Characterization of Prokaryotic Ku DNA Binding Properties

Characterization of Prokaryotic Ku DNA Binding Properties

By Lucas Koechlin, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University Master of Science (2020)
Hamilton, Ontario
(Biochemistry)

Title: Characterization of Prokaryotic Ku DNA Binding Properties
Author: Lucas Koechlin, B.Sc. (University of Western Ontario)
Supervisor: Dr. Sara Andres
Number of Pages: 85

Abstract:

DNA damage occurs to all living things; its subsequent repair is a crucial component of life. The most dangerous, and potentially most useful form of DNA damage is the double strand break (DSB). A DSB is defined by breaks occurring to both sugar phosphate backbones in close enough proximity that they lead to the separation of the two pieces of the DNA. This type of damage will kill the cell if left unrepaired. It is the most lethal type of DNA damage.

Most living organisms have also developed ways to take advantage of DSBs through their repair systems, primarily as a means of introducing genetic variation. There are two primary DSB repair pathways across life: homologous recombination (HR) and non-homologous end-joining (NHEJ). The focus of this work is NHEJ. NHEJ is known as "error-prone" because it does not use a homologous template and can introduce small addition or deletion mutations during the repair process. This pathway has been extensively studied in eukaryotes and is known as the primary form of DSB repair in mammalian cells, however the prokaryotic NHEJ system was more recently identified and as a result, a void of information surrounds it.

NHEJ is comprised of 3 core steps: DSB recognition and binding, DNA end processing, and ligation. In the eukaryotic version of NHEJ these 3 steps involve a plethora of factors; conversely, in the prokaryotic version, the same functionality is accomplished by just 2 proteins, bacterial Ku and LigD. The focus of this research is Ku: the DNA end-binding protein responsible for identifying the DSB, binding and protecting the DNA end, as well as recruiting LigD to the break. Ku is composed of 2 domains, the first of which is predicted to be highly homologous to eukaryotic Ku's equivalent domain; this is the core domain which forms a ring-like structure that DNA threads through. The second is completely unique to bacterial Ku, it is the C-terminal domain, which can further be split into 2 sub-domains, the minimal C-terminus, and the extended C-terminus. The sub-domains are defined by their level of conservation across bacterial species, with the minimal C-terminus being highly conserved, while the extended C-terminus is highly variable. Using DNA-binding assays and several mutant constructs which affect the C-terminal domain, I show that this C-terminus is unexpectedly responsible for destabilizing the Ku-DNA interaction. This observation leads me to hypothesize that maintaining a weak interaction with DNA is important for Ku because of the other proteins which need access to the DNA (e.g. replicative helicase). While Ku is bound, it could be capable of blocking regions of DNA, in turn blocking other vital cellular processes like replication. Ku maintaining a lower affinity for DNA should facilitate Ku displacement by other proteins. A tighter binding would restrict Ku's freedom to move on DNA making it more likely to inhibit other critical pathways. To better understand Ku, I attempted to solve the Ku structure using X-ray crystallography, and was able to achieve crystals of Ku, however diffraction was too limited for a structure. Another way to investigate the validity of my proposed model is to use a biophysical approach with atomic force microscopy (AFM) to visualize protein-DNA complexes. The initial work has established key controls for future Ku-DNA AFM work by imaging and analyzing Ku on its own. Interest in bacterial NHEJ is two-fold from the antimicrobial perspective: NHEJ is a highly mutagenic pathway, so it serves as a proverbial well for differentiation and thus the development of antimicrobial resistance (AMR); NHEJ is very important in bacteria that enter a stationary phase due to their lack of a homologous piece of DNA for HR. Thus, NHEJ inhibition

could be useful for slowing bacterial evolution and potentially as a treatment for infections such a *Mycobacterium tuberculosis*, which is known to lie dormant in host macrophages for long periods of time. To investigate the viability of NHEJ inhibition, I had begun the process of creating Δ*ku* strains of *Pseudomonas aeruginosa* to simulate Ku inhibition under various conditions. This Ku project is the focus of the first two chapters, however, during my Master's degree I participated in 2 other major projects. The third chapter details a bacterial DNA damage tolerance pathway, which similarly is highly mutagenic and poorly characterized: the ImuABC translesion synthesis polymerase complex. The fourth and final chapter details the work for a Journal of Visualized Experiments article meant to highlight the benefits of AFM as a means of studying protein-DNA interactions.

Acknowledgements:

I would like to thank my supervisor, Dr. Sara Andres for the tremendous guidance, wisdom, and support you've provided me over these past years as well as the patience you've shown with me. I would also like to thank my committee members, Dr. Giuseppe Melacini and Dr. Marie Elliot for their valuable feedback and suggestions.

Thank you to Cody Caba, Pardis Balari, Jose Rascon, Amro ElRafie, Prakhar Shah, and Niki Esfahanian for their contributions to the Ku project. The Ku project involved a great deal of support from a great group of undergraduates as well as an ex-graduate student.

Thank you to my fellow lab members Dana Sowa and Monica Warner for keeping me company, offering a hand when I needed it, providing an ear to bounce ideas off of, and the fun atmosphere they brought with them to the lab.

Thanks to the staff of SIBYLS! SAXS data was collected at SIBYLS which is supported by the DOE-BER IDAT DE-AC02-05CH11231 and NIGMS ALS-ENABLE (P30 GM124169 and S10OD018483).

Finally, thank you Mom, Dad, William, and Akshaya for the constant love and support.

Table of Contents:

Figure List:

Table List:

Abbreviations:

AFM	Atomic Force Microscopy
AMR	Antimicrobial Resistance
BER	Base Excision Repair
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
DPC	DNA-Protein Crosslink
DSB	Double Strand Break
dsDNA	Double-Stranded DNA
EMSA	Electrophoretic Mobility Shift Assay
HR	Homologous Recombination
MMR	Mismatch Repair
MRSA	Methicillin-Resistant *Staphylococcus Aureus*
MW	Molecular Weight
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End-Joining
PDB	Protein Data Bank
PEG	Polyethylene Glycol
rNTP	ribonucleotide Triphosphate
ROS	Reactive Oxygen Species
SAXS	Small Angle X-ray Scattering
SEC	Size Exclusion Chromatography
SOE	Splicing by Overlap Extension
SSB	Single Strand Break
ssDNA	Single-Stranded DNA
TLS	Translesion synthesis
WT	Wild-Type

Declaration of Academic Achievement:

All experimental content within this thesis is the work of the author, Lucas Koechin, except for the following:

Figure 2.3 – created by Cody Caba

The following students collected the EMSA data: Cody Caba, Pardis Balari, Jose Rascon, Amro ElRafie

The SAXS data was processed by Prakhar Shah

Chapter 1: Introduction

DNA double strand break repair:

DNA is one of the core components of life as we know it. Maintaining the integrity of a genome is crucial for living organisms. Genomes across all life are constantly acquiring different types of damage from a variety of sources, the most catastrophic of which is the double strand break (DSB). A DSB is defined as a break of the sugar phosphate backbone in both DNA strands, in close enough proximity that the base pairing interactions cannot hold the duplex together. This is typically defined as approximately 10bp, however it can be longer[1,2]. DSBs arise from a variety of sources, both endogenous and exogenous. Primary sources of endogenous damage arise from failed replication, such as when replication machinery is blocked by lesions or other DNA associated proteins, leading to collapsed forks. Other endogenous sources include reactive oxygen species (ROS) from regular cellular metabolism. These attack DNA and lead to a variety of types of damage that both directly introduce DSBs, as well as create lesions that disrupt replication, which ultimately lead to DSBs[3–5]. On the other hand, exogenous sources encompass a wide range of potential mutagens like ionizing radiation, which can directly damage the DNA or produce ROS that lead to DSBs. Common chemotherapeutics are also sources of DSB damage, including drugs like bleomycin, which attack DNA and directly cause damage[3,6,7]. Across life, two primary repair pathways remedy this lethal DNA damage: homologous recombination (HR) and non-homologous end-joining (NHEJ) (Fig. 1.1).

DSBs are a truly double-edged sword across the various forms of life, causing cell death if left unrepaired while simultaneously being used to provide different benefits in the diverse domains of life[3,8,9]. A common theme to the usefulness of DSBs across life, is their role in the generation of variation. In mammals for example, the basis of the immune system relies on the introduction and resolution of DSBs in a process called V(D)J recombination[10]. V(D)J recombination is the process by which the immune cells produce a vast array of diverse antigen receptors that allow the immune system to be exceptionally adaptive and capable of responding to almost any foreign entity in the body. This process occurs in developing lymphocytes and involves a recombinase, RAG, that identifies specific DNA sequences coined recombination signal sequences located at antigen receptor loci[11]. The recombinase introduces double strand breaks in these regions and shuttles the products into NHEJ to be rejoined in such a way that large chunks of the chromosome are lost, bringing together one "variable" segment (V), one "diversity" segment (D) and one "joining" segment (J) to the constant domain of the antigen receptor[11]. The numerous potential combinations of VDJ regions allow a smaller section of the genome to encode a vast and diverse array of receptors. This process takes place through NHEJ, which is a highly mutagenic pathway, which makes the repair process further increase the possible variation through a high rate of introduced mutation[10,11]. With the use of induced DSBs, there is also a greater risk to the genomic integrity. As a result, V(D)J recombination has been associated with many lymphomas where mutations to oncogenes leading to cancer can be directly traced back to a part of the process[12]. The DSBs are specifically associated with a high degree of chromosomal translocations, a very significant type of mutation. Beyond this system, DSB repair through eukaryotic life is still linked to mutagenesis, implicating it in afflictions such as cancer and genetic disorders[13]. Although DSBs are known to cause cancer, they are also one of our primary tools for fighting cancer, with several chemotherapeutic agents exerting their effect through DNA damage (e.g. Bleomycin) or by inhibiting DSB repair proteins (e.g. Resveratrol)[14]. The expedited growth rate of cancer cells makes them more prone to DNA damage, as a result, the cancer is often more reliant on DNA repair than regular cells as well as being more sensitive to elevated levels of DNA damage caused by chemotherapeutics.

For prokaryotes, many of the same factors come into play. Like eukaryotes, the integrity of their genome must be maintained for survival, and a DSB is still lethal. Taking advantage of this, humans have found a class of drugs called quinolones, which induce DSBs[15]. Quinolones are a direct source of DSBs as an antibiotic, however, multiple other antibiotics have been implicated in producing DNA damage through ROS production[16]. This indicates DSB repair is heavily involved in bacterial drug tolerance. In addition to this, prokaryotes also tend to use DSB repair systems as a mechanism of "inducible evolution"[8,9]. Much in the same way DSB repair systems are associated with mutations in eukaryotes, prokaryotic DSB repair is also associated with higher rates of mutation, however in bacteria this is often advantageous. This "inducible evolution" facilitates bacterial survival in stressful conditions, allowing bacteria to gain new traits, which provides a means of overcoming the stress. In the context of infection, this means antibiotic use drives the development of antimicrobial resistance (AMR) using DNA repair pathways[17]. NHEJ, which is the focus of this paper, is one such pathway. NHEJ is linked to induced evolution both through mutational rate, and through an improved ability to incorporate foreign DNA into the host genome[8,18]. These DSB repair systems will be expanded upon in greater detail throughout this chapter, however it is clear DSB repair is intricately woven into all forms of life in a critical way.

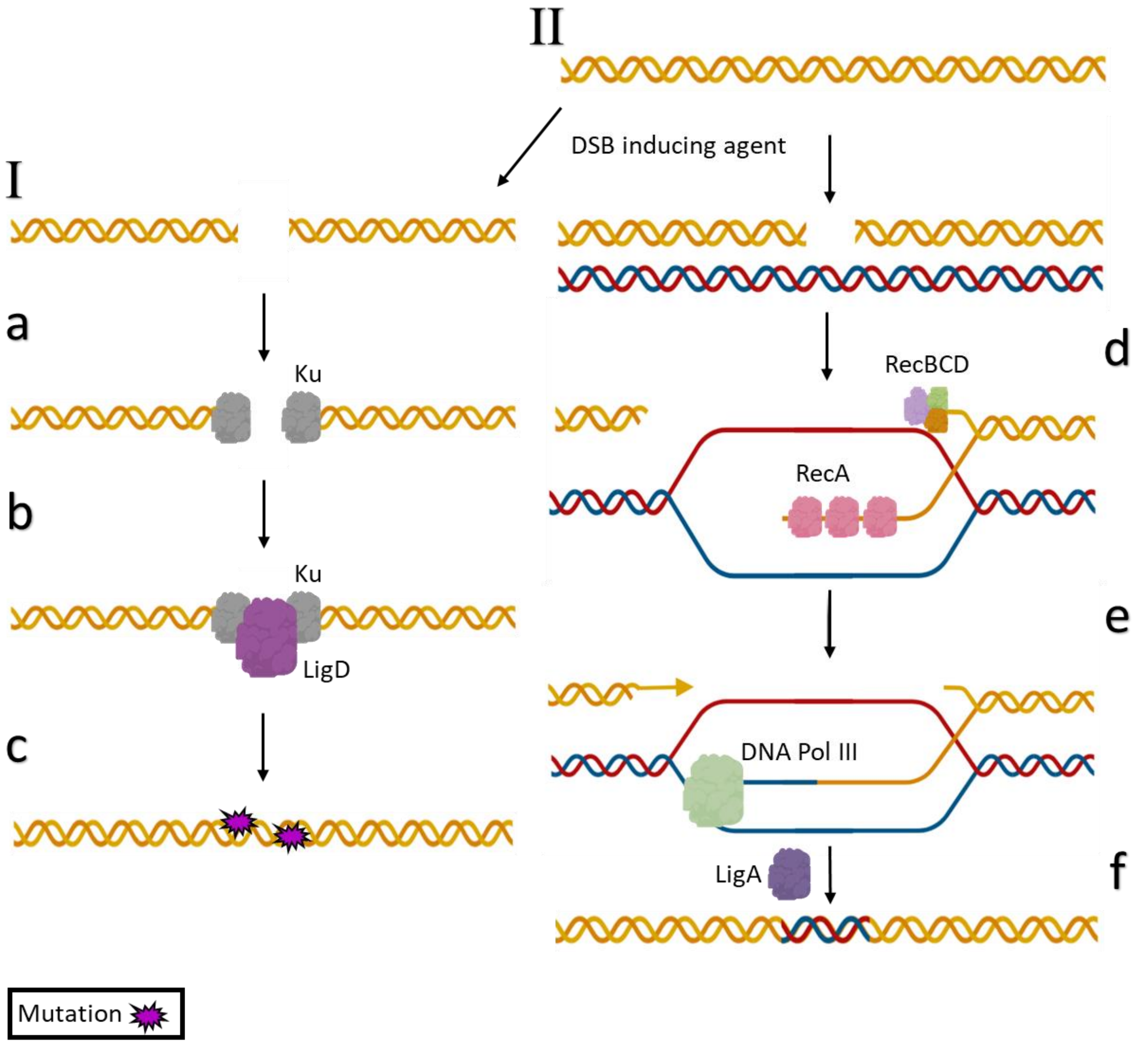


Figure 1.1. *A comparison of prokaryotic non-homologous end-joining (NHEJ) and homologous recombination (HR).* The pathways are indicated by roman numerals: (I) Non-homologous end-joining is composed of 3 primary steps: (a) end recognition of the double strand break by Ku, which subsequently binds to protect the free ends from degradation; (b) recruitment of LigD a multifunctional ligase, which processes the DSB ends by removing any potentially damaged bases and unbiasedly adding nucleotides through its nuclease and polymerase domains respectively leading to small additions and deletions; and (c) ligation of the DNA backbones and dissociation of the protein repair complex, leaving a repaired, but potentially mutated double stranded DNA. (II) Homologous recombination is composed of 3 steps: (d) DSB recognition by RecBCD followed by end resection and loading of RecA; (e) RecA mediated strand invasion of a homologous chromosome and recruitment of a high fidelity, replicative DNA polymerase, PolIII; and (f) the re-annealing of overlapping DNA ends followed by backbone ligation from LigA.

