done using a single Mg^{2+} ion. The prediction of the whole metal-protein complex had an ipTM of 0.91. The confidence of the metal binding site was very high with a pIDDT >0.9 and the position of the Mg^{2+} had a pIDDT between 0.7 and 0.9.*
B: *Alignment of archaeal Hjc and Hje proteins from Sulfurisphaera tokodaii (SULTO), Saccharolobus solfataricus (SACSO), Archaeoglobus fulgidus (ARCFU), and P. furiosus (PYRFU).*
The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014).

3.3.2 Catalytic Site of DV-Hjc

Archaeal Hjc proteins have a conserved motif of two aspartic acid residues, one glutamic acid, and a lysine residue, which forms the catalytic site of those proteins (EXnDXmEXK) (Kvaratskhelia et al., 2000). Based on sequence alignment of the DV-Hjc protein with characterised archaeal Hjc proteins this motif is also present in DV-Hjc. In DV-Hjc the residues of this motif correspond to Glu10, Asp42, Glu55, and Lys57 (Figure 31). Based on the presence of this motif DV-Hjc protein belongs to the large PD-(D/E)XK phosphodiesterase superfamily (Steczkiewicz et al., 2012). This protein family mainly consists of nuclease proteins, including type II restriction endonucleases, Holliday junction resolvases, and the UPF0102 domain proteins (Steczkiewicz et al., 2012). Aside from the active site residues proteins belonging to this family have low sequence similarity (Steczkiewicz et al., 2012). Like many other PD-(D/E)XK family proteins, DV-Hjc and other UPF0102 proteins do not have the proline residue before the aspartic acid residue, as highlighted in Figure 31 (A). In UPF0102 and YraN proteins the proline residue is substituted by a leucine or isoleucine residue. The proline residue is present in most archaeal Hjc proteins, apart from the *P. furiosus* Hjc, where it is substituted by valine. This agrees with the suggestion from Steczkiewicz et al. (2012), that proline can be substituted by hydrophobic side chains. As shown in Figure 31 (B), the substitution of proline to leucine in DV-Hjc does not affect the position of the other active site residues. The leucine residue is probably involved in the formation of the hydrophobic core of the protein.

The DV-Hjc active site residues superimpose well with those from archaeal Hjc proteins, as shown in the example of the active site superimposition of *P. furiosus* and *A. fulgidus* with DV-Hjc (Figure 31, B, i and ii). The DV-Hjc Glu13, Asp42, Glu55, and Lys57 correspond to the Glu-9, Asp-33, Glu-46, and Lys-48 in *P. furiosus* and *A. fulgidus* (Komori et al., 2000; Nishino, Komori, Tsuchiya, et al., 2001). Biochemical assays (as described in Section 3.4.3) confirmed that this is true for Asp42, which is required for the catalytic activity of the DV-Hjc protein. The activity of the DV-Hjc^{D42A} mutant had over 50% activity reduction compared to the wild type. Therefore, we can be certain that this residue is involved in the cleavage of DNA by DV-Hjc. The other predicted active site residues (Glu13, Glu55, and Lys57) were not mutated to determine whether they are required for the activity of the DV-Hjc protein in the scope of this project. However, based on the sequence alignments and structure superposition with characterised archaeal Hjc proteins it is highly likely to form the active site of the protein (Figure 31, A and B)(Bond et al., 2001; Komori et al., 2000; Kvaratskhelia et al., 2000; Middleton et al., 2004; Nishino, Komori, Ishino, et al., 2001). The same active site residues can also be found in the EcoRV restriction endonuclease. The residues Glu12, Asp42, Glu55, and Lys57 in DV-Hjc are equivalent to Glu45, Asp74, Asp90, and Lys92 in EcoRV (Vipond et al., 1995).

Glu13 shows alternative conformations in the solved structure of DV-Hjc and is shown as such. This may be due to the lack of a metal ion or substrate in the active site. The glutamic acid in this position is not part of the PD-(D/E)XK motif, but it is highly conserved across archaeal Hjc proteins and UPF0102 family proteins and is required for nucleolytic activity.

The lysine residue in the active site of DV-Hjc is a single basic amino acid, Lys57. This residue is part of the motif in PD-(D/E)XK phosphodiesterase family proteins. It is highly conserved and multiple studies have shown that it is required for the activity of homologous proteins (Komori et al., 2000; Kvaratskhelia et al., 2000; Nishino, Komori, Ishino, et al., 2001; Vipond et al., 1995). The lysine residue is predicted to play a role in stabilising the transition state during phosphodiester bond cleavage (Kvaratskhelia et al., 2000). In two of the UPF0102 family proteins shown here, this residue is replaced by arginine, which due to the similar properties may be able to fulfil the same function as lysine in this position.

An electrostatic potential surface of DV-Hjc shows that the acidic side chains in the active site form a negatively charged binding pocket (Figure 31, C, i). Figure 31 (C, ii) shows the acidic residues cluster together at one part of the active site, while the lysine is located at the other side. The active site is located in the centre of the positively charged predicted binding surface, suggesting that this positively charged surface binds the DNA and places it in the active site allowing interaction of the residues in the active site with the DNA.

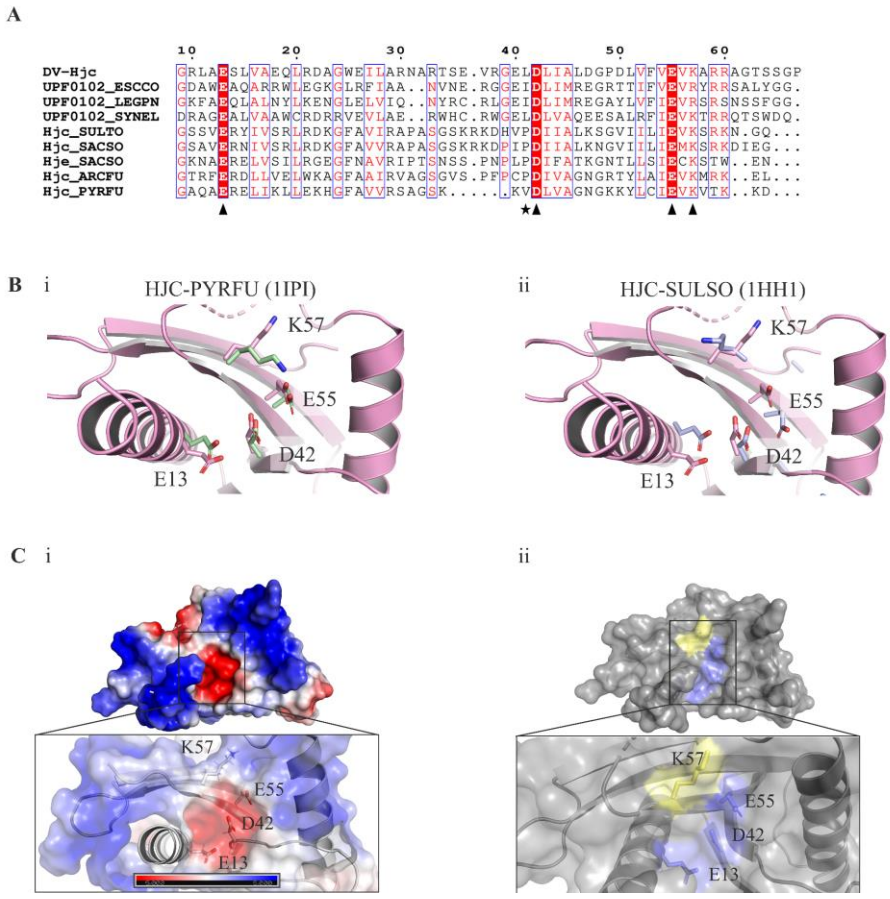


Figure 31: Active site of DV-Hjc.
***A:** Amino acid sequence alignment of DV-Hjc with YraN proteins from E. coli (ESCCO), Legionella pneumophila (LEGPN), and Synechococcus elongatus (SYNEL), and archaeal Hjc proteins from Sulfurisphaera tokodaii (SULTO), Saccharolobus solfataricus (SACSO), Archaeoglobus fulgidus (ARCFU), and P. furiosus (PYRFU).*
***B:** Active site of DV-Hjc showing the active site residues Glu13, Asp42, Glu55, and Lys57 superimposed onto those of P. furiosus (PYRFU) (pdb ID: 1IPI) (i) and Sulfurisphaera tokodaii (SULTO) (pdb ID: 1HH1)(ii).*
***C:** i: Electrostatic surface potential of DV-Hjc highlighting the position of the active site and catalytic residues in the active site. ii: position of the catalytic residues in the DV-Hjc structure showing the basic side chain in yellow and acidic side chains in blue.*
The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014).

The active site of nucleases greatly differs between different enzymes (summarised in Table 8). However, a common feature is a cluster of acidic residues with one or two basic side chains, mostly lysine. The same active site organisation is present in DV-Hjc, which suggests its function as a nuclease protein.

Table 8: Summary of the active site diversity of nuclease enzymes.

Protein, Organism	**Active Site Residues**	**Reference**
DV-Hjc	Glu12, Asp42, Glu55, Lys57	
Archaeal Hjc, *P. furiosus*	Glu9, Arg10, Arg25, Asp33, Glu46, Lys48	(Komori et al., 2000)

RecU, *B. subtilis*	Glu36, Lys56, Asp88, Asp99, Glu101, Lys103	(McGregor et al., 2005)
RusA	Asp70, Asp72, Lys76, Asp91	(Rafferty et al., 2003a)
RuvC, *E. coli*	Asp7, Glu66, Asp138, Asp141	(Saito et al., 1995)
EcoRV	Glu45, Asp74, Asp90, Lys92	(Vipond et al., 1995)

3.3.3 Prediction of Dimer Formation by DV-Hjc

As mentioned above DV-Hjc belongs to the structural family of type II restriction endonucleases. Proteins that belong to this family are predominantly dimeric and some tetrameric (Pingoud et al., 2005). This includes all archaeal Hjc and Hje proteins, that have been characterised to date, which are homodimeric in solution and bind Holliday junctions in dimeric form (Wyatt & West, 2014). Archaeal Hjc dimers have two active sites allowing the enzyme to catalyse the cleavage of the two phosphodiester bonds on either side of the Holliday junction within one DNA binding event (Wyatt & West, 2014). The dimer interface is mainly formed by interactions of hydrophobic residues in the β-sheets at the dimer interface (Figure 32, B) (Bond et al., 2001; Komori et al., 2000; Nishino, Komori, Tsuchiya, et al., 2001). As shown in Figure 32 (B), in the *P. furiosus* Hjc protein the two aromatic residues Phe68 and Phe72 are also involved in the dimer formation through cation-π interactions and mutation of those residues to alanine residues showed a reduction in dimer formation (Komori et al., 2000).

Because it has a Type II restriction enzyme-like structure and its homology to archaeal Hjc proteins, the dimer formation of DV-Hjc was explored. This was done using AlphaFold3 predictions since the DV-Hjc protein was present in monomeric conformation in all crystal structures obtained during this project (Abramson et al., 2024). Multimers from two to four protein units were predicted using AlphaFold3 and the predicted dimeric Hjc had the highest ipTM with 0.18. This low score indicates that it is likely a failed prediction in terms of the position of the two complexes. This is further confirmed by the predicted dimer interface of this model between the α1 α-helices of the two monomers. The predicted dimer interface consists of interactions between Ser14, Glu19 and Asp22. The main interactions at the dimer interface would therefore be hydrogen bonds between the serine and the acidic residues as indicated in Figure 32 (A). The AlphaFold3 predicted dimer conformation, therefore, appears unlikely.

Comparison of the DV-Hjc protein to archaeal Hjc proteins shows that the phenylalanine residues required for dimer formation are absent in DV-Hjc (Figure 32, C). However, Phe72, which is important for dimer formation in *P. furiosus* is substituted by an aromatic side chain in DV-Hjc, Trp89. As shown

in Figure 32, this conservational amino acid change is also present in the *E. coli* UPF0102 and a conservational amino acid change to tyrosine in the *L. pneumophila* and *S. elongatus*. Additionally, hydrophobic residues on the β-sheets β1 and β2 in the archaeal Hjc proteins are involved in the dimer formation. The amino acid sequence alignment shows that hydrophobic residues on β1 and β2 are conserved across both the UPF0102 proteins and the archaeal Hjc and Hje proteins shown. Overall, dimer formation of DV-Hjc seems likely based on the homology of DV-Hjc to other dimeric proteins, however, a confident prediction of this dimer could not be made and therefore possible dimer formation must be confirmed experimentally.

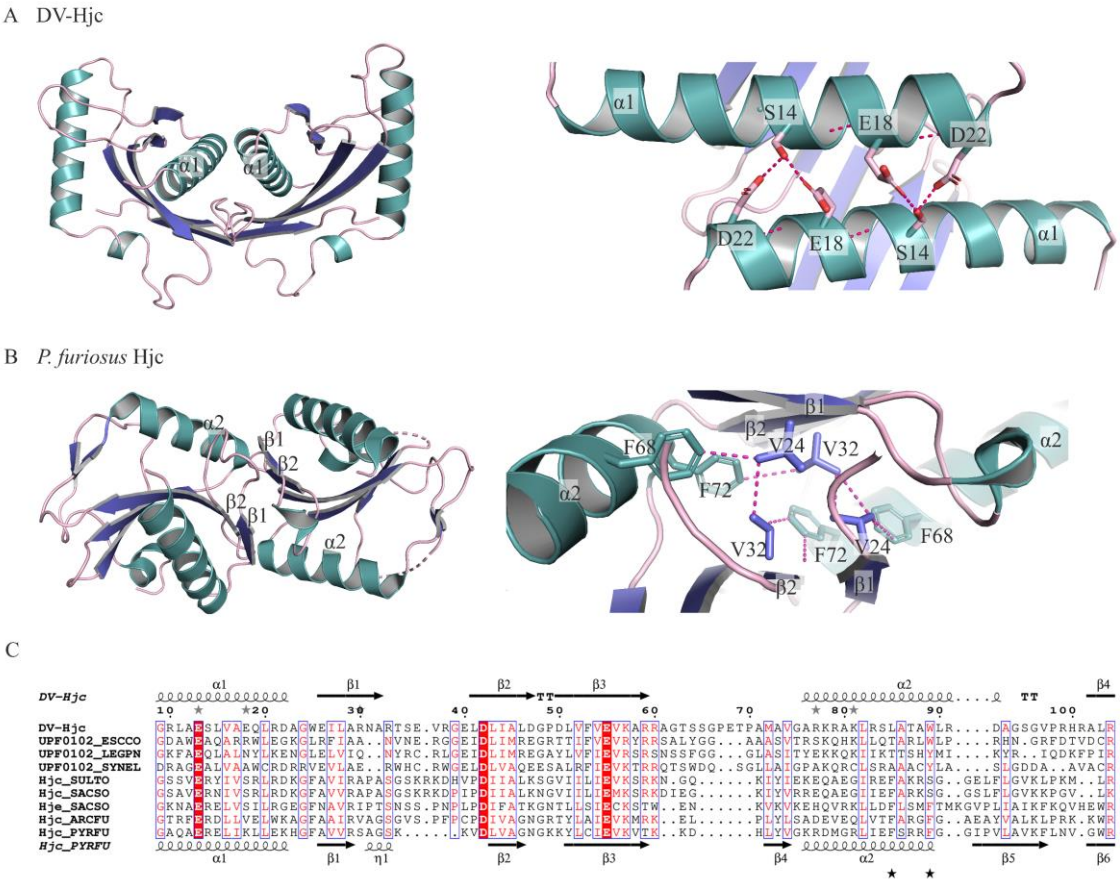


Figure 32: Predicted dimer formation of DV-Hjc.
***A:** AlphaFold 3 predicted DV-Hjc dimer (ipTM: 0.18). The full dimer is shown on the left and a closer look at the dimer interface on the right.*
***B:** P. furiosus Hjc protein dimer (1GEF)(Nishino, Komori, Tsuchiya, et al., 2001). The full dimer is shown on the left and a closer look at the dimer interface on the right.*
***C:** Amino acid sequence alignment of DV-Hjc with bacterial UPF0102 protein sequence and archaeal Holliday junction resolvases. The structural features of DV-Hjc and P. furiosus are shown on the top and bottom of the alignment, respectively. The black stars indicate the position of the residues required for dimer formation in P. furiosus (Komori et al., 2000).*
The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014).

Aside from the archaeal Hjc proteins, other known Holliday junction-resolving enzymes are also active dimers. Those include the RusA endonuclease, RuvC resolvase, RecU resolvase, Ydc2 resolvase, Phage T4 endonuclease VII, and MOC1 resolvase (Ceschini et al., 2001; Gorecka et al., 2013; McGregor et

al., 2005; Rafferty et al., 2003b; Yan et al., 2020). Dimer formation in those proteins is facilitated mostly through hydrophobic interactions at the dimer interface. The predicted DV-Hjc dimer also has some hydrophobic interactions at the predicted dimer interface however, they are not as extensive as those found in similar proteins.

When comparing the dimer formation between the mesophilic RuvC from *E. coli* with that of the thermophilic RuvC from *Thermus thermophilus* the dimer interface is packed more tightly in the *T. thermophilus* RuvC homodimer. There is also an increased number of aromatic amino acids at the dimer interface leading to a more stable dimer. This may contribute to the thermostability of the *T. thermophilus* RuvC compared to its mesophilic homologue. Dimer formation of the T4 endonuclease VII from bacteriophage T4 diverges from that of the Holliday junction resolving enzymes discussed above. Structural studies show that in its dimeric form, the protein is highly intertwined with high levels of domain swapping. The structure of T4 endonuclease VII is highly flexible and can cleave a wide range of DNA substrates, which may correlate with its high flexibility (Biertümpfel et al., 2007; Raaijmakers et al., 2001; Raaijmakers et al., 1999; Solaro et al., 1993). Similarly, the monomers of T7 endonuclease I are highly intertwined forming a dimer with an intensive dimeric interface allowing for high stability of the dimer in solution. The recently characterised GEN1 protein – a unique Holliday junction resolving enzyme – is monomeric in solution and dimerises when binding Holliday junctions (Rass et al., 2010).

Based on similar proteins to DV-Hjc it would appear likely that it forms a dimer. However, based on the lack of the amino acids required for dimer formation, oligomerisation may only occur upon substrate binding, like in the Gen1 protein (Liu et al., 2015). To further characterise DV-Hjc it is important to determine its oligomeric state. For this purpose, size exclusion chromatography experiments and further structural analysis via protein crystallisation and small-angle X-ray scattering will be used.

3.3.4 DNA Binding

DV-Hjc-DNA Co-crystallisation

As shown in Section 3.4.2 DV-Hjc binds to a range of DNA substrates. Trials of DV-Hjc-DNA co-crystallisation did not result in a structure of DV-Hjc with DNA bound. Co-crystallisation was attempted with 10 nt long double-stranded DNA substrates as described in Section 2.4.2. In most conditions, crystal forming conditions crystal formed within two weeks. Crystals were obtained for the crystallisation screening conditions shown in Table 9.

Table 9: Crystallisation conditions from the Hampton Research crystallisation screens PEGxR and Natrix, that showed positive results for DV-Hjc-DNA co-crystallisation.

Crystallisation Screen	Well	Crystallisation solution composition

PEGRx	F1	0.1 M Succinic acid pH 7.0, 0.1 M BICINE pH 8.5, 30% v/v Polyethylene glycol monomethyl ether 550
PEGRx	F5	0.2 M L-Proline, 0.1 M HEPES pH 7.5, 24% w/v Polyethylene glycol 1,500
Natrix	H5	0.08 M Potassium chloride, 0.04 M Sodium cacodylate trihydrate pH 7.0, 60% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride

Based on these results fine screens were set up with the same conditions. Additionally, the pH, precipitant concentration and protein-to-reagent ratios were varied, as detailed in Section 2.4.1. Crystals formed in all conditions shown in Table 9. However, the quality of crystals varied between the conditions, some examples are shown in Figure 33. The 1:2 ratio of protein to mother liquor generally lowered the quality of the crystals and led to the formation of needle clusters (Figure 33). With a 1:1 ratio of protein to mother liquor crystals formed mostly in plate form. Single plates from the conditions indicated by the blue boxes in Figure 33 were taken for X-ray diffraction data collection at the Australian Synchrotron. In the case of clustered plates, they were separated to retrieve single plates. The condition in B2 was the only one that formed a needle-shaped clustered crystal. The layers of the crystal were broken up to retrieve a single needle-shaped crystal from this condition.

The quality of X-ray diffraction data collected from the crystals at the Australian synchrotron was highest from the B2 crystal with a resolution of 2.9Å., however, analysis of the data indicated that the crystal included only protein, and DNA had not co-crystalised. The other crystals did diffract but the resolution was not high enough and the precipitant interfered with the collection of some data. Analysis of the diffraction data from those crystals also indicated that the DNA had not crystallised with the protein. To further optimise the crystallisation of DV-Hjc with DNA different DNA substrates will be tried. Based on the binding activity of DV-Hjc to DNA substrates in electron mobility shift assays (shown in Section 3.4.2), where binding affinity was higher with the longer double-stranded DNA substrates, 20 and 40 nt long double-stranded DNA substrates will be tried for crystallisation instead of the 10 nt DNA substrate tested here. Additionally, the crystallisation solution will be further optimised to achieve a more defined formation of single crystals and reduce the amount of precipitant.

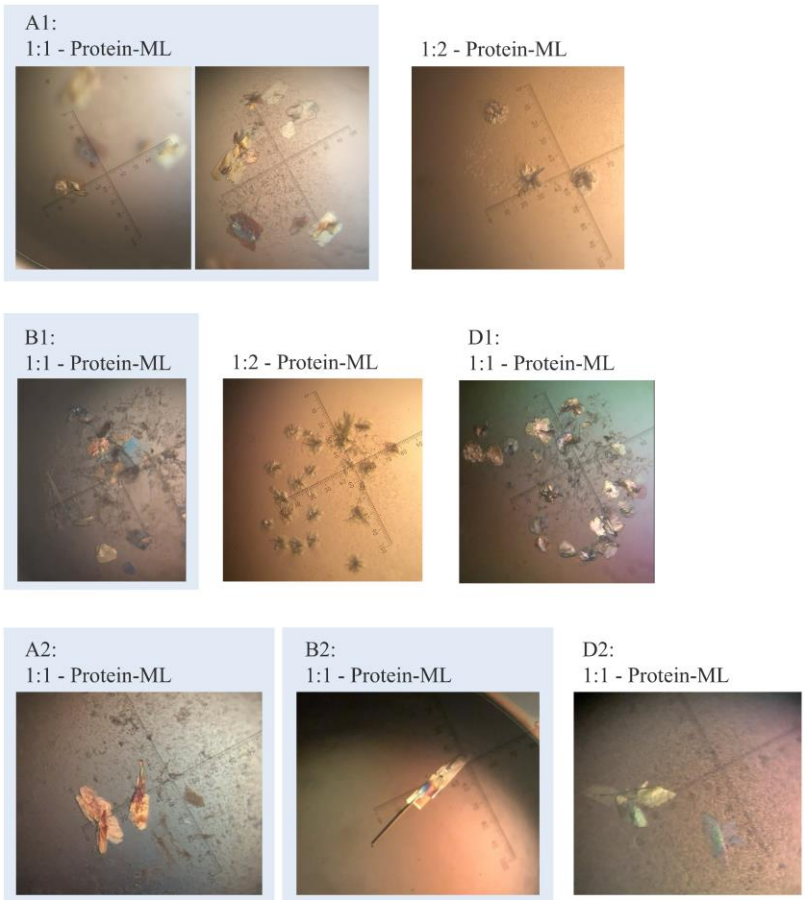


Figure 33: Crystal formation of DV-Hjc with double-stranded DNA in fine screens using the dropping vapour diffusion method. On top of the images, the plate position and protein-to-mother-liquor (ML) ratio are indicated. For the reaction conditions in each well refer to Table 9. The blue background indicates crystals, that were picked for diffraction data collection.

DV-Hjc DNA Binding Prediction

Because a structure of DV-Hjc with DNA could not be obtained during the project. A DNA-DV-Hjc complex structure was predicted using AlphaFold3.

The DV-Hjc protein structure has a large positively charged surface, which is likely the DNA binding surface. The binding of the protein to DNA was shown using electron mobility shift assays with different DNA substrates as explained in more detail in Section 3.4.2.. DNA-DV-Hjc complex co-crystallisation was attempted but no complex structure was obtained. The monomeric protein with 20 nt long double-stranded DNA and a Mg^{2+} ion, shown in Figure 34, had the highest pTM (0.75) and ipTM (0.45) of the constructs tried. The pTM value above 0.5 indicates that the overall predicted fold is likely to be accurate; however, the ipTM value below 0.6 suggests that the position of the complexes in the prediction is incorrect. However, based on the assumption that the large positively charged surface is the DNA binding site the position of the DNA to the DNA binding site was identified correctly as the DNA is placed in the large positively charged surface of the DV-Hjc protein (Figure 34, B). In the predicted structure the DNA is located at the N-terminus of α1 and α2, and the loop region between β1

and β2 (Figure 34, A). The binding is facilitated by interactions between lysine and arginine residues in those regions and the positively charged DNA backbone (Figure 34, C). The Arg5 residue specifically has extensive interactions with the DNA nucleotides in the predicted structure (Figure 34, C, iii). Predicted binding of the DNA at the positively charged surface places the DNA phosphodiester backbone in the active site of the protein and near the Mg^{2+} ion (Figure 34).

There are no structures of archaeal Hjc proteins in a complex with DNA, however, DNA binding has been studied using site-directed mutagenesis in *P. furiosus* (Nishino, Komori, Ishino, et al., 2001). This showed that the N-terminus of the protein is significantly involved in DNA binding, which agrees with the DV-Hjc prediction, in which five arginine (Arg5, Arg7, and Arg9) residues on the N-terminus of the protein have polar interactions with the DNA backbone and the nucleotides as shown in Figure 35 (C, iii). In *P. furiosus* four lysine residues are also required for DNA binding, which are not conserved in the DV-Hjc sequence. In the predicted DV-Hjc-DNA structure complex only a single lysine residue (Lys81) forms a polar contact with the DNA. The main interactions between DV-Hjc and DNA are predicted to be formed by arginine residues, including the three arginine residues at the N-terminus mentioned above and two additional arginine residues (Arg5, Arg7, Arg9, Arg 33, and Arg38) (Figure 34, C, ii). Site-directed mutagenesis of these residues could provide more insight into the DNA binding of DV-Hjc.

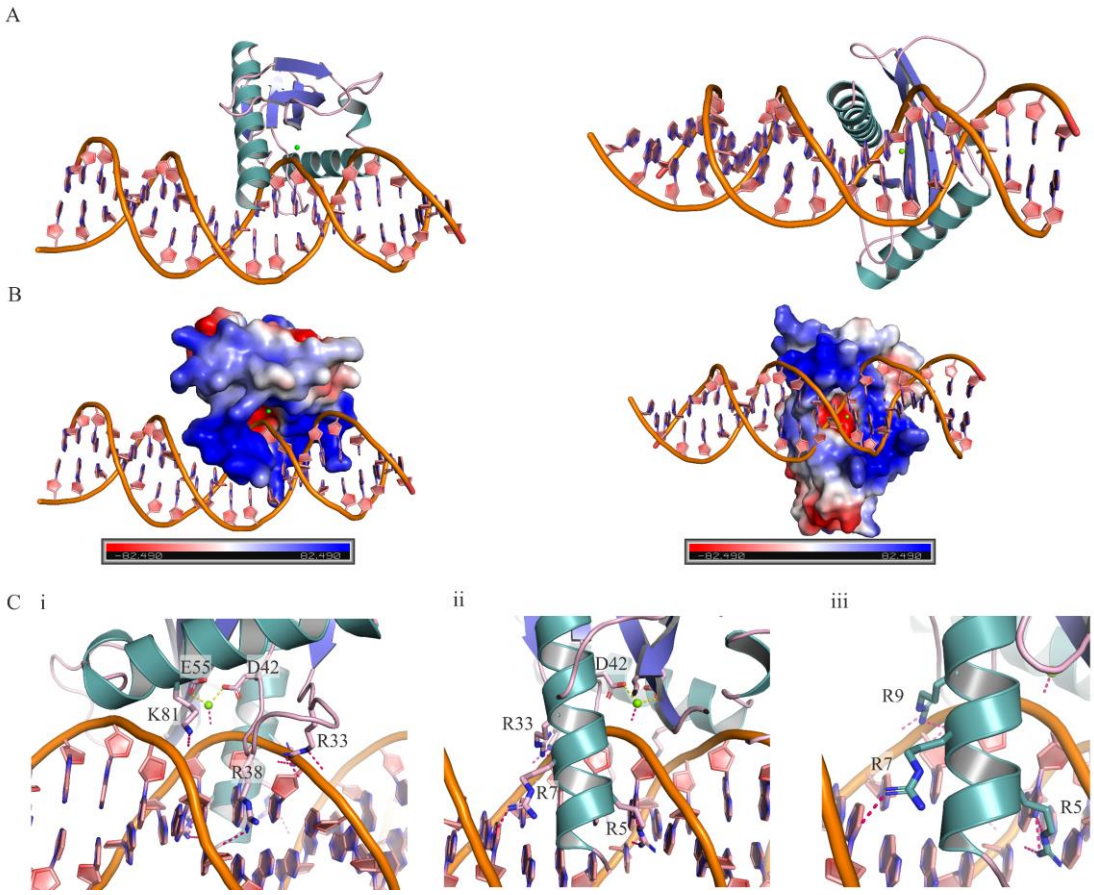


Figure 34: AlphaFold3 predicted monomeric DV-Hjc-DNA complex with a Mg^{2+} ion.
The prediction was done using AlphaFold3 (Abramson et al., 2024).
***A:** Cartoon structure of the predicted DV-Hjc-DNA complex.*
***B:** Electrostatic surface potential view of the predicted DV-Hjc-DNA complex.*
***C:** Close-up view of the DV-Hjc and DNA interface. The side chains of residues predicted to be involved in binding are shown. Polar contact between residues or Mg^{2+} and the DNA are shown in pink and polar contacts between active site residues and the Mg^{2+} ion are shown in yellow.*

3.4 Biochemical Characterisation of DV-Hjc

3.4.1 Thermal Shift Assay

Thermal stability analysis of DV-Hjc was done previously to the commencement of this project by Dr Williamson and was published in Rzoska-Smith; et al. (2023). The thermal stability of DV-Hjc was analysed using differential scanning fluorimetry and showed that the protein's maximum melting temperature was at 50°C at a pH of 6.5 (Figure 35). The melting temperature rapidly decreases when increasing the pH above 7.0. Thermal stability also slightly decreased when the pH was lowered below 6.5 however pH levels below 5.5 were not explored (Rzoska-Smith; et al., 2023). This information was used for the optimisation of purification and assay buffers to increase the stability of the protein.

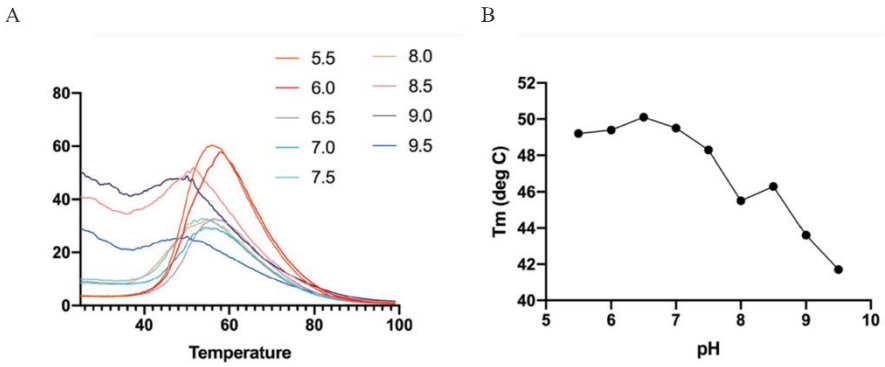


Figure 35: Thermal stability analysis of DV-Hjc using differential scanning fluorimetry.
***A:** Normalised fluorescence signal changer between 30°C and 100°C.*
***B:** Melting temperature (in °C) at the pH range measures (pH 5.5 to 9.5) based on A.*
This figure was taken from Rzoska-Smith; et al. (2023).

3.4.2 DV-Hjc DNA-Binding

DNA binding of DV-Hjc was first shown in Rzoska-Smith; et al. (2023) through electron mobility shift assays of DV-Hjc with double- and single-stranded 20 and 40 nt DNA substrates. The protein was incubated with the DNA substrate at 15°C for 30 min before electrophoresis. The binding of DV-Hjc to different substrates was further explored and optimised during this project as described in the following.

Electrophoretic mobility shift assays (EMSAs) with the DV-Hjc protein showed binding of the protein to a wide range of DNA substrates with more or less affinity. The binding affinity was highest when substrate and protein were incubated on ice for 15 min before electrophoresis. Based on the amount of bound DNA on the higher position compared to the amount of unbound DNA substrate on the lower position of the gel, the protein binds 40 nt long double-stranded DNA with the highest affinity (Figure 36, B), and with slightly less affinity to 40 nt single-stranded and Holliday junction substrates (Figure 36, A and C). Binding affinity was lower with the shorter 20 nt DNA double-stranded and absent for the single-stranded DNA (see Section 7.2.1). High levels of binding were also observed for the 3'trail and splayed DNA substrates and lower levels of binding for the 5'tail and flap 5' and 3' substrates (see 7.2.1). DV-Hjc has an extremely low binding affinity with 40 nt long damaged DNA substrates (8-oxo-G; abasic site; uracil mismatch; uracil match; regular mismatch), which have a centrally placed damage (see 7.2.1). Based on this it appears that DV-Hjc preferably binds the longer 40 nt DNA over the short DNA substrates and has a higher binding affinity with double-stranded substrates over single-stranded substrates. Additionally, it does not bind damaged substrates, which points to the importance of an undistorted DNA double helix for DV-Hjc binding.

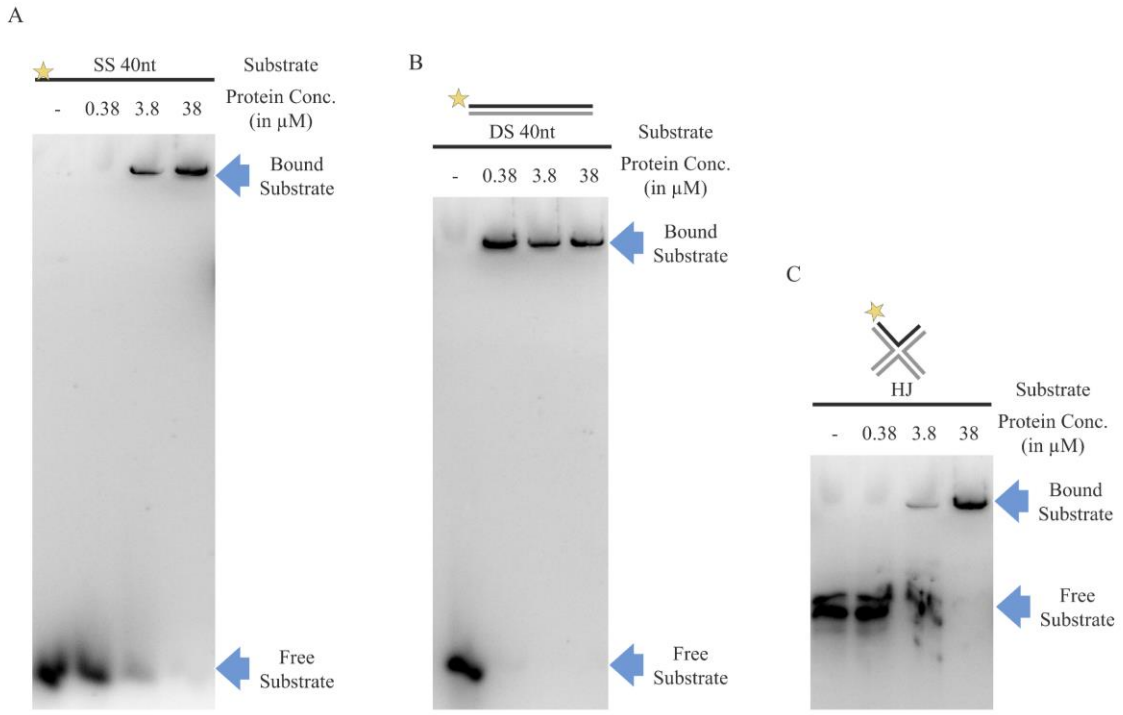


Figure 36: EMSA of DV-Hjc with three substrates as indicated above the gel images.
An illustration of the substrate is shown above the corresponding gel images. The black strand and the star indicate the fluorescently labelled strand of the substrate. The protein concentrations (from 0 to 38 mM) are indicated above the gel images. The position of bound DNA and unbound DNA are indicated by the blue arrows.
***A:** Single-stranded 40 nt DNA substrate **B:** Double-stranded 40 nt DNA substrate **C:** Holliday junction mimicking substrate.*

3.4.3 DV-Hjc Activity Assays

To determine whether DV-Hjc possesses nucleolytic activity with DNA, the purified DV-Hjc was assayed against a range of fluorescently labelled DNA substrates. Assays with the short and long oligonucleotide substrates shown in Section 7.2.2., did not show any cleavage of the substrates by the DV-Hjc protein. While Hardy et al. (2024) showed nuclease activity of a DV-Hjc homologue, the *L. pneumophila* YraN protein, on a 100 nt single-stranded DNA substrate, this activity was not observed with DV-Hjc on a similar 90 nt long single-stranded substrate (Supplementary Figure 3, E). Based on its homology to archaeal Hjc proteins, the activity of DV-Hjc was tested with a Holliday junction mimicking substrate, but no activity was observed with that substrate despite incubation times of up to 14 hours and multiple rounds of optimisation (see Section 7.2.2.).

3.4.4 pUC19 Plasmid-Based Assay

Due to the lack of nuclease activity of DV-Hjc on the synthetic oligonucleotide, substrates the protein was tested with pUC19 plasmid. This assay provides a wider range of DNA sequences than the short oligonucleotides in case DV-Hjc has sequence-specific activity. pUC19 plasmid-based assays use an intact pUC19 plasmid as the substrate for the tested enzyme. After purification, the plasmid is present

in supercoiled form and, due to its compactness, runs as the smallest band on an agarose gel (Figure 37, B). Enzymes can then cleave a single one of the DNA strands producing a nick in the plasmid, which runs as the largest product on an agarose gel (Figure 37, B). Nucleases may also cut both DNA strands creating a linear product. If a single cut is made in both DNA strands the product will run at the size of the plasmid on the agarose gel, which is 2686 bp for the pUC19 plasmid (Figure 37, B). Nucleases that create multiple cuts within the plasmid may create a variety of linear products with different sizes. Preliminary assays of the DV-Hjc protein showed nuclease activity with pUC19 plasmid, which was optimised as described in the following paragraphs.

Metal Ion Preference

Since a majority of nuclease enzymes require a metal ion for their activity, DV-Hjc was first tested with common metal co-factors. To determine the preferred metal ion for DV-Hjc activity, purified protein was assayed with 10 mM of $MgCl_2$, $MnCl_2$, or $ZnCl_2$ and in the absence of a metal ion (Figure 37, A). Without the addition of a divalent metal ion DV-Hjc activity is almost absent, so a metal ion is required for its activity (Figure 37, C). The low-level activity of DV-Hjc in the absence of a metal ion may be due to metal ions present in the purification buffers, which can be carried through the purification. Nuclease activity was highest with the addition of $MnCl_2$, which resulted in a cleavage efficiency of 40%. Due to the precipitation of $MnCl_2$ in one of the reactions, as shown by the smeared band in lane 11 (Figure 37, C), the reaction result was removed from the data for calculation of the activity depicted in Figure 37 (D). The addition of $MgCl_2$ showed nicking of the plasmid to 16% (Figure 37, C and D). The protein was not active with $ZnCl_2$. Because of the inconsistent results caused by precipitation of $MnCl_2$ in reactions, $MnCl_2$ was eventually removed from assays and $MgCl_2$ was used as a metal co-factor to achieve more consistent results.

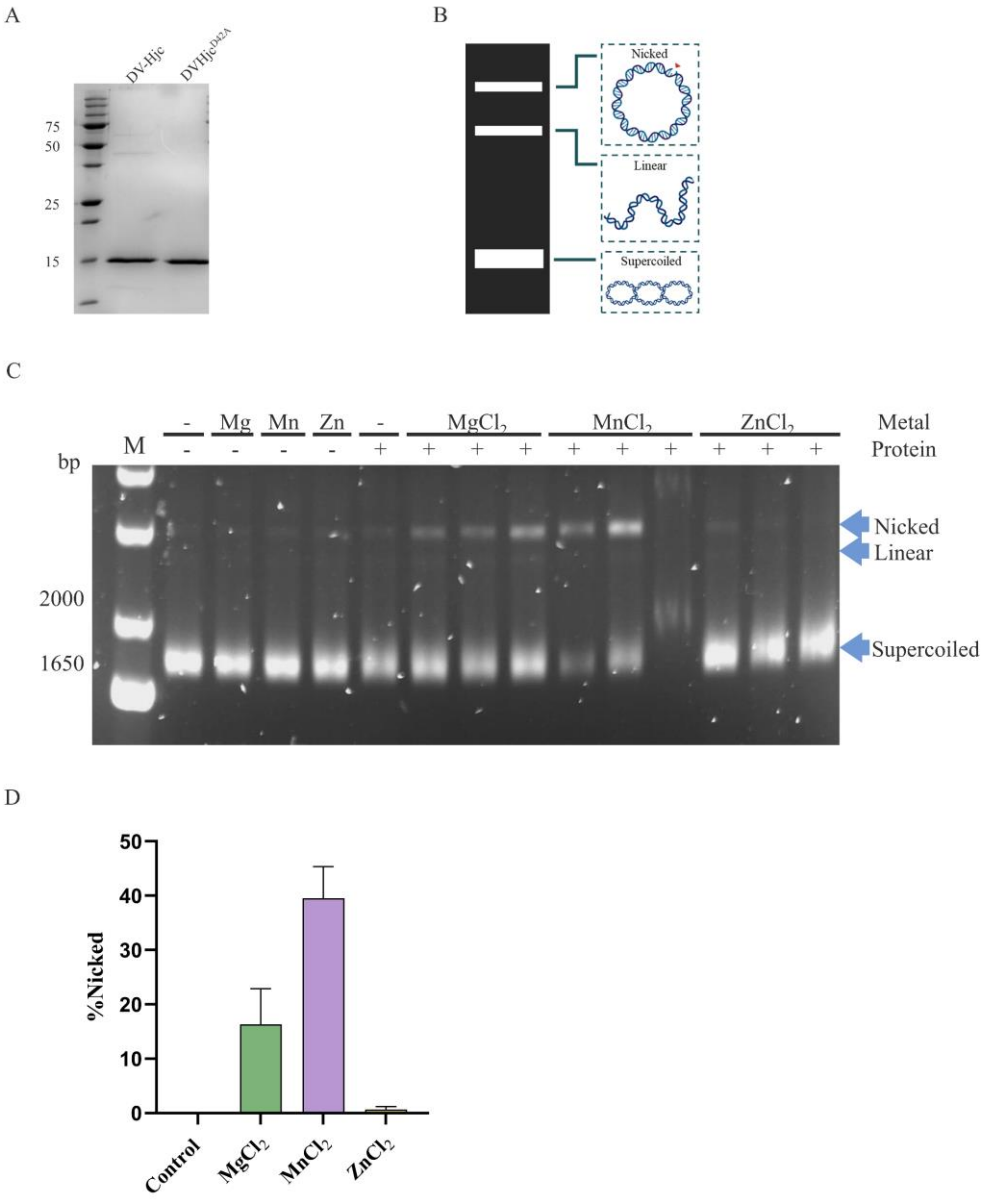


Figure 37: Metal ion testing of DV-Hjc nucleolytic activity with pUC19.
***A:** Purified DV-Hjc and DV-Hjc^{D42A} proteins used in all assays shown. Protein samples were electrophoresed on 15% SDS-PAGE gels. The Precision Plus Protein™ standard was used as a molecular weight indicator.*
***B:** Illustration of the plasmid-based assay results on an agarose gel with depictions of the plasmid on the right. Nicked plasmid runs highest, linear plasmid runs in the middle and supercoiled plasmid runs lowest on the gel. Plasmid is initially present in supercoiled.*
***C:** 1x SYBR Safe™ 1% agarose gel of pUC19 plasmid-based assay with 8µM DV-Hjc protein and 108mg/mL pUC19 plasmid. Reactions were carried out in triplicates. The presence (+) and absence (-) of protein and metal are indicated above the gel. The position of supercoiled plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).*
***D:** Summary of the data shown in C. The Y-axis shows the % nicked product. Four bar graphs are shown representing from left to right the control reaction, $MgCl_2$, $MnCl_2$ and $ZnCl_2$ reactions. The reactions were carried out in triplicates and the average of those is shown here, including standard error bars, except for the $MnCl_2$, where one of the triplicates was removed.*

DV-Hjc pUC19 Plasmid-Based Assay Optimisation

As shown in Figure 37, the activity of DV-Hjc with the 'standard' assay setup in 10 mM Bis-Tris pH 6.5 and 50 mM NaCl only gave a product-to-substrate turnover of 40%. This was partially due to the

instability of the DV-Hjc protein in the purification buffer, which was therefore optimised as described in Section 3.2 by changing the protein into the DV-Hjc-specific storage buffer. Additionally, the reaction conditions were optimised to stabilise DV-Hjc in the reaction and achieve higher activity. The most optimal reaction conditions for the reaction buffer were: 50 mM HEPES (pH 6.5), 50 mM NaCl, 0.1 mg/mL BSA (Bovines Serum albumin), 0.5 mM DTT, and 6% glycerol. The optimised conditions allowed for DV-Hjc reactions to be carried out at higher temperatures further increasing its activity (Figure 38). While at initial conditions protein activity decreased at temperatures above 5°C, the protein was stable and active at temperatures as high as 35°C with the adapted conditions. Additionally, even though the protein concentration was only 4.8 µM in the assay with the adapted conditions, cleavage efficiency reached 97% after just 2 hours at 5°C, while with the standard setup at the same temperature and a protein concentration of 8 µM a cleavage efficiency of only 32% was reached after an incubation period of 12 hours (Figure 38, C). These optimised conditions were used for all following assays of DV-Hjc.

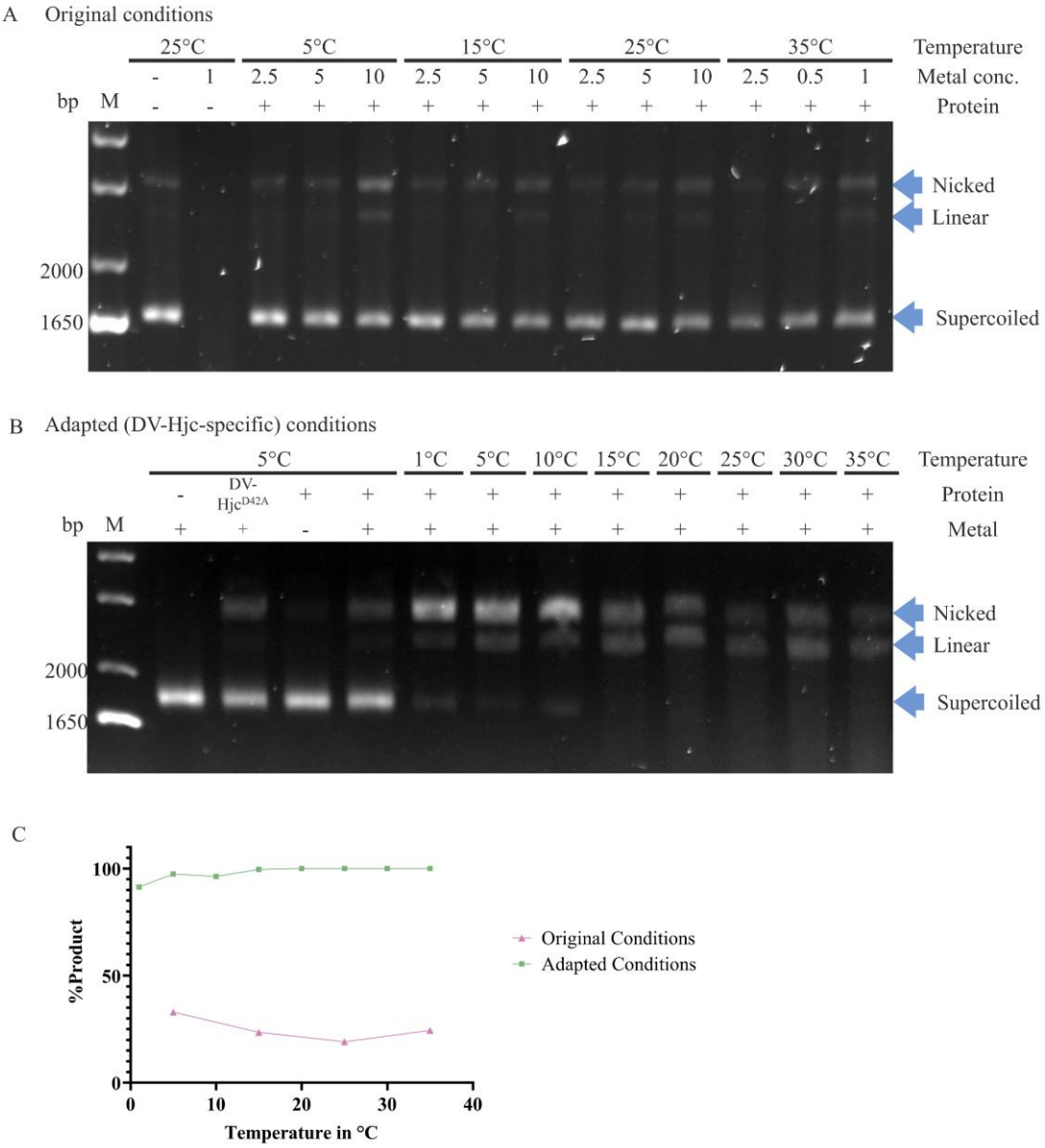


Figure 38: Temperature testing of DV-Hjc nucleolytic activity with pUC19 and assay buffer adjusting.
1x SYBR SafeTM 1% agarose gel of pUC19 plasmid-based assay with 108mg/mL pUC19 plasmid. Indicated above the gel are: presence (+) or absence (-) of protein; metal ion absence (-), presence (+) or the concentration of added metal ion (in mM); temperature (in °C). The position of the supercoiled plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).
***A:** Protein assay using standard conditions to test a temperature gradient from 5°C to 35°C and a metal gradient from 2.5 to 10 mM of $MgCl_2$. The standard buffer at pH 8.0 (see Supplementary Table 11) was used. The reactions were incubated for 12 hours before electrophoresis and had a protein concentration of 8 µM.*
***B:** Protein assay using adapted (DV-Hjc-specific) conditions to test a temperature gradient from 1°C to 35°C with the adjusted DV-Hjc specific buffer at pH 6.5 (see Supplementary Table 12) and 4.8 µM protein. The reactions were incubated for 2 hours before electrophoresis.*
***C:** Summarising graph of the %product (nicked and linearised) of the assays shown above. Keep in mind the differing protein concentrations used for the assays. The 'Original conditions' series shows the results from A, with a protein concentration of 8 µM and the standard assay setup. The 'Adapted conditions' series shows the results from B, with a protein concentration of 4.8 µM and the DV-Hjc specific assay setup.*

To further optimise DV-Hjc activity with the pUC19 plasmid substrate, a range of $MgCl_2$ concentrations were tested. The assay was done with $MgCl_2$ concentrations ranging from 0 to 25 mM at 10°C for 2 hours. As shown in Figure 39, the highest activity level was achieved with $MgCl_2$ added at a concentration of 10 mM with 98% of the substrate turned over into product. 80% of the product was

present as nicked plasmid and 18% as linear plasmid. Increasing the $MgCl_2$ above 10 mM slowly lowers the activity of DV-Hjc. Concentrations above 25 mM need to be tested to determine at which concentration – if any – of $MgCl_2$ will inhibit the activity of DV-Hjc. While the activity was highest at 10 mM $MgCl_2$, the percent linearised plasmid was highest at 1 mM $MgCl_2$ with 31% linearised, followed by 5 mM $MgCl_2$, where 27% of the substrate was linearised. At 1 mM $MgCl_2$, the linear to nicked product ratio was near 1 (Figure 39, B). With increasing metal concentrations, the ratio of linearised to nicked product was lowered to 0.4 at 5 mM $MgCl_2$, and 0.2 at 10 mM $MgCl_2$. The same trend was observed when doing the assay over a narrower concentration gradient from 0 mM to 10 mM $MgCl_2$ concentration in the assay in intervals of 2.5 mM. At a lower concentration of $MgCl_2$ DV-Hjc appears to create a cut in both DNA strands of the plasmid DNA but the total substrate turnover into product (linearised and nicked) is lower compared to the higher $MgCl_2$ concentrations (Figure 39, C).

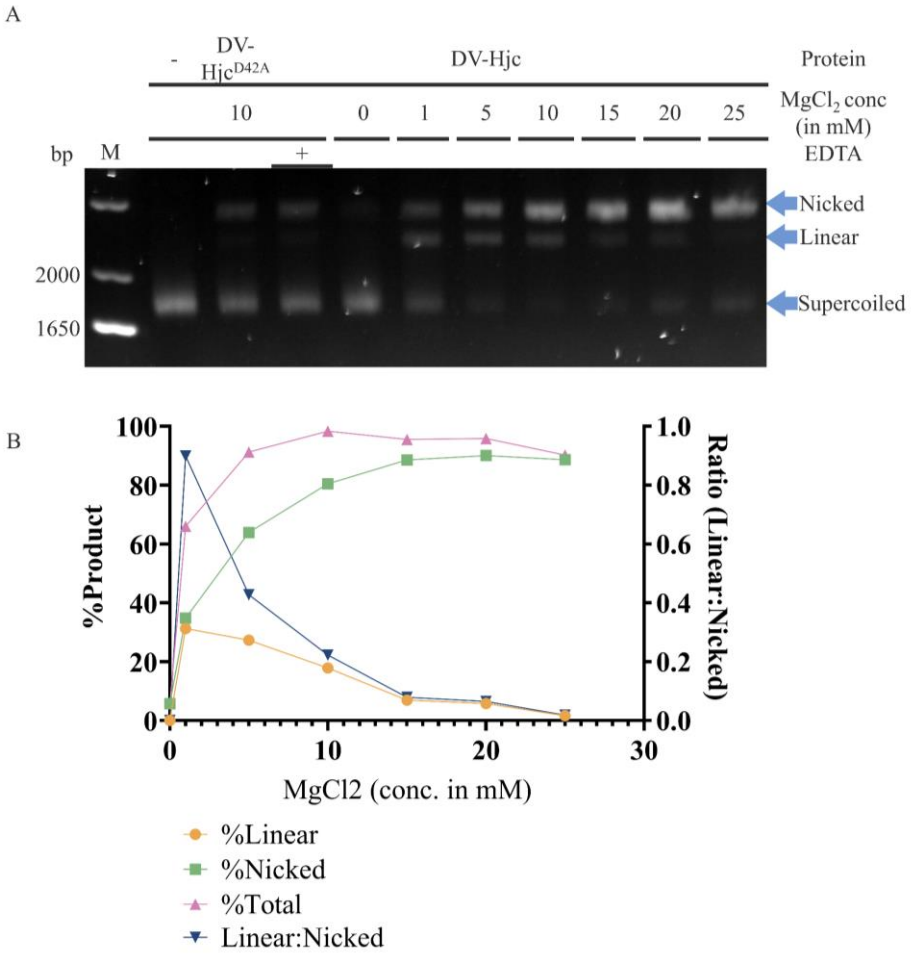


Figure 39: $MgCl_2$ concentration gradient to test the nucleolytic activity of DV-Hjc with pUC19 plasmid and the adapted DV-Hjc specific conditions.
***A:** 1x SYBR SafeTM 1% agarose gel of pUC19 plasmid-based assay. Reactions contained 108mg/mL pUC19 plasmid and 1 µM protein and were incubated for 2 hours at 5°C before electrophoresis. Indicated above the gel are: absence (-) of protein or presence of DV-Hjc or DV-Hjc^{D42A}; $MgCl_2$ concentration (in mM); addition of EDTA (+). The position of the supercoiled*

plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).
B: *Summary of the data from A. The x-axis shows the $MgCl_2$ concentration in mM. Three graphs show the %Linear, %Nicked and %Total product (nicked and linear), which are represented on the left Y-axis showing the % product. One graph shows the linear to nicked DNA ratio represented on the right Y-axis.*

DV-Hjc^{D42A} Mutant Analysis

Based on alignments of the DV-Hjc protein with other proteins belonging to the UPF0102 family and characterised archaeal Hjc proteins, the conserved Asp42 residue was mutated to an Ala to produce an inactive mutant. Mutation of the predicted active site aspartic acid residue (Asp42) reduced the activity of the protein by more than 50% (Figure 40, C). The DV-Hjc WT protein turned over 95% of the pUC19 plasmid substrate into product (linear and nicked), while DV-Hjc^{D42A} turned over only 41%. This confirms that Aps42 in DV-Hjc is important for the activity of the protein, and it was used as an active site mutant in the assays.

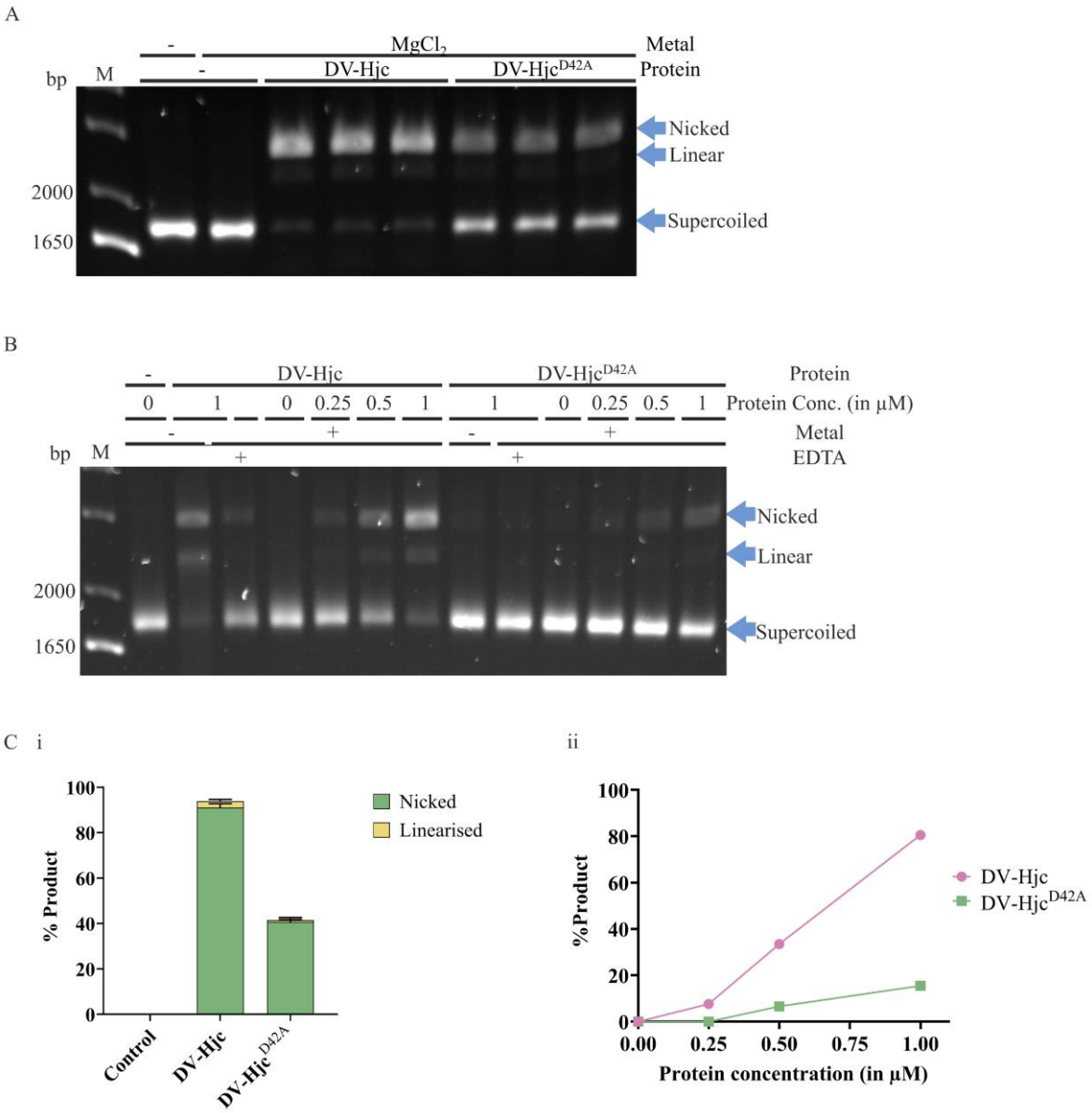


Figure 40: Testing of the DV-Hjc^{D42A} active site mutant of DV-Hjc.
A: 1x SYBR SafeTM 1% agarose gel of pUC19 plasmid-based assay. Reactions contained 108mg/mL pUC19 plasmid and 4.8 µM protein and were incubated for 2 hours at 5°C before electrophoresis. Reactions were carried out in triplicates. Indicated above the gel are: absence (-) of protein or presence of DV-Hjc or DV-Hjc^{D42A}; absence (-) or presence of $MgCl_2$. The position of the supercoiled plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).
B: 1x SYBR SafeTM 1% agarose gel of pUC19 plasmid-based assay. Reactions contained 108mg/mL pUC19 plasmid and 0.25 to 1 µM protein and were incubated for 2 hours at 5°C before electrophoresis. Indicated above the gel are: absence (-) of protein or presence and concentration of DV-Hjc or DV-Hjc^{D42A}; absence (-) or presence of $MgCl_2$, and the presence of EDTA (+). The position of the supercoiled plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).
C: Summary of the data shown in A (i) and B (ii). (i) On the Y-axis the % product is shown which is separated into nicked and linearised product. Three bar graphs are shown representing (from left to right) the control reaction, DV-Hjc reactions and DV-Hjc^{D42A} reactions. The reactions were carried out in triplicates and the average of those is shown here, including standard error bars. (ii) Two graphs show the %product of DV-Hjc and DV-Hjc^{D42A}. The x-axis shows the protein concentration (in µM) and the y-axis shows the %Product.

Time Dependence of DV-Hjc

To compare the turnover of supercoiled pUC19 plasmid into nicked and linear plasmid by DV-Hjc over time a time series was carried out over 8 hours in intervals of 2 hours at 5°C. The results show that after 2 hours 93%, after 4 hours 97% and after 6 hours 100% of the substrate is turned over into product (linear and nicked combined) (Figure 41). When comparing the ratio of the nicked to linear product over time the amount of linear product increases, as shown in Figure 41 (B). Suggesting that DV-Hjc cleavages the second DNA strand over time.

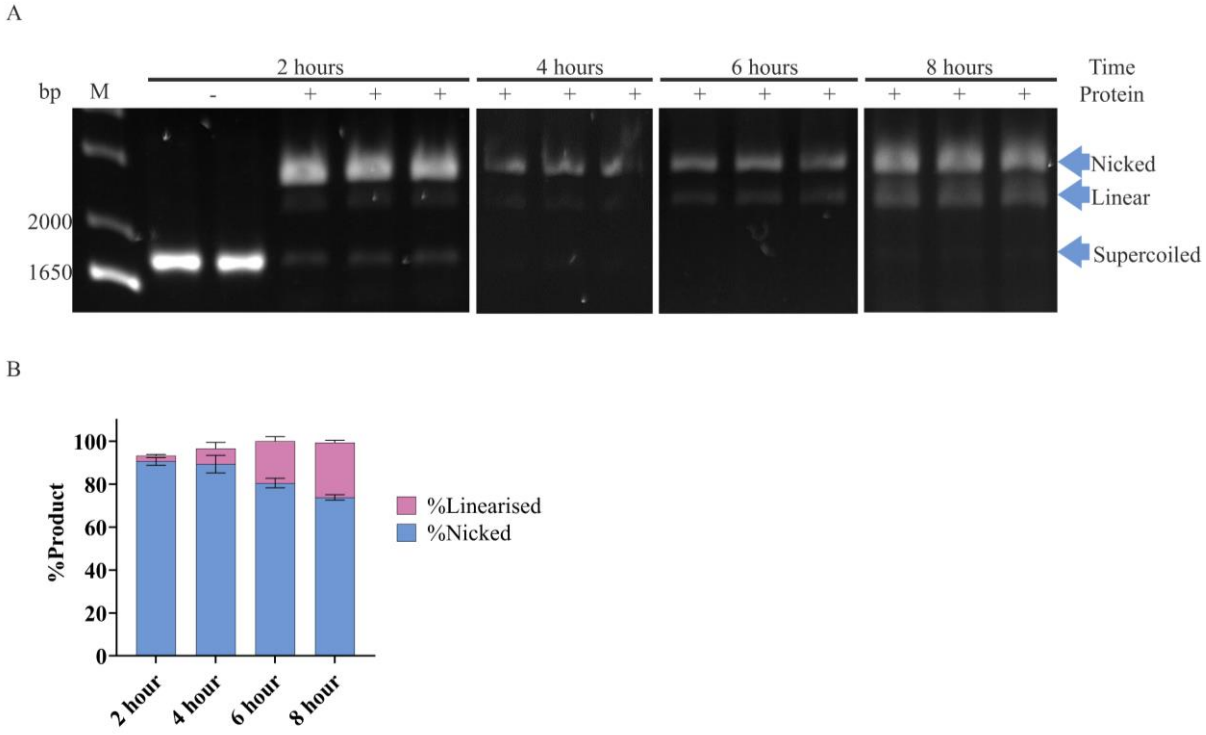


Figure 41: DV-Hjc nucleolytic activity assay over 8 hours.
A: *1x SYBR SafeTM 1% agarose gel of pUC19 plasmid-based assay. Reactions contained 108mg/mL pUC19 plasmid and 1 µM protein and were incubated for 2 to 8 hours at 5°C before electrophoresis. Indicated above the gel are: absence (-) or presence of DV-Hjc, incubation time (in hours). The position of the supercoiled plasmid substrate, nicked plasmid and linearised plasmid are indicated by the blue arrows. A 1 kb+ DNA ladder was used as a molecular standard (M).*
B: *Summary of the data shown in A. On the Y-axis the % product is shown which is separated into nicked and linearised product. Four bar graphs are shown representing (from left to right) the 2-, 4-, 6-, and 8-hour incubation periods. The reactions were carried out in triplicates and the average of those is shown here, including standard error bars.*

3.4.5 Proposed DNA Cleavage Mechanism of DV-Hjc

The biochemical and structural characterisation of the DV-Hjc protein, supported by predictions made using AlphaFold3, suggests that DV-Hjc is a nuclease enzyme with activity on double-stranded DNA. It uses a single Mg^{2+} ion, which is positioned in its active site by two acidic residues. The active site of DV-Hjc consists of Glu13, Asp42, Glu55, and Lys57, resembling the active site of archaeal Hjc

proteins. However, little is known about the specific catalytic mechanisms of archaeal Hjc proteins, largely due to a lack of structural data.

As discussed above the active site lysine residue is conserved across a range of metallonucleases, including archaeal Holliday junction nucleases, and some Type II restriction endonucleases. While it was initially thought that the lysine in the active site stabilises the unstable transition state during the cleavage of the phosphate backbone, studies of EcoRV, suggest otherwise because of its position in the active site (Horton et al., 1998). Further evidence shows that, apart from a few nuclease enzymes such as BamHI, lysine cannot typically be replaced by an acidic side chain (Cowan, 2004). This suggests that the lysine residue is more likely to be involved in the stabilisation and orientation of the attacking hydroxyl ion. This is often done by the deprotonation of water by a general base, i.e. the lysine residue (Yang, 2011). In the case of EcoRV, the lysine also donates a second hydrogen bond to the 3'phosphate.