Homologous Recombination

HR, featured on the right in figure 1.1 above, is often referred to as the error-free repair pathway for DSB's due to an association with a very low mutational rate[19]. It is broadly broken down into 3 generic steps across life: the recognition and resection of the DNA DSB; strand invasion of a homologous piece of DNA; and finally, resolution of the complex formed by the extension of 1 or both strands of the broken piece of DNA on the template DNA which leads to intertwined molecules (in the case both strands are extended, this is known as a Holliday junction)[19]. The first step involves the initial recognition of a DSB followed by a resection of the 5' ended strand to leave single-stranded 3' over-hangs, which are capable of invading homologous DNA. In the model of HR for *Escherichia coli* (Fig. 1.1), the recognition/resection of the DSB is carried out by the RecBCD nuclease complex. In bacteria, the protein which fulfills the role of resecting the break is highly variable, however all of them have 5' to 3' nuclease activity, a feature conserved across life[19,20]. The second step involves binding of the 3' overhang single-stranded DNA (ssDNA) by a ssDNA binding protein (in the model RecA), which forms a nucleoprotein filament that coats the DNA. The nucleoprotein filament facilitates a homology search along an intact piece of DNA until a homologous region is identified, at which time, the ssDNA binding protein facilitates strand invasion[19–22]. Unlike RecBCD, RecA is strongly conserved in bacteria and shares homology with the eukaryotic equivalent Rad51, which is also highly conserved amongst the eukaryotic kingdom[20,21]. Once the ssDNA has invaded a homologous DNA strand, it is extended by a high-fidelity polymerase using the homologous strand as a template, adding to the overhangs until they contain significant overlap. The third and final resolution step can take place in one of two ways. The simpler method is the synthesis dependent strand annealing pathway that is outlined in figure 1.1, where a DNA overhang is extended until it overlaps the 2nd overhang. This strand then dissociates from the template DNA, and anneals to the other free broken end, becoming the template to extend the other broken end. Once the gaps are filled in, the ends are ligated back together. The second outcome is the double strand break repair pathway. This entails both overhangs being extended on the alternate strands of the homologous DNA, forming a Holliday junction. The Holliday junction can be resolved by a cross-over or non-crossover event[19].

The use of HR is limited by cell cycle since HR requires a homologous copy of the DNA to be repaired. Most species of bacteria spend most of their life in exponential growth, where a second copy of the genome is almost always present to serve as a DNA template for HR[23]. There is also evidence that HR may occur using a similar segment of DNA from the same chromosome if a second copy of the genome isn't present, however, this requires the break to have occurred within a region that is highly similar to another region of the genome[23]. Some bacteria, however, enter dormant/non-dividing states, such as *Bacillus subtilis*, which undergoes sporulation, allowing it to survive harsh conditions for up to millions of years[24,25]. These bacteria are often equipped with alternative pathways to HR, primarily NHEJ, which fulfills a much larger role during these dormant states due to the lack of a second copy of the chromosome[24,25].

Non-homologous end-joining

Once thought to be unique to eukaryotic organisms, *in silico* analysis identified homologues to NHEJ components in prokaryotes, which facilitated the discovery of NHEJ systems in approximately 20-25% of bacterial species[26–28]. In mammalian cells, NHEJ is the primary source of DSB repair and, as such, the system has been studied in depth. In prokaryotes, however, due to its later discovery and non-ubiquitous nature, NHEJ has been characterized in lesser detail[29,30].

NHEJ, featured on the left of figure 1.1, is often referred to as the error-prone repair pathway. Like HR, NHEJ can be broken down into 3 key steps: damage recognition and DNA end binding, DNA end processing, and ligation[18,29,30]. During the damage recognition step, the Ku protein binds the available free double stranded ends initiating NHEJ. Then either one or multiple enzymes which have the ability to add, or to remove, bases are recruited to "process" the DNA ends. These processing enzymes remove damaged or mismatched bases while inserting new bases to create small overlapping regions. Finally, a ligase enzyme is recruited to seal the DNA backbone completing the repair. NHEJ is known as error-prone because of a propensity for incorrectly repairing the DNA, resulting in mutations in as high as 50% of repair events[18,29,30]. Unlike HR, which requires a second copy of the genome, NHEJ can ligate a diverse array of ends back together without the need for a template DNA. It is the lack of a template that partially facilitates the high mutation rate[29,30]. The second contributing factor to the mutation rate is the DNA end processing step. Processing involves both polymerase and nuclease activity to add and/or remove bases unbiasedly. This functionality is what allows the system to handle a diverse array of DSBs, as different sources of damage can produce different DNA ends, often leaving them "unligatable". By processing the ends, NHEJ machinery returns the DNA to a state where microhomologies of as few as 2-4bp exist, facilitating ligation at the cost of small insertion and/or deletion mutations[18,29,30]. These mutations can lead to new traits if they are in genes or other functional regions of DNA. For example, as discussed above, mutations in a human may lead to the development of cancer if an oncogene is affected, while in bacteria, this may lead to a new AMR trait if a drug target becomes mutated[13,18]. Below, the eukaryotic and prokaryotic NHEJ systems are broken down in greater detail to contrast their essential proteins.

Eukaryotic NHEJ

Eukaryotic cells span a wide range of growth rates and as such, tend to vary on their reliance on NHEJ. Lower complexity eukaryotes, such as *Saccharomyces cerevisiae,* tend to rely more heavily on HR because they spend larger proportions of time in the S and G2 phase of the cell cycle where there is a homologous piece of DNA available for HR[19]. Higher complexity eukaryotes like mammals, on the other hand, tend to lean more on NHEJ since they spend more time in G_1 and G_0, so it is less likely for a homologous DNA template to be available[19]. This divide between eukaryotes has also been linked to the amount of repetitive sequence within the genome. As the amount of repetitive DNA goes up, the reliance on HR decreases because repetitive genomes reduce the efficiency of HR, increasing the risk that an incorrect template DNA will be used[19]. The first step of NHEJ involves the recognition of a DSB by the Ku heterodimer, a protein complex consisting of Ku70 and Ku80 (named as such according to their approximate molecular weight (MW))[29,31]. One Ku heterodimer binds each free DNA end quickly after a break occurs due to its high affinity for DNA. This protects the DNA ends from endogenous nucleases, which would degrade the DNA[32–34]. Ku binding is accomplished by the "core" domain, a ring like structure that threads DNA through its center. DNA binding sites are present on both Ku70 and Ku80 and coordinate binding using the major and minor grooves of DNA[32]. The Ku complex then becomes the binding hub for a plethora of other proteins involved in the repair process[29,32]. The first protein to associate with Ku is DNA-Protein Kinase catalytic subunit (DNA-PKcs), which, together with Ku, forms DNA-Protein kinase (DNA-PK), a kinase that undergoes autophosphorylation[29,32,35]. This complex is also responsible for the phosphorylation of other NHEJ proteins. In many instances, however, this phosphorylation has been demonstrated as unnecessary for repair both *in vitro*, and *in vivo*[32]. Ku also bridges the gap between DNA ends and in doing so, stabilizes the gap for the rest of the repair process[36]. Outside of these primary repair roles, there is also evidence that Ku is involved in the end-processing step that follows, demonstrating an enzymatic ability to remove abasic sites from DNA[37].

The end processing step follows the break recognition and end binding steps. This begins with the nuclease Artemis, the primary nuclease for eukaryotic NHEJ. Artemis is activated by phosphorylation from autophosphorylated DNA-PK[10,29,38]. Artemis also has a DNA-PK independent ssDNA exonuclease activity, but following activation by DNA-PK, it gains an endonuclease activity able to cut DNA at ssDNA-dsDNA interfaces, which allows for the removal of DNA overhangs, gaps and loops[10,29]. These DNA structures may arise as the result of initial damage, or due to mismatched bases at the ends; in both instances, they may impair ligation and thus require removal[29,39]. Artemis is however, not needed for every type of DNA damage. The source of the damage appears to play a key role in determining whether Artemis is used during NHEJ repair. For example, Artemis appears to have a role in the repair of damage caused by the topoisomerase II inhibitor etoposide, but not to the topoisomerase II inhibitor ICRF-193, which has a different mechanism of action[40]. As Artemis is not always used, it has been postulated that several other nucleases may be involved in DNA end processing for NHEJ including: FEN1, DNA2, APLF, and EXO1[29,41,42]. This is only half of the end processing. Additional processing includes the addition of nucleotides. This is accomplished in mammalian cells by Pol μ and Pol λ[43,44]. These polymerases are recruited to the break site through an interaction with Ku; both are promiscuous in nature, capable of inserting dNTP's as well as rNTP's, and able to function in a template dependent or independent fashion[43–46]. They do however, differ in their preference for the DNA end structure and their proficiency in template dependent/independent base addition. Pol μ is more capable of template independent addition, and accounts for most addition when no complementarity between strands is present. Pol λ conversely accounts for more of the addition when there is some base-pairing between overhangs[47]. Pol μ is more likely to introduce mutations as a result of its proficiency at template independent additions as well as its higher proficiency in using a mispaired template[47]. Together, both end resection and end extension serve as alternate and complementary ways in which the cell generates the microhomologies needed for ligation[29].

The final step in eukaryotic NHEJ repair is strand ligation, which is carried out by Ligase IV (LigIV), a eukaryotic ligase, in conjunction with the scaffolding proteins XRCC4, XLF, and PAXX. The PAXX protein interacts with Ku, localizing it to the break, while XRCC4 and XLF together, interact with the DNA[48–50]. The XRCC4-XLF components align the LigIV-XRCC4-

XLF complex and each of the DNA ends for ligation[48,50]. The formation of this complete complex stimulates the function of Lig IV[49]. The ligase then carries out ligation in comparable fashion to other DNA ligases. By using a series of three steps involving nucleotidyl transfer reactions, LigIV harvests the energy from ATP to chemically link a 3'-OH and 5'-PO_4, thus completing the repair process[51].

Prokaryotic NHEJ

As previously mentioned, prokaryotic NHEJ is found in approximately 20-25% of bacteria dispersed randomly throughout the kingdom, with a minor trend towards NHEJ-containing bacteria having slower growth rates, higher GC content, and overall larger genomes[28]. Unlike eukaryotes, most prokaryotes spend the majority of their lives in exponential growth where cells are constantly dividing, meaning a second copy of the chromosome is almost always available[20]. This means HR predominates in prokaryotes, but NHEJ is still critical to cell survival. NHEJ is particularly important for sporulating bacteria, or any bacteria that enter a dormant phase of life, like a spore[18,24,25]. During these non-dividing states, HR would have to rely on similar areas of the same chromosome for repair; while this does occur, studies using NHEJ knockout strains of bacteria that sporulate or enter dormancy, such as *B. subtilis*, demonstrate a significant sensitization to DNA damaging sources such as ionizing radiation, when compared to both wild-type and HR null mutants[24,25]. Figure 1.1 features the prokaryotic NHEJ system and demonstrates the "simpler" nature of repair compared to eukaryotic NHEJ, based on the current understanding of the system. Prokaryotic NHEJ contains only two proteins, compared to the many listed above in eukaryotic NHEJ[18,26,30]. Although there may only be two components, NHEJ proceeds through the same three steps. The two components are a bacterial Ku and a multifunctional ligase, LigD, which are discussed in greater detail below.

Ku

The prokaryotic NHEJ system employs a Ku protein, similar to eukaryotes. In contrast to eukaryotes, Ku functions as a homodimer instead of a heterodimer[26]. Sequence analysis of bacterial Ku demonstrates a high level of predicted homology to the core domain of Ku70/80, suggesting bacterial Ku forms a similar ring structure to thread DNA. Bacterial NHEJ systems were discovered by identifying probable Ku homologues in bacteria[26]. Unlike eukaryotic Ku though, prokaryotic Ku lacks domains key to interacting with all the various components of eukaryotic NHEJ, instead, there is only a unique C-terminal domain, unlike anything on Ku70/80[27]. This C-terminal domain is further broken down to a minimal C-terminus, which is highly conserved across bacterial species, as well as an extended C-terminus, which is highly variable across bacterial species (Fig. 1.2).