The hypothesis for DV-Hjc is that the DNA is bound by lysine and arginine residues in the large positively charged binding surface, which orients the DNA phosphodiester backbone in the active site of the protein. In the active site, a single Mg^{2+} ion is oriented by Asp42 and Glu55. The lysine residue in the DV-Hjc active site activates the nucleophile via deprotonation of water and then orients the hydroxyl group for nucleophilic attack on the scissile phosphate (Figure 42). This results in the formation of an unstable penta-covalent intermediate, which is stabilised by the Mg^{2+}. Finally, the leaving group is protonated, potentially by a metal-ion-coordinated water molecule, leading to a nick in the double-stranded DNA strand. The role of Glu13 was not determined and to determine its function further investigation is required, which could be done by site-directed mutagenesis. The predicted mechanism is indicated in Figure 42.

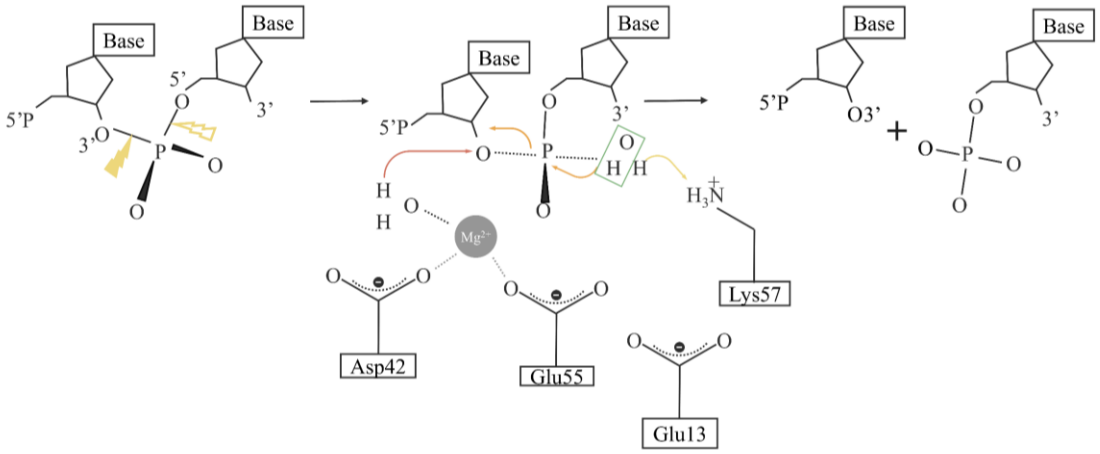


Figure 42: Predicted phosphodiester bond cleavage mechanisms by DV-Hjc.

To more accurately determine the activity of the DV-Hjc protein a structure of the DV-Hjc protein bound to a DNA substrate will be required. This will give insight into the position of the DNA strand in the DV-Hjc active site and may also provide information about the quaternary structure of the protein. This may be achieved by comparing analytical size exclusion data of the apoprotein and the protein-DNA complex of DV-Hjc with different types of DNA substrates. Additionally, a crystal structure of DV-Hjc bound to DNA substrates would be desirable or structural data of the protein-DNA complex in solution via small angle X-ray scattering

3.5 *In Vivo* Characterisation of the *E. coli* UPF0102 Protein

To determine the function of the UPF0102 protein family in cells *in vivo*, cells from the Keio knockout collection were grown under DNA-damaging conditions. The Keio knockout collection is a collection of *E. coli* K-12 cells, which have single-gene knockouts of every single non-essential gene in the *E. coli* genome (Baba et al., 2006). The *yraN* gene knockout strains were used for this purpose. UV radiation and H_2O_2 treatment were used as DNA-damaging stressors since they both induce high levels of DNA damage in cells as discussed in Section 1.1.2..

3.5.1 H_2O_2 Treatment

H_2O_2 causes the formation of reactive oxygen species in cells most importantly hydroxyl radicals, which can cause a variety of DNA damages, including double-stranded breaks, and oxidised nucleotide bases, such as 8-oxo-G (Fasnacht & Polacek, 2021).

For the H_2O_2 treatment, the *ruvC* gene knockout was used as a positive control. This knockout has a higher sensitivity to H_2O_2 compared to other knockouts (Dhawale et al., 2021). Additionally, it codes for the Holliday junction endonuclease RuvC, which shares some common functional features with the DV-Hjc protein as discussed above. The WT (BW25113), the *yraN* knockout (JW3117) and the *ruvC* knockout (JW1852) were grown with and without 3 mM H_2O_2 for 90 min. After 90 min the H_2O_2 was removed from the cells via centrifugation and growth was continued.

The WT and *yraN* knockout growth curve based on OD_{600} shows that under normal conditions, they grow at similar rates. During their linear growth phase, the WT and *yraN* knockout had a slope of $3.85*10^{-3}$ and $3.71*10^{-3}$, respectively (Figure 43, A). Compared to those the growth rate of the *ruvC* knockout is lower with a slope of $3.24*10^{-3}$. If this is compared to the growth with 3 mM H_2O_2, growth is reduced in all three strains with the treatment, with a slope of $2.54*10^{-3}$ for the WT, $3.12*10^{-3}$ for the *yraN* knockout, and $2.12*10^{-3}$ for the *ruvC* knockout. The highest reduction in the slope of the growth

curve can be observed for the WT with a reduction of 1.13, followed by the *ruvC* knockout with 1.12 and the *yraN* knockout with 0.59.

Following H_2O_2 treatment of 90 min cultures were also plated to record their CFU/mL. The average CFU/mL for each strain (treated and untreated) is shown in Figure 43 (B). The CFU/mL of untreated *yraN* knockouts was much higher compared to that of the other two strains. The survival of each strain was determined using the treated and untreated CFU/mL of each strain. The survival was lowest in the *yraN* knockouts, followed by the *ruvC* knockout and the WT *E. coli* (Figure 43, C). The fold-difference between the survival of the knockout strains and the WT showed the largest difference between the *yraN* knockout and the WT (Figure 43, D). As shown in Figure 43 (B), the CFU/mL of the untreated *yraN* knockouts is much higher compared to that of the WT and RuvC knockout, even though cells were treated with H_2O_2 at the same OD_{600}. The colony morphology between the *yraN* knockout also differed from the other strains, with much smaller colonies. The high CFU/mL of the untreated *yraN* knockout strain may have affected the survival calculation shown in Figure 43 (C). Therefore, the H_2O_2 treatment and specifically the OD_{600}, at which the H_2O_2 treatment is started, requires optimisation to achieve the same CFU/mL for the untreated strains. This would give a stronger basis for a more accurate comparison of the three strains.

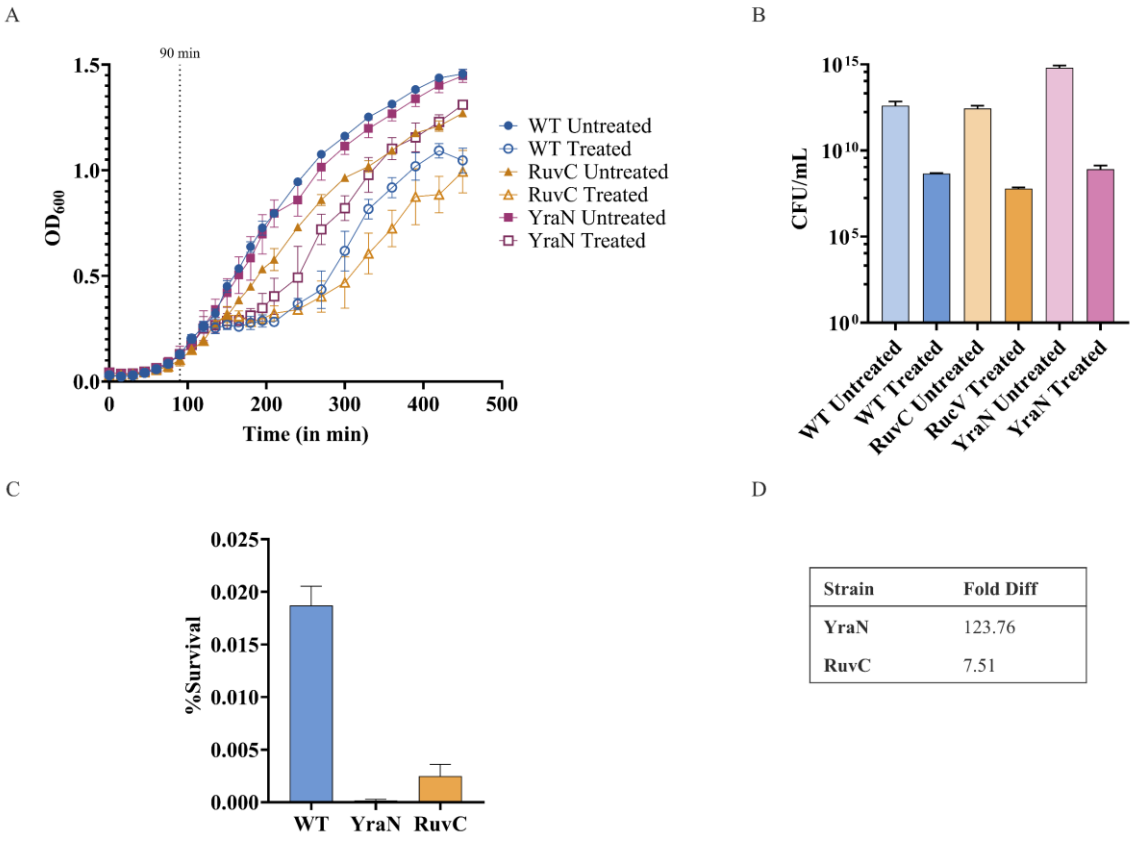


Figure 43: 90-minute 3 mM H_2O_2 treatment of E. coli strains (WT, yraN knockout and ruvC knockout).
***A:** Growth kinetics of the stains in LB medium, with and without 3 mM H_2O_2 treatment. The H_2O_2 was removed after 90 min (as indicated by the dashed line). The graphs show the OD_{600} of each culture (y-axis) over a time span of 450 min (x-axis).*
***B:** Bar graph showing the CFU/mL of the three strains grown on LB agarose plates after 90 min of growth under H_2O_2 treatment, or growth without treatment.*
***C:** Survival of the three strains after growth with H_2O_2 treatment.*
***D:** Fold-difference between the %survival of the WT E. coli and the yraN and ruvC gene knockout strains.*

Reactive oxygen species produce a range of DNA lesions including double-stranded DNA breaks and base modification. As discussed above different DNA repair pathways are responsible for the repair of those breaks. Most importantly homologous recombination for the repair of double-stranded DNA breaks and nucleotide or base excision repair pathways for the repair of base lesions. Previous studies of H_2O_2 treatment of *E. coli* knockouts showed a reduction in the survival of knockouts from different DNA repair pathways. Treatment of nucleotide excision repair pathway protein-deficient cells with 3 mM H_2O_2 for 90 min had lower survival than the WT strain (Dhawale et al., 2021; Dhawale & Rath, 2019). This was also shown for the RecFJQ-mediated double-stranded repair pathway under oxidative stress (Dhawale et al., 2021). Holliday junctions are an intermediate during double-stranded break repair in bacteria. Their resolution is important for the survival of cells after damage repair. *E. coli* deficient in the Holliday junction resolving enzymes RuvABC and RecA show higher sensitivity to oxidative stress (Dhawale et al., 2021). This highlights the importance of functional DNA repair during oxidative stress. The reduction in survival of UPF0102 protein-deficient *E. coli* may suggest its involvement in a

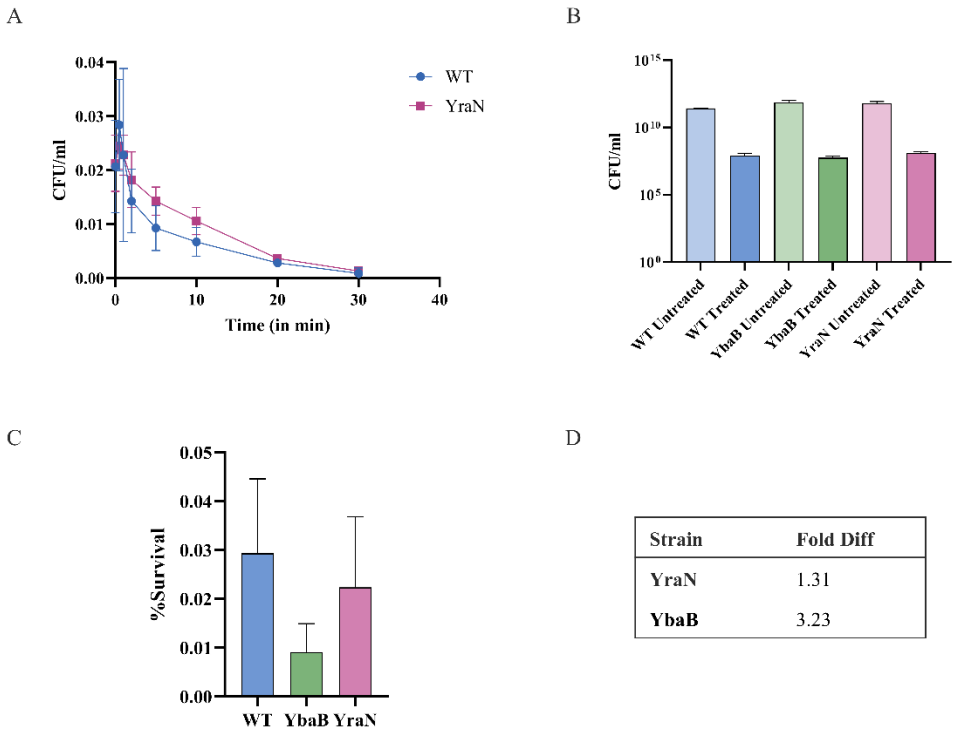
DNA repair pathway that is important for the repair of reactive oxygen species-induced damage. However, as mentioned above to more accurately determine the importance of the YraN (UPF0102) protein for H_2O_2 survival the experiment requires further optimisation.

3.5.2 UV Treatment of UPF0102 Deficient *E. coli* Cells

To test whether the YraN protein may help in UV-radiation-associated damages. YraN knockout cells from the Keio collection were treated with high levels of UV radiation (Baba et al., 2006). Growth curves of the YraN knockout and the WT *E. coli* cells show no great difference (Figure 44, A). The variability between the CFU/mL was large specifically at the start of the time series.

The CFU/mL of three 10 Jm^{-2} UV-treated and untreated *E. coli* strains (YraB, YraN and WT), were used to determine the survival of the cells (Figure 44, B and C). The fold-difference between the survival of the WT and the *yraN* and *ybaB* knockouts was 1.31 and 3.23, respectively (Figure 44, D). The difference between the WT and *yraN* knockout was not large enough to make conclusions about the importance of the protein in UV-induced damage repair. The results do however show a difference in the survival between the *yraN* and *ybaB* knockouts. The survival of the *ybaB* knockout was lower compared to that of the *yraN* knockout. This suggests that the YraN protein is less important for UV-radiation survival than the YbaB knockout.

A comparable study by Sargentini et al. (2016) treated *E. coli* isogenic knockouts with 75 Jm^{2-} resulting in $4.13*10^{-5}$ survival of the *ybaB* knockout and a 976-fold difference compared to the WT cells. This suggests that a higher UV dose may be required to determine significant differences in the survival between the WT and knockout strains of *E. coli*. A UV radiation dose over 10 Jm^{-2} was not possible with the instrument used here. However, a new instrument for UV treatment has been purchased and will be used to repeat this experiment with higher UV radiation levels. However, the results did not show whether YraN is important for UV radiation survival. Hardy et al. (2024) determined the growth of *yraN* knockouts from *L. pneumophia* and *A. baylyi* and the WT, comparison of them showed no difference between the growth of the knockouts and the wild type.



Strain	Fold Diff
YraN	1.31
YbaB	3.23

Figure 44: UV radiation treatment of E. coli strains (WT, yraN knockout and ybaB knockouts).
***A:** Growth kinetics of the stains in LB medium, with UV-radiation over 30 min.*
***B:** Bar graph showing the CFU/mL of the three strains grown on LB agarose plates after 10 Jm^{-2} UV irradiation treatment.*
***C:** Survival of the three strains after treatment with 10 Jm^{2-}UV irradiation treatment.*
***D:** Fold-difference between the % survival of the WT E. coli and the yraN and ybaB gene knockout strains*

3.6 Conclusion and Future Directions

The DV-Hjc protein is a homologue of the archaeal Hjc proteins. Structurally the protein is most similar to the *P. furiosus* Hjc. However, DV-Hjc is missing the aromatic side chains for dimer formation and the two arginine residues (Arg10 and Arg25), that are important for the catalytic function of *P. furiosus* Hjc. While these residues are not part of the nuclease centre of Hjc, they are important for Holliday junction cleavage by the protein (Komori et al., 2000; Nishino, Komori, Ishino, et al., 2001). DV-Hjc has residues required for DNA binding and binds double-stranded DNA and Holliday junction substrates with similar high affinity. However, the protein does not cleave the Holliday junction substrate tested here. It does however cleave plasmid DNA, which suggests that DV-Hjc is either specific to linear coiled DNA or a DNA sequence. Other known Holliday junction resolvases, like the RuvC protein show sequence-specific activity on Holliday junctions (Górecka et al., 2019). The activity of DV-Hjc with DNA and the requirement of the UPF0102 protein in *E. coli* during oxidative stress, shown with H_2O_2 treatment of YraN knockouts, suggests that this protein may be required for a DNA repair process for repair of reactive oxygen-induced DNA damages. DV-Hjc is a homologue of the YraN protein, which was recently shown to be involved in a homologous recombination event during

natural transformation in *L. pneumophila* and *S. elongatus*. Its structure and active site are similar to archaeal Hjc proteins. Based on this the hypothesis is that DV-Hjc is involved in the resolution of Holliday junctions in a homologous recombination repair process. However, DV-Hjc did not show activity with Holliday junction substrates.

The absence of activity with Holliday junction substrates may be due to sequence-specific cleavage of the protein. This will be addressed by narrowing down the sequence at which the protein cleaves the pUC19 plasmid. Additionally, the structural characterisation of a DV-Hjc-DNA complex by crystallisation and SAXS analysis may provide further insight into the activity of the protein. This could be further verified by using site-directed mutagenesis of the predicted DNA binding and active site residues.

To determine the role of the UPF0102 protein *in vivo*, experiments with *E. coli* from the Keio collection need to be further optimised as mentioned above. Additionally, as the UPF0102 protein, is present in a range of bacterial strains, its role may also be explored in other bacterial strains in a similar way as was demonstrated here. To confirm whether the protein is required in DNA repair processes other treatments, like cisplatin treatment, could be explored.

4 CHAPTER FOUR: DV1-1 NUCLEASE

4.1 DV1-1 Protein Background

Many known DNA repair proteins have multiple functionally different domains. A well-known multidomain DNA repair protein is the LigD protein. LigD as described above is involved in the non-homologous end-joining repair pathway for the repair of double-stranded breaks (Pitcher et al., 2007). Double-stranded breaks are the most deleterious type of DNA damage, which can be fixed by non-homologous end joining as an alternative double-stranded break repair pathway to homologous recombination (Bertrand et al., 2019; Pitcher et al., 2007; Shuman & Glickman, 2007). In non-homologous end joining the ends of the double-stranded break are bound by the Ku protein and the ends are then processed and sealed by a LigD ligase (Bertrand et al., 2019). This process can result in error-free repair of double-stranded breaks, but due to the lack of a homologous template for repair errors may be introduced.

LigD ligase and its function in non-homologous end joining was first identified in *M. tuberculosis*, but it has since been identified in a range of other bacteria, including *Bacillus subtilis*, *Agrobacterium tumefaciens* and *Pseudomonas aeruginosa*, (de Vega, 2013; Weller et al., 2002; Zhu & Shuman, 2005; Zhu & Shuman, 2007). LigD is a unique ligase protein, which can have an ATP-dependent DNA ligase, phosphodiesterase and PrimPol domain and exhibits functions of all three domains (Pitcher et al., 2005) (Figure 45, A and B). Some LigD proteins have only the ATP-dependent DNA ligase and polymerase domains, like the LigD protein from *B. subtilis* (de Vega, 2013) (Figure 45, C). As shown in Figure 45, the domain order can vary between the LigD proteins from different organisms. The multiple domains of LigD allow it to carry out the same functions, which in eukaryotic organisms are carried out by many separate proteins (Pitcher et al., 2007).

s

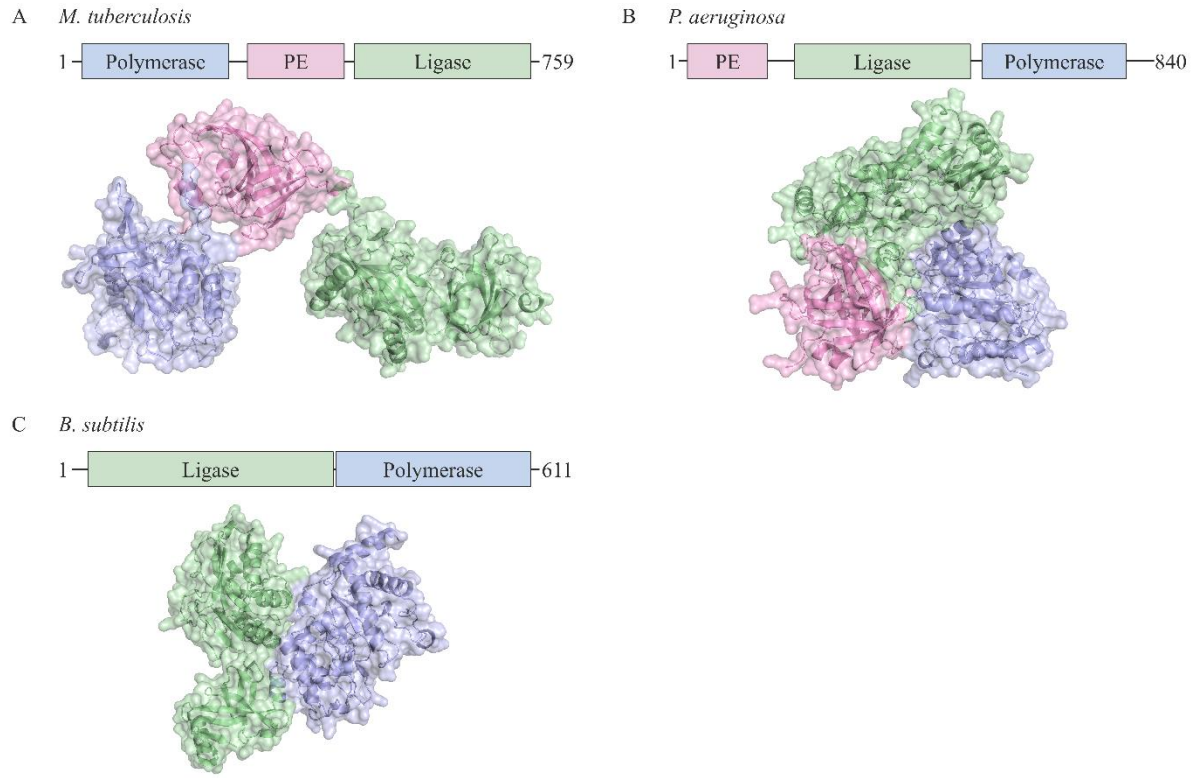


Figure 45: The multidomain protein LigD.
The structures of these proteins were predicted using AlphaFold3. The ligase domain is shown in green, the polymerase domain in blue, and the nuclease domain in pink.
***A:** M. tuberculosis LigD. **B:** P. aeruginosa LigD. **C:** B. subtilis LigD.*

4.1.1 Identification of a Unique Nuclease-Ligase Fusion Protein from Antarctic Dry Valley Metagenomes

As described in Rzoska-Smith; et al. (2023) metagenomes from the Dry Valleys were mined for protein sequences of unique DNA repair enzymes, including the search for an N-terminal DNA-binding domain, that is found in the LigB class DNA_ligase_A_N (PF04675). Multiple sequences with this domain were found in the Dry Valley metagenomes. A sequence similarity network of those sequences with LigB type ligase domain containing UniRef sequences (PF04675) at a 50% threshold formed four large clusters. Most sequences from the Dry Valley metagenomes formed a cluster with the UniRef sequences (Figure 46, A, i, iii, and iv). One cluster contained fewer UniRef sequences than those (Figure 46, A, ii). A hmmer search of the sequences in this cluster revealed that 13 sequences contain an N-terminal Zn-dependent metallo-hydrolase RNA specificity domain RMMBL (PF07521) fused to the LigB ligase DNA-binding domain (Figure 46, B).

As shown in Figure 46 (C), there is little synteny between the clusters in which this unique DNA ligase protein is located. In contig Ga0136611_10000860, the DNA ligase protein sits in a gene cluster with

other predicted DNA repair and replication enzymes, including a RecA-domain protein, two error-prone DNA polymerases, and an AP endonuclease (Rzoska-Smith; et al., 2023). Based on its uniqueness and its position in a cluster with other DNA repair and replication enzymes the nuclease-ligase protein (Gene ID: Ga0136611_1000086013) (DV1-1 from here on) is predicted to be involved in a unique DNA repair pathway and was chosen for further characterisation.

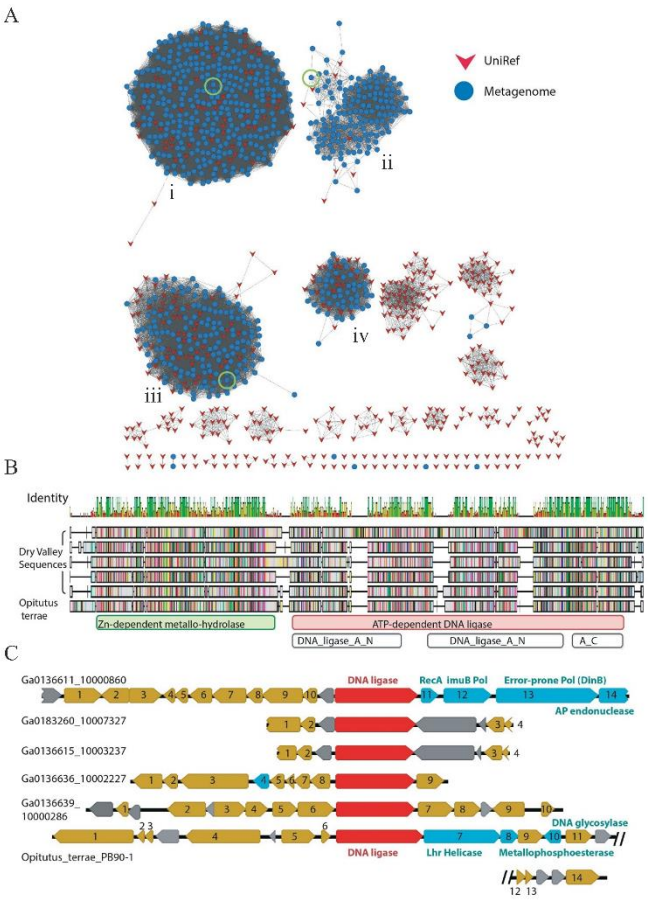


Figure 46: Identification of the DV1-1 protein from Dry Valley metagenomes.
A: *Sequence similarity network of hmmer search results to LigB-type DNA ligase proteins at a 50% identity threshold. The blue coloured circles indicate nodes from Dry Valley metagenome sequences and the red arrows indicate nodes from UniRef50 sequences. The sequence of the DV1-1 protein is indicated by the green circle in cluster ii.*
B: *Alignment of Dry Valley metagenome sequences with a N-terminal Zn-dependent metallo-hydrolase domain and a C-terminal ATP-dependent DNA ligase domain with the sequence of the Opititus terrae DNA ligase gene OTER_RS15935 from the genome sequence NC_010571_Opitutus_terrae_PB90-1. The results from Pfam/Interpro sequence searches of the sequences are shown below the alignment.*
C: *Genomic context of Dry Valley contigs with a homologue of the DV1-1 ligase-nuclease fusion protein. The ligase-nuclease fusion protein is shown in red. Predicted DNA repair proteins are shown in blue. Hypothetical proteins are shown in grey. All other proteins are shown in yellow.*
This figure was taken from Rzoska-Smith; et al. (2023).

4.1.2 Full-length DV1-1 Protein

A protein BLAST search with the DV1-1 protein showed that homologues are present in multiple bacteria belonging to the Verrucomicrobiota phylum (Altschul et al., 1997; Altschul et al., 2005). This aligns with the IMG (IMG/M: https://img.jgi.doe.gov/m/) prediction which places the DV1-1 gene in

the genome of *Chthoniobacter flavus*, which also belongs to the Verrucomicrobiota phylum (Chen et al., 2023). A similar ligase-nuclease fusion protein is present in *Opititus terrae* but the function of such a ligase-nuclease fusion protein has not yet been characterised (Ejaz & Shuman, 2018). An InterPro scran places the N-terminus of this protein in the protein superfamily of Metallo-hydrolase/oxidoreductase (SSF56281) and the C-terminus in the protein family of ATP-dependent ligases (IPR050191) (Jones et al., 2014). Therefore, the N- and C-terminal domains are named DV1-1 NucDom and DV1-1 LigDom, respectively (referred to as NucDom and LigDom from here on).

As mentioned earlier, the DV1-1 protein is co-located in a gene cluster with other DNA repair proteins. These include a predicted RecA protein, an imuB polymerase, an error-prone polymerase and an AP endonuclease protein. The location of these proteins in the same operon suggests that they are involved in the same DNA processing process (Kountz & Balskus, 2021). Based on this the hypothesis is that DV1-1 is a DNA processing enzyme.

Full-length DV1-1 Protein Structure

The structure of DV1-1 has not yet been determined; therefore, an AlphaFold3 predicted structure was used for the structural characterisation of the DV1-1 protein. The whole protein has a mass of 105 kDa, and the NucDom and LigDom have a mass of 44.3 kDa and 61.5 kDa, respectively. The predicted structure of DV1-1 is a mixed α-β-protein with a globular shape. The NucDom and LigDom are located close to each other and connect at the DNA-binding domain and the oligonucleotide/oligosaccharide-binding of LigDom and an α-helix from the MBL-domain of NucDom. The interactions between the two domains are mainly polar ones. The two domains are linked by a 37-residue-long linker region, which is a relatively long linker region compared to linker regions of other proteins (Reddy Chichili et al., 2013). The flexibility of the linker may be determined through small-angle X-ray scattering (SAXS) of the whole DV1-1.

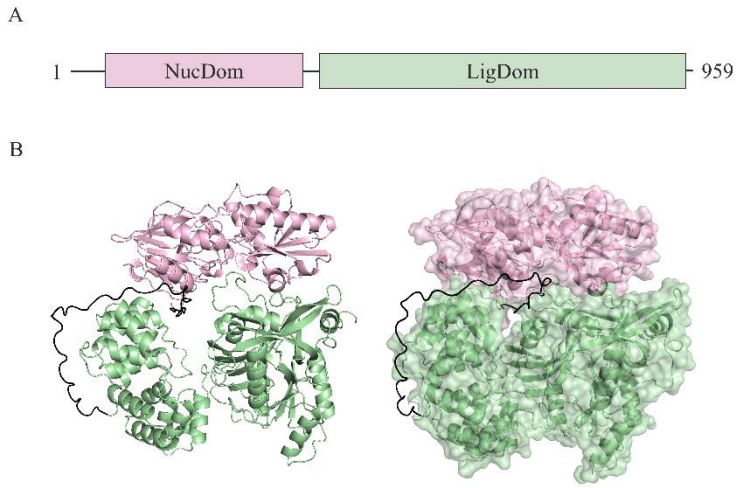


Figure 47: DV1-1 ligase-nuclease fusion protein.
***A:** Graphical depiction of DV1-1 amino acid sequence (left to right: N- to C-terminus) showing the nuclease domain (NucDom) in pink and the ligase domain (LigDom) in green.*
***B:** Cartoon figure of DV1-1 showing NucDom in pink and LigDom in green. The linker region is shown in black.*
This figure was adapted from Rzoska-Smith (2023).

Purification and Biochemical Characterisation of DV1-1

Biochemical characterisation of the full-length DV1-1 protein was done previous to the commencement of this project by my co-supervisor Dr Rzoska-Smith (Rzoska-Smith, 2023). While initial purification of the full-length DV1-1 protein construct was unsuccessful, truncation of the N-terminus allowed for the production of soluble protein (Rzoska-Smith, 2023). This truncation was based on a sequence alignment of the initial DV1-1 construct with homologous proteins, which lack this N-terminal region (Figure 48, A). In the predicted structure of DV1-1, this N-terminal region is an unstructured loop region (Figure 48). The start of the protein may have been annotated wrongly initially, which is supported by the increased solubility of the protein after the removal of the N-terminus.

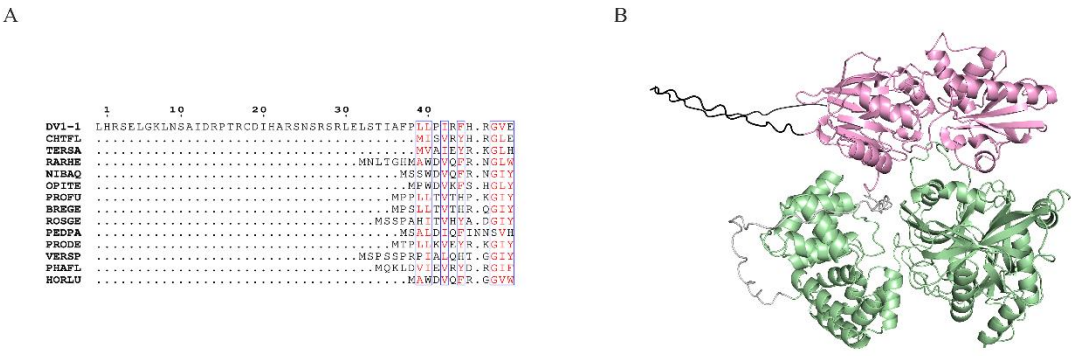


Figure 48: Truncation of the N-terminal region of the DV1-1 protein, which led to soluble expression of the full-length protein.
***A:** Amino acid sequence alignment of DV1-1 with homologous proteins.*
***B:** AlphaFold3 predicted the structure of the DV1-1 full-length protein. The truncated N-terminal region is shown in black. The NucDom is shown in pink and the LigDom is shown in green.*

The protein sequence alignment was generated using ESPript 3.0 (Robert & Gouet, 2014).

The MBP-tagged full-length DV1-1 protein showed both nuclease activity and ligase activity with the same substrates as the NucDom and LigDom alone, but activity was reduced for both activities in the full-length DV1-1 compared to the separate domains (Rzoska-Smith, 2023). The nucleolytic activity of DV1-1 was tested against a range of DNA substrates and the highest activity was observed with the abasic site substrate (Rzoska-Smith, 2023). NucDom had the highest activity at 30°C with almost 60% cleavage efficiency (Figure 49, A). The nucleolytic activity of DV1-1 creates a specific cut at the abasic site of the substrate. The ligase activity of the protein was tested against nicked DNA substrate, which resulted in a low activity of around 13% ligation of the substrate at 50°C after an 8-hour incubation (Figure 49, B) (Rzoska-Smith, 2023).

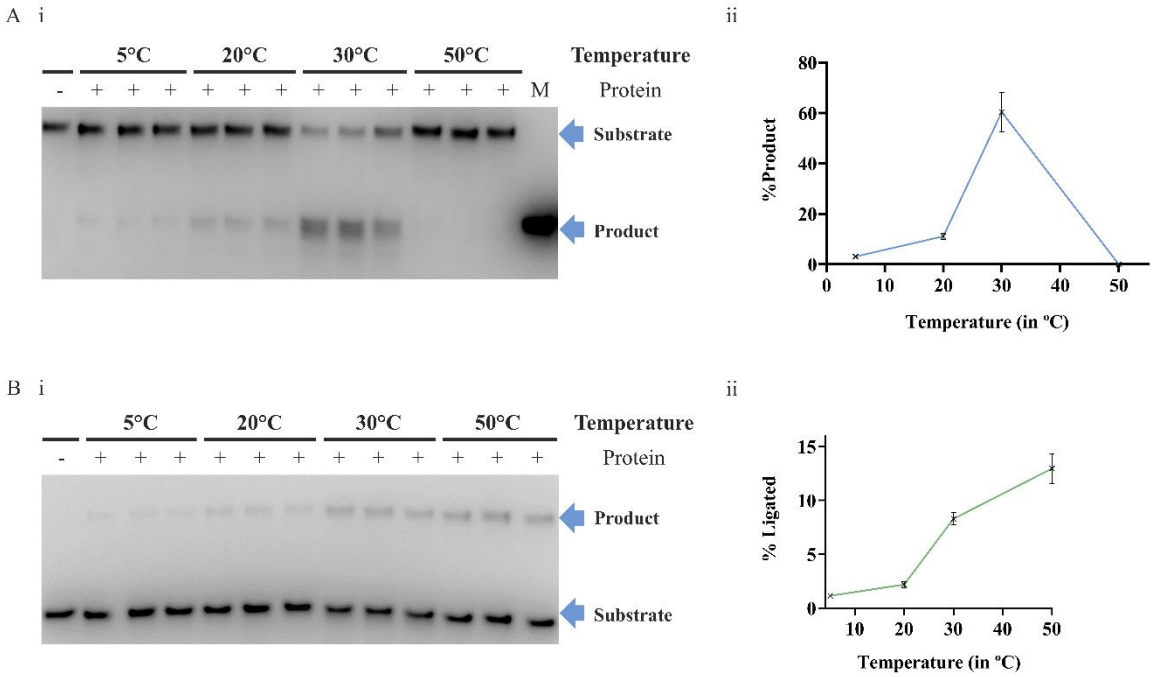


Figure 49: The DV1-1 full-length protein nucleolytic and ligation activity with abasic site substrate and nicked substrate over a temperature gradient.
***A:** Nucleolytic activity of DV1-1 on abasic site DNA substrate. The assay was done with the standard conditions for 8 hours with 10 mM $MgCl_2$. The presence (+) or absence of 1.5 mg/mL DV1-1 protein and incubation temperatures are indicated above the gel. Results were visualised on a Urea-PAGE gel (i) and results were quantified and are shown on the graph with the %Product against the temperature (ii).*
***B:** Ligation activity of DV1-1 on nicked DNA substrate. The assay was done using the standard assay conditions with 1 mM APT and 10 mM $MgCl_2$ over 8 hours. The presence (+) or absence of 1.5 mg/mL DV1-1 protein and incubation temperatures are indicated above the gel. Results were visualised on a Urea-PAGE gel (i) and results were quantified and are shown on the graph with the %Ligated product against the temperature (ii).*
This figure was adapted from Rzoska-Smith (2023).

The DV1-1 NucDom and LigDom were expressed separately to increase the protein purification yield and to characterise the protein domains separately (Figure 50). Based on the AlphaFold structure of the full-length protein, the constructs were designed, and parts of the linker region were removed. The N-

terminal region of NucDom and the C-terminal region of the LigDom were kept as in the original construct (Rzoska-Smith, 2023; Rzoska-Smith; et al., 2023).

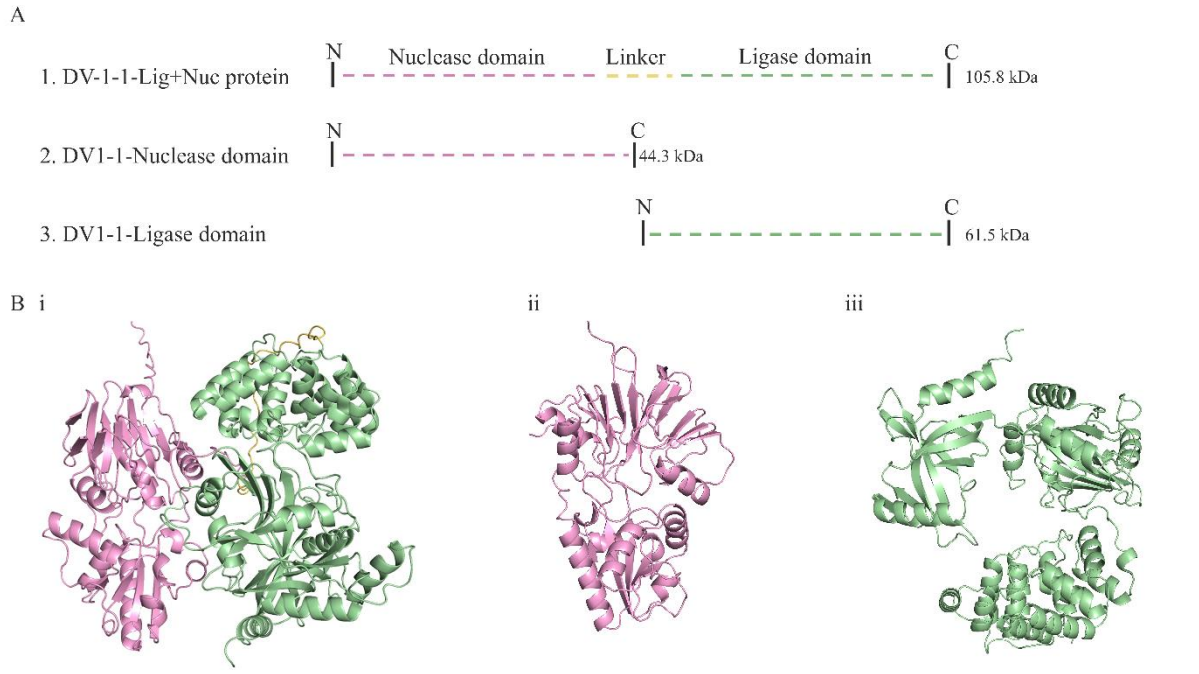


Figure 50: Separation of the DV1-1 into the nuclease domain (NucDom) and ligase domain (LigDom).
***A:** Graphical depiction of the amino acid sequence showing the nuclease domain (pink), linker (yellow), and ligase domain (green) from N- to C-terminus. 1. shows the full-length protein. 2. shows the nuclease domain. 3. shows the ligase domain. The protein size in kDa of each protein is indicated on the right.*
***B:** AlphaFold3 predicted full-length DV1-1 protein (i), nuclease domain in pink (ii) and ligase domain in green (iii) (Abramson et al., 2024).*
This figure was adapted from Rzoska-Smith (2023).

4.1.3 DV1-1 LigDom

DNA ligases are part of the enzyme group of nucleotidyl transferases and they catalyse the reaction that joins two DNA fragments (Pergolizzi et al., 2016). They are essential in all living organisms due to their role in DNA repair, replication and recombination. The ligation of nicks in the DNA backbone is often one of the final steps during DNA repair pathways. Ligation is carried out by a conserved lysine residue, which transfers an adenylated group from a co-factor to the 5'-phosphate end of the DNA. The adenylated end is then joined to the 3'OH of the neighbouring nucleotide (Pergolizzi et al., 2016). The co-factor used for adenylation is either NAD^+ or ATP and they often use a metal co-factor for catalysis.

DNA ligases are useful tools in biotechnological applications, including recombinant DNA technology, ligase chain reaction and DNA sequencing methods (Chambers & Patrick, 2015). The bacteriophage T4 DNA ligase is the most commonly used ligase in biotechnological applications (Chambers & Patrick, 2015). Additionally, due to their essential roles in bacteria, ligases are potential targets for antibiotic treatment (van Eijk et al., 2017). Therefore, ligases are an extensively studied enzyme group.

Characterisation of the DV1-1 LigDom protein was done by Dr Rzoska-Smith previous to the start of this project (Rzoska-Smith, 2023). The DV1-1 LigDom based on its structure and sequence is a LigB-type ligase protein (Rzoska-Smith, 2023; Rzoska-Smith; et al., 2023). The structure of DV1-1 LigDom has not yet been determined. Crystallisation of the protein has been attempted and protein crystals were produced using the hanging-drop diffusion method. However, the resolution of MX2 data collected at the Australian Synchrotron was too low for further processing. Therefore, a predicted structure was used for the structural analysis of DV1-1 LigDom by Rzoska-Smith (2023). The LigDom structure resembles that of a LigB-type ligase protein with a DNA-binding domain at the N-terminus, a central nucleotidyltransferase (NTase) adenylation domain, and a C-terminal oligonucleotide-binding fold (Rzoska-Smith et al., 2023; Williamson et al., 2016).

The DV1-1 LigDom was recombinantly expressed in *E. coli* (DE3) Origami cells and purified as described in Rzoska-Smith et al. (2023). LigDom is monomeric in solution and has a melting temperature of 45°C and a pH optimum between 7 and 8 (Rzoska-Smith, 2023). As demonstrated by Rzoska-Smith (2023) the DV1-1 ligase domain (LigDom) is an ATP/ADP-dependent ligase that adenylates nicked DNA substrates. LigDom requires additional metal co-factors for activity and can use both Mg^{2+} and Mn^{2+}. The protein has the highest ligation activity with both metal co-factors between 25°C and 30°C. LigDom ligates nicked DNA substrates with an 80% turnover from the substrate to the ligated product after an 8-hour incubation, which is the highest ligation efficiency observed with LigDom and all the tested substrates.

For a detailed characterisation of the full-length DV1-1 protein and DV1-1 LigDom see Rzoska-Smith (2023) and Rzoska-Smith; et al. (2023).

4.2 Results – DV1-1 NucDom

4.2.1 DV1-1 NucDom Purification

As described by Rzoska-Smith (2023), purification of the original NucDom construct yielded very low amounts of soluble protein. Therefore, the protein sequence was revisited to increase the protein solubility. Based on amino acid sequence alignments with similar protein sequences, the NucDom sequence was truncated at the N- and C-terminus. The N-terminal region is the same as the predicted wrongly annotated region of the full-length DV1-1 protein and the C-terminal region is part of the long linger between NucDom and LigDom, which might contribute to the insolubility of the protein. In the predicted structure of NucDom, these regions correspond to unstructured regions of the protein and are likely to have no function. Removal of these regions increased soluble protein yield allowing for the

characterisation of the protein, which was established prior to the start of this project through collaboratively work by myself and my co-supervisor, Dr Rzoska-Smith (Rzoska-Smith, 2023).

NucDom was recombinantly expressed in *E. coli* Origami cells as described in Section 2.1.2. The cells were lysed, and the crude lysate was applied to an IMAC column. The protein eluted at approximately 0.25 M imidazole. The presence of NucDom was confirmed via SDS-PAGE electrophoresis (Figure 51) and protein-containing fractions were pooled and changed into TEV-protease suitable buffer. The protein was incubated with TEV protease overnight, which resulted in approximately 80% cleavage efficiency, as shown in (Figure 51, B, ii). The cleaved protein was applied to an IMAC column for reverse IMAC (Figure 51, B). A large amount of NucDom eluted in the flow through with an over-expressed 50 kDa *E. coli* contaminant, which carried through from the IMAC elution. One peak contained clean NucDom protein and the fractions corresponding to this peak were further purified using an S200 gel filtration chromatography column. The protein eluted after approximately 90 mL eluent had passed through the column (Figure 51, C). The protein-containing fractions, as determined by SDS-PAGE electrophoresis, were pooled and up-concentrated (Figure 51, C, ii). The presence of NucDom was further confirmed by LC-MS/MS of the protein band in Figure 51 (C, ii).

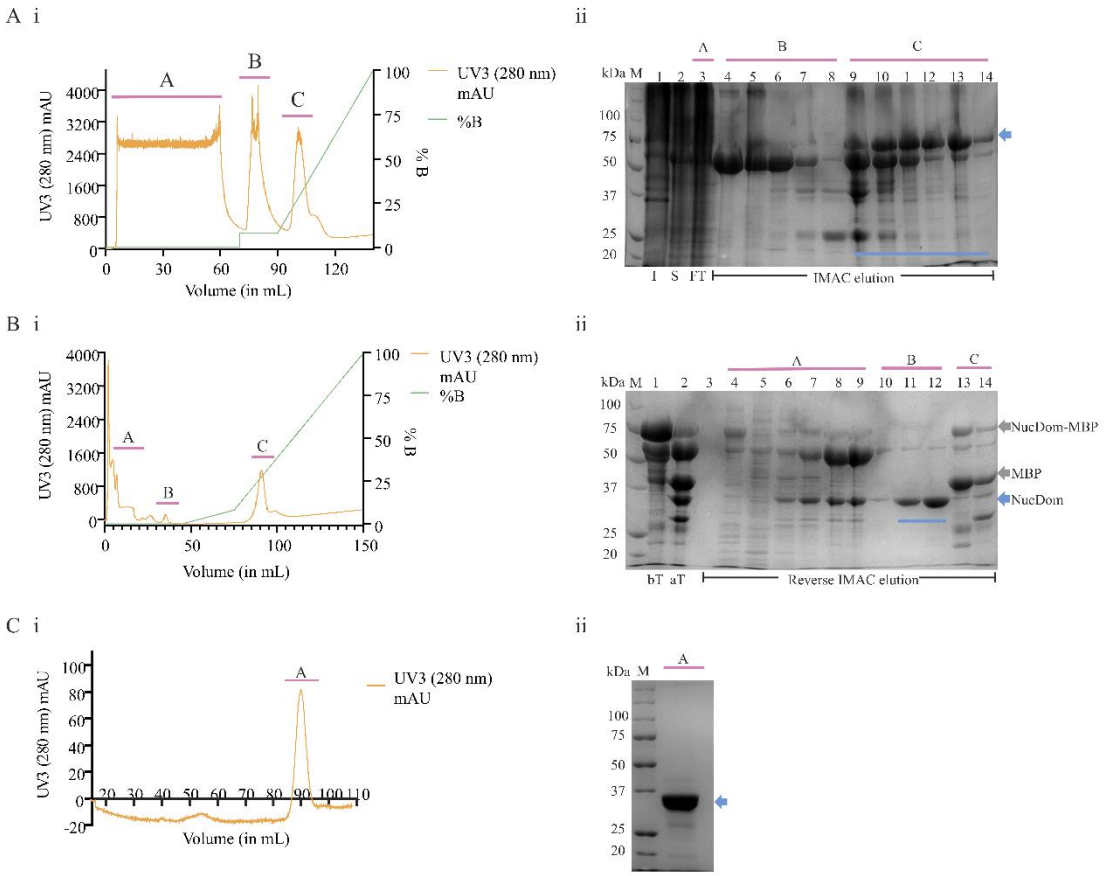


Figure 51: NucDom protein purification chromatograms and SDS-PAGE gels.

All SDS-PAGE gels have the Precision Plus Protein™ standard as a molecular weight indicator (M). Blue arrows indicate the position of NucDom on the gel. Blue lines indicate, which fractions were taken into the next step.
***A:** i: Chromatogram of the IMAC elution of MBP-tagged protein. ii: 12% SDS-PAGE gel, showing the insoluble proteins (I) in the first lane and crude soluble proteins (S) in the second lane. Lane 3 contains the flow through (FT). Lanes 4 to 16 are fractions from the IMAC elution. The pink underlined A, B, and C indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***B:** Chromatogram of the reverse IMAC elution of untagged NucDom. ii: 12% SDS-PAGE gel showing the pooled protein from before and after TEV incubation in lanes 1 and 2, respectively (BT and AT, respectively) (ii). Lanes 3 to 14 are fractions from the reverse IMAC elution. The pink underlined A, B and C indicate the position of the chromatogram peaks on the SDS-PAGE gel.*
***C:** Gel filtration chromatogram. ii:15% SDS-PAGE gel with fractions from the gel filtration elution. The pink underlined A indicates the position of the chromatogram peak on the SDS-PAGE gel.*

4.2.2 DV1-1 NucDom Protein Structure

Crystallisation Attempts

To determine the structure of NucDom crystallisation of the protein was attempted. Initially, crystallisation screens were used to determine conditions that favour the crystal formation of NucDom. NucDom protein at a concentration of 181 µM was used for crystal screens with apo-protein and protein-DNA complexes at a 1:1 ratio of protein to mother liquor from the PEGxR and NATRIX Hampton Research crystallisation screens. NucDom protein with DNA formed crystals in the conditions shown in Table 10 and the most promising crystals are shown in Figure 52. These conditions were used and further adapted by varying the pH and precipitant concentration for crystallisation in hanging drop diffusion; however, these did not result in the formation of any protein crystals. Therefore, the NucDom structure predicted by AlphaFold3 will be used for the structural characterisation of the protein (Abramson et al., 2024).

Table 10: Crystallisation screening conditions of DV1-1 NucDom (181 µM) with DNA.
Conditions are from the Hampton Research Natrix crystallisation screen.

Position	Condition
F5	0.08 M Potassium chloride, 0.02 M Barium chloride dihydrate, 0.04 M Sodium cacodylate trihydrate pH 6.0, 40% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride
F7	0.08 M Sodium chloride, 0.04 M Sodium cacodylate trihydrate pH 6.0, 45% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride
F10	0.08 M Potassium chloride, 0.04 M Sodium cacodylate trihydrate pH 6.0, 55% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride
G10	0.04 M Lithium chloride, 0.08 M Strontium chloride hexahydrate, 0.02 M Magnesium chloride hexahydrate, 0.04 M Sodium cacodylate trihydrate pH 7.0, 30% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride
H5	0.08 M Potassium chloride, 0.04 M Sodium cacodylate trihydrate pH 7.0, 60% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.012 M Spermine tetrahydrochloride

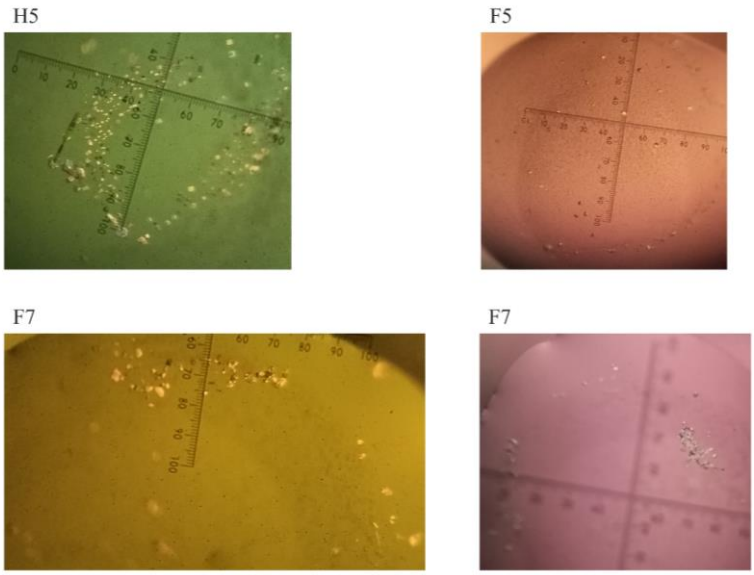


Figure 52: NucDom crystallisation screening results.
The number and letter above the images refer to the Hampton Research Natrix screening conditions shown in Table 10.

NucDom Structure Predictions

Based on an InterPro search the NucDom belongs to the Ribonuclease Z/Hydroxyacylglutathione hydrolase-like protein family (IPR036866) and has a Zn-dependent metallo-hydrolase, RNA specificity domain (IPR011108). NucDom has been identified as an MBL-β-CASP family protein. The MBL (metallo-β-lactamase) fold superfamily, is a large and diverse group of enzymes that share a common αββα fold (González, 2021). They can be found in all domains of life and proteins from this superfamily fulfil a wide range of functions, including multiple phosphoesterases, glyoxalases, quorum-quenching lactonases, organophosphorus hydrolases, and persulfate dioxygenases (González, 2021). The MBL-fold is usually the catalytic domain of the protein with the conserved site for metal binding and Zn, Mn, and Fe are often found in the active sites of MBL-fold proteins (González, 2021). Within the MBL-fold protein superfamily sits the β-CASP protein family, to which the NucDom belongs. The β-CASP (*CPSF*, Artemis, SNM1, PSO2) protein family was first identified in the SNM1 and PSO2 proteins from *Saccharomyces cerevisiae*. Homologues of these have been identified in vertebrates, which are called SNM1A, SNM1B/Apollo, and SNM1C/Artemis (Isabelle Callebaut et al., 2002). β-CASP family proteins are involved in DNA repair and RNA processing (I. Callebaut et al., 2002). The β-CASP domain can be found in proteins from all domains of life, but human β-CASP family proteins – SNM1A, SNM1B/Apollo and SNM1C/Artemis – are the most well-characterised ones, because they are promising cancer therapy targets (Yosaatmadja et al., 2021).

NucDom has the typical MBL and β-CASP domains of β-CASP family proteins as shown in Figure 53. The β-CASP domain consists of a five-stranded β-sheet with parallel β-sheets flanked by seven α-helices surrounding (Figure 53, A). The MBL domain has a typical α-β/β-α sandwich fold with two

mixed β-sheets flanked by three α-helices. (Figure 53). Both the N-terminal (residues 1 to 137) and C-terminal (residues 275 to 325) regions form the MBL-fold domain of the protein and the β-CASP domain is located in between the MBL-fold domain (residues 138 to 274).

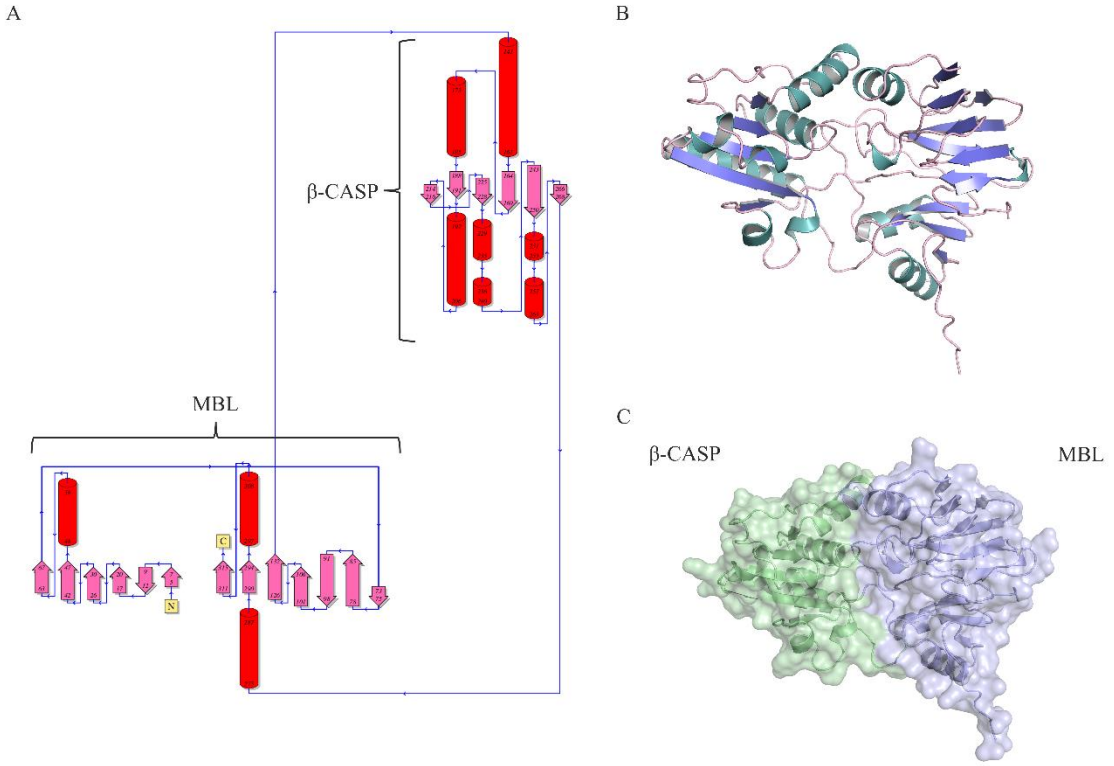


Figure 53: The MBL and β-CASP domain of NucDom.
The AlphaFold3-generated structure was used to generate these images (Abramson et al., 2024).
***A:** Topology of NucDom showing the MBL and β-CASP domain. Topology diagrams were generated in PDBsum (Laskowski, 2022).*
***B:** Cartoon representation of NucDom. Showing α-helices in green, β-sheets in blue and loop regions in pink.*
***C:** Cartoon and surface representations of NucDom highlighting the position of the β-CASP domain in green and the MBP-fold domain in blue.*

Structural superimposition of NucDom with the structures of other β-CASP family proteins shows a moderate level of structural conservation between the proteins (Figure 54). The organisation of the domains is conserved across the proteins and the MBL-fold shows high structural similarity. The β-CASP domain appears less conserved. The sequences of the proteins also show low levels of conservation from the human proteins to the bacterial NucDom with around 20% sequence identity between NucDom and the human β-CASP family proteins shown in Figure 54.

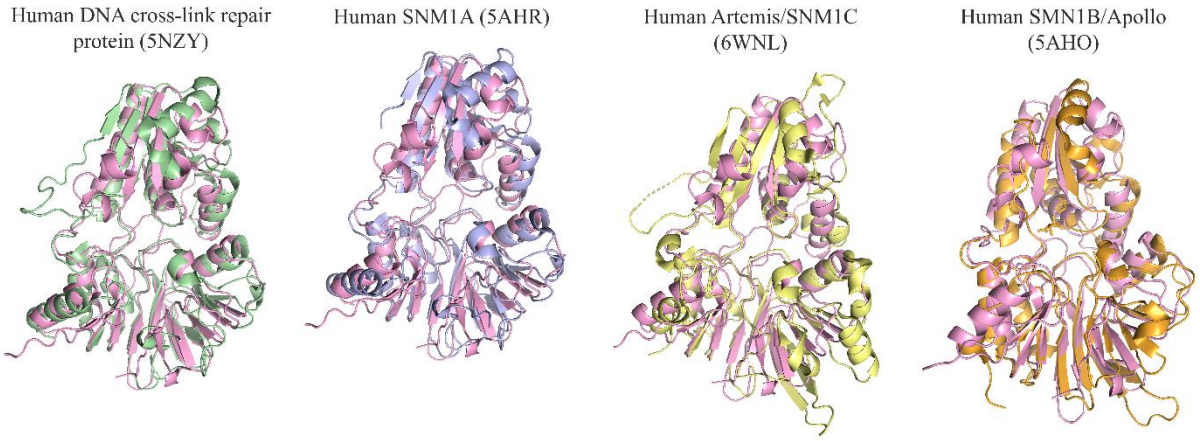


Figure 54: Structure superimposition of the predicted NucDom structure (in pink) with DNA processing β-CASP family proteins.
From left to right: human DNA cross-link repair protein in green (pdb ID: 5NZY) (Newman et al., 2017), human SNM1A in blue (pdb ID: 5AHR) (Allerston et al., 2015), human Artemis/SNM1C in yellow(pdb ID: 6WNL) (Karim et al., 2020), and human SNM1B/Apollo in orange (pdb ID: 5AHO) (Allerston et al., 2015).

DV1-1 NucDom Metal Binding and Active Site

NucDom as discussed above belongs to the MBL-β-CASP family. MBL-fold family proteins share four common motifs (motif 1-4), which form the active site of MBL-fold protein families (motif 1 [Asp], motif 2 [HxHxDH], motif 3 [His], and motif 4 [Asp]). These motifs are also conserved in proteins from the β-CASP family (I. Callebaut et al., 2002). Alignments of NucDom with the Human SNM1C, SNM1A, and SNM1B exonucleases, the pre-mRNA 3'-end processing endonuclease (CPSF3) and the Human tRNase Z (ELAC1) show that these motifs are conserved across nucleases from the β-CASP family (Figure 55, A). The NucDom Motif 2 corresponds to His32, His34, Asp36, and His37, motif 1 corresponds to Asp20, motif 3 to His87 and motif 4 to D108. In the predicted NucDom structure the motifs are located in the same region and show a similar conformation as in other β-CASP family proteins, as shown on superimpositions of NucDom with the CPSF73, SNM1A, and SNM1C proteins. Residues in the active site of these known β-CASP align with the location of these residues in the predicted structure of NucDom as shown in Figure 55 (B). The main function of these conserved residues is the orientation of metal ions in the active site.

Mutations in this motif cause a reduction or abolishment of the nuclease activity in MBL-fold nuclease proteins, which shows that these residues are important for the activity of the protein (Allerston et al., 2015). For example, in the human Artemis/SNM1C protein mutation of the D37A, H33A, and H35D which correspond to Asp36, His32 and His34 in NucDom, cause a complete loss of exonuclease function of the protein (Yosaatmadja et al., 2021).

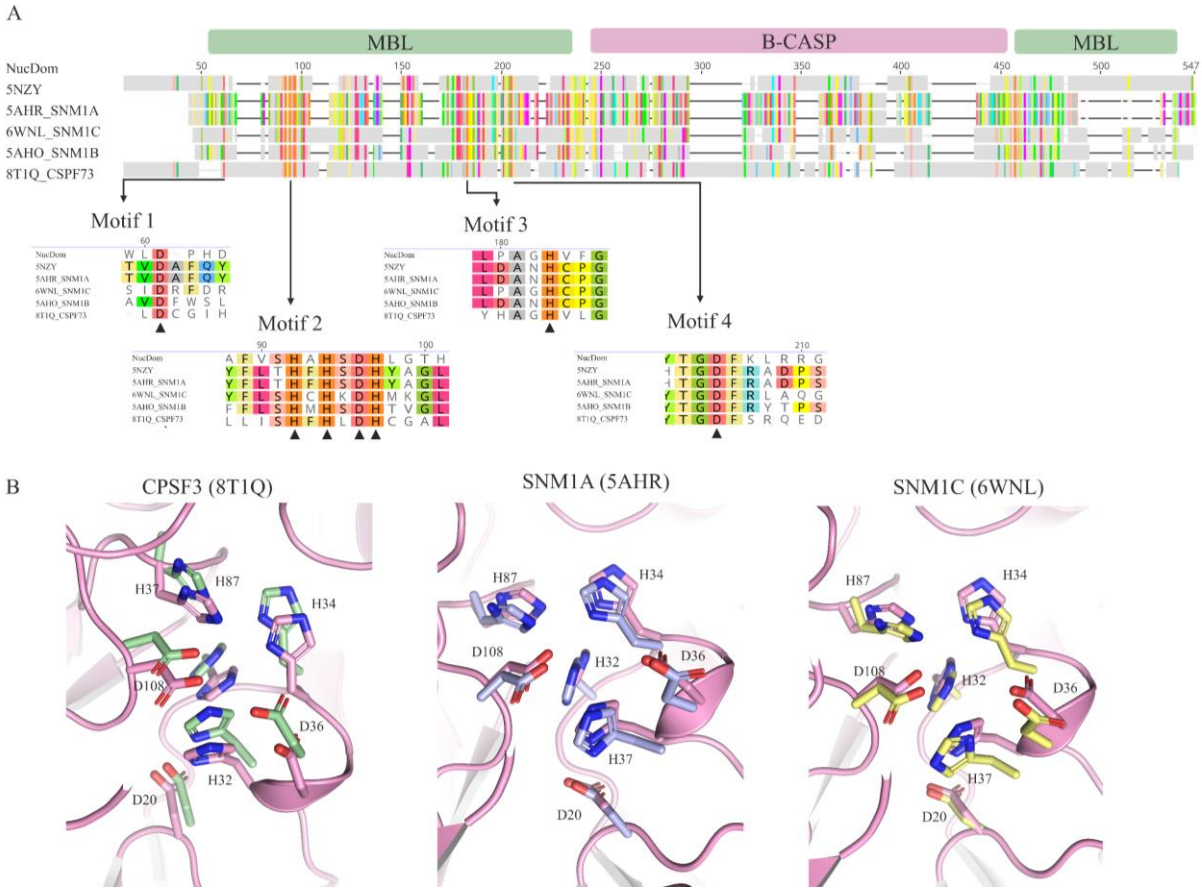


Figure 55: NucDom predicted catalytic site.
***A:** Amino acid sequence alignment of NucDom with β-CASP nucleases (Larkin et al., 2007; Madeira et al., 2024). The MBL-fold specific motifs 1-4 are highlighted and the black triangles indicate to amino acids of each motif. Motif 1 = Asp, motif 2 = HxHxDH, motif 3 = His and motif 4 = Asp.*
***B:** Superimposition of the predicted NucDom active site in pink with the CPSF3 in green (pdb ID: 8T1Q), SNM1A in blue (pdb ID: 5AHR), and SNM1C in yellow (pdb ID: 6WNL) protein active site (Allerston et al., 2015; Karim et al., 2020; Tao et al., 2024). The side chains of residues from motifs 1-4 are shown and labelled.*

In addition to the MBL-fold-specific motifs, β-CASP family proteins have three motifs, that are much less conserved than the MBL motifs 1-4 (Isabelle Callebaut et al., 2002). Those are motif A [Asp/Glu], motif B [His] and motif C [His/Val]. Motifs A to C are located near the active site of the proteins as shown in Figure 56. Sequence and structural alignment of NucDom with β-CASP family proteins show that these motifs are conserved in NucDom. Motif A is Glu131, motif B is His273, and motif C is His295. Regarding these motifs, NucDom is therefore more similar to the CSPF73 protein.