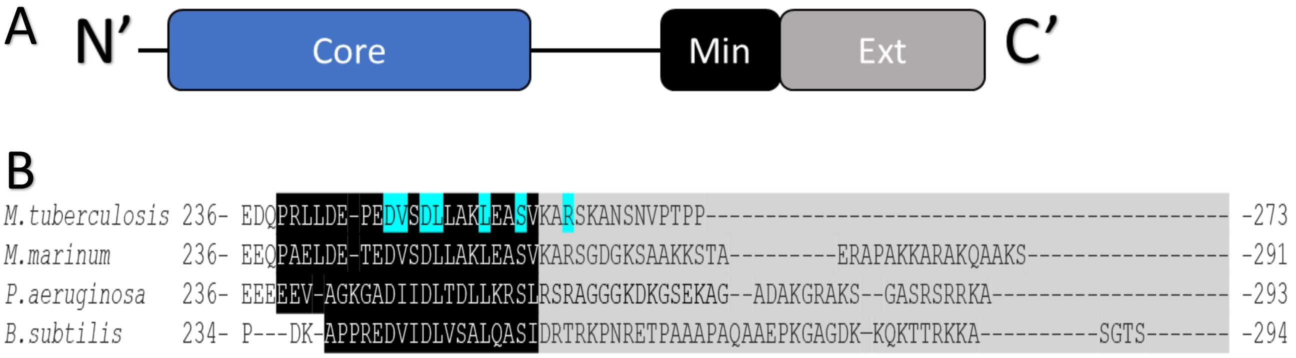


Figure 1.2. *Domains of M. tuberculosis Ku.* A) Domain architecture of the Ku protein consisting of a core domain (blue), a minimal C-terminus (Black), and an extended C-terminus (grey). B) Highlighted in black is the minimal C-terminus, while grey is the extended C-terminus. The 7 blue letters are conserved residues, which were mutated to investigate their relation to the DNA binding properties of Ku. Alignment was carried out by Clustal Omega[52].

The Ku C-terminus has been linked to a variety of functions in different bacteria, such as facilitating the binding of undamaged dsDNA in *B. subtilis* and *Mycobacterium smegmatis*, while in *Pseudomonas aeruginosa*, the C-terminus is linked to Ku's ability to bridge DNA[27,53–55]. These findings suggest that the C-terminus interacts with DNA, but may also recruit other NHEJ components, similar to the non-core domains of Ku70/80. The Ku C-terminus appears to be essential for LigD binding, specifically the minimal C-terminus[27,55]. How this domain interacts with DNA and how it interacts with LigD remain poorly understood. In fact, the entire interaction with DNA remains poorly characterized, although it is hypothesized to behave similarly to Ku70/80. Part of the issue in understanding these interactions is a direct result of the lack of a protein structure.

Like Ku70/80, bacterial Ku serves the same three roles of damage recognition, end recognition, and a recruitment factor for the other NHEJ proteins. However, less is understood about bacterial Ku's affinity for DNA, and thus about how quickly Ku is recruited to breaks. If homology modeling predictions are correct in that bacterial Ku threads onto DNA, the role of both recognition and end protection likely remain the same across the two kingdoms[26]. Unique compared to its eukaryotic counterpart, prokaryotic Ku has a completely different protein-interaction domain. Truncations of prokaryotic Ku have demonstrated the minimal C-terminus is involved in binding both DNA and LigD, however, the mechanisms behind these interactions are not completely understood[27,55]. Finally, similar to Ku70/80, evidence suggests that Ku may also be partially involved in NHEJ processing as the lyase ability to remove abasic sites is conserved[53].

LigD

Ligase D or LigD is unlike any individual enzyme in the eukaryotic system. It is composed of 3 domains: a polymerase domain (Pol), a nuclease domain (Nuc), and a ligase domain (Lig), which together encompass all the activities of the eukaryotic NHEJ proteins not named Ku[56,57](Fig. 1.3).

Figure 1.3. *Domain arrangement of M. tuberculosis LigD.* LigD is a multifunctional enzyme composed of an N-terminal polymerase domain (Pol), a central nuclease domain (Nuc) and a C-terminal ligase domain (Lig)

The Nuc domain functions in a similar capacity to Artemis, however it differs in that LigD only has 3'-ribonuclease and 3'-phosphatase activity. LigD Nuc domain has none of the endonuclease activity that is used by Artemis to remove bubbles, loops, and other potential structural flaws at dsDNA to ssDNA interfaces; instead, LigD is forced to rely on exonuclease activity[57]. The Pol domain fulfills the roles of Pol μ and Pol λ, and is likewise able to add nucleotides in a template-dependent or independent manner. It may use both dNTPs and rNTPs. Like Pol μ and Pol λ, the LigD Pol domain demonstrates a preference for rNTPs which it incorporates at a faster rate[58,59]. The Pol domain preference for incorrect insertions leads to a lower fidelity. More specifically, rNTP insertion yields a much higher mismatch rate. This, in combination with the template-independent capabilities of LigD Pol domain, leads to LigD being a very low fidelity polymerase, which is highly prone to error[58,59]. Unlike the eukaryotic polymerases, the LigD polymerase may also act as a primase. It is still unclear how this is used in NHEJ, but it has been shown that this function is accomplished using an overlapping active-site with the polymerase function[56,60]. The function of the Pol domain appears to extend beyond enzymatic activity. Purified individual LigD domains were used to demonstrate that the Pol domain interacts with Ku in a DNA dependent manner although this interaction isn't fully understood[56,57]. Data suggests LigD Pol domain also fills a role similar to PAXX, by facilitating the interaction with damage sensing Ku[56,57]. Together, these two domains accomplish the same processing functions as the multiple polymerases and nucleases in eukaryotes. The final domain in LigD is the Lig domain, fulfilling the same role as LigIV. The Lig domain uses ATP to power the chemical linkage of a 3'-OH and 5'-PO_4 using a similar set of reactions as LigIV, ultimately sealing the strands back together[56,61].

Additional NHEJ contributors

Although the two definitive components of prokaryotic NHEJ are described above, additional research suggests there may be other proteins involved as part of the regular repair pathway, or like a few proteins in eukaryotic NHEJ, used under specific circumstances, primarily dependent on the type of DNA end damage. One such protein is UvrD1, a DNA-dependent ATPase and helicase related to other UvrD helicases that are involved in nucleotide excision repair (NER) as well as mismatch repair (MMR)[62]. UvrD1 was identified in a yeast two-hybrid screen using *M. tuberculosis* Ku as bait. Following this, Δ*uvrD1 Mycobacterium smegmatis* strains were shown to demonstrate DSB sensitivity using I-SceI endonuclease induced DSBs. Finally, *M. tuberculosis* Ku was shown to stimulate both *M. smegmatis* and *M. tuberculosis* UvrD1 helicase activity *in vitro*[63]. Additional studies suggest that helicases may occasionally also be involved in eukaryotic NHEJ. For example, Sgs1 has been linked to end processing in eukaryotic NHEJ, which works alongside nucleases[42]. This sort of example in eukaryotes supports the possibility that UvrD1 is involved in bacterial NHEJ. Sir2 is another protein identified by TAP screening, which also used Ku as bait. It is a deacetylase that interacts with both Ku and LigD. Like UvrD1, Sir2 deletion in *M. smegmatis* resulted in NHEJ deficient cells as demonstrated by plasmid linearization assays. This phenotype was extended to a significant sensitivity to ionizing radiation in stationary cells, consistent with known effects of Ku or LigD deficient bacteria in dormant states[24,25,64]. The exact role of Sir2 in NHEJ remains unclear; it is believed that Sir2 is involved in the post-translational regulation of Ku. Ku is acetylated at K29 in an inversely proportional way to NHEJ activity. This suggests Sir2 may be responsible for the deacetylation that increases NHEJ activity, however, this interaction has yet to be characterized[65]. Together, these two proteins demonstrate there remains much to be learned about bacterial NHEJ, as the interactions of additional proteins with Ku suggests additional factors have yet to be uncovered.

The growing AMR crisis

Bacteria have impacted human health for as long as humans have existed. Some bacteria are beneficial while others cause serious problems for human health. It is only recently that humans have harnessed the powerful and intricate arsenal of weapons known as antibiotics. The focus of this thesis is to improve our understanding of bacteria, exposing vulnerabilities to target in the development of new antibiotics.

AMR bacteria pose a significant and escalating risk to global health. Current models predict that AMR bacterial related deaths will increase by more than 14 times the 2014 rate of 700,000 yearly fatalities, to 10,000,000 per year by 2050. Global estimates put the economic impact at more than $100,000,000,000,000 due to temporary or permanent reductions of the global workforce related to infection[66–68]. Despite these staggering numbers, these models are likely under-representative of the severity of the potential crisis; they exclude difficult to predict, indirect factors such as secondary medical effects associated with modern procedures dependent on antibiotics, and the potential effects related to human behaviour such as decreased travel, as people avoid areas which experience outbreaks[67].

The AMR crisis exists because bacteria continue to evolve new resistance mechanisms at a rate that far surpasses our ability to treat them. Combine this rapid evolution with the fact that fewer companies are actively researching and/or pursuing new antibiotics[69], and we have a healthcare disaster in the making[69–71]. In addition, antibiotic use, and over-use, drives AMR by exerting a selective pressure on bacteria. This means that as AMR worsens, there will be increased antibiotic use, leading to an increase in the development of further AMR, thus exacerbating the problem[70,72]. This possibility highlights an immediate need for new solutions to AMR, whether it be novel bactericidal or bacteriostatic compounds, ways to revitalize old antibiotics, such as synergistic drug relationships, or new treatment approaches like phage therapy. Solutions are needed sooner than later.

At first glance, this would seem to suggest that new classes of antibiotics should be prioritized, and to some extent this is true; however, a whole new approach is needed as is emphasized by the review from Tyers & Wright (2019)[73] and Silver (2007)[74]. These reviews highlight key changes the research community should be making to address the growing issue of AMR. Notably, drug development should focus on either drugs which have multiple targets, or drugs that work synergistically with other existing, or yet to be discovered, antimicrobials. Contrary to these approaches, recent focus has prioritized the design of antimicrobials that target individual gene products that are known to be essential[74]. An important trait of these new directions is that they tend to make it harder for resistance to develop, due to the multiple mutations often required to confer resistance. Antibiotics that have multiple targets are an optimal route to take, because even when resistance develops to one of the targets, antibiotics still exert an effect through another mechanism(s). This means that for a bacterium to overcome an antibiotic like this, multiple mutations providing resistance, or at least an increased tolerance, to all of the effects of the drug, would need to occur. The more targets a drug has, the less likely a bacterium develops resistance to all of its effects simultaneously and survives treatment, decreasing the chance of AMR development[73,74]. In much the same way, treating infections with drug cocktails can have a similar effect. If each component of the cocktail is acting on a different component of the cell, and doing so in a growth inhibiting manner, then this once again increases the likelihood that multiple mutations would be needed for the bacterium to survive[73,74]. The most effective drug cocktails take advantage of synergistic effects between components, i.e. the effects of the drugs in the cocktail benefit one another, improving their overall efficacy and slowing AMR development. Both methods address AMR development by decreasing the likelihood that 1 mutation could provide immunity or enough mutations could occur simultaneously that the bacterium could survive. This is what makes DNA repair such an interesting target; rather than reduce the number of possible safe outcomes for a bacterium, targeting DNA repair for inhibition effectively targets the primary means for a bacteria to develop AMR traits. This link between DNA repair inhibition and AMR development is well established with several studies demonstrating a variety of DNA repair pathways are linked to AMR development, as these repair pathways often act as sources of mutations[8,9,17,18]. Along with this promising feature of DNA repair inhibition, several studies have demonstrated that inhibiting repair pathways, such as HR, would synergize with existing drugs[75,76].

Antibiotics cause DNA DSBs

Antibiotics have numerous, distinct mechanisms of action, however only one definitively exerts its primary killing mechanism through the introduction of DNA damage: quinolones[77]. This is a class of drug that functions through inhibition of the two bacterial type II topoisomerases: DNA gyrase and topoisomerase IV[78,79]. These enzymes are involved in the supercoiling and relaxing of DNA and are able to induce transient DSBs by nicking and then repairing the DNA backbone[80]. The general model for quinolones involves stabilizing a topoisomerase-DNA-quinolone complex[81]. This complex, coined the "cleavage complex", causes several problems for the cell; specifically of note is the introduction of double strand breaks[82]. This leads to an overwhelming number of DSBs within the cell, and ultimately cell death[82]. Initially, it was believed that the cleavage complex fixed on the DNA, collided with replication machinery and caused replication fork collapse leading to DSBs[82]. This was later ruled out when replication inhibited bacteria still experienced the lethal induced DSB effects of quinolones; as a result, the exact mechanism of action is still poorly understood[72,83].

A second class of antibiotic, aminocoumarins, also inhibit DNA gyrase and Topo IV[83]. These drugs, however, do not form cleavage complexes that lead to DSBs; instead, they interfere with the ATPase activity of DNA gyrase and topo IV, which affects the amount of DNA supercoiling[84]. Interestingly, aminocoumarins have the opposite effect on DNA repair systems when compared to quinolones, despite acting on the same target proteins. Quinolones activate RecA dependent repair pathways through their introduction of DSBs. This DNA repair activation is strongly inhibited by aminocoumarins, although how is not fully understood[84].

Outside of the primary DNA damaging effects of quinolones, a second theory exists, suggesting that many antibiotics produce a secondary killing mechanism through the production of DNA damaging ROS[85]. ROS are highly volatile oxygen free radicals, which react with proteins, lipids, and importantly, DNA. Examples of ROS include H_2O_2, O_2^-, and *OH, the most damaging of which is *OH or the hydroxyl radical. The *OH molecules are extremely reactive, often limited only by diffusion, and are responsible for the worst types of cellular damage. When it comes to DNA specifically, As a result of the high reactivity of *OH, it may react with many of the atoms in either the base or the sugar moiety. Because of this, as many as 20 possible oxidized DNA products have been observed linked to *OH[86]. Among this range of outcomes are multiple backbone breaks, which lead to either single strand breaks (SSB) or DSBs, as well as DNA-protein crosslinks (DPCs), which occur when the *OH reacts with DNA in a way that primes the DNA to react with nearby proteins. DPCs will block replication, as they are large structural impediments. Stalled replication is a secondary mechanism that leads to DSBs, when replication forks fail. The *OH may also produce a wide array of lesions, such as intra-strand crosslinks or abasic sites, opening up yet more doors for DSB repair to become involved when replication forks collapse[86,87]. Some of the first work supporting this hypothesis came from Becerra & Albesa (2002)[88] , where quinolones were found to induce oxidative stress through the production of free radicals. Furthering this theory, Waddell *et al.* (2004)[89] identified multiple oxidative stress markers by examining micro-array data of *Mycobacterium tuberculosis* following exposure to several different antibiotics. Since these initial findings, numerous papers have been published on this topic that both support[85,90,91] and contest[92] the theory that ROS generation produces a secondary killing mechanism for many antibiotics.