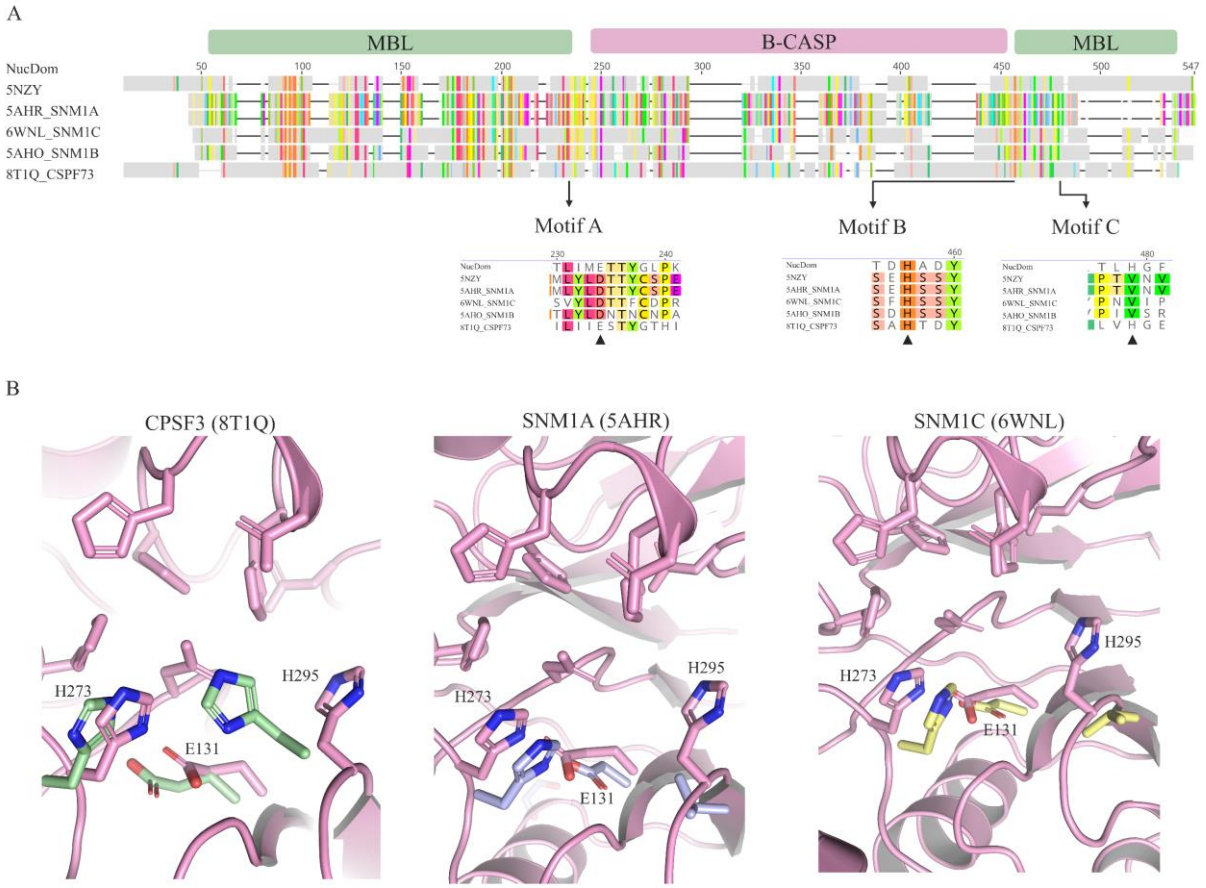


Figure 56: NucDom predicted catalytic site.
***A:** Amino acid sequence alignment of NucDom with β-CASP nucleases (Larkin et al., 2007; Madeira et al., 2024). The MBL-fold specific motifs A-C are highlighted, and the amino acids of each motif are indicated by the black triangles. Motif A = Asp/Glu, motif B = His, and motif C = Val/His.*
***B:** Superimposition of the predicted NucDom active site in pink with the CPSF3 in green (pdb ID: 8T1Q), SNM1A in blue (pdb ID: 5AHR), and SNM1C in yellow (pdb ID: 6WNL) protein active site (Allerston et al., 2015; Karim et al., 2020; Tao et al., 2024). The side chains of residues from motifs A-C are shown and labelled.*

Predicted Metal-Binding Site of NucDom

Activity assays of NucDom showed that Zn^{2+}, Mg^{2+} or Mn^{2+} is required for NucDom activity. To determine, where the metal ion is positioned in the NucDom structure, it was predicted using AlphaFold Server with varying numbers of metal ions used for the prediction (Abramson et al., 2024). The AlphaFold3 prediction places two Mg^{2+} ions in the predicted active site in the MBL domain of NucDom. Based on their predicted position in the active site His32, His34, His38, His87 and Asp108 appear to be involved in the coordination of the metal ion (Figure 57, B). In the homologous structure of DV1-1 metal ions are positioned in the same position in the active site of the proteins. Additionally, water molecules in the active site of β-CASP family proteins have been shown to help with the orientation of metal ions.

Other β-CASP family proteins have one or two metal ions in their active site. SNM1A is the only β-CASP family protein with a single zinc ion in its active site. The SNM1B/Apollo, CSPF73 and Artemis

have a nickel and an iron ion, two zinc ions, and two zinc ions in their active sites, respectively. Based on its structural similarity to these proteins NucDom may be a one or two metal ion proteins. As shown in Figure 57 (B, i), a third metal ion was placed near the interface of the β-CASP and MBL-domain of the protein. However, no residues for metal orientation are present at this site, suggesting that this prediction is likely not correct, which is supported by a pLDDT between 50 and 70 at the position of this metal ion, indicating low confidence in this position.

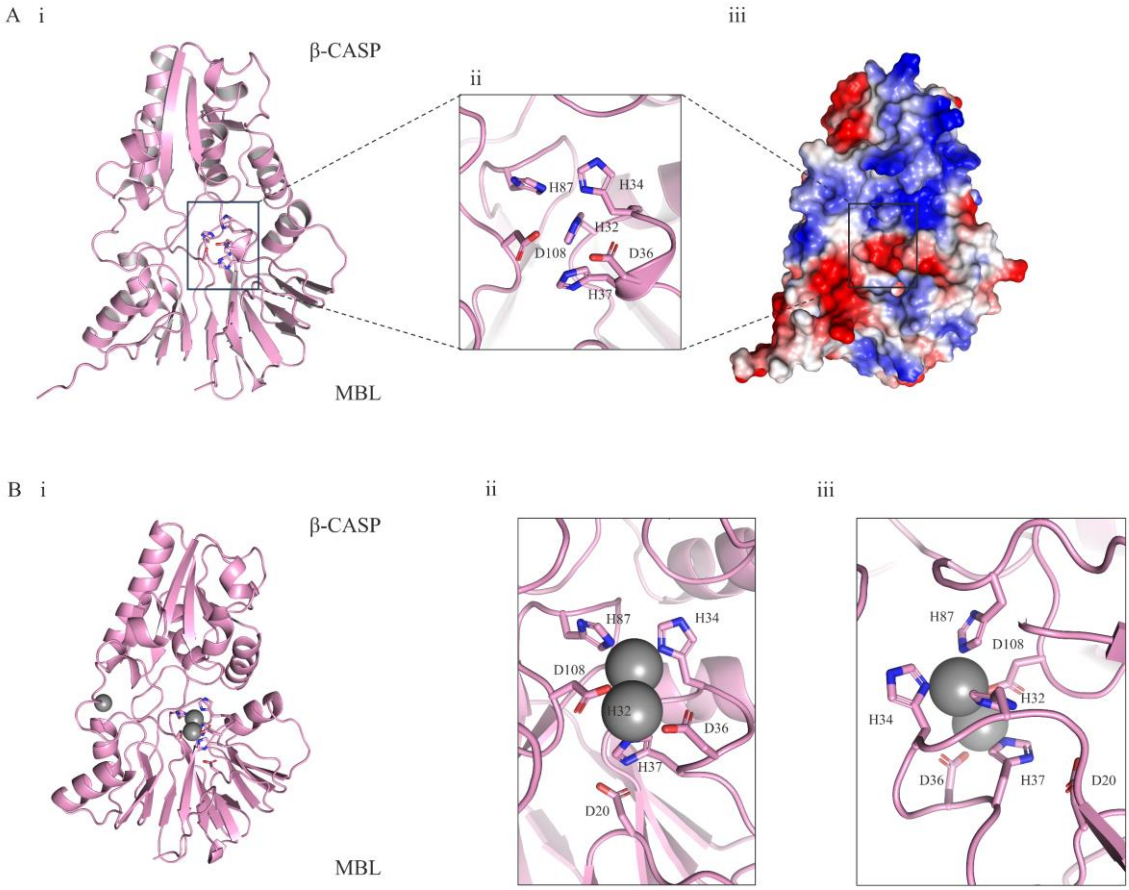


Figure 57: Position of the NucDom active site and orientation of metal ions in the active site.
A: *AlphaFold3 predicted structure of NucDom shown in cartoon representation (i) and the electrostatic surface potential surface showing negatively charged areas in red and positively charged areas in blue (iii)(Abramson et al., 2024). The active site is highlighted in each by a black box, which is shown more closely showing the side chains of the conserved residues from motifs 1-4 (ii).*
B: *AlphaFold3 predicted structure of NucDom with three Mg^{2+} ions bound (i) (Abramson et al., 2024). Two ions bind to the active site of the protein and are located near the conserved residues from motifs 1-4 (ii and iii).*

An active site mutant of NucDom was designed as a negative control for assays. The first mutant was designed prior to the start of this project by Dr Rzoska-Smith, which was a D36A and H37A mutant ($NucDom^{D36A,\ H37A}$) of the protein in the hope of disrupting metal ion binding in the active site of the protein (Rzoska-Smith, 2023). However, while the activity of this mutant was slightly reduced it was not reduced enough for this mutant to be used as a negative control (Rzoska-Smith, 2023). This may be due to the activity of the second metal ion in the active site of NucDom, which is oriented by His87, Asp108, and His34, as predicted by AlphaFold3 (Abramson et al., 2024). Therefore, a second attempt

was made by mutating the His34 and Asp36 to alanine. Mutation of the corresponding residues showed reduced activity in the SNM1C/Artemis, and SNM1B/Apollo (Baddock et al., 2021; Pannicke et al., 2004; Sengerová et al., 2012; van der Burg et al., 2007). Mutation of the same residues in SNM1A caused fully abolished activity with magnesium, however with manganese residual amounts activity was observed (Sengerová et al., 2012). This was hypothesised to be due to tighter binding of manganese to the protein's active site. The NucDom$^{H34A,\ D36A}$ mutant has been successfully cloned into the expression vectors pHMGWA and pDEST17 and has been transformed into the multiple *E. coli* strains for recombinant expression, which will be done in the future to complete the initial biochemical characterisation of this protein.

4.2.3 DV1-1 NucDom Biochemical Characterisation

Thermal Stability

The thermal stability of NucDom was determined via differential scanning fluorimetry. As shown in Figure 58, the thermal melt with 1µL and 2 µL of NucDom showed similar T_m values of 59°C and 60°C.

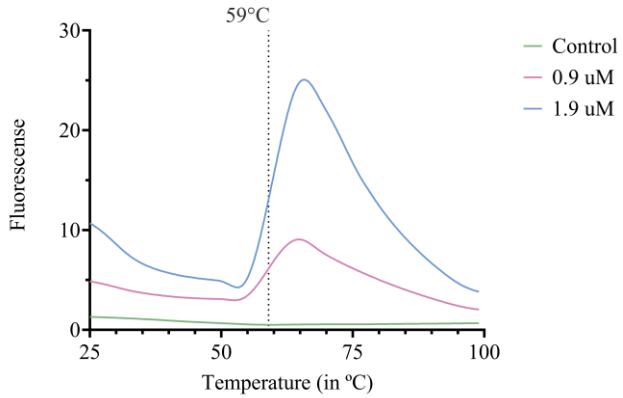


Figure 58: Results of differential scanning fluorimetry using SYPRO orange with NucDom protein.
The differential scanning fluorimetry experiment was carried out with 0.9 µM (pink) and 1.9 µM (blue) of protein and a negative control (green). A graph of the fluorescence over temperatures from 25°C to 100°C is shown. The T_m based on the max d(fluorescence)/d(T) is indicated by the dotted line.

The Activity of NucDom with Synthetic DNA Substrates

The difference between endo- and exonuclease activity can be observed when electrophoresing the DNA substrates on denaturing Urea PAGE gels after processing by the nuclease enzymes (Figure 59). If the fluorescently labelled strand of the substrate is cut by nuclease activity this results in shorter fluorescently labelled products. During electrophoresis, these smaller nuclease-processed labelled strands run below the uncut fluorescently labelled strand of the substrate. In the substrate used here, the damage is placed in the centre of the 40 nt long fluorescently labelled strand, so that cleavage at the

damaged site results in a 20 nt long fluorescently labelled product strand. Therefore, damage-specific endonuclease activity results in a single 20 nt long product band on denaturing Urea-PAGE electrophoresis gels. Exonucleic activity causes gradual degradation of the DNA substrate, which results in fluorescently labelled DNA strands with varying lengths below 39 nt (Figure 59, B, i). When the product is electrophoresed on denaturing Urea-PAGE gels this results in multiple bands below the substrate band on the gel (Figure 59, B, ii).

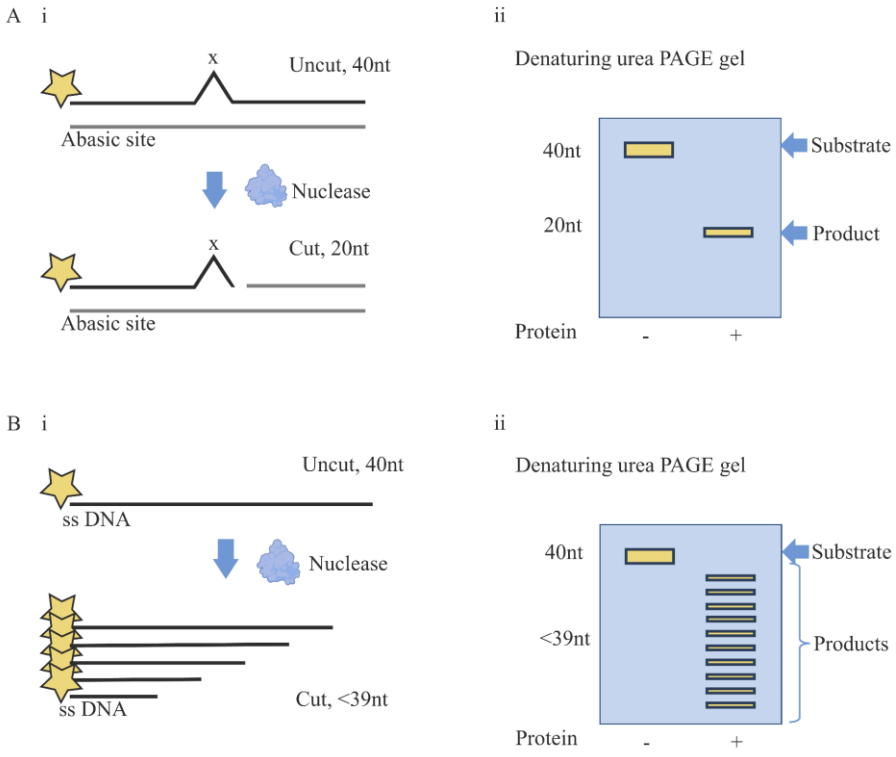


Figure 59: Illustration of a nuclease assay showing endonuclease and exonuclease activity.
A: *Damage-specific endonuclease activity. i: Abasic substrate is cut specifically at the site of the abasic damage (indicated by 'x') by a nuclease. The abasic damage is located on the fluorescently labelled DNA strands (indicated by the yellow star). ii: after denaturation of the product, it will run at 20 nt on a Urea-PAGE gel producing a single band on the gel. The substrate is larger than the product due to nuclease activity and will run higher on the gel. The substrate and product are indicated by the blue arrows on the right. The size of the product and substrate are indicated on the left. The presence (+) or absence (-) of the protein is indicated below the gel.*
B: *Unspecific exonuclease activity shown on a single-stranded DNA substrate. i: The single-stranded DNA substrate is cut into strands of differing lengths by exonuclease activity. This results in DNA fragments of different lengths. These fragments will electrophorese at different sizes on a Urea-PAGE gel, producing a ladder of products on the gel (ii). Substrate and product positions are indicated by the blue arrow and bracket on the right. The size of the product and substrate are indicated on the left. The presence (+) or absence (-) of the protein is indicated below the gel.*
This figure was adapted from Stelzer et al. (2024).

NucDom showed both endo-and exonuclease activity with multiple DNA substrates during preliminary assays. This was demonstrated by assaying the protein with synthetic oligonucleotide substrates as described in Stelzer et al. (2024). Damage-specific endonuclease activity was seen with uracil-mismatch and abasic site substrates. NucDom cleaved those substrates specifically at the damaged sites leading to the production of a 20 nt long product, as shown in Figure 60. Exonuclease activity was seen with single- and double-stranded DNA substrates, including 3' and 5' flapped, and splayed DNA substrates.

NucDom Temperature Dependence

To further characterise NucDom activity, it was tested at temperatures ranging from -40°C to 80°C with an 8-hour incubation period. No protein activity was observed at temperatures above 65°C. This aligns with the thermal stability data of NucDom which showed that the protein has a melting temperature of approximately 60°C. The highest activity of the protein was seen at 35°C with over 98% activity. Activity at below-freezing temperatures was observed but was low with around 3% at both -20 and -40°C.

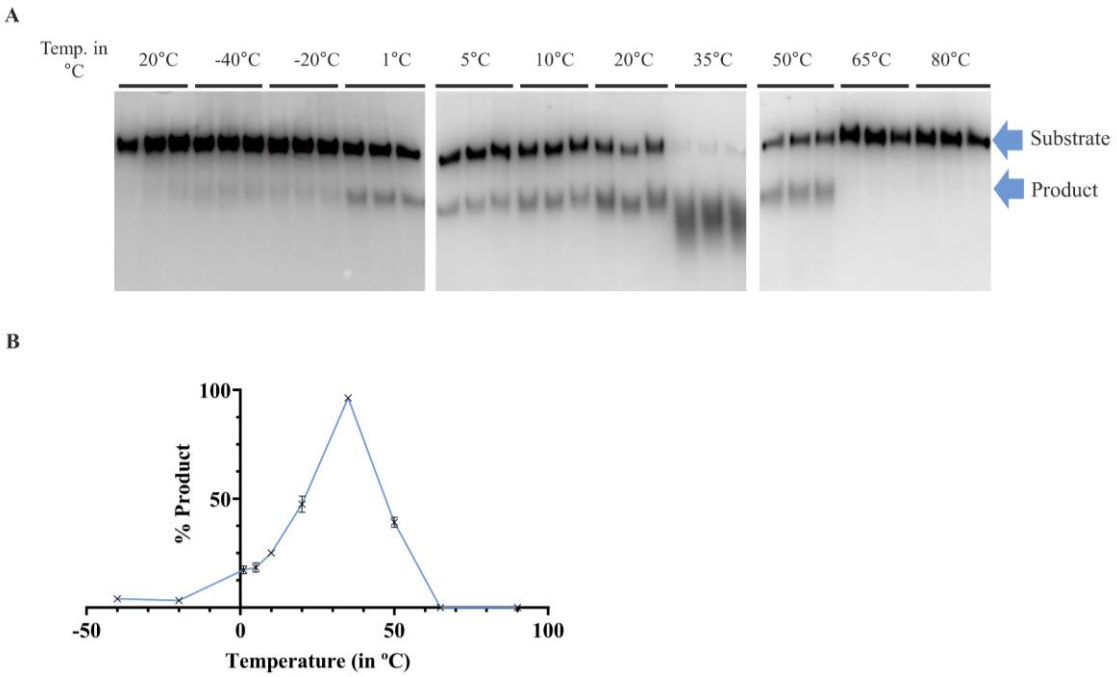


Figure 60: NucDom activity with abasic site substrate over a temperature gradient.
The assay was carried out using the standard assay conditions with protein at 10.423μM and 10 mM $MgCl_2$ and an incubation period of 8 hours. Temperatures from -40°C to 80°C were tested. The reactions were done in triplicates.
***A:** Denaturing Urea-PAGE gel showing the assay results. The incubation temperature (in °C) is indicated above the gel. The position of the substrate and product are indicated on the right. Lanes 1-3 show the results from a no-protein control, which was incubated at 20°C.*
***B:** Graphical summary of the data shown in A. The Y-axis shows the % Product over the temperature range shown on the X-axis. The average %product of the triplicates is shown with the corresponding standard error shown in black.*

NucDom was active at extremely low temperatures. The ability to cleave DNA at below 5°C, even though activity is higher at ambient temperatures, makes the protein a psychrotolerant protein. The optimum temperature of NucDom aligns with that of LigDom which is also at 30°C. However, in the full-length protein ligation activity was highest at 50°C, while nucleolytic activity remained at its highest at 30°C (Figure 49) (Rzoska-Smith, 2023).

Time Dependence of NucDom

The nucleolytic activity of NucDom was further characterised with the abasic site substrate over 16 hours with manganese and magnesium as metal ion co-factors. This showed that the protein initially cuts DNA with an abasic site at the site of the damage, causing a 20nt-long product. This product is then further degraded by the exonuclease activity of NucDom (Figure 61, A). With the undamaged double-stranded DNA substrate, the protein has only exonuclease activity which is shown by the different size products on the Urea-PAGE gel (Figure 61, B). There is no specific cutting with the double-stranded DNA substrate, so the product is present at sizes <40 nt, whereas with the abasic site substrate cutting occurs initially at the abasic site and the substrate is then further degraded into products <20 nt long.

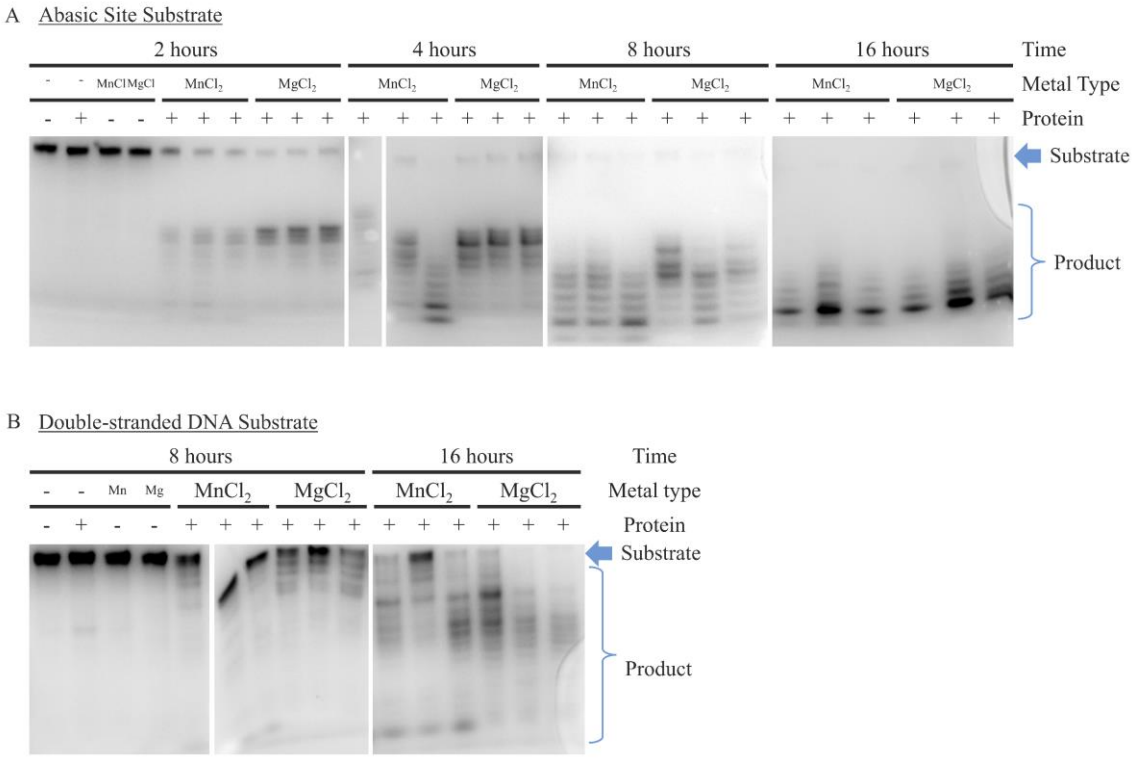


Figure 61: Time series assay of NucDom showing the difference between endonuclease and exonuclease activity of the protein. The assay was carried out at 25°C with 10.423μM and varying incubation times. Manganese or magnesium were used at 10 mM.
***A:** Assay with the abasic site substrate. The incubation times (2, 4, 8, and 16 hours) are indicated above the gel. The absence (-) or presence of 10 mM manganese ($MnCl_2$) or magnesium ($MgCl_2$) and the presence (+) or absence (-) of protein are indicated above the gel.*
***B:** Assay with double-stranded DNA substrate. The incubation times (8 and 16 hours) are indicated above the gel. The absence (-) or presence of 10 mM manganese ($MnCl_2$) or magnesium ($MgCl_2$) and the presence (+) or absence (-) of protein are indicated above the gel.*

Its activity on DNA substrates makes NucDom the first β-CASP nuclease protein from bacteria with DNA degrading activity rather than RNA. DNA processing β-CASP family proteins have only been identified in eukaryotic organisms so far (Isabelle Callebaut et al., 2002). The SNM1A/Artemis protein

is involved in V(D)J recombination and DNA repair like non-homologous end joining. Like NucDom this protein has both exo- and endonuclease activity. In the case of the Artemis protein, endonuclease activity is activated by DNA-PKcs (Goodarzi et al., 2006; Ma et al., 2002). As demonstrated here the endonuclease versus exonuclease activity of NucDom appears to be determined by the substrate and no additional factors are required for endonuclease activity.

Metal Dependence of NucDom

The human β-CASP nuclease proteins SNM1A and SNM1B can use manganese, magnesium, zinc and calcium for nucleolytic activity (Sengerová et al., 2012). Activity with zinc was lower than with manganese, magnesium and calcium, and activity with zinc is highest at low concentrations (0.01 mg/mL), while higher concentrations have an inhibitory effect on those proteins. This is similar to the results with NucDom, where activity with 0.15 mM zinc showed activity above the higher concentrations and at 1 mM zinc inhibited the activity of NucDom (Figure 62).

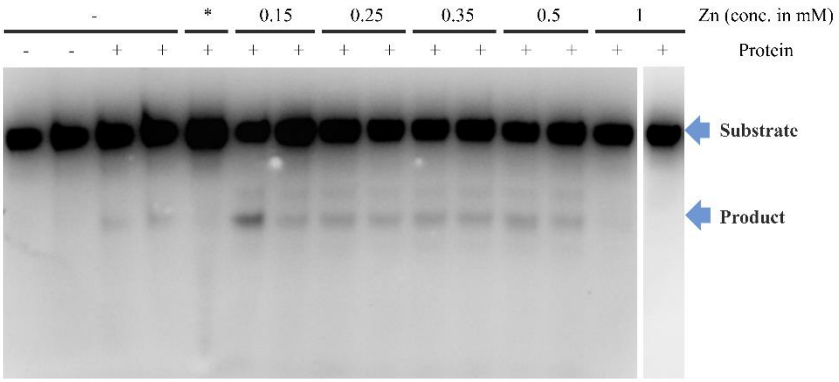


Figure 62: NucDom zinc gradient assay with abasic site substrate.
NucDom at a final concentration of 10.423 µM was tested with abasic site substrate at standard assay conditions. The reactions were incubated at 25°C for 8 hours and varying amounts of $ZnCl_2$ as indicated above the gel. The presence (+) or absence (-) of protein is indicated above the gel. The position of the substrate and product are indicated by the blue arrows. '' indicates where EDTA was added at 10 mM instead of $ZnCl_2$.*

4.2.4 DV1-1 NucDom DNA-Binding

Initial tries to show NucDom binding with DNA substrates were unsuccessful. However, an additional formaldehyde cross-linking step in the electron mobility shift assay successfully showed DNA binding by NucDom (Hoffman et al., 2015). This involves a short incubation of the DNA and substrate after the initial incubation period as described in Section 2.2.4. Binding was tested with double-stranded DNA, the abasic site, uracil mismatch and A/C mismatch DNA substrates. NucDom had the highest binding affinity with the abasic site substrate (Figure 63), this is consistent with the specific activity seen with this substrate. Binding was also observed with the double-stranded, uracil mismatch and A/C mismatch

substrate. To ensure that the formaldehyde cross-linking was not producing a false positive, the assay was tested with BSA (Bovine serum albumin) protein, which did not show binding (Supplementary Figure 4).

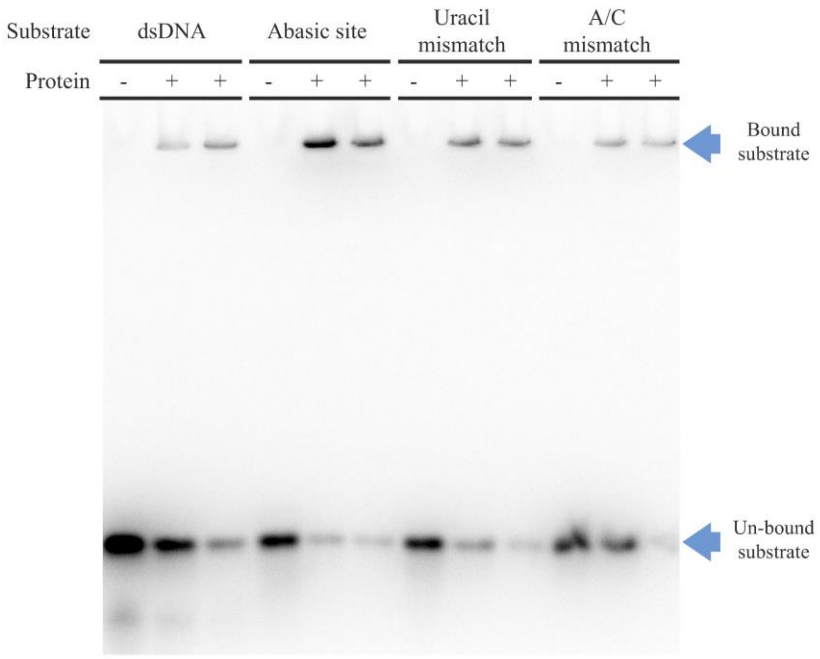


Figure 63: Electron mobility shift assay of NucDom with four DNA substrates.
The reactions were incubated with 12 µM NucDom protein and the substrate at 10°C for 30 min before a 10-minute incubation with 1% (v/v) formaldehyde. The presence (+) or absence (-) of protein and the substrate type are indicated above the gel. The positions of the un-bound substrate and the bound substrate are indicated by the blue arrows.

4.3 Conclusion and Future Directions

The DV1-1 protein is a unique nuclease-ligase fusion protein, which is the first one of its kind to be characterised (Rzoska-Smith, 2023). It is located in a gene cluster with other DNA repair enzymes, that are likely to be involved in the same process. Therefore, it would be desirable to purify and characterise those proteins too, to get insight into this process.

The NucDom protein is a β-CASP family protein with structural similarity to characterised eukaryotic DNA processing β-CASP family proteins. The protein has the conserved MBL-family motifs 1-4 and the β-CASP family protein-specific motifs A-C. In other members of the β-CASP protein family residues from the motifs 1-4 were shown to be required for nucleolytic activity of the protein (Baddock et al., 2021). Bacterial and archaeal β-CASP proteins that have been characterised to date are mostly RNA processing enzymes, for example, the *Thermus thermophilus* TTHA0252 protein, which is thought to be a functional homologue of the bacterial RNase E protein and the *B. subtilis* RNase J (Dominski et al., 2013; Ishikawa et al., 2006; Newman et al., 2011). This makes NucDom the first

bacterial β-CASP protein that is active on DNA. Many β-CASP proteins have both endo- and exonuclease activity, similar to what was observed with NucDom on different DNA substrates (Dominski et al., 2013).

Biochemical characterisation of the DV1-1 full-length protein, LigDom and NucDom is at an advanced state and this project has contributed to the characterisation of NucDom specifically by characterising its binding affinity and nucleolytic activity with DNA substrates, design of an active site mutant, and crystallisation attempts. To complete the biochemical characterisation of NucDom, the active site mutant must be purified and characterised. This mutant has been designed and will be recombinantly expressed and assayed. Furthermore, structural analysis of the DV1-1 protein, LigDom and NucDom would be desirable to understand the interactions between the two protein domains and their interactions with DNA. Crystallisation of the proteins will continue to be attempted and further optimised based on crystallisation screens. However, as the proteins appear difficult to crystallise, alternative methods, like small angle X-ray diffraction (SAX) and CryoEM could be used, which can be carried out with proteins in solution.

5 CONCLUSION

Analysis of metagenomes from the harsh Dry Valley environment revealed many DNA repair-associated genes. As demonstrated here and in (Rzoska-Smith, 2023), the main difficulty when characterising those proteins is protein production. Many of these novel DNA repair enzymes show low expression levels in the vectors and expression systems trialled. Their analysis here, in Rzoska-Smith (2023) and in Rzoska-Smith et al. (2023) has demonstrated that these proteins – as predicted – process DNA. Multiple LigB-type ligases with DNA ligation activity were identified in the Dry Valley metagenomes. A nuclease homolog to NucS nucleases was identified, which showed specific endonuclease activity on several damaged DNA substrates (abasic, uracil match, uracil mismatch and A/C mismatch) (Rzoska-Smith, 2023). To further characterise the DNA repair processes and DNA repair enzymes in Dry Valley organisms, additional enzymes need to be characterised. The characterisation of multiple DNA repair proteins from a single operon, like that of the DV1-1 protein, would be desirable to identify potentially novel DNA repair processes. Attempts of this have been unsuccessful so far due to difficulties with soluble protein expression. Therefore, different protein production strategies may be required, like the use of a different expression system.

The presented work describes two nuclease enzymes found in the Dry Valley metagenomes, an archaeal Hjc homologue (DV-Hjc) and a β-CASP protein family nuclease (NucDom), that is part of a nuclease-ligase fusion protein (DV1-1). While the basics of their functions were identified during this project, both proteins required further characterisation to fully understand their function. Both proteins showed nucleolytic activity with DNA substrates at low temperatures. Proteins with activity at low temperatures are promising biotechnological tools and additional proteins with low-temperature activity could be discovered and characterised in a similar manner as DV-Hjc and NucDom. The optimum temperatures were observed at median temperatures (around 35°C) in DV-Hjc and NucDom, which makes these proteins psychrotolerant. So far no truly psychrophilic DNA repair enzymes have been identified from the Dry Valley metagenomes, which may be linked to the temperature fluctuations in the environment, and the ability of some organisms to grow at both cold and ambient temperatures. Their activity at low temperatures may indicate that some DNA repair processes in Dry Valley organisms are active at cold temperatures, highlighting the importance of DNA repair in this environment.

6 REFERENCES

Abbotts, R., & Wilson, D. M. (2017). Coordination of DNA single strand break repair. *Free Radical Biology and Medicine, 107*, 228-244. https://doi.org/https://doi.org/10.1016/j.freeradbiomed.2016.11.039

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., . . . Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature, 630*(8016), 493-500. https://doi.org/10.1038/s41586-024-07487-w

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D, 66*(2), 213-221. https://doi.org/doi:10.1107/S0907444909052925

Adrio, J. L., & Demain, A. L. (2014). Microbial enzymes: tools for biotechnological processes. *Biomolecules, 4*(1), 117-139. https://doi.org/10.3390/biom4010117

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D, 68*(4), 352-367. https://doi.org/doi:10.1107/S0907444912001308

Aghajari, N., Feller, G., Gerday, C., & Haser, R. (1998). Structures of the psychrophilic Alteromonas haloplanctis α-amylase give insights into cold adaptation at a molecular level. *Structure, 6*(12), 1503-1516. https://doi.org/10.1016/S0969-2126(98)00149-X

Albarracín, V. H., Pathak, G. P., Douki, T., Cadet, J., Borsarelli, C. D., Gärtner, W., & Farias, M. E. (2012). Extremophilic Acinetobacter strains from high-altitude lakes in Argentinean Puna: remarkable UV-B resistance and efficient DNA damage repair. *Orig Life Evol Biosph, 42*(2-3), 201-221. https://doi.org/10.1007/s11084-012-9276-3

Alekseeva, I. V., & Kuznetsov, N. A. (2023). Historical Aspects of Restriction Endonucleases as Intelligent Scissors for Genetic Engineering. *Fermentation, 9*(10), 874. https://www.mdpi.com/2311-5637/9/10/874

Aliyu, H., De Maayer, P., Sjöling, S., & Cowan, D. A. (2017). Metagenomic Analysis of Low-Temperature Environments. In R. Margesin (Ed.), *Psychrophiles: From Biodiversity to Biotechnology* (pp. 389-421). Springer International Publishing. https://doi.org/10.1007/978-3-319-57057-0_16

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., & Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. *The EMBO journal, 16*(14), 4467-4476. https://doi.org/https://doi.org/10.1093/emboj/16.14.4467

Allerston, C. K., Lee, S. Y., Newman, J. A., Schofield, C. J., McHugh, P. J., & Gileadi, O. (2015). The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res, 43*(22), 11047-11060. https://doi.org/10.1093/nar/gkv1256

Almeida, L. C. d., Calil, F. A., Machado-Neto, J. A., & Costa-Lotufo, L. V. (2021). DNA damaging agents and DNA repair: From carcinogenesis to cancer therapy. *Cancer Genetics, 252-253*, 6-24. https://doi.org/https://doi.org/10.1016/j.cancergen.2020.12.002

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, *25*(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., & Yu, Y. K. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS J*, *272*(20), 5101-5109. https://doi.org/10.1111/j.1742-4658.2005.04945.x

Amare, B., Mo, A., Khan, N., Sowa, D. J., Warner, M. M., Tetenych, A., & Andres, S. N. (2021). LigD: A Structural Guide to the Multi-Tool of Bacterial Non-Homologous End Joining [Review]. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.787709

Aragao, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N. P., Ericsson, D. J., Gee, C. L., Macedo, S., Mudie, N., Panjikar, S., Price, J. R., Riboldi-Tunnicliffe, A., Rostan, R., Williamson, R., & Caradoc-Davies, T. T. (2018). MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation*, *25*(3), 885-891. https://doi.org/doi:10.1107/S1600577518003120

Aravind, L., & Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res*, *11*(8), 1365-1374. https://doi.org/10.1101/gr.181001

Ariaeenejad, S., Zolfaghari, B., Sadeghian Motahar, S. F., Kavousi, K., Maleki, M., Roy, S., & Hosseini Salekdeh, G. (2021). Highly Efficient Computationally Derived Novel Metagenome α-Amylase With Robust Stability Under Extreme Denaturing Conditions. *Front Microbiol*, *12*, 713125. https://doi.org/10.3389/fmicb.2021.713125

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from E. coli. *Cell*, *78*(6), 1063-1072.

Au, K. G., Welsh, K., & Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *Journal of Biological Chemistry*, *267*(17), 12142-12148. https://doi.org/10.1016/S0021-9258(19)49816-5

Ayala-del-Río Héctor, L., Chain Patrick, S., Grzymski Joseph, J., Ponder Monica, A., Ivanova, N., Bergholz Peter, W., Di Bartolo, G., Hauser, L., Land, M., Bakermans, C., Rodrigues, D., Klappenbach, J., Zarka, D., Larimer, F., Richardson, P., Murray, A., Thomashow, M., & Tiedje James, M. (2010). The Genome Sequence of Psychrobacter arcticus 273-4, a Psychroactive Siberian Permafrost Bacterium, Reveals Mechanisms for Adaptation to Low-Temperature Growth. *Applied and Environmental Microbiology*, *76*(7), 2304-2312. https://doi.org/10.1128/AEM.02101-09

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, *2*, 2006 0008. https://doi.org/10.1038/msb4100050

Baddock, H. T., Newman, J. A., Yosaatmadja, Y., Bielinski, M., Schofield, C. J., Gileadi, O., & McHugh, P. J. (2021). A phosphate binding pocket is a key determinant of exo- versus endo-nucleolytic activity in the SNM1 nuclease family. *Nucleic Acids Res*, *49*(16), 9294-9309. https://doi.org/10.1093/nar/gkab692

Banaś, A. K., Zgłobicki, P., Kowalska, E., Bażant, A., Dziga, D., & Strzałka, W. (2020). All You Need Is Light. Photorepair of UV-Induced Pyrimidine Dimers. *Genes (Basel)*, *11*(11). https://doi.org/10.3390/genes11111304

Banneville, A.-S., Bouthier de la Tour, C., De Bonis, S., Hognon, C., Colletier, J.-P., Teulon, J.-M., Le Roy, A., Pellequer, J.-L., Monari, A., Dehez, F., Confalonieri, F., Servant, P., & Timmins, J. (2022). Structural and functional characterization of DdrC, a novel DNA damage-induced

nucleoid associated protein involved in DNA compaction. *Nucleic acids research*, *50*(13), 7680-7696. https://doi.org/10.1093/nar/gkac563

Bar Dolev, M., Braslavsky, I., & Davies, P. L. (2016). Ice-Binding Proteins and Their Function. *Annu Rev Biochem*, *85*, 515-542. https://doi.org/10.1146/annurev-biochem-060815-014546

Beese, L. S., & Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo j*, *10*(1), 25-33. https://doi.org/10.1002/j.1460-2075.1991.tb07917.x

Bennett, R. J., Dunderdale, H. J., & West, S. C. (1993). Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell*, *74*(6), 1021-1031. https://doi.org/10.1016/0092-8674(93)90724-5

Bentchikou, E., Servant, P., Coste, G., & Sommer, S. (2010). A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in Deinococcus radiodurans. *PLoS Genet*, *6*(1), e1000774. https://doi.org/10.1371/journal.pgen.1000774

Bertrand, C., Thibessard, A., Bruand, C., Lecointe, F., & Leblond, P. (2019). Bacterial NHEJ: a never ending story. *Mol Microbiol*, *111*(5), 1139-1151. https://doi.org/10.1111/mmi.14218

Bétermier, M., Bertrand, P., & Lopez, B. S. (2014). Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet*, *10*(1), e1004086. https://doi.org/10.1371/journal.pgen.1004086

Bhargava, R., Onyango, D. O., & Stark, J. M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet*, *32*(9), 566-575. https://doi.org/10.1016/j.tig.2016.06.007

Biertümpfel, C., Yang, W., & Suck, D. (2007). Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature*, *449*(7162), 616-620. https://doi.org/10.1038/nature06152

Billi, D., Friedmann, E. I., Hofer, K. G., Caiola, M. G., & Ocampo-Friedmann, R. (2000). Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium Chroococcidiopsis. *Appl Environ Microbiol*, *66*(4), 1489-1492. https://doi.org/10.1128/aem.66.4.1489-1492.2000

Bockheim, J. G. (1997). Properties and Classification of Cold Desert Soils from Antarctica. *Soil Science Society of America Journal*, *61*(1), 224-231. https://doi.org/10.2136/sssaj1997.03615995006100010031x

Bockheim, J. G., Campbell, I. B., & McLeod, M. (2007). Permafrost distribution and active-layer depths in the McMurdo Dry Valleys, Antarctica. *Permafrost and Periglacial Processes*, *18*(3), 217-227. https://doi.org/https://doi.org/10.1002/ppp.588

Bodell, W. J., & Singer, B. (1979). Influence of hydrogen bonding in DNA and polynucleotides on reaction of nitrogens and oxygens toward ethylnitrosourea. *Biochemistry*, *18*(13), 2860-2863. https://doi.org/10.1021/bi00580a029

Bond, C. S., Kvaratskhelia, M., Richard, D., White, M. F., & Hunter, W. N. (2001). Structure of Hjc, a Holliday junction resolvase, from Sulfolobus solfataricus. *Proceedings of the National Academy of Sciences*, *98*(10), 5509-5514. https://doi.org/doi:10.1073/pnas.091613398

Borrego-Soto, G., Ortiz-López, R., & Rojas-Martínez, A. (2015). Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer. *Genet Mol Biol*, *38*(4), 420-432. https://doi.org/10.1590/s1415-475738420150019

Bowater, R., & Doherty, A. J. (2006). Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet*, *2*(2), e8. https://doi.org/10.1371/journal.pgen.0020008

Brettel, K., & Byrdin, M. (2010). Reaction mechanisms of DNA photolyase. *Current Opinion in Structural Biology*, *20*(6), 693-701. https://doi.org/https://doi.org/10.1016/j.sbi.2010.07.003

Breyer, W. A., & Matthews, B. W. (2000). Structure of Escherichia coli exonuclease I suggests how processivity is achieved. *Nature Structural Biology*, *7*(12), 1125-1128. https://doi.org/10.1038/81978

Brooks, S. C., Adhikary, S., Rubinson, E. H., & Eichman, B. F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochim Biophys Acta*, *1834*(1), 247-271. https://doi.org/10.1016/j.bbapap.2012.10.005

Brutlag, D., Atkinson, M. R., Setlow, P., & Kornberg, A. (1969). An active fragment of DNA polymerase produced by proteolytic cleavage. *Biochemical and biophysical research communications*, *37*(6), 982-989.

Byrne, R. T., Klingele, A. J., Cabot, E. L., Schackwitz, W. S., Martin, J. A., Martin, J., Wang, Z., Wood, E. A., Pennacchio, C., Pennacchio, L. A., Perna, N. T., Battista, J. R., & Cox, M. M. (2014). Evolution of extreme resistance to ionizing radiation via genetic adaptation of DNA repair. *Elife*, *3*, e01322. https://doi.org/10.7554/eLife.01322

Cadet, J., Davies, K. J. A., Medeiros, M. H. G., & Wagner, P. J. R. (2017). Formation and repair of oxidatively generated damage in cellular DNA. *Free Radical Biology and Medicine*, *107*, 13-34. https://doi.org/https://doi.org/10.1016/j.freeradbiomed.2016.12.049

Cadet, J., Douki, T., Gasparutto, D., Ravanat, J.-L., & Wagner, J. R. (2012). Oxidatively Generated Nucleobase Modifications in Isolated and Cellular DNA. In *Encyclopedia of Radicals in Chemistry, Biology and Materials*. https://doi.org/https://doi.org/10.1002/9781119953678.rad038

Cadet, J., & Wagner, J. R. (2013). DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb Perspect Biol*, *5*(2). https://doi.org/10.1101/cshperspect.a012559

Callebaut, I., Moshous, D., Mornon, J. P., & de Villartay, J. P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res*, *30*(16), 3592-3601. https://doi.org/10.1093/nar/gkf470

Callebaut, I., Moshous, D., Mornon, J. P., & de Villartay, J. P. (2002). Metallo‐β‐lactamase fold within nucleic acids processing enzymes: the β‐CASP family. *Nucleic acids research*, *30*(16), 3592-3601. https://doi.org/10.1093/nar/gkf470

Cannan, W. J., & Pederson, D. S. (2016). Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *J Cell Physiol*, *231*(1), 3-14. https://doi.org/10.1002/jcp.25048

Carolis, C., Koehler, C., Sauter, C., Basquin, J., Suck, D., & Toeroe, I. (2009). *Crystal structures of Holliday junction resolvases from Archaeoglobus fulgidus bound to DNA substrate*.

Caron, P. R., Kushner, S. R., & Grossman, L. (1985). Involvement of helicase II (uvrD gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex. *Proc Natl Acad Sci U S A*, *82*(15), 4925-4929. https://doi.org/10.1073/pnas.82.15.4925

Cary, S. C., McDonald, I. R., Barrett, J. E., & Cowan, D. A. (2010). On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat Rev Microbiol*, *8*(2), 129-138. https://doi.org/10.1038/nrmicro2281

Casanueva, A., Tuffin, M., Cary, C., & Cowan, D. A. (2010). Molecular adaptations to psychrophily: the impact of 'omic' technologies. *Trends in Microbiology*, *18*(8), 374-381. https://doi.org/https://doi.org/10.1016/j.tim.2010.05.002

Cavicchioli, R. (2016). On the concept of a psychrophile. *The ISME Journal*, *10*(4), 793-795. https://doi.org/10.1038/ismej.2015.160

Ceschini, S., Keeley, A., McAlister, M. S. B., Oram, M., Phelan, J., Pearl, L. H., Tsaneva, I. R., & Barrett, T. E. (2001). Crystal structure of the fission yeast mitochondrial Holliday junction

resolvase Ydc2. *The EMBO journal*, *20*(23), 6601-6611. https://doi.org/https://doi.org/10.1093/emboj/20.23.6601

Chakarov, S., Petkova, R., Russev, G. C., & Zhelev, N. (2014). DNA damage and mutation. Types of DNA damage. *BioDiscovery*, *11*. https://doi.org/10.7750/BioDiscovery.2014.11.1

Chambers, C. R., & Patrick, W. M. (2015). Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology. *Archaea*, *2015*, 170571. https://doi.org/10.1155/2015/170571

Chan, G. L., Doetsch, P. W., & Haseltine, W. A. (1985). Cyclobutane pyrimidine dimers and (6-4) photoproducts block polymerization by DNA polymerase I. *Biochemistry*, *24*(21), 5723-5728. https://doi.org/10.1021/bi00342a006

Chan, Y., Van Nostrand, J. D., Zhou, J., Pointing, S. B., & Farrell, R. L. (2013). Functional ecology of an Antarctic Dry Valley. *Proc Natl Acad Sci U S A*, *110*(22), 8990-8995. https://doi.org/10.1073/pnas.1300643110

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*, *18*(8), 495-506. https://doi.org/10.1038/nrm.2017.48

Chapman, J. R., Taylor, M. R., & Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*, *47*(4), 497-510.

Chen, H.-W., Ruan, B., Yu, M., Wang, J.-d., & Julin, D. A. (1997). The RecD Subunit of the RecBCD Enzyme from Escherichia coli Is a Single-stranded DNA-dependent ATPase *. *Journal of Biological Chemistry*, *272*(15), 10072-10079. https://doi.org/10.1074/jbc.272.15.10072

Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, Stephan J., Webb, C., Wu, D., Varghese, Neha J., Reddy, T. B. K., Mukherjee, S., Ovchinnikova, G., Nolan, M., Seshadri, R., Roux, S., Visel, A., Woyke, T., . . . Ivanova, Natalia N. (2023). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic acids research*, *51*(D1), D723-D732. https://doi.org/10.1093/nar/gkac976

Chou, K. M., & Cheng, Y. C. (2003). The exonuclease activity of human apurinic/apyrimidinic endonuclease (APE1). Biochemical properties and inhibition by the natural dinucleotide Gp4G. *J Biol Chem*, *278*(20), 18289-18296. https://doi.org/10.1074/jbc.M212143200

Collins, A. R. (2009). Oxidative DNA Damage. In M. Schwab (Ed.), *Encyclopedia of Cancer* (pp. 2209-2211). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-47648-1_4306

Collins, T., & Margesin, R. (2019). Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. *Applied Microbiology and Biotechnology*, *103*(7), 2857-2871. https://doi.org/10.1007/s00253-019-09659-5

Collins, T., Meuwis, M.-A., Gerday, C., & Feller, G. (2003). Activity, Stability and Flexibility in Glycosidases Adapted to Extreme Thermal Environments. *Journal of Molecular Biology*, *328*(2), 419-428. https://doi.org/https://doi.org/10.1016/S0022-2836(03)00287-0

Confalonieri, F., & Sommer, S. (2011). Bacterial and archaeal resistance to ionizing radiation. Journal of Physics: Conference Series,

Cooper, D. L., Lahue, R. S., & Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry*, *268*(16), 11823-11829. https://doi.org/https://doi.org/10.1016/S0021-9258(19)50274-5

Cooper, J. P., & Hagerman, P. J. (1989). Geometry of a branched DNA structure in solution. *Proc Natl Acad Sci U S A*, *86*(19), 7336-7340. https://doi.org/10.1073/pnas.86.19.7336

Couvé, S., Ishchenko Alexander, A., Fedorova Olga, S., Ramanculov Erlan, M., Laval, J., & Saparbaev, M. (2013). Direct DNA Lesion Reversal and Excision Repair in Escherichia coli. *EcoSal Plus*, *5*(2), 10.1128/ecosalplus.1127.1122.1124. https://doi.org/10.1128/ecosalplus.7.2.4

Cowan, D. A., Russell, N. J., Mamais, A., & Sheppard, D. M. (2002). Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass. *Extremophiles*, *6*(5), 431-436. https://doi.org/10.1007/s00792-002-0276-5

Cowan, J. (2004). Role of metal ions in promoting DNA binding and cleavage by restriction endonucleases. *Restriction endonucleases*, 339-360.

Cox, M. M., & Battista, J. R. (2005). Deinococcus radiodurans — the consummate survivor. *Nature Reviews Microbiology*, *3*(11), 882-892. https://doi.org/10.1038/nrmicro1264

D'Amico, S., Collins, T., Marx, J. C., Feller, G., Gerday, C., & Gerday, C. (2006). Psychrophilic microorganisms: challenges for life. *EMBO reports*, *7*(4), 385-389-389. https://doi.org/https://doi.org/10.1038/sj.embor.7400662

David, S. S., O'Shea, V. L., & Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature*, *447*(7147), 941-950. https://doi.org/10.1038/nature05978

Davies, K. J. A. (1995). Oxidative stress: the paradox of aerobic life. *Biochemical Society Symposia*, *61*, 1-31. https://doi.org/10.1042/bss0610001

de Jager, T. L., Cockrell, A. E., & Du Plessis, S. S. (2017). Ultraviolet Light Induced Generation of Reactive Oxygen Species. *Adv Exp Med Biol*, *996*, 15-23. https://doi.org/10.1007/978-3-319-56017-5_2

de Vega, M. (2013). The Minimal Bacillus subtilis Nonhomologous End Joining Repair Machinery. *PLoS One*, *8*(5), e64232. https://doi.org/10.1371/journal.pone.0064232

Deming, J. W., & Young, J. N. (2017). The Role of Exopolysaccharides in Microbial Adaptation to Cold Habitats. In R. Margesin (Ed.), *Psychrophiles: From Biodiversity to Biotechnology* (pp. 259-284). Springer International Publishing. https://doi.org/10.1007/978-3-319-57057-0_12

Derbyshire, V., Grindley, N. D., & Joyce, C. M. (1991). The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. *Embo j*, *10*(1), 17-24. https://doi.org/10.1002/j.1460-2075.1991.tb07916.x

Devasagayam, T. P., Steenken, S., Obendorf, M. S., Schulz, W. A., & Sies, H. (1991). Formation of 8-hydroxy(deoxy)guanosine and generation of strand breaks at guanine residues in DNA by singlet oxygen. *Biochemistry*, *30*(25), 6283-6289. https://doi.org/10.1021/bi00239a029

Dextraze, M. E., Gantchev, T., Girouard, S., & Hunting, D. (2010). DNA interstrand cross-links induced by ionizing radiation: an unsung lesion. *Mutat Res*, *704*(1-3), 101-107. https://doi.org/10.1016/j.mrrev.2009.12.007

Dhawale, A., Bindal, G., Rath, D., & Rath, A. (2021). DNA repair pathways important for the survival of Escherichia coli to hydrogen peroxide mediated killing. *Gene*, *768*, 145297. https://doi.org/10.1016/j.gene.2020.145297

Dhawale, A., & Rath, A. (2019). *Nucleotide excision repair is important for survival of hydrogen peroxide mediated killing in *Escherichia coli**. https://www.researchpublish.com/papers/nucleotide-excision-repair-is-important-for-survival-of-hydrogen-peroxide-mediated-killing-in-escherichia-coli

Di Felice, F., Micheli, G., & Camilloni, G. (2019). Restriction enzymes and their use in molecular biology: An overview. *Journal of biosciences*, *44*(2), 38.

Dianov, G., & Lindahl, T. (1994). Reconstitution of the DNA base excision-repair pathway. *Curr Biol*, *4*(12), 1069-1076. https://doi.org/10.1016/s0960-9822(00)00245-1

Dillingham, M. S., Spies, M., & Kowalczykowski, S. C. (2003). RecBCD enzyme is a bipolar DNA helicase. *Nature*, *423*(6942), 893-897. https://doi.org/10.1038/nature01673

Dinglay, S., Trewick, S. C., Lindahl, T., & Sedgwick, B. (2000). Defective processing of methylated single-stranded DNA by E. coli AlkB mutants. *Genes Dev*, *14*(16), 2097-2105.

Doherty, A. J., Jackson, S. P., & Weller, G. R. (2001). Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett*, *500*(3), 186-188. https://doi.org/10.1016/s0014-5793(01)02589-3

Dominski, Z., Carpousis, A. J., & Clouet-d'Orval, B. (2013). Emergence of the β-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim Biophys Acta*, *1829*(6-7), 532-551. https://doi.org/10.1016/j.bbagrm.2013.01.010

Dronkert, M. L. G., & Kanaar, R. (2001). Repair of DNA interstrand cross-links. *Mutation Research/DNA Repair*, *486*(4), 217-247. https://doi.org/https://doi.org/10.1016/S0921-8777(01)00092-1

Dupureur, C. M. (2008a). An integrated look at metallonuclease mechanism. *Current Chemical Biology*, *2*(2), 159-173.

Dupureur, C. M. (2008b). Roles of metal ions in nucleases. *Current Opinion in Chemical Biology*, *12*(2), 250-255. https://doi.org/https://doi.org/10.1016/j.cbpa.2008.01.012

Ejaz, A., & Shuman, S. (2018). Characterization of Lhr-Core DNA helicase and manganese- dependent DNA nuclease components of a bacterial gene cluster encoding nucleic acid repair enzymes. *J Biol Chem*, *293*(45), 17491-17504. https://doi.org/10.1074/jbc.RA118.005296

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D*, *66*(4), 486-501. https://doi.org/doi:10.1107/S0907444910007493

Enderle, J., Dorn, A., & Puchta, H. (2019). DNA- and DNA-Protein-Crosslink Repair in Plants. *International Journal of Molecular Sciences*, *20*(17).

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, *357*(2), 289-298. https://doi.org/https://doi.org/10.1016/j.ab.2006.07.027

Fasnacht, M., & Polacek, N. (2021). Oxidative Stress in Bacteria and the Central Dogma of Molecular Biology. *Front Mol Biosci*, *8*, 671037. https://doi.org/10.3389/fmolb.2021.671037

Feller, G. (2013). Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica (Cairo)*, *2013*, 512840. https://doi.org/10.1155/2013/512840

Ferrer, M., Beloqui, A., Timmis, K. N., & Golyshin, P. N. (2009). Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol*, *16*(1-2), 109-123. https://doi.org/10.1159/000142898

Fijalkowska, I. J., & Schaaper, R. M. (1996). Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. *Proceedings of the National Academy of Sciences*, *93*(7), 2856-2861. https://doi.org/doi:10.1073/pnas.93.7.2856

Fijalkowska, I. J., Schaaper, R. M., & Jonczyk, P. (2012). DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. *FEMS Microbiology Reviews*, *36*(6), 1105-1121. https://doi.org/10.1111/j.1574-6976.2012.00338.x

Fountain, A. G., Nylen, T. H., Monaghan, A., Basagic, H. J., & Bromwich, D. (2010). Snow in the McMurdo Dry Valleys, Antarctica. *International Journal of Climatology*, *30*(5), 633-642. https://doi.org/https://doi.org/10.1002/joc.1933

Fukui, K. (2010). DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids*, *2010*. https://doi.org/10.4061/2010/260512

Fuxreiter, M., Luo, N., Jedlovszky, P., Simon, I., & Osman, R. (2002). Role of Base Flipping in Specific Recognition of Damaged DNA by Repair Enzymes. *Journal of Molecular Biology*, *323*(5), 823-834. https://doi.org/https://doi.org/10.1016/S0022-2836(02)00999-3

Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C., & Feller, G. (2003). Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J Biol Chem*, *278*(39), 37015-37023. https://doi.org/10.1074/jbc.M305142200

Georlette, D., Jónsson, Z. O., Van Petegem, F., Chessa, J. P., Van Beeumen, J., Hübscher, U., & Gerday, C. (2000). A DNA ligase from the psychrophile Pseudoalteromonas haloplanktis gives insights into the adaptation of proteins to low temperatures. *European Journal of Biochemistry*, *267*(12), 3502-3512. https://doi.org/https://doi.org/10.1046/j.1432-1327.2000.01377.x

Gibrat, J. F., Madej, T., & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Curr Opin Struct Biol*, *6*(3), 377-385. https://doi.org/10.1016/s0959-440x(96)80058-3

Gilbert, J. A., Hill, P. J., Dodd, C. E. R., & Laybourn-Parry, J. (2004). Demonstration of antifreeze protein activity in Antarctic lake bacteria. *Microbiology*, *150*(1), 171-180. https://doi.org/https://doi.org/10.1099/mic.0.26610-0

Goedken, E. R., & Marqusee, S. (2001). Co-crystal of Escherichia coli RNase HI with Mn2+ ions reveals two divalent metals bound in the active site. *J Biol Chem*, *276*(10), 7266-7271. https://doi.org/10.1074/jbc.M009626200

González, J. M. (2021). Visualizing the superfamily of metallo-β-lactamases through sequence similarity network neighborhood connectivity analysis. *Heliyon*, *7*(1), e05867. https://doi.org/10.1016/j.heliyon.2020.e05867

Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., Härer, C., Marchetti, C., Morrice, N., Jeggo, P. A., & Lees-Miller, S. P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *Embo j*, *25*(16), 3880-3889. https://doi.org/10.1038/sj.emboj.7601255

Goordial, J., Davila, A., Greer, C. W., Cannam, R., DiRuggiero, J., McKay, C. P., & Whyte, L. G. (2017). Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper-arid polar desert. *Environ Microbiol*, *19*(2), 443-458. https://doi.org/10.1111/1462-2920.13353

Goordial, J., Raymond-Bouchard, I., Zolotarov, Y., de Bethencourt, L., Ronholm, J., Shapiro, N., Woyke, T., Stromvik, M., Greer, C. W., Bakermans, C., & Whyte, L. (2015). Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile Rhodococcus sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. *FEMS Microbiology Ecology*, *92*(2). https://doi.org/10.1093/femsec/fiv154

Goosen, N., & Moolenaar, G. F. (2008). Repair of UV damage in bacteria. *DNA Repair (Amst)*, *7*(3), 353-379. https://doi.org/10.1016/j.dnarep.2007.09.002

Gorecka, K. M., Komorowska, W., & Nowotny, M. (2013). Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic acids research*, *41*(21), 9945-9955.

Górecka, K. M., Krepl, M., Szlachcic, A., Poznański, J., Šponer, J., & Nowotny, M. (2019). RuvC uses dynamic probing of the Holliday junction to achieve sequence specificity and efficient resolution. *Nature Communications*, *10*(1), 4102. https://doi.org/10.1038/s41467-019-11900-8

Grilley, M., Griffith, J., & Modrich, P. (1993). Bidirectional excision in methyl-directed mismatch repair. *Journal of Biological Chemistry*, *268*(16), 11830-11837. https://doi.org/https://doi.org/10.1016/S0021-9258(19)50275-7

Grilley, M., Welsh, K. M., Su, S. S., & Modrich, P. (1989). Isolation and Characterization of the Escherichia coli mutL Gene Product. *Journal of Biological Chemistry*, *264*(2), 1000-1004. https://doi.org/https://doi.org/10.1016/S0021-9258(19)85043-3

Grollman, A. P., & Moriya, M. (1993). Mutagenesis by 8-oxoguanine: an enemy within. *Trends in Genetics*, *9*(7), 246-249. https://doi.org/https://doi.org/10.1016/0168-9525(93)90089-Z

Groth, P., Ausländer, S., Majumder, M. M., Schultz, N., Johansson, F., Petermann, E., & Helleday, T. (2010). Methylated DNA Causes a Physical Block to Replication Forks Independently of Damage Signalling, O6-Methylguanine or DNA Single-Strand Breaks and Results in DNA Damage. *Journal of Molecular Biology*, *402*(1), 70-82. https://doi.org/https://doi.org/10.1016/j.jmb.2010.07.010

Gutierrez, R., & O'Connor, T. R. (2021). DNA direct reversal repair and alkylating agent drug resistance. *Cancer Drug Resist*, *4*(2), 414-423. https://doi.org/10.20517/cdr.2020.113

Hadden, J. M., Déclais, A.-C., Carr, S. B., Lilley, D. M. J., & Phillips, S. E. V. (2007). The structural basis of Holliday junction resolution by T7 endonuclease I. *Nature*, *449*(7162), 621-624. https://doi.org/10.1038/nature06158

Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L., & Dixon, N. E. (2002). Structural basis for proofreading during replication of the Escherichia coli chromosome. *Structure*, *10*(4), 535-546. https://doi.org/10.1016/s0969-2126(02)00738-4

Hamilton, H. L., & Dillard, J. P. (2006). Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Molecular Microbiology*, *59*(2), 376-385. https://doi.org/https://doi.org/10.1111/j.1365-2958.2005.04964.x

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev*, *68*(4), 669-685. https://doi.org/10.1128/mmbr.68.4.669-685.2004

Hardie, M. E., & Murray, V. (2018). The sequence preference of DNA cleavage by T4 endonuclease VII. *Biochimie*, *146*, 1-13. https://doi.org/https://doi.org/10.1016/j.biochi.2017.11.002

Hardy, L., Plantade, J., Morales, V., Mazzamuro, F., Rocha, E. P. C., Polard, P., & Charpentier, X. (2024). YraN is a helicase-associated nuclease fostering extended recombination events by natural transformation. *Preprint*. https://doi.org/10.1101/2024.02.06.579203

Harris, D. R., Tanaka, M., Saveliev, S. V., Jolivet, E., Earl, A. M., Cox, M. M., & Battista, J. R. (2004). Preserving Genome Integrity: The DdrA Protein of Deinococcus radiodurans R1. *PLOS Biology*, *2*(10), e304. https://doi.org/10.1371/journal.pbio.0020304

Hassan, F. M. N., & Gupta, R. S. (2018). Novel Sequence Features of DNA Repair Genes/Proteins from Deinococcus Species Implicated in Protection from Oxidatively Generated Damage. *Genes*, *9*(3), 149. https://www.mdpi.com/2073-4425/9/3/149

Heldmann, J. L., Pollard, W., McKay, C. P., Marinova, M. M., Davila, A., Williams, K. E., Lacelle, D., & Andersen, D. T. (2013). The high elevation Dry Valleys in Antarctica as analog sites for subsurface ice on Mars. *Planetary and Space Science*, *85*, 53-58. https://doi.org/https://doi.org/10.1016/j.pss.2013.05.019

Hill-Perkins, M., Jones, M. D., & Karran, P. (1986). Site-specific mutagenesis in vivo by single methylated or deaminated purine bases. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *162*(2), 153-163. https://doi.org/https://doi.org/10.1016/0027-5107(86)90081-3

Hoffman, E. A., Frey, B. L., Smith, L. M., & Auble, D. T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem*, *290*(44), 26404-26411. https://doi.org/10.1074/jbc.R115.651679

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research*, *5*(2), 282-304. https://doi.org/10.1017/S0016672300001233

Horowitz, N. H., Cameron, R. E., & Hubbard, J. S. (1972). Microbiology of the Dry Valleys of Antarctica. *Science*, *176*(4032), 242-245. https://doi.org/doi:10.1126/science.176.4032.242

Horton, N. C., Newberry, K. J., & Perona, J. J. (1998). Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc Natl Acad Sci U S A*, *95*(23), 13489-13494. https://doi.org/10.1073/pnas.95.23.13489

Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P., & Tainer, J. A. (1999). Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, *98*(3), 397-408. https://doi.org/10.1016/s0092-8674(00)81968-6

Hsia, K.-C., Li, C.-L., & Yuan, H. S. (2005). Structural and functional insight into sugar-nonspecific nucleases in host defense. *Current Opinion in Structural Biology*, *15*(1), 126-134. https://doi.org/https://doi.org/10.1016/j.sbi.2005.01.015

Husain, I., Van Houten, B., Thomas, D. C., Abdel-Monem, M., & Sancar, A. (1985). Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proceedings of the National Academy of Sciences*, *82*(20), 6774-6778. https://doi.org/doi:10.1073/pnas.82.20.6774

Huston, A. L., Methe, B., & Deming, J. W. (2004). Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile Colwellia psychrerythraea strain 34H. *Applied and Environmental Microbiology*, *70*(6), 3321-3328.