Recently published data, in contradiction of the antibiotic-induced ROS theory, comes from Rowe *et al.* (2020)[93], where it was demonstrated that ROS produced from macrophages during an immune response actually slowed cellular metabolism and dramatically improved the tolerance to antibiotics of a *Staphylococcus aureus* infection. To show this, the experiments used mouse strains with either normal macrophages or mutated respiratory burst null macrophages. Their results indicate that when macrophages lose the ability to attack bacteria with ROS, the infecting bacteria are more sensitive to antibiotics than the bacteria infecting normal macrophages. Rowe *et al.* (2020)[93] argue that these results serve as evidence that ROS counteract antibiotics rather than synergize with them. The weakness of these conclusions though, is that they were drawn from work done by eliminating host production of ROS. In the case of respiratory burst, cells are extracellularly exposed to primarily H_2O_2, whereas antibiotics produce intracellular ROS with an increased variety of species[86,87]. These key differences may explain the observations that macrophage ROS production desensitized *S. aureus* to antibiotics while not precluding antibiotic ROS production as a means of killing bacteria, since different aspects of the cell are being affected by varied reactive compounds. This limits the conclusion that ROS species counteract antibiotic activity as the cellular exposure to the radicals produced was not biologically equivalent. Not long before, Hong *et al.* (2019)[16] developed an assay in which they show that the first-generation quinolone, nalidixic acid, produces a secondary mechanism of killing through the production of ROS. Nalidixic acid may be entirely removed from cells following quinolone exposure, thus, its primary effects may be separated from secondary residual effects, which was confirmed by mass-spectrometry. Taking advantage of this, they demonstrated that nalidixic acid produced killing effects long after it had been washed away, which could be counteracted through the addition of ROS protective agents (e.g. bipyridyl)[16]. Following removal of the antibiotic, increases in fluorescence were observed for a ROS-sensing dye, suggestive of a self-amplifying effect of ROS production. This observation was expanded to include other antibiotics from a variety of classes. It is worth noting that in all cases, the ROS effects appear to be self-amplifying. Using a YFP-tagged RecN to visualize DNA damage, Hong *et al.* (2019)[16] showed nalidixic acid exposure led to immediate DNA damage as expected. Using their assay, Hong *et al.* (2019)[16] demonstrated that ROS protective agents were able to increase cell survival without having any impact on the observed amount of DNA damage. This suggests that the primary effects of nalidixic acid (induced DSBs) were not responsible for any of the observed cell death, instead indicating that ROS plays the primary role. This phenotype is lost when tested in HR deficient mutants, thus demonstrating that DNA repair compensates for the induced DNA damage from nalidixic acid. These studies were expanded to include a second quinolone, oxolinic acid, validating that HR is compensating for the DNA damage inflicted by these two quinolones[16].

DNA repair as a tolerance mechanism

Previous studies have demonstrated that HR is an antibiotic tolerance mechanism. Hong *et al.* (2019)[16] demonstrated that both *recA* and *recBC* deletions sensitized *E. coli* to the DNA damaging effects of quinolones. All three proteins are involved in HR. RecA is also associated with the SOS-response and DNA damage cascade. Similar observations have been shown for a range of quinolones in different bacterial species where deletion of any of RecABC (or homologs of RecBC) results in a significant sensitization of the bacteria [94–96]. These findings are further supported by Wilkinson *et al.* (2016)[76] through their investigation of the phage encoded protein, Gam, a RecBCD inhibitor, which blocks DNA binding. Gam expression not only improved the efficacy of ciprofloxacin, but it was able to revert several clinical strains of quinolone-resistant *Klebsiella pneumoniae* to quinolone-susceptible[76].

Amundsen *et al.*(2009)[94] extended the importance of DNA repair to the infectivity of *Heliobacter pylori*, a stomach pathogen. This study used a library of *H. pylori* mutants of AddAB (functional homolog of RecBC) to demonstrate both that AddAB is crucial for the response to ciprofloxacin, increasing the expected survivability, and that all mutants were marginally less capable of colonizing the gut of a mouse. By using point mutations, this study isolated the 2 primary roles of AddAB as a nuclease and helicase, then demonstrated that DNA repair required at least partial activity of both roles. Amundsen *et al.*(2009)[94] hypothesized based on their results that DNA repair was necessary to cope with a host immune response[94]. This aligns with what is known about the consequences of ROS exposure for a cell[85,97,98] and supports a link between DNA repair and ROS. This link is more definitively drawn by Foti *et al.* (2012)[99], where by using *E. coli* Δ*mutT,* not only were cells sensitized to quinolones, but also to aminoglycosides and beta-lactams in an oxidized nucleotide dependent manner. *E. coli* Δ*mutT* was used because MutT is a nucleotide sanitizer that breaks down oxidized nucleotides before they can be mis-incorporated into the genome. When oxidized nucleotides are incorporated into the genome, they both directly and indirectly result in DSBs. The oxidized nucleotides act as lesions interrupting replication which may directly cause DSBs. Indirectly, this type of damage is generally repaired via base excision repair (BER), however, when too many BER events happen in close proximity, there is a significant risk of DSBs resulting from the nicks to the backbone needed for BER to proceed normally[99]. Thus, Foti *et al.* (2012)[99] show that oxidized nucleotides are elevated following antibiotic exposure to 3 different classes of antibiotics (β-Lactams, quinolones, and aminoglycosides) and that the fitness of the bacteria is directly tied to the presence or absence of MutT indicating the antibiotics exert a killing effect through oxidized nucleotides[99]. Complementary to this, it was demonstrated the overproduced MutT was protective, desensitizing bacteria to antibiotics. The study was expanded to include HR deficient mutants (Δ*recA* and Δ*recB*) which further sensitized the bacteria to damage from oxidized nucleotides. This implicates DSB generation as a mechanism of action for all 3 antibiotics, and links this DSB generation to the production of ROS[99].

This evidence all supports a strong hypothesis: DNA DSB repair inhibition synergizes with existing antibiotics. The most definitive evidence though, comes from Lim *et al.*(2019)[75] who detail the design of a small molecule inhibitor of the RecBCD/AddAB complex hypothesized to function through interaction with the DNA-enzyme complex. Using a plasmid-based GFP assay, where GFP is under the *recA* promoter, they demonstrated that a designer small molecule inhibits the SOS-response. The molecule was based around a quinolone scaffold, however, it demonstrated no quinolone activity, instead pulling the AddAB complex out of lysate in pull-down assays. Notably, this study demonstrated significant synergy between ciprofloxacin, a commonly used clinical quinolone, and their lead compound in the killing of clinically relevant Methicillin-resistant *S. aureus* (MRSA). Interestingly, neither mitomycin C nor hydrogen peroxide display synergistic effects with other DNA damaging conditions. For mitomycin C, this can be attributed to the fact that the primary repair mechanism is linked to alternative repair pathways. Mitomycin C produces DNA intra-strand crosslinks, an issue usually resolved by nucleotide excision repair (NER)[100,101]. The limited results surrounding hydrogen peroxide though, are quite puzzling, as this would appear to disagree with other studies[16,99]. However, in a similar manner to Rowe *et al.* (2020)[93], this is extracellular hydrogen peroxide exposure. This is very similar to a macrophage respiratory burst, and so it appears to support a theory that extracellular vs intracellular ROS exposure plays a crucial role in the effect on the cell[75]. This model helps explain the conflicting observations that antibiotic tolerance is tied to macrophage respiratory burst via ROS exposure and that antibiotic killing occurs, at least partially, due to ROS induction by the antibiotic[16,93].

The discovery of a RecBC/AddAB inhibitor has already proven tremendously promising and informative. This inhibitor has demonstrated the promise of DSB repair inhibition as a means to synergize with quinolones, dramatically improving their efficacy. An inhibitor such as this should, in theory, mimic the results seen in Δ*recA* and Δ*recBC.* In such strains, alternative antibiotics such as β-Lactams and aminoglycosides, which don't act primarily through DNA damage, demonstrate a DSB dependent sensitivity to these non-quinolone antibiotics. Thus, it is expected that an AddAB/RecBC inhibitor would also demonstrate synergy with the β-Lactams and aminoglycosides however this connection hasn't been explored[16,99]. This definitively demonstrates that not only is DNA repair a means by which bacteria tolerate antibiotics, but that there is indeed promise for improving the efficacy of the current arsenal of drugs with antibiotic adjuvants designed to target DNA repair.

DNA repair and mutagenesis

Genetic change is a primary mechanism for bacterial development of antibiotic resistance[102]. There is a deep connection between multiple different DNA pathways and genomic change[18,103–108]. This, together with the promise of synergy with current drugs, makes DNA repair inhibition a tantalizing target in the fight against AMR.

Several links exist between DNA repair, DNA damage tolerance, and genomic change. The most prominent link exists through induced mutagenesis. Not all repair pathways are created equal. Some, as HR mentioned above, have a high fidelity through use of a homologous chromosome for near error-free replication[109]. This same pathway, however, has also been linked with an inducible mutator phenotype following excessive activation of RecA. The excessive activation leads to upregulation of several genes involved in the SOS-response including error-prone polymerases, which funnel into, and reduce the fidelity of, HR[110]. These error-prone polymerases belong to the Y-family polymerases involved in a separate a damage tolerance pathway called translesion synthesis (TLS). TLS components are regulated by RecA as most other components of the DNA damage SOS-response in bacteria[104,105,108,110]. TLS is a process by which low fidelity polymerases are recruited to "lesions" on the DNA, which are defined as anything that the replicative polymerase cannot replicate past. These specialized polymerases have a more relaxed structure, which allows them to replicate past lesions, because they are able to insert any nucleotide, regardless of the template[72,89,103,105,108]. NHEJ, as discussed previously, does not require the use of a homologous chromosome for repair, and it is highly mutagenic due to the end-processing of LigD, which tends to introduce small addition and/or deletion mutations[18]. Together, these three systems are all potential generators of AMR through mutagenesis.

Numerous studies demonstrate that an SOS response after antibiotic exposure leads to both an increased rate of mutagenesis and AMR development across many bacterial species such as: *E. coli, S. aureus, Vibrio cholerae*, and *M. tuberculosis*[72,108,110–115]. This phenotype is directly tied to ROS produced from antibiotic exposure[111] and several genes in both TLS and HR pathways, indicative of a multifaceted approach to induced evolution. Interestingly, two entirely different antibiotic classes, specifically the antimicrobials novobiocin and amikacin, are able to relieve this inducible evolution phenotype through a reduction of available RecA[84,116]. This demonstrates promise for DNA repair inhibition as a means to slow AMR development.

NHEJ is also linked to AMR development in response to antibiotic use, particularly in stationary phase bacteria, when bacteria lack a second copy of the chromosome needed for HR, meaning

NHEJ is responsible for a greater portion of DSB repair[18,117]. In conjunction, NHEJ also promotes the integration of foreign DNA into the bacterial chromosome[8,106,118]. This is known as horizontal gene transfer, posing the second major link between DNA repair and AMR development. Similar evidence exists for a link to the SOS-response and to HR[9,103,119]. Beaber *et al.*(2004)[103] demonstrated that the SOS-response facilitates a direct, antibiotic-induced increase in recombination frequency of a multidrug resistance cassette into *Vibrio cholerae*. This effectively implicates antibiotics in stimulating the spread of AMR between bacteria in a DNA repair dependent manner. Development of AMR by HR and the SOS response is driven primarily by the release of mobile elements[103,119] and the sharing of genetic material in closely related species[9], whereas AMR caused by NHEJ, occurs because of an improved ability to incorporate distantly related or completely random DNA into the chromosome[8,118]. Meta data demonstrate that bacterial organisms that have NHEJ machinery are far more likely to have received new genetic material compared to bacteria lacking NHEJ[106].

The relationship between DNA repair and horizontal gene transfer between distantly or closely related species, demonstrates just how powerful of a tool these repair systems actually are. These are mechanisms for spreading AMR traits throughout bacterial communities, which are also driven by antibiotic use. Expanding on this, the findings of Prudhomme *et al.* (2006)[120] and Slager *et al.* (2014)[121] both demonstrate that antibiotic exposure improves a cell's ability to uptake new DNA and Overballe-Petersen *et al.* (2013)[122] demonstrate that bacteria (*Acinobacter baylyi* in the case of this study) are able to uptake ancient, damaged, fragmented DNA, and incorporate it into the genome. Collectively, this suggests that bacteria become almost sponge-like for genetic material in response to antibiotics, absorbing any available DNA with an improved ability to incorporate even damaged DNA into their genomes in a DNA repair dependent manner.

Together, the outcome of DNA repair and damage tolerance mechanisms causing genomic change, demonstrate that DNA repair systems are involved in a form of inducible evolution for bacteria as a means to overcome environmental stressors, including antibiotics. Extending this idea, work by Valencia *et al.* (2017)[116] and Schroder *et al.* (2013)[84] both demonstrate drugs that reduce DSB repair through an undetermined mechanism via an effect on RecA (amikacin and novobiocin), that were able to halt genomic change associated to ciprofloxacin treatment. If harnessed, this possibility of inhibiting DNA repair, and its impact on genomic change, may be used to slow the global AMR crisis and provide more time to develop a larger arsenal of antibiotics. It would also improve individual patient outcomes when taken in conjunction with existing antibiotics by reducing the likelihood of an AMR trait developing *de novo* during treatment[17]. This theory requires more research though, as the mechanism by which two very different drug classes (amikacin, an aminoglycoside, and novobiocin, an aminocoumarin), inhibit the SOS response when used with ciprofloxacin is not understood. Understanding the mechanism of this inhibition could expand our ability to inhibit AMR development through DNA repair. Unfortunately, these promising interactions have only been observed for ciprofloxacin. Ideally DNA repair inhibition wouldn't be reliant on a separate specific drug, instead it should work with the entire spectrum of antibiotics[84,116]. Beyond these 2 examples of inducible evolution inhibition, both bacterial NHEJ and TLS pathways are promising targets. Given their pronounced mutational frequencies, both systems are intricately linked to the concept of inducible evolution[8,102].