Ilmberger, N., Meske, D., Juergensen, J., Schulte, M., Barthen, P., Rabausch, U., Angelov, A., Mientus, M., Liebl, W., Schmitz, R. A., & Streit, W. R. (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids—linking thermostability and halotolerance. *Applied Microbiology and Biotechnology*, *95*(1), 135-146. https://doi.org/10.1007/s00253-011-3732-2

Imhof, M., & Schlötterer, C. (2001). Fitness effects of advantageous mutations in evolving Escherichia coli populations. *Proceedings of the National Academy of Sciences*, *98*(3), 1113-1117. https://doi.org/doi:10.1073/pnas.98.3.1113

Isaksen, G. V., Åqvist, J., & Brandsdal, B. O. (2014). Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PLoS Comput Biol*, *10*(8), e1003813. https://doi.org/10.1371/journal.pcbi.1003813

Isaksen, G. V., Åqvist, J., & Brandsdal, B. O. (2016). Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proc Natl Acad Sci U S A*, *113*(28), 7822-7827. https://doi.org/10.1073/pnas.1605237113

Ishikawa, H., Nakagawa, N., Kuramitsu, S., & Masui, R. (2006). Crystal Structure of TTHA0252 from Thermus thermophilus HB8, a RNA Degradation Protein of the Metallo-β-lactamase Superfamily. *The Journal of Biochemistry*, *140*(4), 535-542. https://doi.org/10.1093/jb/mvj183

Ivanov, I., Tainer, J. A., & McCammon, J. A. (2007). Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proceedings of the National Academy of Sciences*, *104*(5), 1465-1470. https://doi.org/10.1073/pnas.0603468104

Iyer, R. R., Pluciennik, A., Burdett, V., & Modrich, P. L. (2006). DNA Mismatch Repair: Functions and Mechanisms. *Chemical reviews*, *106*(2), 302-323. https://doi.org/10.1021/cr0404794

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu031

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J.,

Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D*, *66*(2), 125-132. https://doi.org/doi:10.1107/S0907444909047337

Kaina, B. (2003). DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling. *Biochemical Pharmacology*, *66*(8), 1547-1554. https://doi.org/https://doi.org/10.1016/S0006-2952(03)00510-0

Karim, M. F., Liu, S., Laciak, A. R., Volk, L., Koszelak-Rosenblum, M., Lieber, M. R., Wu, M., Curtis, R., Huang, N. N., Carr, G., & Zhu, G. (2020). Structural analysis of the catalytic domain of Artemis endonuclease/SNM1C reveals distinct structural features. *Journal of Biological Chemistry*, *295*(35), 12368-12377. https://doi.org/10.1074/jbc.RA120.014136

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1992). Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution. *Journal of Molecular Biology*, *223*(4), 1029-1052.

Kciuk, M., Marciniak, B., Mojzych, M., & Kontek, R. (2020). Focus on UV-Induced DNA Damage and Repair-Disease Relevance and Protective Strategies. *Int J Mol Sci*, *21*(19). https://doi.org/10.3390/ijms21197264

Kibota, T. T., & Lynch, M. (1996). Estimate of the genomic mutation rate deleterious to overall fitness in E. coll. *Nature*, *381*(6584), 694-696. https://doi.org/10.1038/381694a0

Kim, S. H. (2018). Chapter Nine - TIRF-Based Single-Molecule Detection of the RecA Presynaptic Filament Dynamics. In M. Spies & A. Malkova (Eds.), *Methods in Enzymology* (Vol. 600, pp. 233-253). Academic Press. https://doi.org/https://doi.org/10.1016/bs.mie.2017.11.012

Klenow, H., & Henningsen, I. (1970). Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. *Proceedings of the National Academy of Sciences*, *65*(1), 168-175.

Koh, H. Y., Park, H., Lee, J. H., Han, S. J., Sohn, Y. C., & Lee, S. G. (2017). Proteomic and transcriptomic investigations on cold-responsive properties of the psychrophilic Antarctic bacterium sychrobacter sp. PAMC 21119 at subzero temperatures. *Environmental Microbiology*, *19*(2), 628-644. https://doi.org/https://doi.org/10.1111/1462-2920.13578

Komori, K., Sakae, S., Daiyasu, H., Toh, H., Morikawa, K., Shinagawa, H., & Ishino, Y. (2000). Mutational analysis of the Pyrococcus furiosus holliday junction resolvase hjc revealed functionally important residues for dimer formation, junction DNA binding, and cleavage activities. *J Biol Chem*, *275*(51), 40385-40391. https://doi.org/10.1074/jbc.M006294200

Kountz, D. J., & Balskus, E. P. (2021). Leveraging Microbial Genomes and Genomic Context for Chemical Discovery. *Acc Chem Res*, *54*(13), 2788-2797. https://doi.org/10.1021/acs.accounts.1c00100

Kow, Y. W. (2002). Repair of deaminated bases in DNA12 1Guest Editor: Miral Dizdaroglu 2This article is part of a series of reviews on "Oxidative DNA Damage and Repair." The full list of papers may be found on the homepage of the journal. *Free Radical Biology and Medicine*, *33*(7), 886-893. https://doi.org/https://doi.org/10.1016/S0891-5849(02)00902-4

Kowalczykowski, S. C. (2015). An overview of the molecular mechanisms of recombinational DNA repair. *Cold Spring Harbor Perspectives in Biology*, *7*(11), a016410.

Krokan, H. E., & Bjørås, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol*, *5*(4), a012583. https://doi.org/10.1101/cshperspect.a012583

Krokan, H. E., Standal, R., & Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochem J*, *325 (Pt 1)*(Pt 1), 1-16. https://doi.org/10.1042/bj3250001

Kunkel, T. A., & Erie, D. A. (2005). DNA mismatch repair. *Annu Rev Biochem*, *74*, 681-710. https://doi.org/10.1146/annurev.biochem.74.082803.133243

Kurth, D., Belfiore, C., Gorriti, M. F., Cortez, N., Farias, M. E., & Albarracín, V. H. (2015). Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile Acinetobacter sp. Ver3. *Front Microbiol*, *6*, 328. https://doi.org/10.3389/fmicb.2015.00328

Kvaratskhelia, M., Wardleworth, B. N., Norman, D. G., & White, M. F. (2000). A Conserved Nuclease Domain in the Archaeal Holliday Junction Resolving Enzyme Hjc *. *Journal of Biological Chemistry*, *275*(33), 25540-25546. https://doi.org/10.1074/jbc.M003420200

Kvaratskhelia, M., & White, M. F. (2000). Two Holliday junction resolving enzymes in Sulfolobus solfataricus. *J Mol Biol*, *297*(4), 923-932. https://doi.org/10.1006/jmbi.2000.3624

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

Laskowski, R. A. (2022). PDBsum: A standalone program for generating PDBsum analyses. *Protein Science*, *31*(12), e4473. https://doi.org/https://doi.org/10.1002/pro.4473

Le, P. T., Makhalanyane, T. P., Guerrero, L. D., Vikram, S., Van de Peer, Y., & Cowan, D. A. (2016). Comparative Metagenomic Analysis Reveals Mechanisms for Stress Response in Hypoliths from Extreme Hyperarid Deserts. *Genome Biology and Evolution*, *8*(9), 2737-2747. https://doi.org/10.1093/gbe/evw189

Lee, C. K., Laughlin, D. C., Bottos, E. M., Caruso, T., Joy, K., Barrett, J. E., Brabyn, L., Nielsen, U. N., Adams, B. J., & Wall, D. H. (2019). Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. *Communications biology*, *2*(1), 1-10.

Lee, K. C., Caruso, T., Archer, S. D. J., Gillman, L. N., Lau, M. C. Y., Cary, S. C., Lee, C. K., & Pointing, S. B. (2018). Stochastic and Deterministic Effects of a Moisture Gradient on Soil Microbial Communities in the McMurdo Dry Valleys of Antarctica [Original Research]. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02619

Lehman, I. (2003). Discovery of DNA polymerase. *Journal of Biological Chemistry*, *278*(37), 34733-34738.

Lehtinen, D. A., & Perrino, F. W. (2004). Dysfunctional proofreading in the Escherichia coli DNA polymerase III core. *Biochem J*, *384*(Pt 2), 337-348. https://doi.org/10.1042/bj20040660

Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research*, *18*(1), 85-98. https://doi.org/10.1038/cr.2007.115

Li, M., & Wilson, D. M., 3rd. (2014). Human apurinic/apyrimidinic endonuclease 1. *Antioxid Redox Signal*, *20*(4), 678-707. https://doi.org/10.1089/ars.2013.5492

Li, X., & Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research*, *18*(1), 99-113. https://doi.org/10.1038/cr.2008.1

Lieber, M. R. (1998). Pathological and physiological double-strand breaks: roles in cancer, aging, and the immune system. *The American journal of pathology*, *153*(5), 1323-1332.

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, *79*, 181-211. https://doi.org/10.1146/annurev.biochem.052308.093131

Lilley, D. M. J. (2017). Holliday junction-resolving enzymes-structures and mechanisms. *FEBS Lett*, *591*(8), 1073-1082. https://doi.org/10.1002/1873-3468.12529

Lilley, D. M. J., & Clegg, R. M. (1993). The structure of branched DNA species. *Quarterly Reviews of Biophysics*, *26*(2), 131-175. https://doi.org/10.1017/S0033583500004054

Lin, Y.-L., Elias, Y., & Huang, R. H. (2005). Structural and Mutational Studies of the Catalytic Domain of Colicin E5: A tRNA-Specific Ribonuclease. *Biochemistry*, *44*(31), 10494-10500. https://doi.org/10.1021/bi050749s

Lindahl, T. (1974). An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues. *Proceedings of the National Academy of Sciences*, *71*(9), 3649-3653. https://doi.org/doi:10.1073/pnas.71.9.3649

Liu, Y., Freeman, Alasdair D. J., Déclais, A.-C., Wilson, Timothy J., Gartner, A., & Lilley, David M. J. (2015). Crystal Structure of a Eukaryotic GEN1 Resolving Enzyme Bound to DNA. *Cell Reports*, *13*(11), 2565-2575. https://doi.org/https://doi.org/10.1016/j.celrep.2015.11.042

Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research*, *42*(1), 3-19.

Lovett, S. T. (2004). Exonucleases, Bacterial. In W. J. Lennarz & M. D. Lane (Eds.), *Encyclopedia of Biological Chemistry* (pp. 66-72). Elsevier. https://doi.org/https://doi.org/10.1016/B0-12-443710-9/00217-9

Lu, M., Guo, Q., Seeman, N. C., & Kallenbach, N. R. (1991). Parallel and antiparallel Holliday junctions differ in structure and stability. *J Mol Biol*, *221*(4), 1419-1432. https://doi.org/10.1016/0022-2836(91)90942-y

Ma, Y., Pannicke, U., Schwarz, K., & Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*, *108*(6), 781-794. https://doi.org/10.1016/s0092-8674(02)00671-2

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic acids research*, *52*(W1), W521-W525. https://doi.org/10.1093/nar/gkae241

Matos, J., & West, S. C. (2014). Holliday junction resolution: regulation in space and time. *DNA Repair (Amst)*, *19*(100), 176-181. https://doi.org/10.1016/j.dnarep.2014.03.013

McCallum, N., Berger-Bächi, B., & Senn, M. M. (2010). Regulation of antibiotic resistance in Staphylococcus aureus. *International Journal of Medical Microbiology*, *300*(2), 118-129. https://doi.org/https://doi.org/10.1016/j.ijmm.2009.08.015

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658-674. https://doi.org/doi:10.1107/S0021889807021206

McGregor, N., Ayora, S., Sedelnikova, S., Carrasco, B., Alonso, J. C., Thaw, P., & Rafferty, J. (2005). The Structure of Bacillus subtilis RecU Holliday Junction Resolvase and Its Role in Substrate Selection and Sequence-Specific Cleavage. *Structure*, *13*(9), 1341-1351. https://doi.org/https://doi.org/10.1016/j.str.2005.05.011

Meikrantz, W., Bergom, M. A., Memisoglu, A., & Samson, L. (1998). O6-alkylguanine DNA lesions trigger apoptosis. *Carcinogenesis*, *19*(2), 369-372. https://doi.org/10.1093/carcin/19.2.369

Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., & Doxey, A. C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic acids research*, *47*(9), 4442-4448. https://doi.org/10.1093/nar/gkz246

Methé, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., Moult, J., Madupu, R., Nelson, W. C., Dodson, R. J., Brinkac, L. M., Daugherty, S. C., Durkin, A. S., DeBoy, R. T., Kolonay, J. F., Sullivan, S. A., Zhou, L., Davidsen, T. M., Wu, M., . . . Fraser, C. M. (2005). The psychrophilic lifestyle as revealed by the genome sequence of Colwellia psychrerythraea

34H through genomic and proteomic analyses. *Proc Natl Acad Sci U S A*, *102*(31), 10913-10918. https://doi.org/10.1073/pnas.0504766102

Metpally, R. P., & Reddy, B. V. (2009). Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC genomics*, *10*, 11. https://doi.org/10.1186/1471-2164-10-11

Michetti, D., Brandsdal, B. O., Bon, D., Isaksen, G. V., Tiberti, M., & Papaleo, E. (2017). A comparative study of cold- and warm-adapted Endonucleases A using sequence analyses and molecular dynamics simulations. *PLoS One*, *12*(2), e0169586. https://doi.org/10.1371/journal.pone.0169586

Middleton, C. L., Parker, J. L., Richard, D. J., White, M. F., & Bond, C. S. (2004). Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic acids research*, *32*(18), 5442-5451. https://doi.org/10.1093/nar/gkh869

Mielecki, D., Wrzesiński, M., & Grzesiuk, E. (2015). Inducible repair of alkylated DNA in microorganisms. *Mutation Research/Reviews in Mutation Research*, *763*, 294-305. https://doi.org/https://doi.org/10.1016/j.mrrev.2014.12.001

Modrich, P. (2016). Mechanisms in E. coli and Human Mismatch Repair (Nobel Lecture). *Angew Chem Int Ed Engl*, *55*(30), 8490-8501. https://doi.org/10.1002/anie.201601412

Mol, C. D., Hosfield, D. J., & Tainer, J. A. (2000). Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3′ ends justify the means. *Mutation Research/DNA Repair*, *460*(3-4), 211-229.

Mol, C. D., Kuo, C. F., Thayer, M. M., Cunningham, R. P., & Tainer, J. A. (1995). Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature*, *374*(6520), 381-386. https://doi.org/10.1038/374381a0

Munoz, P. A., Marquez, S. L., Gonzalez-Nilo, F. D., Marquez-Miranda, V., & Blamey, J. M. (2017). Structure and application of antifreeze proteins from Antarctic bacteria. *Microb Cell Fact*, *16*(1), 138. https://doi.org/10.1186/s12934-017-0737-2

Murchie, A. I. H., Clegg, R. M., Krtzing, E. v., Duckett, D. R., Diekmann, S., & Lilley, D. M. J. (1989). Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. *Nature*, *341*(6244), 763-766. https://doi.org/10.1038/341763a0

Musilova, M., Wright, G., Ward, J. M., & Dartnell, L. R. (2015). Isolation of Radiation-Resistant Bacteria from Mars Analog Antarctic Dry Valleys by Preselection, and the Correlation between Radiation and Desiccation Resistance. *Astrobiology*, *15*(12), 1076-1090. https://doi.org/10.1089/ast.2014.1278

Mykytczuk, N. C., Foote, S. J., Omelon, C. R., Southam, G., Greer, C. W., & Whyte, L. G. (2013). Bacterial growth at -15 °C; molecular insights from the permafrost bacterium Planococcus halocryophilus Or1. *ISME J*, *7*(6), 1211-1226. https://doi.org/10.1038/ismej.2013.8

Nagata, Y., Mashimo, K., Kawata, M., & Yamamoto, K. (2002). The Roles of Klenow Processing and Flap Processing Activities of DNA Polymerase I in Chromosome Instability in Escherichia coli K12 Strains. *Genetics*, *160*(1), 13-23. https://doi.org/10.1093/genetics/160.1.13

Nakabeppu, Y., Kajitani, K., Sakamoto, K., Yamaguchi, H., & Tsuchimoto, D. (2006). MTH1, an oxidized purine nucleoside triphosphatase, prevents the cytotoxicity and neurotoxicity of oxidized purine nucleotides. *DNA Repair*, *5*(7), 761-772. https://doi.org/https://doi.org/10.1016/j.dnarep.2006.03.003

Nam, N. N., Do, H. D. K., Loan Trinh, K. T., & Lee, N. Y. (2023). Metagenomics: An Effective Approach for Exploring Microbial Diversity and Functions. *Foods*, *12*(11). https://doi.org/10.3390/foods12112140

Neeley, W. L., & Essigmann, J. M. (2006). Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem Res Toxicol*, *19*(4), 491-505. https://doi.org/10.1021/tx0600043

Newman, J. A., Aitkenhead, H., Kupinska, K., Burgess-Brown, N. A., Talon, R., Krojer, T., von Delft, F., Arrowsmith, C. H., Edwards, A., Bountra, C., & Gileadi, O. (2017). *Crystal structure of DNA cross-link repair protein 1A in complex with Cefotaxime*.

Newman, J. A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C. R., & Lewis, R. J. (2011). Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure*, *19*(9), 1241-1251. https://doi.org/10.1016/j.str.2011.06.017

Niederberger, T. D., Bottos, E. M., Sohm, J. A., Gunderson, T., Parker, A., Coyne, K. J., Capone, D. G., Carpenter, E. J., & Cary, S. C. (2019). Rapid Microbial Dynamics in Response to an Induced Wetting Event in Antarctic Dry Valley Soils. *Front Microbiol*, *10*, 621. https://doi.org/10.3389/fmicb.2019.00621

Niederberger, T. D., McDonald, I. R., Hacker, A. L., Soo, R. M., Barrett, J. E., Wall, D. H., & Cary, S. C. (2008). Microbial community composition in soils of Northern Victoria Land, Antarctica. *Environ Microbiol*, *10*(7), 1713-1724. https://doi.org/10.1111/j.1462-2920.2008.01593.x

Nieminuszczy, J., & Grzesiuk, E. (2007). Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA. *Acta Biochim Pol*, *54*(3), 459-468.

Nishino, T., Komori, K., Ishino, Y., & Morikawa, K. (2001). Dissection of the regional roles of the archaeal Holliday junction resolvase Hjc by structural and mutational analyses. *Journal of Biological Chemistry*, *276*(38), 35735-35740.

Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., & Morikawa, K. (2001). Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure*, *9*(3), 197-204.

Nishino, T., & Morikawa, K. (2002). Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene*, *21*(58), 9022-9032. https://doi.org/10.1038/sj.onc.1206135

Noll, D. M., Mason, T. M., & Miller, P. S. (2006). Formation and Repair of Interstrand Cross-Links in DNA. *Chemical reviews*, *106*(2), 277-301. https://doi.org/10.1021/cr040478b

Norais, C. A., Chitteni-Pattu, S., Wood, E. A., Inman, R. B., & Cox, M. M. (2009). DdrB protein, an alternative Deinococcus radiodurans SSB induced by ionizing radiation. *J Biol Chem*, *284*(32), 21402-21411. https://doi.org/10.1074/jbc.M109.010454

Obryk, M. K., Doran, P. T., Fountain, A. G., Myers, M., & McKay, C. P. (2020). Climate From the McMurdo Dry Valleys, Antarctica, 1986–2017: Surface Air Temperature Trends and Redefined Summer Season. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2019JD032180. https://doi.org/https://doi.org/10.1029/2019JD032180

Okazaki, R., Arisawa, M., & Sugino, A. (1971). Slow Joining of Newly Replicated DNA Chains in DNA Polymerase I-Deficient Escherichia coli Mutants[*]. *Proceedings of the National Academy of Sciences*, *68*(12), 2954-2957. https://doi.org/doi:10.1073/pnas.68.12.2954

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985). Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature*, *313*(6005), 762-766. https://doi.org/10.1038/313762a0

Orren, D. K., & Sancar, A. (1990). Formation and enzymatic properties of the UvrB.DNA complex. *J Biol Chem*, *265*(26), 15796-15803.

Osipiuk, J., Skarina, T., Kagan, O., Savchenko, A., Edwards, A. M., & Joachimiak, A. (2009). *X-ray crystal structure of protein RPA0323 of unknown function from Rhodopseudomonas palustris*. Midwest Center for Structural Genomics (MCSG).

Ott, E., Kawaguchi, Y., Kölbl, D., Rabbow, E., Rettberg, P., Mora, M., Moissl-Eichinger, C., Weckwerth, W., Yamagishi, A., & Milojevic, T. (2020). Molecular repertoire of Deinococcus radiodurans after 1 year of exposure outside the International Space Station within the Tanpopo mission. *Microbiome*, *8*(1), 150. https://doi.org/10.1186/s40168-020-00927-5

Öz, R., Wang, J. L., Guerois, R., Goyal, G., Kk, S., Ropars, V., Sharma, R., Koca, F., Charbonnier, J.-B., Modesti, M., Strick, T. R., & Westerlund, F. (2021). Dynamics of Ku and bacterial non-homologous end-joining characterized using single DNA molecule analysis. *Nucleic acids research*, *49*(5), 2629-2641. https://doi.org/10.1093/nar/gkab083

Pannicke, U., Ma, Y., Hopfner, K. P., Niewolik, D., Lieber, M. R., & Schwarz, K. (2004). Functional and biochemical dissection of the structure-specific nuclease ARTEMIS. *Embo j*, *23*(9), 1987-1997. https://doi.org/10.1038/sj.emboj.7600206

Pavlov, Y. I., Shcherbakova, P. V., & Rogozin, I. B. (2006). Roles of DNA Polymerases in Replication, Repair, and Recombination in Eukaryotes. In K. W. Jeon (Ed.), *International Review of Cytology* (Vol. 255, pp. 41-132). Academic Press. https://doi.org/https://doi.org/10.1016/S0074-7696(06)55002-8

Pearl, L. H. (2000). Structure and function in the uracil-DNA glycosylase superfamily. *Mutation Research/DNA Repair*, *460*(3), 165-181. https://doi.org/https://doi.org/10.1016/S0921-8777(00)00025-2

Pergolizzi, G., Wagner, G. K., & Bowater, R. P. (2016). Biochemical and Structural Characterisation of DNA Ligases from Bacteria and Archaea. *Biosci Rep*, *36*(5), 00391. https://doi.org/10.1042/bsr20160003

Perona, J. J., & Martin, A. M. (1997). Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis. *J Mol Biol*, *273*(1), 207-225. https://doi.org/10.1006/jmbi.1997.1315

Pessoa, T. B. A., Rezende, R. P., Marques, E. d. L. S., Pirovani, C. P., dos Santos, T. F., dos Santos Gonçalves, A. C., Romano, C. C., Dotivo, N. C., Freitas, A. C. O., Salay, L. C., & Dias, J. C. T. (2017). Metagenomic alkaline protease from mangrove sediment. *Journal of Basic Microbiology*, *57*(11), 962-973. https://doi.org/https://doi.org/10.1002/jobm.201700159

Pfeiffer, P., Goedecke, W., & Obe, G. (2000). Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis*, *15*(4), 289-302. https://doi.org/10.1093/mutage/15.4.289

Phillips, R. J., Hickleton, D. C., Boehmer, P. E., & Emmerson, P. T. (1997). The RecB protein of Escherichia coli translocates along single-stranded DNA in the 3′ to 5′ direction: a proposed ratchet mechanism. *Molecular and General Genetics MGG*, *254*(3), 319-329. https://doi.org/10.1007/PL00008605

Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cellular and molecular life sciences*, *62*, 685-707.

Pingoud, A., & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res*, *29*(18), 3705-3727. https://doi.org/10.1093/nar/29.18.3705

Pingoud, A., Wilson, G. G., & Wende, W. (2014). Type II restriction endonucleases—a historical perspective and more. *Nucleic acids research*, *42*(12), 7489-7527. https://doi.org/10.1093/nar/gku447

Pitcher, R. S., Brissett, N. C., & Doherty, A. J. (2007). Nonhomologous End-Joining in Bacteria: A Microbial Perspective. *Annual Review of Microbiology*, *61*(Volume 61, 2007), 259-282. https://doi.org/https://doi.org/10.1146/annurev.micro.61.080706.093354

Pitcher, R. S., Tonkin, L. M., Green, A. J., & Doherty, A. J. (2005). Domain Structure of a NHEJ DNA Repair Ligase from Mycobacterium tuberculosis. *Journal of Molecular Biology*, *351*(3), 531-544. https://doi.org/https://doi.org/10.1016/j.jmb.2005.06.038

Pointing, S. B., Chan, Y., Lacap, D. C., Lau, M. C. Y., Jurgens, J. A., & Farrell, R. L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. *Proceedings of the National Academy of Sciences*, *106*(47), 19964-19969. https://doi.org/10.1073/pnas.0908274106

Portero, L. R., Alonso-Reyes, D. G., Zannier, F., Vazquez, M. P., Farías, M. E., Gärtner, W., & Albarracín, V. H. (2019). Photolyases and Cryptochromes in UV-resistant Bacteria from High-altitude Andean Lakes. *Photochemistry and Photobiology*, *95*(1), 315-330. https://doi.org/https://doi.org/10.1111/php.13061

Pukkila, P. J., Peterson, J., Herman, G., Modrich, P., & Meselson, M. (1983). EFFECTS OF HIGH LEVELS OF DNA ADENINE METHYLATION ON METHYL-DIRECTED MISMATCH REPAIR IN ESCHERICHIA COLI. *Genetics*, *104*(4), 571-582. https://doi.org/10.1093/genetics/104.4.571

Raaijmakers, H., Törö, I., Birkenbihl, R., Kemper, B., & Suck, D. (2001). Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage1 1Edited by K. Morikawa. *Journal of Molecular Biology*, *308*(2), 311-323. https://doi.org/https://doi.org/10.1006/jmbi.2001.4592

Raaijmakers, H., Vix, O., Törõ, I., Golz, S., Kemper, B., & Suck, D. (1999). X ‐ ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain ‐ swapped dimer architecture. *The EMBO journal*, *18*(6), 1447-1458-1458. https://doi.org/https://doi.org/10.1093/emboj/18.6.1447

Rafferty, J. B., Bolt, E. L., Muranova, T. A., Sedelnikova, S. E., Leonard, P., Pasquo, A., Baker, P. J., Rice, D. W., Sharples, G. J., & Lloyd, R. G. (2003a). The Structure of Escherichia coli RusA Endonuclease Reveals a New Holliday Junction DNA Binding Fold. *Structure*, *11*(12), 1557-1567. https://doi.org/10.1016/j.str.2003.11.004

Rafferty, J. B., Bolt, E. L., Muranova, T. A., Sedelnikova, S. E., Leonard, P., Pasquo, A., Baker, P. J., Rice, D. W., Sharples, G. J., & Lloyd, R. G. (2003b). The Structure of Escherichia coli RusA Endonuclease Reveals a New Holliday Junction DNA Binding Fold. *Structure*, *11*(12), 1557-1567. https://doi.org/https://doi.org/10.1016/j.str.2003.11.004

Rass, U., Compton, S. A., Matos, J., Singleton, M. R., Ip, S. C., Blanco, M. G., Griffith, J. D., & West, S. C. (2010). Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev*, *24*(14), 1559-1569. https://doi.org/10.1101/gad.585310

Raymond-Bouchard, I., Goordial, J., Zolotarov, Y., Ronholm, J., Stromvik, M., Bakermans, C., & Whyte, L. G. (2018). Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. *FEMS Microbiology Ecology*, *94*(4). https://doi.org/10.1093/femsec/fiy023

Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of protein-protein interactions. *Protein Sci*, *22*(2), 153-167. https://doi.org/10.1002/pro.2206

Reyes, G. X., Schmidt, T. T., Kolodner, R. D., & Hombauer, H. (2015). New insights into the mechanism of DNA mismatch repair. *Chromosoma*, *124*(4), 443-462. https://doi.org/10.1007/s00412-015-0514-0

Rice, P., & Steitz, T. (1994). Refinement of γ δ resolvase reveals a strikingly flexible molecule. *Structure*, *2*(5), 371-384.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, *42*(W1), W320-W324. https://doi.org/10.1093/nar/gku316

Rodrigues, D. F., & Tiedje, J. M. (2008). Coping with our cold planet. *Appl Environ Microbiol*, *74*(6), 1677-1686. https://doi.org/10.1128/aem.02000-07

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. *Annu Rev Biochem*, *79*, 233-269. https://doi.org/10.1146/annurev-biochem-060408-091030

Rosenquist, T. A., Zharkov, D. O., & Grollman, A. P. (1997). Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase. *Proc Natl Acad Sci U S A*, *94*(14), 7429-7434. https://doi.org/10.1073/pnas.94.14.7429

Rossetti, G., Dans, P. D., Gomez-Pinto, I., Ivani, I., Gonzalez, C., & Orozco, M. (2015). The structural impact of DNA mismatches. *Nucleic Acids Res*, *43*(8), 4309-4321. https://doi.org/10.1093/nar/gkv254

Rzoska-Smith, E. (2023). *Repair eXtreme-DNA repair proteins from Antarctic extremophiles* The University of Waikato].

Rzoska-Smith, E., Stelzer, R., Monterio, M., Cary, S. C., & Williamson, A. (2023). DNA repair enzymes of the Antarctic Dry Valley metagenome [Original Research]. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1156817

Rzoska-Smith;, E., Stelzer;, R., Monteiro;, M. R., Cary;, S. C., & Williamson, A. (2023). DNA Repair Enzymes of the Antarctic Dry Valley Metagenome. *Frontiers in Microbiology*, *14*, in press. https://doi.org/10.3389/fmicb.2023.1156817

Saito, A., Iwasaki, H., Ariyoshi, M., Morikawa, K., & Shinagawa, H. (1995). Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase. *Proc Natl Acad Sci U S A*, *92*(16), 7470-7474. https://doi.org/10.1073/pnas.92.16.7470

Sapienza, P. J., Rosenberg, J. M., & Jen-Jacobson, L. (2007). Structural and thermodynamic basis for enhanced DNA binding by a promiscuous mutant EcoRI endonuclease. *Structure*, *15*(11), 1368-1382. https://doi.org/10.1016/j.str.2007.09.014

Sarai, N., Kagawa, W., Kurumizaka, H., & Yokoyama, S. (2007). *Crystal Structure of Holliday Junction Resolvase ST1444*.

Sargentini, N. J., Gularte, N. P., & Hudman, D. A. (2016). Screen for genes involved in radiation survival of Escherichia coli and construction of a reference database. *Mutat Res*, *793-794*, 1-14. https://doi.org/10.1016/j.mrfmmm.2016.10.001

Sasnauskas, G., Connolly, B. A., Halford, S. E., & Siksnys, V. (2007). Site-specific DNA transesterification catalyzed by a restriction enzyme. *Proc Natl Acad Sci U S A*, *104*(7), 2115-2120. https://doi.org/10.1073/pnas.0608689104

Schrödinger, L. *The PyMOL Molecular Graphics System*. In (Version Version 2.5.4)

Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., & Lindahl, T. (2007). Repair of alkylated DNA: Recent advances. *DNA Repair*, *6*(4), 429-442. https://doi.org/https://doi.org/10.1016/j.dnarep.2006.10.005

Sedgwick, B., & Lindahl, T. (2002). Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene*, *21*(58), 8886-8894. https://doi.org/10.1038/sj.onc.1205998

Selent, U., Rüter, T., Köhler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschläger, T., Wolfes, H., Peters, F., & Pingoud, A. (1992). A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease EcoRV. *Biochemistry*, *31*(20), 4808-4815. https://doi.org/10.1021/bi00135a010

Selvam, K., Duncan, J. R., Tanaka, M., & Battista, J. R. (2013). DdrA, DdrD, and PprA: Components of UV and Mitomycin C Resistance in Deinococcus radiodurans R1. *PLoS One*, *8*(7), e69007. https://doi.org/10.1371/journal.pone.0069007

Sengerová, B., Allerston, C. K., Abu, M., Lee, S. Y., Hartley, J., Kiakos, K., Schofield, C. J., Hartley, J. A., Gileadi, O., & McHugh, P. J. (2012). Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J Biol Chem*, *287*(31), 26254-26267. https://doi.org/10.1074/jbc.M112.367243

Shah Punatar, R., Martin, M. J., Wyatt, H. D. M., Chan, Y. W., & West, S. C. (2017). Resolution of single and double Holliday junction recombination intermediates by GEN1. *Proceedings of the National Academy of Sciences*, *114*(3), 443-450. https://doi.org/10.1073/pnas.1619790114

Shi, K., Moeller, N. H., Banerjee, S., McCann, J. L., Carpenter, M. A., Yin, L., Moorthy, R., Orellana, K., Harki, D. A., Harris, R. S., & Aihara, H. (2021). Structural basis for recognition of distinct deaminated DNA lesions by endonuclease Q. *Proc Natl Acad Sci U S A*, *118*(10). https://doi.org/10.1073/pnas.2021120118

Shida, T., Noda, M., & Sekiguchi, J. (1996). Cleavage of Single- and Double-Stranded DNAs Containing an Abasic Residue by Escherichia Coli Exonuclease III (AP Endonuclease VI). *Nucleic acids research*, *24*(22), 4572-4576. https://doi.org/10.1093/nar/24.22.4572

Shih, Y.-M., Cooke, M. S., Pan, C.-H., Chao, M.-R., & Hu, C.-W. (2019). Clinical relevance of guanine-derived urinary biomarkers of oxidative stress, determined by LC-MS/MS. *Redox Biology*, *20*, 556-565. https://doi.org/https://doi.org/10.1016/j.redox.2018.11.016

Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews Microbiology*, *5*(11), 852-861. https://doi.org/10.1038/nrmicro1768

Siddiqui, K. S., Feller, G., D'Amico, S., Gerday, C., Giaquinto, L., & Cavicchioli, R. (2005). The active site is the least stable structure in the unfolding pathway of a multidomain cold-adapted alpha-amylase. *J Bacteriol*, *187*(17), 6197-6205. https://doi.org/10.1128/jb.187.17.6197-6205.2005

Singh, J., Su, L., & Snow, E. T. (1996). Replication across O6-Methylguanine by Human DNA Polymerase β in Vitro: INSIGHTS INTO THE FUTILE CYTOTOXIC REPAIR AND MUTAGENESIS OF O6-METHYLGUANINE*. *Journal of Biological Chemistry*, *271*(45), 28391-28398. https://doi.org/https://doi.org/10.1074/jbc.271.45.28391

Slade, D., Lindner, A. B., Paul, G., & Radman, M. (2009). Recombination and replication in DNA repair of heavily irradiated Deinococcus radiodurans. *Cell*, *136*(6), 1044-1055. https://doi.org/10.1016/j.cell.2009.01.018

Smith, J. J., Tow, L. A., Stafford, W., Cary, C., & Cowan, D. A. (2006). Bacterial diversity in three different Antarctic Cold Desert mineral soils. *Microb Ecol*, *51*(4), 413-421. https://doi.org/10.1007/s00248-006-9022-3

Sočan, J., Isaksen, G. V., Brandsdal, B. O., & Åqvist, J. (2019). Towards Rational Computational Engineering of Psychrophilic Enzymes. *Sci Rep*, *9*(1), 19147. https://doi.org/10.1038/s41598-019-55697-4

Solaro, P. C., Birkenkamp, K., Pfeiffer, P., & Kemper, B. (1993). Endonuclease VII of Phage T4 Triggers Mismatch Correction in Vitro. *Journal of Molecular Biology*, *230*(3), 868-877. https://doi.org/https://doi.org/10.1006/jmbi.1993.1207

Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., & Ginalski, K. (2012). Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic acids research*, *40*(15), 7016-7045. https://doi.org/10.1093/nar/gks382

Steele, H. L., Jaeger, K.-E., Daniel, R., & Streit, W. R. (2008). Advances in Recovery of Novel Biocatalysts from Metagenomes. *Journal of Molecular Microbiology and Biotechnology*, *16*(1-2), 25-37. https://doi.org/10.1159/000142892

Stelzer, R., Rzoska-Smith, E., Gundesø, S., Rothweiler, U., & Williamson, A. (2024). Using Modified Synthetic Oligonucleotides to Assay Nucleic Acid-Metabolizing Enzymes. *JoVE*(209), e66930. https://doi.org/doi:10.3791/66930

Strahl, H., & Errington, J. (2017). Bacterial Membranes: Structure, Domains, and Function. *Annual Review of Microbiology*, *71*(Volume 71, 2017), 519-538. https://doi.org/https://doi.org/10.1146/annurev-micro-102215-095630

Straume, D., Stamsås, G. A., & Håvarstein, L. S. (2015). Natural transformation and genome evolution in Streptococcus pneumoniae. *Infection, Genetics and Evolution*, *33*, 371-380. https://doi.org/https://doi.org/10.1016/j.meegid.2014.10.020

Su, S. S., & Modrich, P. (1986). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proceedings of the National Academy of Sciences*, *83*(14), 5057-5061. https://doi.org/doi:10.1073/pnas.83.14.5057

Szabla, R., Li, M., Warner, V., Song, Y., & Junop, M. (2024). DdrC, a unique DNA repair factor from D. radiodurans, senses and stabilizes DNA breaks through a novel lesion-recognition mechanism. *Nucleic acids research*, gkae635. https://doi.org/10.1093/nar/gkae635

Taira, K., Kaneto, S., Nakano, K., Watanabe, S., Takahashi, E., Arimoto, S., Okamoto, K., Schaaper, R. M., Negishi, K., & Negishi, T. (2013). Distinct pathways for repairing mutagenic lesions induced by methylating and ethylating agents. *Mutagenesis*, *28*(3), 341-350. https://doi.org/10.1093/mutage/get010

Takeshita, M., Chang, C. N., Johnson, F., Will, S., & Grollman, A. P. (1987). Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J Biol Chem*, *262*(21), 10171-10179.