Project Summary

The thesis presented herein focuses on improving our understanding of the bacterial Ku protein from a biochemical and structural perspective. As a key component in non-homologous end-joining, Ku is an optimal target for small molecule inhibition to combat AMR for two reasons: first, as discussed above, evidence suggests DNA repair inhibition synergizes with modern antibiotics[16,75,76]; and second, DNA repair acts as a proverbial well for differentiation through the introduction of mutations into the genome, which is particularly relevant to NHEJ as a highly mutagenic pathway[18,107,110–115].

As iterated above, in the search for new antimicrobials, focus needs to be altered to find antimicrobials that make it difficult for bacteria to develop resistance. Both Tyers & Wright (2019)[73] and Silver (2007)[74] highlight ways to do this using molecules that affect multiple targets separately, bind at a site involving 2 or more proteins, or work synergistically with another existing drug. As evidenced by Lim *et al.* in (2019), designing an inhibitor of HR works synergistically with an existing drug class of fluoroquinolones[75]. The majority of the literature currently focuses on HR and the SOS-pathway when studying a link between antibiotics and DNA repair, while leaving a void regarding NHEJ's involvement. This void is significant because HR inhibition has already shown promise, yet it is not always biologically relevant. The requirement of HR for a homologous template excludes it as an option during dormant, non-dividing states. As such, various studies have demonstrated that dormant or sporulated bacteria have an increased reliance on NHEJ[24,25]. This suggests NHEJ inhibition may be critical to treat bacterial infections during these states. Dormant bacteria are also often very hard to kill, so infections like *M. tuberculosis*, which can lie dormant in host macrophage cells for years, become increasingly difficult to treat[123]. NHEJ inhibition may also be of further interest as a way to synergize with HR inhibition, as NHEJ has been shown to compensate for HR deficiencies in some instances[30]. Even when not considering the potential of DNA repair inhibition to kill bacteria as a treatment, inhibiting DNA repair carries tremendous appeal as a means of slowing the rate of AMR development. This is particularly true of NHEJ, which is linked to a very high mutational rate of up to 50%[18], and a high rate of horizontal gene transfer that leads to an increased likelihood of integrating distantly related foreign DNA[106]. These traits make NHEJ inhibition an optimal target. However, before drug design, the void of information surrounding NHEJ must be filled. That is why this project will take a step in that direction by focusing on characterizing bacterial Ku from a structural and biochemical perspective with an emphasis on its interaction with DNA.

Chapter 2: Understanding the Ku-DNA interaction

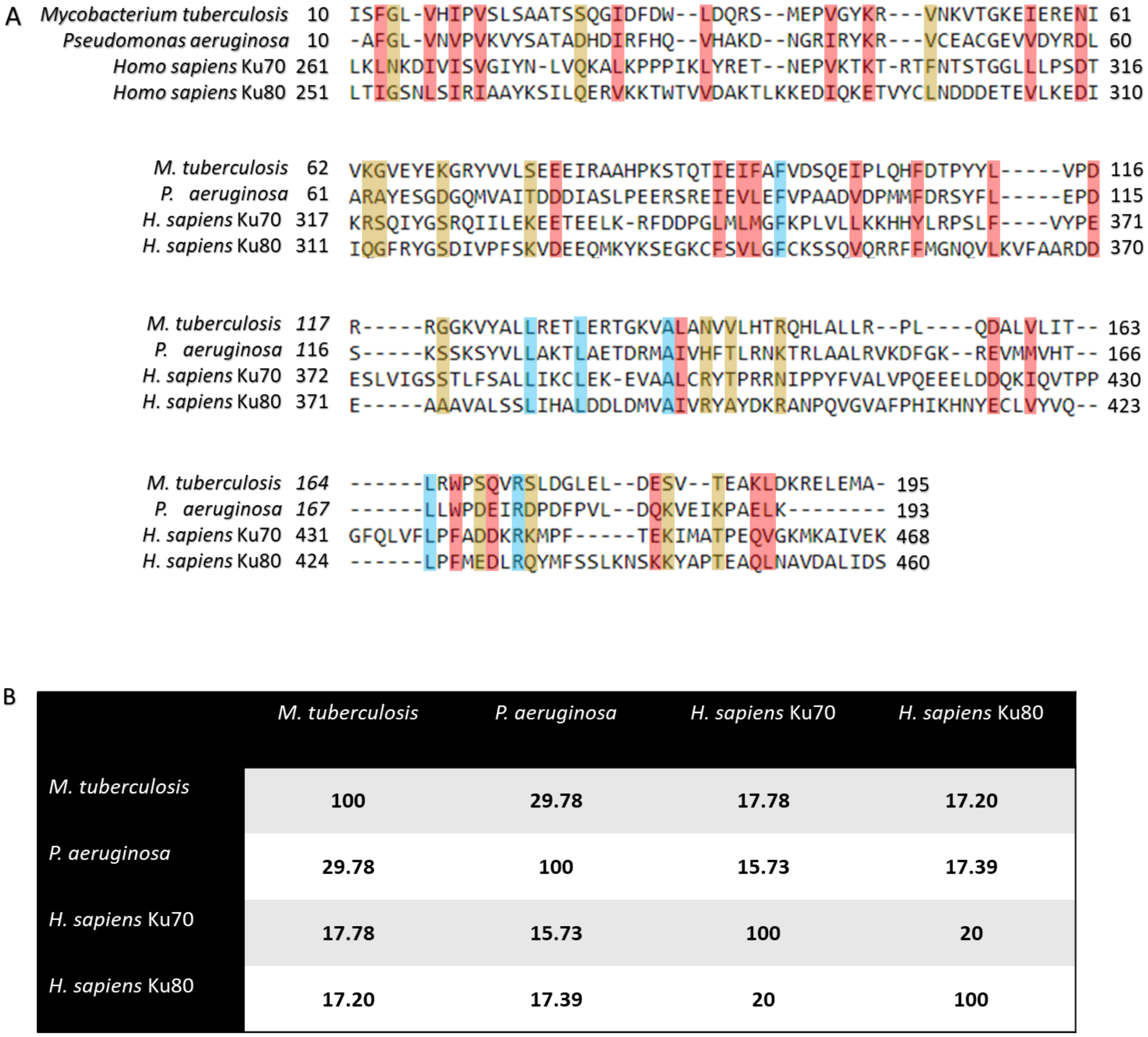
Project overview:

Project rationale:

The NHEJ pathway is a highly mutagenic and poorly understood pathway in bacteria for repair of the devastating and lethal DNA DSB[26]. The combination of high mutagenicity and NHEJ's role as a critical DNA repair pathway makes NHEJ inhibition a promising target for antibiotic design. However, a dearth of information exists for the mechanics of NHEJ in prokaryotes. The reason Ku is specifically intriguing as a drug target is multifaceted. Firstly, it is essential to the repair process by acting both as the recruitment hub, and as the DNA binding component[26,27]. Not only is Ku essential, but its role necessitates multiple interaction interfaces, which could all potentially serve as inhibition targets, because it interacts both with DNA and other proteins[27]. Adding to this, Ku is needed in every NHEJ repair instance, unlike the two processing domains of LigD, which may or may not be necessary, thus Ku inhibition would likely lead to total NHEJ inhibition rather than partial suppression[28]. Finally, bacterial Ku contains a unique C-terminal domain, not seen in other eukaryotic homologs, which would limit the likelihood of off-target effects if targeted. The goal of this thesis is to investigate one of these Ku interfaces, by improving our understanding of the molecular mechanisms of Ku's interaction with DNA. I achieved this by multiple mechanisms: first, by investigating Ku DNA-binding efficiency using a library of Ku mutants that demonstrated the C-terminal domain regulates DNA binding affinity; second, by a preliminary structural investigation, which produced protein crystals that will be further optimized for a crystal structure; and finally, by carrying out an initial biophysical analysis of Ku using AFM that determined the AFM volume of Ku, which will be used further in characterizing Ku behaviour in the presence of DNA. The work presented herein is centered on understanding the specific roles of Ku and its functional domains and the important traits of these domains, which facilitate their roles. In doing so, these studies will set the foundation for targeted inhibitor design.

Background:

Non-homologous end-joining, discussed in detail in chapter 1, is a DSB repair process which attaches two broken DNA ends back together without the need for a DNA template. The Ku protein is essential to this system in both the eukaryotic and prokaryotic versions of NHEJ, and although the core domain appears to be conserved, the peripheral Ku structural domains are entirely different (Fig. 1.2B). The structure of human Ku is solved, but the prokaryotic structure remains undetermined. Thus, assumptions that are made using human Ku with regard to its bacterial counterpart, may be misleading[26,31]. Even the core domain, which was used to identify bacterial NHEJ[26], displays higher levels of homology between bacterial species than with eukaryotic Ku (Fig. 2.1). This discrepancy suggests there may be key differences between prokaryotic and eukaryotic Ku in structure and function yet to be elucidated.



	M. tuberculosis	*P. aeruginosa*	*H. sapiens* Ku70	*H. sapiens* Ku80
M. tuberculosis	**100**	**29.78**	**17.78**	**17.20**
P. aeruginosa	**29.78**	**100**	**15.73**	**17.39**
H. sapiens Ku70	**17.78**	**15.73**	**100**	**20**
H. sapiens Ku80	**17.20**	**17.39**	**20**	**100**

Figure 2.1. *Conservation of human and bacterial Ku core domain.* A) Sequence alignment for the core domain of *M.tuberculosis* and *P.aeruginosa* Ku with Ku70 and Ku80 from *H. sapiens*. Highlighted residues are conserved. Gold are mildly conserved, red are highly conserved and blue are identical. B) Sequence similarity between pairs of Ku core domains. Similarities are represented as percentages. Alignments and calculated sequence similarities were carried out by Clustal Omega[52].

The core domain of prokaryotic Ku is accompanied by a unique C-terminal domain, which is split into two sub domains, the extended and the minimal C-terminus based on the former being highly variable, and the latter highly conserved amongst bacterial species[27]. Figure 1.2B highlights the conservation between 4 different bacterial Ku C-terminal domains. The C-terminal domain is of particular interest because it is unique from its eukaryotic counterpart, as well as it is linked to a range of different functions. Functions include being the recruitment hub for LigD[27,55], bridging multiple DNA ends[55], and binding undamaged DNA[53,54]. Increasing the intrigue associated with the C-terminus, these functions are not consistent across species, a phenomenon currently not fully understood.

The means by which this small domain accomplishes the diversity of roles that have been attributed to it may be related to a predicted level of disorder within the C-terminus. Figure 2.2 shows the disorder prediction for Ku (D^2P^2,[124]), which displays a high level of predicted disorder towards the C-terminus. In all 4 bacterial species the C-terminal domain is predicted to be heavily disordered. While the minimal C-terminal domain may have some structure, the entirety of every extended C-terminal domain is predicted to be fully disordered. Disordered proteins are known for their ability to interact with an array of other molecules, serving a diverse set of functions analogous to what is known about the C-terminal domain in Ku[125]. Since the C-terminal domain has been linked to a wide array of roles, if an inhibitor were designed to target this region, it should impair Ku's function in multiple ways. Considering that the C-terminal domain is unique to bacteria, it may be used to inhibit bacterial Ku in a highly specific manner. Together this suggests that the C-terminal domain is a promising region for Ku inhibition. However, without a structure and a better understanding of how this domain is involved in Ku function, such as how it interacts with DNA and LigD, the Ku C-terminal domain cannot be exploited as an antimicrobial target.

As an essential component of NHEJ[25], Ku is an optimal target for NHEJ inhibition. Since NHEJ is associated with a high degree of genetic change[18,30], a trait linked to the development of AMR, and a role in DSB repair, notably increased in dormant bacteria[24,25], NHEJ is a promising target in the fight against AMR. Thus, Ku is an ideal target to battle AMR, however before it may be targeted, it must be characterized in greater detail.

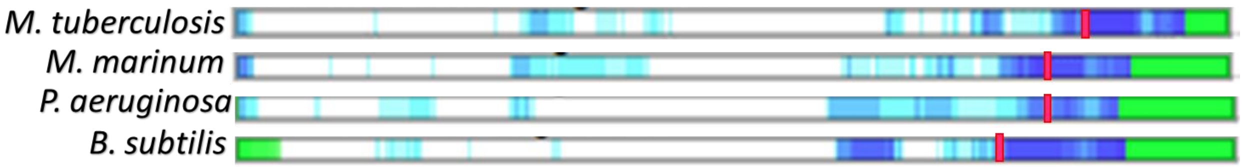


Figure 2.2. *Disordered region prediction for bacterial Ku*. The bars represent the entirety of each bacteria's Ku gene; the colour gradient represents order to disorder prediction across D^2P^2's pooled prediction software. White is unanimously predicted to be ordered while green represents unanimously predicted disorder. In between are the blues, lighter blue was predicted to be disordered by fewer algorithms than darker blue regions. The red lines indicate the start of the C-terminal domain. Disorder prediction analysis by D^2P^2 online software[124].

Research Synthesis Overview

The C-terminal domain of Ku is shrouded in mystery, as it is associated with a range of functions that differ amongst bacteria. To better understand the role of the C-terminal domain, I investigated its part in the fundamental Ku function of DNA binding. For my work, I used Ku from *M. tuberculosis* for two reasons: first, because *M. tuberculosis* is a global problem pathogen responsible for 1.5 million annual deaths as of 2018; and second, because *M. tuberculosis* is known to enter dormant non-dividing states, suggesting NHEJ may be particularly important to *M. tuberculosis* survival[18,126]. To study Ku-DNA binding affinity, I incubated a range of Ku concentrations with a set concentration of fluorescently labelled dsDNA, and analyzed the electrophoretic mobility shift of DNA alone compared to Ku-DNA complexes, to establish a dissociation constant (K_d). Insight into the role of the minimal and extended C-terminal domains, as well as important residues in the DNA binding function of Ku were assessed by using a library of Ku mutants, including truncations of either the extended C-terminus or the entire C-terminus, as well as point mutations at highly conserved residues (Fig. 1.2B). These studies were complemented with initial structural work in X-ray crystallography and small angle X-ray scattering (SAXS), each providing different resolutions of Ku-DNA interactions. X-ray crystallography provides a higher resolution structure, while SAXS provides insight on dynamic regions at lower resolution. Initial biophysical insights into the Ku interaction with DNA through atomic force microscopy (AFM) were planned, however at this time, only preliminary data has been obtained. Finally, I began creating a library of *P. aeruginosa* mutants, including Δ*ku*, Δ*ligD*, Δ*recB,* and Δ*ku*Δ*recB* to study how NHEJ deletion strains compared to HR deletion strains. Each strain will be tested in a variety of DNA damaging stress conditions such as ciprofloxacin or UV exposure, as well as with other antibiotics to investigate the role of DSB repair as a potential tolerance mechanism to all antibiotics, related to possible ROS production[16,99]. By deleting both NHEJ proteins individually, the goal is to establish whether inhibition of Ku or LigD is more effective at blocking NHEJ repair. Deleting *recB* would provide a strain deficient in HR, to compare the outcomes of NHEJ inhibition to HR inhibition, and examine the potential synergistic effects of inhibiting both primary DSB repair pathways in the double mutant strain. Additionally, *ku* deletion strains can be complemented with the library of *ku* mutants to identify phenotypes associated with specific functions of the Ku C-terminal domain. Unfortunately, COVID-19 has prevented some of this work from extending beyond preliminary results for this thesis.

Results

Ku DNA binding properties

To examine the DNA binding capabilities of Ku, and understand the important features of Ku for this interaction, electrophoretic mobility shift assays (EMSAs) were conducted using wildtype Ku, the Ku truncations, and the Ku point mutations, purified in figure 2.3. The domain architecture for each construct can be seen in figure 2.4E. Point mutations were based on conserved residues of the minimal C-terminus or extended C-terminus. Truncations of the C-terminal domain include ΔExt C-term, which is the Ku core domain plus the minimal C-terminus, and ΔC-term, which is the core domain of Ku (Fig. 2.4E). All proteins were expressed in *E. coli* and purified using nickel affinity chromatography, anion exchange chromatography, and finally size-exclusion chromatography. These studies determine the DNA binding affinity for each protein by measuring the ratio of bound DNA to unbound DNA across a range of protein concentrations. Representative results of DNA binding for each protein variant are seen in figure 2.4. As the Ku concentration is increased, there is a shift in mobility of the fluorescently-labelled DNA from a fast mobility (lower band) to a slower mobility (higher band), indicative of Ku binding DNA. All assays were run in triplicate, with the average fraction of DNA bound plotted for each variant (Fig. 2.4B,C). The calculated K_d for each protein is graphed in figure 2.4D, while figure 2.4E provides a visual representation of each protein beside the calculated K_d. This data shows that the highest K_d, and thus the weakest binding, actually belongs to WT Ku at $K_d = 3891 \pm 1669$nM. Some of the point mutations appears to have a small effect on DNA binding, however these were found to be statistically insignificant, limiting the conclusions to be made from L251A $K_d = 1302 \pm 193$nM and S258A $K_d = 1337 \pm 288$nM. Interestingly both truncations seemed to have similar effects to one another. The ΔExt C-term Ku ($K_d = 612 \pm 99$nM) and the ΔC-term Ku ($K_d = 508 \pm 228$nM) both lowered the K_d to similar degrees when compared to WT, as did Ku L255A ($K_d = 613 \pm 88$nM). These 3 Ku mutants all have K_d values ~6-8 fold lower than WT Ku, indicating a tighter Ku-DNA interaction, suggesting that components of the C-terminal domain destabilize the Ku-DNA complex, rather than stabilize Ku-DNA binding as hypothesized. Both mutated charged residues, D250A and R262A, produced the largest increase in affinity for DNA, D250A ($K_d = 142 \pm 8$nM) and R262A ($K_d = 135 \pm 12$nM), with an ~27-29-fold increase. These results are significant for multiple reasons: first, since R262 is in the extended C-terminal domain, and D250 is in the minimal C-terminal domain, it conclusively shows that the entire C-terminal domain is involved in binding DNA. The effect of these two mutations also provides insight on the interaction surface between Ku and DNA. Since the mutation away from charged residues leads to an increase in affinity, it may be hypothesized that these residues are in a non-polar pocket, such as the major or minor grooves of the DNA. Within the grooves of DNA lie the non-polar bases, and these would form unfavourable interactions with either aspartate or arginine due to their charged particles, destabilizing the interaction, while the switch to alanine would reverse the effect, forming favourable interactions with the non-polar bases. This model is supported by knowledge of eukaryotic Ku binding DNA, which coordinates its position on DNA using the grooves of the DNA[31]. Finally, these results suggest that there may also be residues within the C-terminus which are aiding the Ku-DNA interaction, balancing out the overly negative effects seen in relation to the two polar residues, D250 and R262. This conclusion is based on the observation that these 2 polar to non-polar mutations lead to a bigger increase in DNA binding affinity than removing the entire C-terminus. Thus, it is likely that in removing the entire C-terminus, some positive interactions are being lost along with the negative ones like D250 and R262 explaining why D250A ($K_d = 142 \pm 8$nM) and R262A ($K_d = 135 \pm 12$nM) affinities for DNA are so much higher than ΔC-term Ku ($K_d = 508 \pm 228$nM).

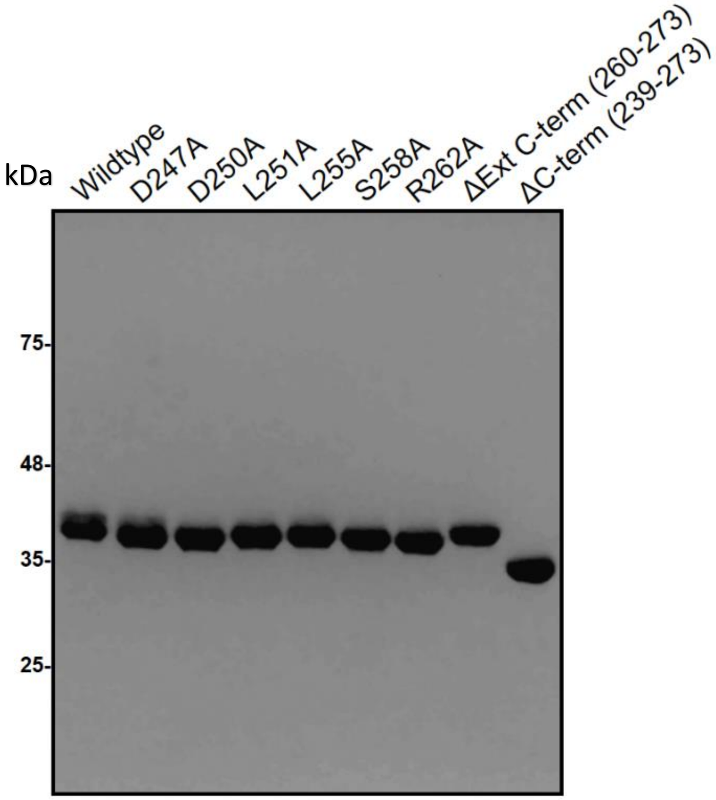
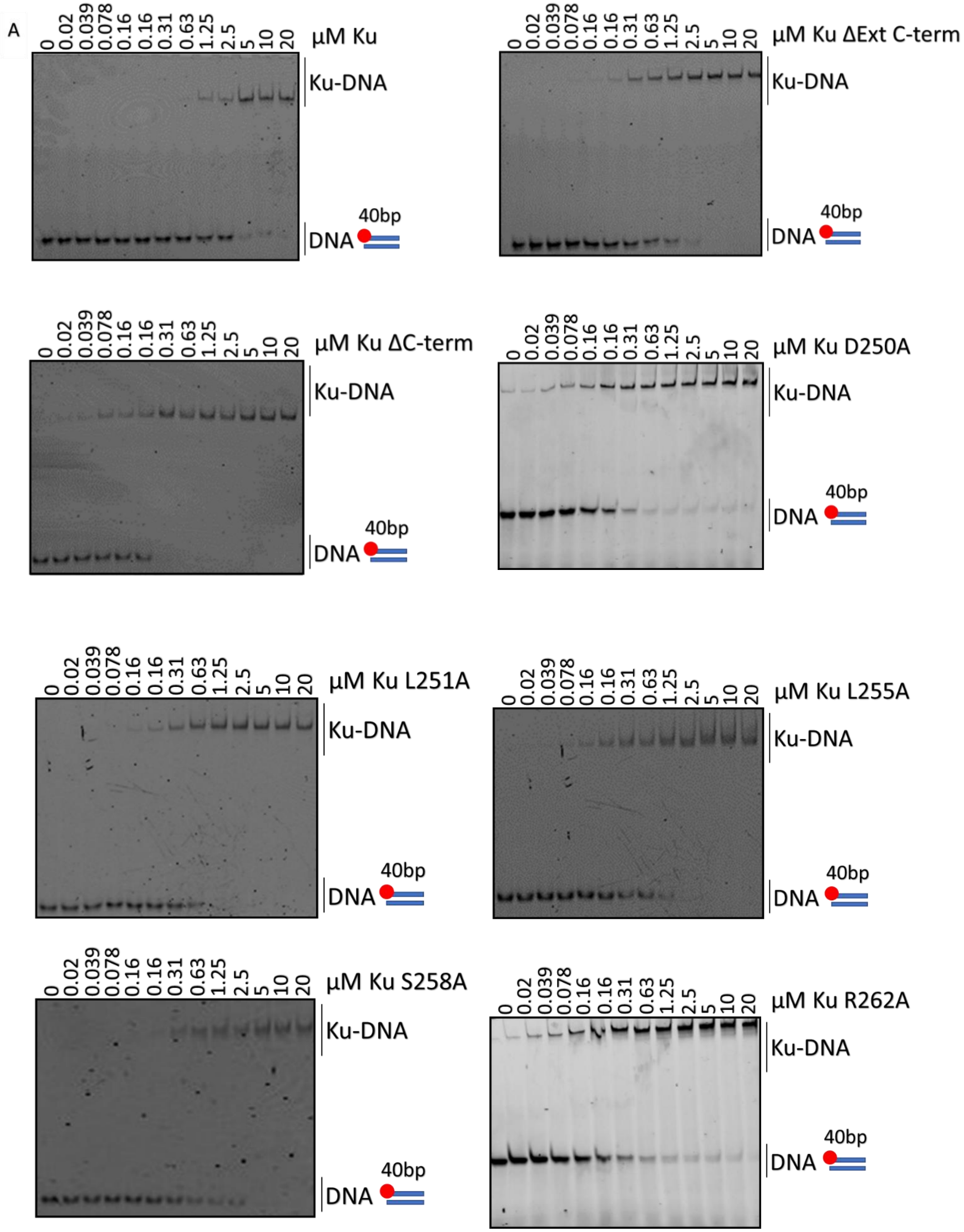


Figure 2.3. *Purified M. tuberculosis Ku used throughout this study*. Molecular mass markers are featured on the left. Lanes are labelled with the Ku construct purified (Figure courtesy of Cody Caba).


A
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku ΔExt C-term
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku ΔC-term
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku D250A
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku L251A
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku L255A
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku S258A
Ku-DNA
DNA
40bp
0 0.02 0.039 0.078 0.16 0.16 0.31 0.63 1.25 2.5 5 10 20
µM Ku R262A
Ku-DNA
DNA
40bp