Tano, K., Shiota, S., Collier, J., Foote, R. S., & Mitra, S. (1990). Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. *Proc Natl Acad Sci U S A*, *87*(2), 686-690. https://doi.org/10.1073/pnas.87.2.686

Tao, Y., Budhipramono, A., Huang, J., Fang, M., Xie, S., Kim, J., Khivansara, V., Dominski, Z., Tong, L., De Brabander, J. K., & Nijhawan, D. (2024). Anticancer benzoxaboroles block pre-mRNA processing by directly inhibiting CPSF3. *Cell Chemical Biology*, *31*(1), 139-149.e114. https://doi.org/10.1016/j.chembiol.2023.10.019

Thompson, P. S., & Cortez, D. (2020). New insights into abasic site repair and tolerance. *DNA Repair*, *90*, 102866. https://doi.org/https://doi.org/10.1016/j.dnarep.2020.102866

Tiao, G., Lee, C. K., McDonald, I. R., Cowan, D. A., & Cary, S. C. (2012). Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys. *Nature Communications*, *3*(1), 660. https://doi.org/10.1038/ncomms1645

Timmins, J., & Moe, E. (2016). A Decade of Biochemical and Structural Studies of the DNA Repair Machinery of Deinococcus radiodurans: Major Findings, Functional and Mechanistic Insight and Challenges. *Computational and Structural Biotechnology Journal*, *14*, 168-176. https://doi.org/https://doi.org/10.1016/j.csbj.2016.04.001

Tribelli, P. M., & López, N. I. (2018). Reporting Key Features in Cold-Adapted Bacteria. *Life*, *8*(1), 8. https://www.mdpi.com/2075-1729/8/1/8

van der Burg, M., Verkaik, N. S., den Dekker, A. T., Barendregt, B. H., Pico-Knijnenburg, I., Tezcan, I., vanDongen, J. J. M., & van Gent, D. C. (2007). Defective Artemis nuclease is characterized by coding joints with microhomology in long palindromic-nucleotide stretches. *European Journal of Immunology*, *37*(12), 3522-3528. https://doi.org/https://doi.org/10.1002/eji.200737624

van Eijk, E., Wittekoek, B., Kuijper, E. J., & Smits, W. K. (2017). DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. *Journal of Antimicrobial Chemotherapy*, *72*(5), 1275-1284. https://doi.org/10.1093/jac/dkw548

van Loon, B., Markkanen, E., & Hübscher, U. (2010). Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair*, *9*(6), 604-616. https://doi.org/https://doi.org/10.1016/j.dnarep.2010.03.004

Verhoeven, E. E. A., Wyman, C., Moolenaar, G. F., & Goosen, N. (2002). The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *The EMBO journal*, *21*(15), 4196-4205. https://doi.org/https://doi.org/10.1093/emboj/cdf396

Vipond, I. B., Baldwin, G. S., & Halford, S. E. (1995). Divalent metal ions at the active sites of the EcoRV and EcoRI restriction endonucleases. *Biochemistry*, *34*(2), 697-704. https://doi.org/10.1021/bi00002a037

Wagner, R., Jr., & Meselson, M. (1976). Repair tracts in mismatched DNA heteroduplexes. *Proc Natl Acad Sci U S A*, *73*(11), 4135-4139. https://doi.org/10.1073/pnas.73.11.4135

Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S., & Masui, R. (2010). Structure of RecJ Exonuclease Defines Its Specificity for Single-stranded DNA *. *Journal of Biological Chemistry*, *285*(13), 9762-9769. https://doi.org/10.1074/jbc.M109.096487

Wallace, S. S. (2014). Base excision repair: a critical player in many games. *DNA Repair (Amst)*, *19*, 14-26. https://doi.org/10.1016/j.dnarep.2014.03.030

Wallon, G., Lovett, S. T., Magyar, C., Svingor, A., Szilagyi, A., Zàvodszky, P., Ringe, D., & Petsko, G. A. (1997). Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium Vibrio sp. I5 suggest reasons for thermal instability. *Protein Eng*, *10*(6), 665-672. https://doi.org/10.1093/protein/10.6.665

Wang, P., & Wang, Y. (2018). Cytotoxic and mutagenic properties of O6-alkyl-2′ -deoxyguanosine lesions in Escherichia coli cells. *Journal of Biological Chemistry*, *293*(39), 15033-15042. https://doi.org/https://doi.org/10.1074/jbc.RA118.004676

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737-738. https://doi.org/10.1038/171737a0

Wei, S. T., Lacap-Bugler, D. C., Lau, M. C., Caruso, T., Rao, S., de Los Rios, A., Archer, S. K., Chiu, J. M., Higgins, C., Van Nostrand, J. D., Zhou, J., Hopkins, D. W., & Pointing, S. B. (2016). Taxonomic and Functional Diversity of Soil and Hypolithic Microbial Communities in Miers Valley, McMurdo Dry Valleys, Antarctica. *Front Microbiol*, *7*, 1642. https://doi.org/10.3389/fmicb.2016.01642

Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., di Fagagna, F. d. A., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P., & Doherty, A. J. (2002). Identification of a DNA Nonhomologous End-Joining Complex in Bacteria. *Science*, *297*(5587), 1686-1689. https://doi.org/10.1126/science.1074584

White, M. F., & Lilley, D. M. J. (1997). The resolving enzyme CCE1 of yeast opens the structure of the four-way DNA junction 1 1 Edited by M. Yaniv. *Journal of Molecular Biology*, *266*(1), 122-134. https://doi.org/https://doi.org/10.1006/jmbi.1996.0795

White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J. J., Lam, P., McDonald, L., . . . Fraser, C. M. (1999). Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. *Science*, *286*(5444), 1571-1577. https://doi.org/10.1126/science.286.5444.1571

Williamson, A., Hjerde, E., & Kahlke, T. (2016). Analysis of the distribution and evolution of the ATP-dependent DNA ligases of bacteria delineates a distinct phylogenetic group 'Lig E'. *Molecular Microbiology*, *99*(2), 274-290. https://doi.org/https://doi.org/10.1111/mmi.13229

Wilson, D. M., 3rd. (2003). Properties of and substrate determinants for the exonuclease activity of human apurinic endonuclease Ape1. *J Mol Biol*, *330*(5), 1027-1037. https://doi.org/10.1016/s0022-2836(03)00712-5

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R., Heathman, S., Bryan, R., Martin, P., Petratos, K., & Wilson, K. (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non ‐ cognate DNA fragments. *The EMBO journal*, *12*(5), 1781-1795.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D*, *67*(4), 235-242. https://doi.org/doi:10.1107/S0907444910045749

Wlostowski, A. N., Gooseff, M. N., McKnight, D. M., Jaros, C., & Lyons, W. B. (2016). Patterns of hydrologic connectivity in the McMurdo Dry Valleys, Antarctica: a synthesis of 20 years of hydrologic data. *Hydrological Processes*, *30*(17), 2958-2975. https://doi.org/https://doi.org/10.1002/hyp.10818

Wolf, F. I., & Cittadini, A. (2003). Chemistry and biochemistry of magnesium. *Molecular Aspects of Medicine*, *24*(1), 3-9. https://doi.org/https://doi.org/10.1016/S0098-2997(02)00087-0

Wozniak, K. J., & Simmons, L. A. (2022). Bacterial DNA excision repair pathways. *Nat Rev Microbiol*, *20*(8), 465-477. https://doi.org/10.1038/s41579-022-00694-0

Wyatt, H. D., & West, S. C. (2014). Holliday junction resolvases. *Cold Spring Harb Perspect Biol*, *6*(9), a023192. https://doi.org/10.1101/cshperspect.a023192

Yan, J., Hong, S., Guan, Z., He, W., Zhang, D., & Yin, P. (2020). Structural insights into sequence-dependent Holliday junction resolution by the chloroplast resolvase MOC1. *Nature Communications*, *11*(1), 1417.

Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., & Bai, L. (2017). Discovery and characterizaton of a novel lipase with transesterification activity from hot spring metagenomic library. *Biotechnology Reports*, *14*, 27-33. https://doi.org/https://doi.org/10.1016/j.btre.2016.12.007

Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys*, *44*(1), 1-93. https://doi.org/10.1017/s0033583510000181

Yang, W., Chen, W. Y., Wang, H., Ho, J. W., Huang, J. D., Woo, P. C., Lau, S. K., Yuen, K. Y., Zhang, Q., Zhou, W., Bartlam, M., Watt, R. M., & Rao, Z. (2011). Structural and functional insight into the mechanism of an alkaline exonuclease from Laribacter hongkongensis. *Nucleic Acids Res*, *39*(22), 9803-9819. https://doi.org/10.1093/nar/gkr660

Yeeles, J. T. P., & Dillingham, M. S. (2010). The processing of double-stranded DNA breaks for recombinational repair by helicase–nuclease complexes. *DNA Repair*, *9*(3), 276-285. https://doi.org/https://doi.org/10.1016/j.dnarep.2009.12.016

Yi, C., & He, C. (2013). DNA repair by reversal of DNA damage. *Cold Spring Harb Perspect Biol*, *5*(1), a012575. https://doi.org/10.1101/cshperspect.a012575

Yonekura, S., Nakamura, N., Yonei, S., & Zhang-Akiyama, Q. M. (2009). Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J Radiat Res*, *50*(1), 19-26. https://doi.org/10.1269/jrr.08080

Yosaatmadja, Y., Baddock, H. T., Newman, J. A., Bielinski, M., Gavard, A. E., Mukhopadhyay, S. M. M., Dannerfjord, A. A., Schofield, C. J., McHugh, P. J., & Gileadi, O. (2021). Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Res*, *49*(16), 9310-9326. https://doi.org/10.1093/nar/gkab693

Yu, W., Zhang, L., Wei, Q., & Shao, A. (2019). O(6)-Methylguanine-DNA Methyltransferase (MGMT): Challenges and New Opportunities in Glioma Chemotherapy. *Front Oncol*, *9*, 1547. https://doi.org/10.3389/fonc.2019.01547

Yung, C. C. M., Chan, Y., Lacap, D. C., Pérez-Ortega, S., de los Rios-Murillo, A., Lee, C. K., Cary, S. C., & Pointing, S. B. (2014). Characterization of Chasmoendolithic Community in Miers Valley, McMurdo Dry Valleys, Antarctica. *Microbial Ecology*, *68*(2), 351-359. https://doi.org/10.1007/s00248-014-0412-7

Yusof, N. (2018). Advances of radiation sterilisation in tissue banking. *Cell and Tissue Banking*, *19*(2), 175-186. https://doi.org/10.1007/s10561-017-9651-4

Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A. B., & Radman, M. (2006). Reassembly of shattered chromosomes in Deinococcus radiodurans. *Nature*, *443*(7111), 569-573. https://doi.org/10.1038/nature05160

Zhang, X., Zhou, Z., Dai, L., Chao, Y., Liu, Z., Huang, M., Qu, Q., & Lin, Z. (2023). Cryo-EM structure of the RuvAB-Holliday junction intermediate complex from Pseudomonas aeruginosa [Original Research]. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1139106

Zhao, H. (2005). Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids. *Journal of Molecular Catalysis B: Enzymatic*, *37*(1), 16-25. https://doi.org/https://doi.org/10.1016/j.molcatb.2005.08.007

Zhu, H., & Shuman, S. (2005). A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). *J Biol Chem*, *280*(1), 418-427. https://doi.org/10.1074/jbc.M410110200

Zhu, H., & Shuman, S. (2007). Characterization of Agrobacterium tumefaciens DNA ligases C and D. *Nucleic acids research*, *35*(11), 3631-3645. https://doi.org/10.1093/nar/gkm145

7 SUPPLEMENTARY

7.1 Supplementary Methods

7.1.1 Bacterial Strains and Plasmids used

Supplementary Table 1: Optimised expression conditions for the DV-Hjc and DV1-1 NucDom protein.

	Expression conditions			**Purification conditions**	
	Vector	*E. coli* Strain	Temperature	TEV cleavage	Gel filtration column
DV-Hjc	pHMGWA	Origami™ (DE3)	15°C	Yes	S75
DV1-1 NucDom	pHMGWA	BL21(DE3) pLysS	30°C	Yes and No*	S200

*as mentioned with the results

Supplementary Table 2: Bacterial strains used in this study.

***E. coli* Strain**	**Use in this study**
DH5α	Plasmid propagation
Origami™ (DE3)	Recombinant protein expression
BL21(DE3) pLysS	Recombinant protein expression
K-12	Negative control for the *in vivo* nuclease assay
K-12 *yraN* gene knockout (JW3117)	Single-gene deletion of the *yraN* gene in *E. coli* K-12 used for the *in vivo* nuclease assay
K-12 *ybaB* gene knockout (JW0460)	Single-gene deletion of the *ybaB* gene in E. coli K-12 used for the in vivo nuclease assay
K-12 *ruvC* gene knockout (JW1852)	Single-gene deletion of the *ruvC* gene in E. coli K-12 used for the in vivo nuclease assay

Supplementary Table 3: Plasmids used in this study.

Plasmid	**Description and use in this study**	**Antibiotic resistance**

pENTR	Entry clone containing the genes of interest	Kanamycin
pHMGWA	Gene expression vector with an N-terminal 6xHis-MBP tag Recombinant expression of the DV-Hjc and DV1-1 NucDom protein	Ampicillin
pProEx	Bacterial gene expression with an N-Terminal 6xHis tag Used here to express the DV-Hjc protein in *E. coli* K-12 *yraN* gene knockout cells	Ampicillin
pUC19	Cloning vector Used for the pUC19 plasmid-based assay	Ampicillin

7.1.2 Transformation

Transformation of chemically competent cells was carried out by adding 2 µL of the plasmid to thawed competent cells and incubating on ice for 30 min. This is followed by 'heat shock' at 42°C for 45 sec. The cells are put back on ice for 2 min before adding 1 mL of SOC media (20 mg/L Tryptone, 5 g/L Yeast Extract, 0.5 g/L NaCl, 2.5 mM KCl, pH 7.5) and incubating for 1 hour at 37°C with 350 rpm shaking. The cells are then pelleted and grown on LB agar plates with the appropriate antibiotic for 16 hours at 37°C.

7.1.3 Purification Supplementary:

Purification Buffers

Supplementary Table 4: Lysis Buffer recipe.

Component	**Concentration**
Tris (pH 8)	0.05 M
NaCl	0.75 M
Glycerol	5%
$MgCl_2$	1 mM

Supplementary Table 5: Purification 'Buffer A' recipe.

Component	Concentration
Tris (pH 8)	0.05 M
NaCl	0.75 M
Glycerol	5%
Imidazole	0.01 M

Supplementary Table 6: Purification 'Buffer B' recipe.

Component	Concentration
Tris (pH 8)	0.05 M
NaCl	0.75 M
Glycerol	5%
Imidazole	0.5 M

Supplementary Table 7: Purification 'Buffer C' recipe.

Component	Concentration
Tris (pH 8)	0.05 M
NaCl	0.1 M
Glycerol	5%
DTT	1 mM

Supplementary Table 8: Protein 'Storage Buffer' recipe.

Component	Concentration
Bis-Tris (pH 6.5)	0.05 M
NaCl	0.2 M
Glycerol	10%
DTT	1 mM

7.1.4 Oligonucleotide Assay Supplementary

Oligonucleotide sequences

Supplementary Table 9: Assay oligonucleotide sequences. Shown from 3' to 5'. '/56-FAM/' indicates the location a FAM-label and '/5Phos/' indicates phosphorylation of the 3'end.

Oligonucleotide Name	**Sequence (3' to 5')**
NL1	/56-FAM/AGGCCATGGCTGATATCGCA
NL3	CGACGGAGCTCGAATGCCTATGCGATATCAGCCATGGCCT
NL5	/56-FAM/AGGCCATGGCTGATATCGCATAGGCATTCGAGCTCCGTCG
NL6	/5Phos/CGACGGAGCTCGAATGCCTA
NL7	TGCGATATCAGCCATGGCCT
NL10	CGACGGAGCTCGAATGCCTACGCGATATCAGCCATGGCCT
HJ1	GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC
HJ2	TGGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAATGTAATCGTCAAGCTTTATGCCGTT
HJ3	CAACGGCATAAAGCTTGACGATTACATTGCTAGGACATGCTGTCTAGAGGATCCGACTATCGA
HJ4	ATCGATAGTCGGATCCTCTAGACAGCATGTCCTAGCAAGGCACTGGTAGAATTCGGCAGCGT
HJ5	ATCATAGCTAACATGACTAGTGCGATATCAGCCATGGCCT
HJ6	CTAGTCATGTTAGCTATGAT
MD6	/56-FAM/AGGCCATGGCTGATATCGCA/idSp/AGGCATTCGAGCTCCGTCG
MD9	/56-FAM/AGGCCATGGCTGATATCGCA/ideoxyU/AGGCATTCGAGCTCCGTCG
MD10	CGACGGAGCTCGAATGCCTGTGCGATATCAGCCATGGCCT

Supplementary Table 10: Oligonucleotide combinations for the assay substrate mastermixes.

Substrate Name	3'fluorescently labelled oligonucleotide	Complement oligonucleotide
Double-Strand 40nt	NL5	NL3
Single-strand 40nt	NL5	
3'tail	NL5	NL6
5'tail	NL1	NL3
Double-strand 20nt	NL1	NL6
Single-strand 20nt	NL1	
Flap3'	NL5	HJ6
		HJ5
Flap 5'	NL5	NL7
		HJ5
Splayed	NL5	HJ6
Holliday Junction	HJ1	HJ2
		HJ3
		HJ4
Abasic Site	MD6	NL3
Uracil Mismatch	MD9	MD10
A/C Mismatch	NL5	NL10

7.1.5 Nuclease Assay Supplementary

Supplementary Table 11: 'Standard' assay recipe.

Component	Stock concentration	Final assay concentration
Bis-Tris (pH 6.5)	0.1 M	0.01 M
NaCl	0.5 M	0.05 M

Supplementary Table 12: DV-Hjc-specific buffer recipe.

Component	Stock concentration	Final assay concentration
Hepes (pH 6.5)	0.25 M	0.05 M
NaCl	0.25 M	0.05 M
BSA (Bovine Serum Albumin)	0.5 mg/mL	0.1 mg/mL
DTT (Dithiothreitol)	2.5 mM	0.5 mM
Glycerol	30%	6%

7.1.6 Primer and Protein Sequences

Supplementary Table 13: Primer sequences

Name	Sequence 3' to 5'
p.20240423.1	TTAGAACGCCGCACGCGGATCC
p.20240423.2	GGATCCTCTGCGGCGCGTCAGC
T7 forward	TAATACGACTCACTATAGGG
T7 reverse	GCTAGTTATTGCTCAGCGG

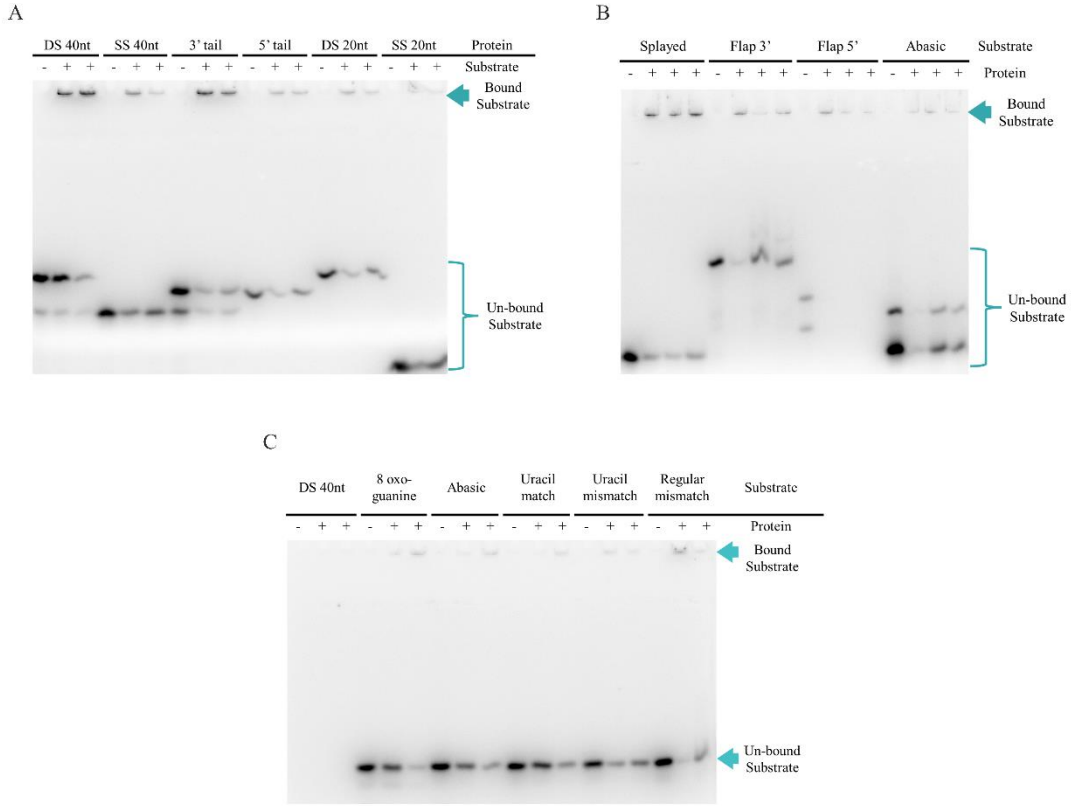
Supplementary Table 14: Amino acid sequence of DV-Hjc, DV-Hjc^{D42A}, DV1-1 NucDom, and DV1-1 NucDom$^{H34A, D36A}$.

Name	Sequence
DV-Hjc	MSAARQRTGRLAESLVAEQLRDAGWEILARNARTSEVRGELDLIALDGPDLV FVEVKARRAGTSSGPETPAMAVGARKRAKLRSLATAWLRDAGSGVPRHRAL RFDVVGIRLDARGRAREYEHLRAAF*
DV-Hjc^{D42A}	MSAARQRTGRLAESLVAEQLRDAGWEILARNARTSEVRGELDLIALAGPDLV FVEVKARRAGTSSGPETPAMAVGARKRAKLRSLATAWLRDAGSGVPRHRAL RFDVVGIRLDARGRAREYEHLRAAF*
DV1-1 NucDom	LLPIRFHRGVELPEQSLWLDPHDPKPFAFVSHAHSDHLGTHAEIITSKGTSALM RERLPGERIEHVLEFDSPATIRGLNVTLLPAGHVFGSAQLFLQTENESLLYTGD FKLRRGLSAEPTGWRHADTLIMETTYGLPKYAMPPTEETLARMIAFCQEAQEE GAVPVLLGYSLGKAQEILCALVQAGLTPMLHGAVWNMTEVYRKLRPDFPCG YERYAAGETAGKVLVCPPSAIRMKMVTQIKQRRVAVLTGWALDPGAIYRYQ

	CDAAFPLTDHADYPDLLRYVELVQPKRVLTLHGFAAEFARDLRERGVEAWA LSEENQLELTL*
DV1-1 NucDom$^{H34A, D36A}$	LLPIRFHRGVELPEQSLWLDPHDPKPFAFVSHAASAHLGTHAEIITSKGTSALM RERLPGERIEHVLEFDSPATIRGLNVTLLPAGHVFGSAQLFLQTENESLLYTGD FKLRRGLSAEPTGWRHADTLIMETTYGLPKYAMPPTEETLARMIAFCQEAQEE GAVPVLLGYSLGKAQEILCALVQAGLTPMLHGAVWNMTEVYRKLRPDFPCG YERYAAGETAGKVLVCPPSAIRMKMVTQIKQRRVAVLTGWALDPGAIYRYQ CDAAFPLTDHADYPDLLRYVELVQPKRVLTLHGFAAEFARDLRERGVEAWA LSEENQLELTL*

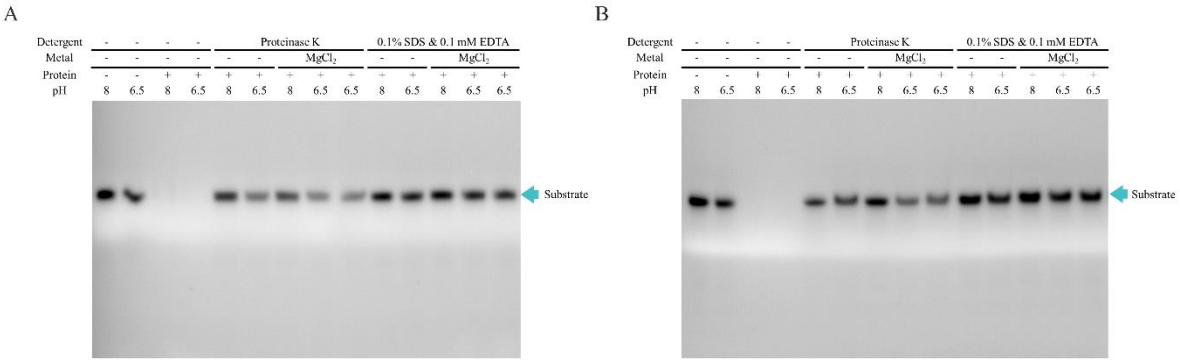
7.2 Supplementary Results – DV-Hjc

7.2.1 Supplementary – DV-Hjc Binding

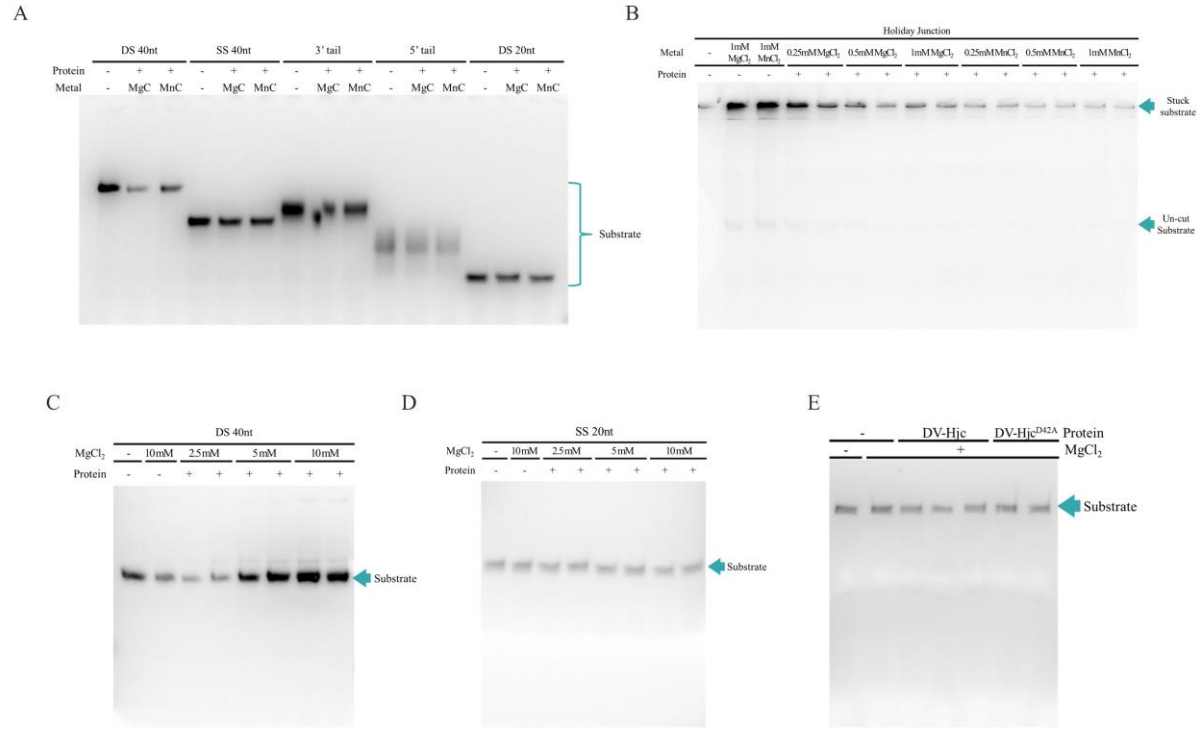


Supplementary Figure 1: EMSA of DV-Hjc with DNA substrates indicated above the gel images.
The presence (+) or absence (-) of the protein at 19 nM are indicated above the gel images. The position of bound DNA and unbound DNA are indicated by the blue arrows and brackets.

7.2.2 Supplementary – DV-Hjc Assays



Supplementary Figure 2: Optimisation of DV-Hjc assays with proteinase K or 0.1% SDS and 0.1mM EDTA treatment. Incubation with proteinase K or SDS and EDTA was done before the reactions were run on the gel to inhibit DV-Hjc binding to the substrate and stopping the substrate from getting stuck in the wells of the gel. The treatment type (proteinase K or 0.1% SDS and 0.1 mM EDTA), presence of $MgCl_2$, presence (+) or absence (-) of the protein (198 nm) and the pH of the reaction are indicated above the gel.



Supplementary Figure 3: DV-Hjc assays with DNA substrates. Protein at a concentration of 198 nm was used and the presence (+) or absence (-) of protein are indicated above the gel. The protein concentrations are indicated above the gel, where no concentration is specified 10 mM of the metal was used. The substrate is indicated above the gel and the location of the substrate is indicated by the blue arrowa.

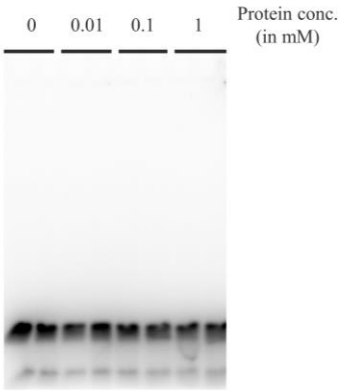
7.2.3 Supplementary – *In vivo* Characterisation of UPF0102 in *E. coli*

Supplementary Table 15: 90-minute 3 mM H2O2 treatment of E. coli strains from the Keio collection including the Wild Type K-12 E. coli, the yraN gene knockout (JW3117) and the ruvC knockout (JW1852). Indicated in red are CFU/mL, which were suspected to be contaminated, which were removed from the final results.

Strain	**CFU/mL (untreated)**	**CFU/mL (H_2O_2 treated)**

Wild-type	7.58E+12	4.85E+08
	2.70E+12	4.70E+08
	2.10E+12	4.20E+08
***yraN* knockout**	4.30E+14	1.30E+09
	8.70E+14	9.00E+08
	6.30E+14	3.00E+08
***ruvC* knockout**	4.20E+12	5.30E+07
	2.20E+12	7.40E+07
	2.00E+12	5.70E+07

Supplementary of DV1-1



Supplementary Figure 4: Negative control for the formaldehyde crosslinking method in electron mobility shift assays using BSA was used as a negative control. The protein concentrations are indicated above the gel. The reactions were electrophoresed on a 20% acrylamide 7 M UREA PAGE gel.