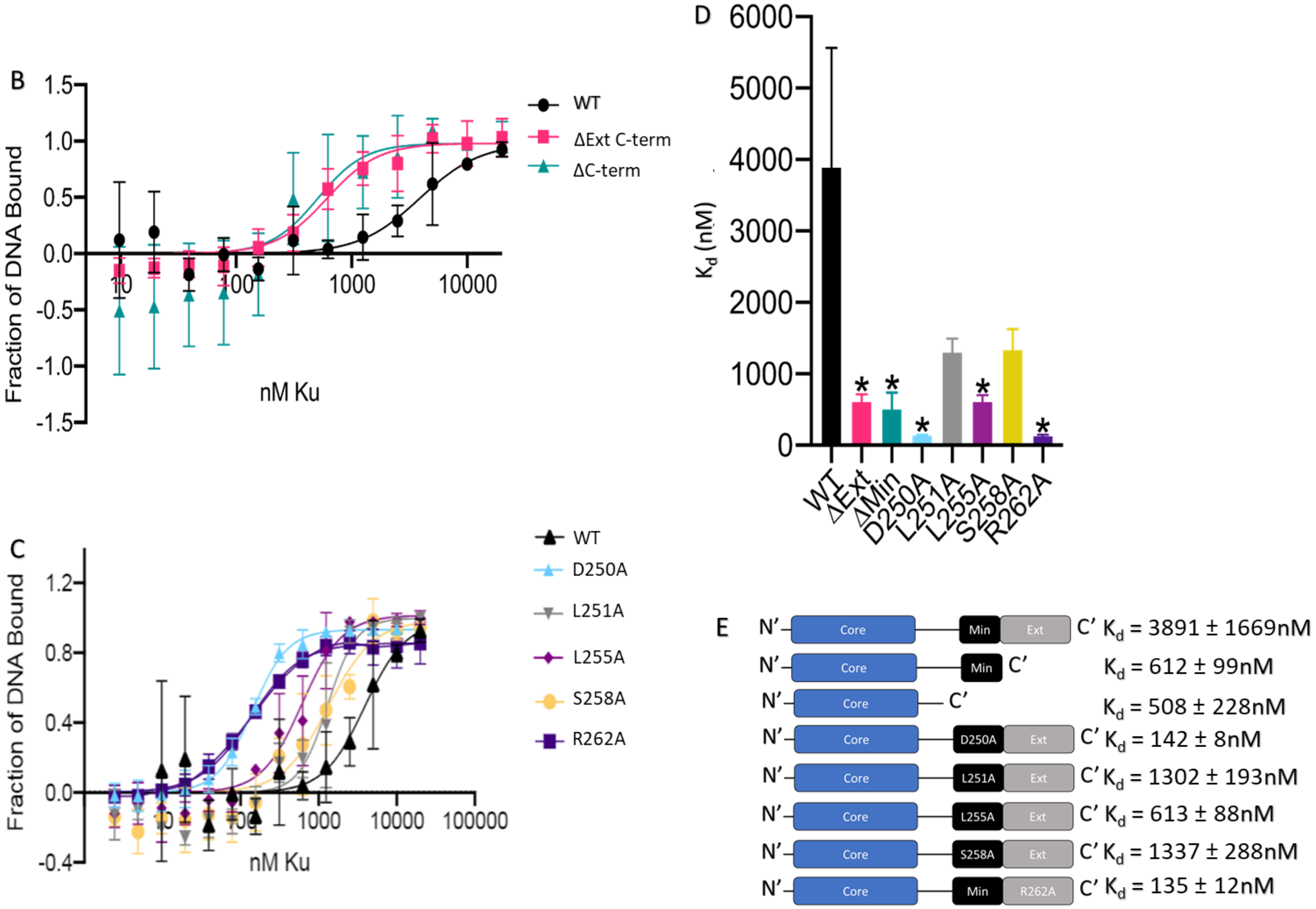


Figure 2.4. *Altering the Ku C-terminal domain stabilizes the Ku-DNA interaction.* A) Representative images of electrophoretic mobility shift assays for bacterial Ku and each of the C-terminal mutants binding 40bp dsDNA B) DNA binding activity of Ku WT and Ku truncations determined using n=3 independent experiments C) DNA binding activity of Ku WT and Ku point mutants determined using n=3 independent experiments D) Calculated Ku-DNA dissociation constant (K_d) on 40bp dsDNA for each Ku construct determined using n=3 independent experiments.*,$p<0.1$ compared to WT (Welch's two-tailed t-test).

Given the unique nature of the C-terminal domain of Ku, compared to eukaryotic Ku, I focused my attention to this region. Surprisingly, these results suggest the C-terminal domain of Ku plays a role in destabilizing its association to DNA. This observation was somewhat unexpected. Previous work demonstrated that the C-terminal domain is important for DNA bridging and binding undamaged dsDNA[53,54]. This indicates that the C-terminus binds DNA in some circumstances, but the results of this assay indicate that regarding the formation of a Ku-DNA complex, the C-terminus destabilizes the interaction. These results partially agree with the previous work of Zhu et al. (2010)[55], who observed a similar feature for the C-terminus in *P. aeruginosa* Ku. The work on *P. aeruginosa* Ku examined DNA binding affinity of a library of Ku truncation mutants and concluded that the C-terminus had destabilizing properties. This conclusion was based on an increased binding affinity seen in their first 4 constructs, the shortest being 1-253 (see figure 1.2B for reference). Contrary to these results for *M. tuberculosis* Ku, the destabilizing effect Zhu et al. (2010)[55] observed in *P. aeruginosa* Ku was limited to the extended C-terminus as well as approximately half of the minimal C-terminus. The 1-240 and smaller truncations demonstrated a significantly decreased DNA binding affinity, vastly different than the significantly increased affinity seen in the very similar *M. tuberculosis* ΔC-term Ku (1-238), which is curious. These differences may be related to using different species, as seen before regarding other traits for Ku that change between species. In this instance, the combination of figure 1.2B and 2.1A provide a near complete comparison of *P. aeruginosa* and *M. tuberculosis* Ku. Figure 2.1B shows there is a high level of homology within the 2 core domains, but figure 1.2B show a lack of homology in the C-terminus, suggesting that observed differences are likely a result of differences in the C-terminal domain. Notably, there is a high number of basic residues in the extended C-terminus of *P. aeruginosa* Ku not present in *M. tuberculosis* Ku, which may be affecting the DNA binding properties. These observations could also be a result of differences in the DNA substrate used. While both studies observed DNA binding by EMSA, this work on *M. tuberculosis* Ku used a short, 40bp dsDNA, while the work in *P. aeruginosa* used linearized plasmid (~50x the size). Regardless, both results support a theory that maintaining a lower affinity for DNA is important for the Ku protein.

The process of NHEJ repair may explain the requirement for a lower affinity for DNA. Based on the models formed from eukaryotic Ku, prokaryotic Ku is predicted to function as a ring threading onto DNA[26,31]. It is currently not understood how Ku is removed from the continuous strand following the ligation step. Whether it is through degradation or some other means though, it could be stuck on repaired DNA for some period of time where it would interfere with other cellular proteins which require access to the DNA (e.g. transcription factors). This evidence supports the theory that by maintaining a low affinity for DNA, it is more likely that other DNA binding proteins will be able to displace Ku from the DNA. More importantly, this model explains the moderation of DNA binding affinity by the minimal and extended C-terminal domains. Based on this model, if the Ku C-terminal-DNA interaction could be interrupted via an inhibitor, then the Ku-DNA complex would become tighter. This may interfere with other cellular processes such as replication or transcription while Ku blocked access to DNA. This would have a predominantly negative effect on the cell, which suggests this may be an interesting site to target for drug design. This type of mechanism has already proven useful as quinolones exert their effect by stabilizing topoisomerase-DNA complexes suggesting Ku inhibition in this way may generate similar effects.

Structural characterization of Ku

The Ku C-terminal domain is critical for regulating DNA binding, but how it does this is unknown. To gain a better understanding of the DNA-binding regulation mechanism of Ku, two structural approaches were used: first X-ray crystallography in an effort to capture a high resolution structure of the protein providing intricate insight into how Ku bind DNA and second, SAXS, which produces lower resolution structures, but is better able to capture dynamic regions of the protein. This method should be more effective for studying the predicted disordered C-terminus. For both X-ray crystallography and SAXS, *M. tuberculosis* WT Ku as well as the two Ku C-terminal truncations, ΔExt C-term and ΔC-term were investigated. Structural disorder is known to prevent protein crystallization[127], therefore Ku truncations lacking the predicted disordered region were used for X-ray crystallography. Disordered regions are known to be very dynamic with a high degree of mobility/flexibility[125,127]. Crystallography requires the protein to adopt the same conformation (or at least nearly identical conformations) to crystallize since crystals are just repeating identical units of the protein. This trait of identical repetition is key to the diffraction of x-rays and x-ray crystallography. The greater the uniformity in a crystal, the stronger the diffraction and higher the resolution. Since disordered proteins are more flexible and mobile, it decreases the likelihood that enough proteins in a crystal will adopt the same conformation, making it very hard to crystallize disordered proteins. Even when a protein is crystallized, a disordered region may still pose an issue if it is adopting a slightly different conformation in each unit cell of the crystal because this will significantly decrease the diffraction quality of the crystal[127].

Ku crystallization and X-ray crystallography

Initial broad screening of WT and both truncated Ku proteins used commercially available screens (Table 2.1) loaded into Intelli 3-well plates (Art Robbins Instruments) coupled with varying concentrations of Ku protein, dependent on the construct. In general, a low, medium and high concentration were screened against each precipitant. Each broad screen was duplicated and incubated at 25ºC or 4ºC; however, it was quickly observed that only screens at 25ºC produced initial crystal hits. Therefore, crystallization trials continued exclusively at 25ºC.

Table 2.1: *Broad screen crystallization trials*. The names of the broad screens are across the top, the proteins down the left. NS means not screened, while a temperature (in Celsius) indicates an intelli 3-well plate plated with the corresponding protein and broad screen, incubated at that temperature. The +20bp DNA is a linear 20bp oligonucleotide while the +HP DNA is a designed oligonucleotide which forms a hairpin secondary structure (see Figure 2.7).

Protein	Qiagen NeXtal Tubes Classics Suite I	Qiagen NeXtal Tubes Classics Suite II	Qiagen PEGS II	Hampton Crystal Screen 1 + 2	Hampton Crystal Screen Lite + Hampton Natrix
WT Ku	25º, 4º	25º, 4º	25º, 4º	NS	NS
Ku ΔExt C-term	25º, 4º	25º, 4º	25º, 4º	25º	25º
Ku ΔC-term	25º, 4º	25º, 4º	25º, 4º	25º	25º
Ku ΔC-term +20bp DNA	25º	25º	25º	25º	25º
Ku ΔC-term +HP DNA	25º	25º	25º	25º	25º

Visual inspection of the trays identified multiple conditions producing some level of crystallization (hits) for both truncation mutants, however no hits were observed for WT Ku. Initial hits were optimized by altering the conditions to produce crystals that diffract at high-resolution. Optimizations included altering the pH of the buffer, the concentration of the precipitant and the salt, and through the use of additive screens, which are screens consisting of small molecules known to aid in crystallization, that are added to the hit condition . The optimized conditions were scaled up to larger drop volumes over larger well volumes and refined to produce single crystals large enough to test for diffraction. The protein crystals for Ku ΔC-term took an average of 1 week from plating until they reached maximal size, while the Ku ΔExt C-term crystals took an average of 8-10 weeks. Figure 2.5 shows representative images of crystals for both Ku truncation constructs that were screened for X-ray diffraction on the McMaster home source. X-ray diffraction was not observed for crystals of either truncation. The lack of an observable diffraction pattern informs that these are likely protein crystals. The alternative, salt crystals, almost always produce a diffraction pattern in an extremely characteristic way with very spaced out diffraction spots resulting from the small unit cell of salt crystals[128]. To further validate the crystals were protein, a crystal violet dye was used (Fig. 2.6). Protein crystals have large solvent channels that the dye is taken up into through diffusion. As the dye penetrates deeper into the crystal it becomes harder to diffuse out, concentrating the dye within the crystal as compared to the surrounding solvent, staining the crystal purple[129]. Therefore, crystals that stain a deep purple are presumed protein.

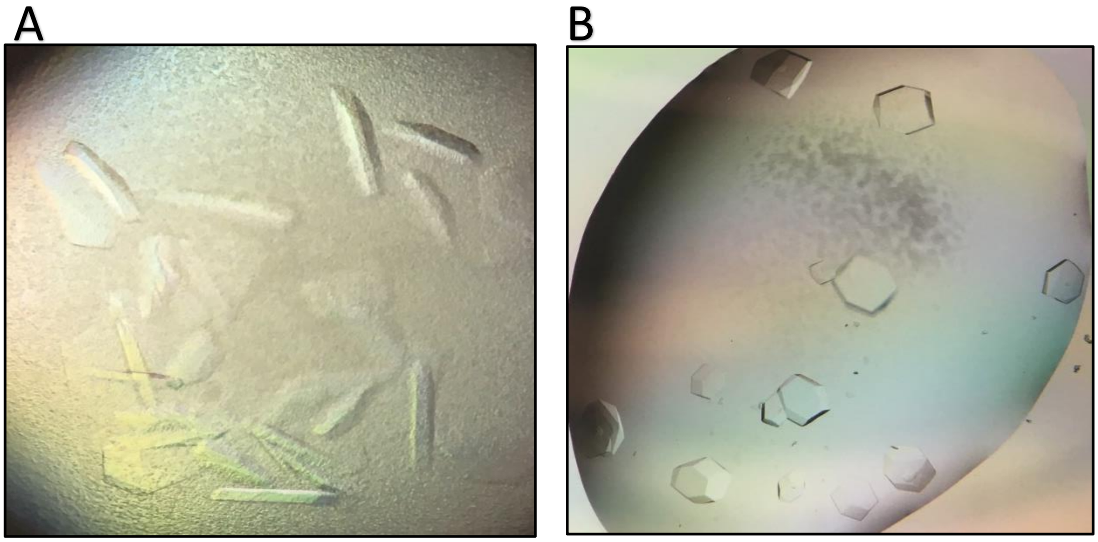


Figure 2.5. *Crystallization of Ku.* A) Crystals of Ku ΔC-term, using 1 μL of 12mg/mL protein mixed with 2μL of 0.1M magnesium sulfate, 0.05M sodium acetate pH 4.5, 0.6M lithium sulfate, and 3% ethylene glycol precipitant. B) Crystals of Ku ΔExt C-term, using 1 μL of 10.9mg/mL protein mixed with 1 μL of 0.01 magnesium acetate, 0.05 cacodylic acid pH 6.5, 1.3M lithium sulfate precipitant.

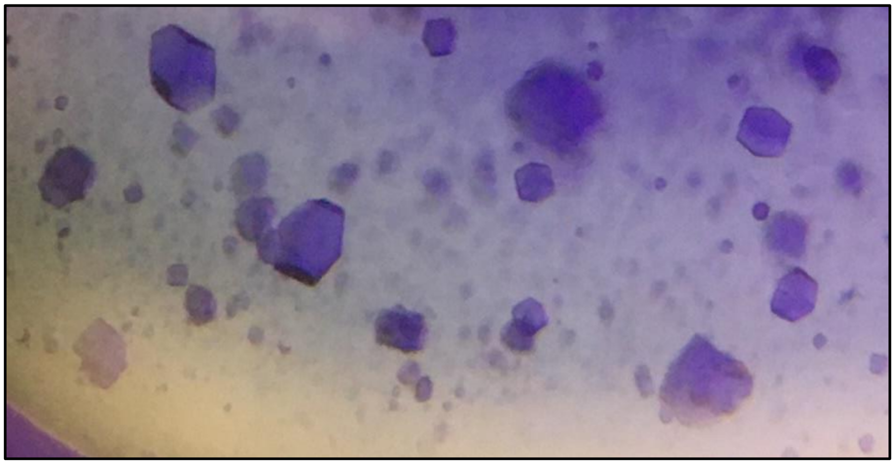


Figure 2.6. *Optimized crystals are Ku protein.* Ku ΔExt C-term crystals stained with 0.2 µL of 0.1% crystal violet. The stain was added to the drop containing the crystals and left to diffuse throughout the drop for 1 hour before images were taken.

One way to improve crystal diffraction quality is by the addition of freezing resistant solvents, called cryoprotectants. When crystals are tested for diffraction, they are flash-frozen in liquid nitrogen and kept at temperature of 100K while screened, in order to offset heat produced by the energy of the X-ray. This flash freezing process will freeze the solvent trapped in the solvent channels of the crystal, which can lead to damage to the crystal lattice and reduces the diffraction quality of the crystal, often leading to no observable diffraction[130].

For the Ku crystals to be tested for diffraction, a variety of cryoprotectants were trialed, including glycerol, an array of polyethylene glycols (PEGs) (750-8000), and ethylene glycol. To add the cryoprotectant, two methods were used: slowly replacing the precipitant that the crystal has grown in with the cryoprotectant over time, or soaking the crystal in the cryoprotectant for different lengths of time, before flash-freezing the crystal in liquid nitrogen[130]. The first method allows for slow replacement of the solvent with cryoprotectant, so as not to cause changes to the crystal packing by rapidly introducing molecules that may alter the crystal packing. It also ensures that all solvent channels are replaced with cryoprotectant. The second method is riskier, because it is done quickly, so not all solvent channels will be filled with the cryoprotectant, and the quick change in precipitant may damage the crystal. Neither method improved crystal diffraction in any of the trials. An alternative method to address cryoprotection involves crystal dehydration[130]. This is a process by which salt is added in increasing concentrations to the well solution of a hanging drop containing protein crystals, over time. To reach equilibrium in the closed system, water diffuses out of the protein to the well solution, lowering the solvent content of the crystals, ideally to a level that freezing will not damage the crystals. Unfortunately, this process was met with the same results as solvent exchange, failing to produce diffraction.

The reason diffraction is impaired by damage to the crystal lattice is that it introduces mosaicity to the crystal. Mosaicity essentially means that the planes within a crystal are disordered. When the solvent expands because of freezing, it can crack the crystal, creating altered orientations of the crystal planes, making the planes disordered. Diffraction relies on perfect crystals where everything is in the same orientation[130]. Another method to reduce the mosaicity introduced by flash freezing is crystal annealing. Crystal annealing involves allowing the crystal to thaw, then flash freezing again, and it has been shown to improve diffraction quality. It is believed that the freeze-thaw cycle allows crystal planes to reorient into their initially ordered state. Several methods exist which either include freeze-thaw cycles before mounting the crystal using liquid nitrogen, or, after mounting the crystal, by blocking the cryostream of the machine for set periods of time[130]. Both genres of this technique were employed for the Ku crystals, with and without overlapping trials with the various cryoprotectants, though these too failed to produce diffraction.

The crystallization process was repeated for Ku ΔC-term but using a DNA substrate. Two DNA substrates were used, the first a linear 20bp oligonucleotide, which had little success, so a second substrate was designed based on the DNA used to solve the structure of the human Ku70/80 heterodimer[31] (Fig. 2.7). This oligonucleotide creates a hairpin secondary structure on one end, to ensure uniform Ku loading. Substrates often increase the rigidity of the protein, which increases the crystal uniformity, thus, this posed a potential solution to the diffraction issue. Ligands stabilize proteins by forming energetically favorable contacts with the surface of the protein; this is what allows the protein to hold onto the ligand. In doing this, the residues involved lose much of their conformational flexibility, which in turn stabilizes the protein. As some residues become "locked in place" through their direct interaction with the ligand, many more residues will indirectly be restricted in their motion[131]. In protein crystallization this decreased conformational flexibility is both beneficial for the formation of crystals, as it increases the likelihood of the crystals being in a similar conformation, as well as often benefitting diffraction for the same reason as diffraction is reliant on the proteins existing in the same conformation[130,131]. Ku crystallized with the new DNA substrate, and the crystals were optimized (Fig. 2.8). However, the Ku-DNA crystals failed to diffract. Similar efforts to improve the diffraction of Ku ΔC-term and ΔExt C-term were taken, employing the use of several cryoprotectants and crystal annealing with the same results.

Figure 2.7. *Hairpin dsDNA construct*. Two oligonucleotides of different lengths (32bp and 19bp) annealed together to produce a hairpin structure. Bases of the hairpin are coloured blue.

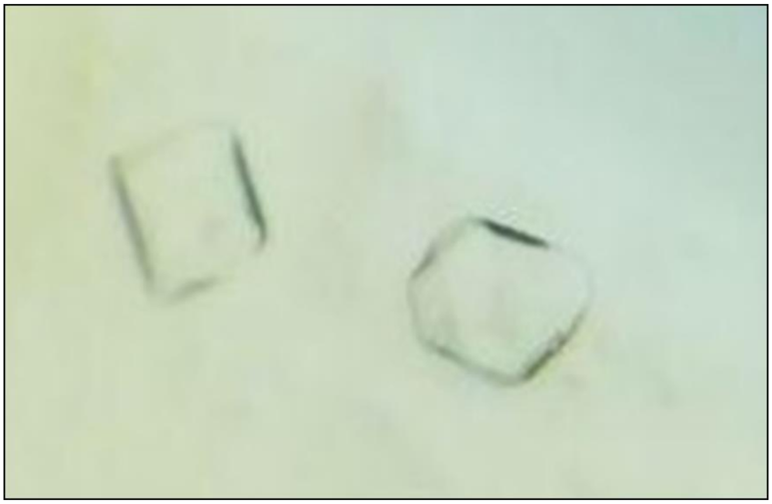


Figure 2.8. *Crystallization of ΔC-term Ku with DNA.* 250nL of 168µM Ku and 100µM DNA substrate mixed with 250nL of 0.05M cadmium sulfate, 0.1M HEPES pH 7.5, and 1M sodium acetate precipitant.

The initial Ku-DNA crystal results are a promising basis for future work, given that crystals of Ku and Ku with DNA were obtained. Additional optimization to reach diffraction-quality crystals include: room temperature diffraction screening, protein cross-linking, and pre-crystallization protease digestion. Room temperature screening avoids the issues associated with freezing the crystals, however it does result in the destruction of the screened crystal before a complete dataset can be obtained. This technique would primarily be useful in identifying whether the cryoprotectants being used and/or the process of freezing, is the problem. The process of cross-linking uses glutaraldehyde to covalently cross-link the protein to itself, increasing crystal uniformity and rigidity. By increasing rigidity, it increases the likelihood that the protein has adopted an identical conformation throughout the crystal, improving diffraction[130]. This is much the same reasoning that truncations of the C-terminal domain were used, although cross-linking would theoretically have an even more drastic effect by stabilizing regions throughout the entire protein. Finally, pre-crystallization protease treatment would involve repeating the entire process, however, it may lead to more stable or less mobile Ku constructs by sampling a random array of possible truncations. More stable constructs both improve crystallization and diffraction. Having achieved crystallization of the protein, these results will form a strong basis for future structural characterization.

These results serve as evidence supporting that the extended C-terminus is disordered. Initial crystallization trials included all three constructs, however no crystallization was seen with WT Ku. The lack of any observed crystallization supports the prediction that the C-terminus is disordered because disordered proteins often impair crystallization[127]. This helps explain how the C-terminal domain has been linked to multiple functions. Multifunctionality is often seen in disordered proteins as the flexibility allows them to adopt a variety of different conformations for different interactions[125]. Although not conclusive evidence, this model aligns well with both the observations of this study and others.

Small angle x-ray scattering analysis of Ku

SAXS data was collected for WT Ku, ΔExt C-term Ku and ΔC-term Ku, as well as WT Ku with DNA and ΔC-term Ku with DNA at the SIBYLS beamline 12.3.1, at the ALS, using the high-throughput mail-in SAXS service[132]. For the DNA substrate, the same hairpin DNA oligonucleotide from crystallography (Fig. 2.7) was used. Initially all data appeared useable, unfortunately upon further analysis, ΔC-term Ku and both samples with DNA produced nonsensical molecular weights suggesting the data set was not accurate. In the case of protein plus DNA samples, the MW weight appeared to more closely match that predicted for the oligonucleotide alone suggesting that the Ku-DNA complex was not stable during data collection, therefore the overall scatter was an average of either protein or DNA, but not the complex. In the future, this issue may be addressed via chemical cross-linking using glutaraldehyde to covalently link the DNA and Ku, stabilizing the complex.

Despite the troubles with Ku-DNA and ΔC-term Ku, WT Ku and ΔExt C-term Ku yielded favorable results (Fig. 2.9). The radius of gyration of both Ku proteins tested were similar (36 and 39Å), suggesting there was no major overall conformational change between the two proteins. This to some extent validates that the DNA-binding effects of the truncation are related to the C-terminal residues themselves and not to an overall introduced conformational change. The maximum particle distance, from the calculated pair-wise distance distribution function (Fig. 2.9B) for WT Ku was 96Å and 77Å for ΔExt C-term Ku. These results are validating, because WT Ku carries an extended C-terminus, but ΔExt C-term Ku lacks a portion of the C-terminus, therefore the maximum particle distance for the truncated mutant should be shorter. Figure 2.9C demonstrates the *ab initio* envelopes of Ku and ΔExt C-term docked with the eukaryotic Ku70/80 structure (PDB 1JEQ). From the docking, the core of Ku can be seen forming a central circular shape, while there is an elongated tip (black arrow Fig. 2.9C), that may be where the prokaryotic Ku C-terminus goes, as this same region is missing from the *ab initio* envelope of the ΔExt C-term alone. These results provide a rough idea of the general shape of the protein and hint at the orientation of the domains. This data also provides good comparisons for future studies where the Ku-DNA complex is analyzed, comparing *ab initio* envelopes would elucidate any major conformational changes associated with binding, particularly as it relates to the C-terminus, which would provide further insight into the orientation of the domains, but more importantly, potentially information as to how the C-terminus functions and how it destabilizes the interaction with DNA.

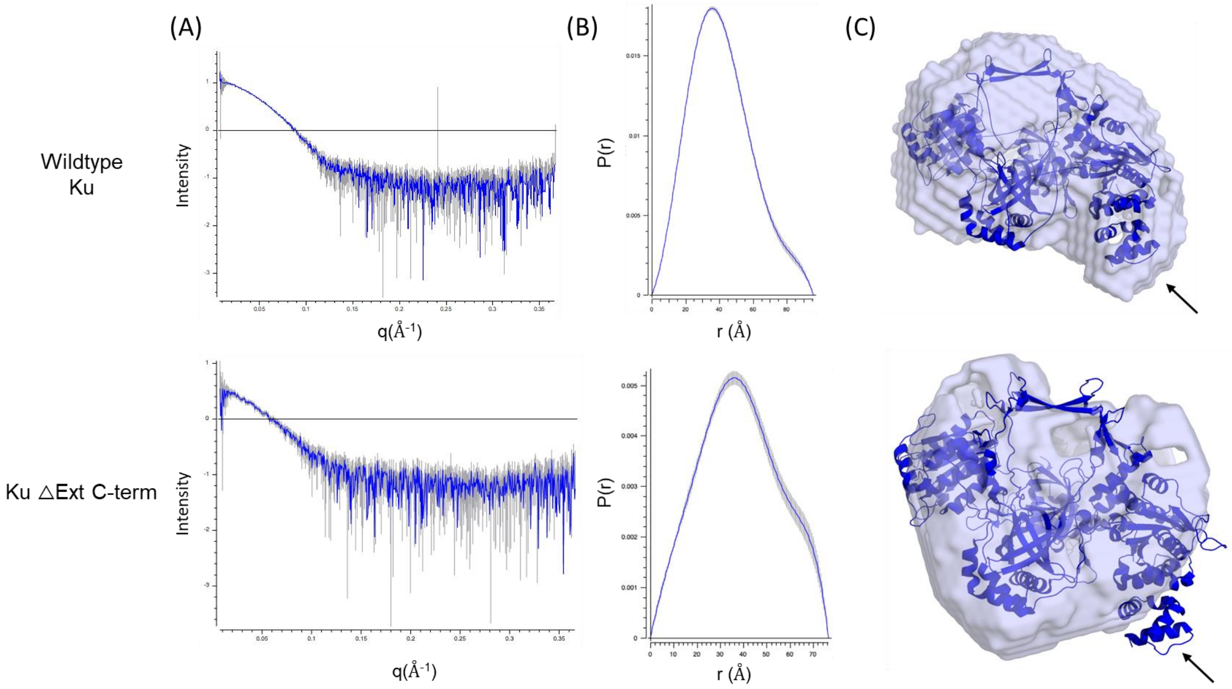


Figure 2.9. *SAXS analysis of M. tuberculosis Ku.* (A) Experimental scattering curves for wildtype Ku (top) and Ku ΔExt C-term (bottom). (B) Real-space pair-wise distance distribution functions for wildtype Ku (top) and Ku ΔExt C-term (bottom). (C) *Ab initio* three-dimensional envelopes (light blue) of wildtype Ku (top) and Ku ΔExt C-term (bottom), calculated from experimental scattering in (A), using Crysol. Crystal structure of wildtype Ku70/80 (PDB 1JEQ) has been modelled into the envelope by SUPCOMB. Black arrow represents proposed C-terminus of Ku. All analysis carried out by Primus, part of the ATSAS suite[133].

Biophysical characterization of Ku

Atomic force microscopy (AFM) is a useful method for characterizing the Ku-DNA interaction from a biophysical perspective. AFM images protein and DNA molecules on the nanometer scale, by creating a topographical map. Figure 2.10A depicts the Ku protein on a mica surface. Each individual peak represents a unit of Ku. Using ImageMetrics image processing software (further discussed in Chapter 4), the volume of each particle was analyzed. Volumes from 1237 particles were pooled and plotted on a frequency of distribution plot (Fig. 2.10B), which was fit with a normalized curve to provide an average volume of 1870nm^3. The average volume of Ku alone is important for analyzing Ku-DNA images as a comparison. The next steps in the AFM characterization of the Ku-DNA interaction are to image Ku in complex with DNA, and DNA on its own. By using the 3 sets of images, a variety of physical measurements can be made. The volume, for example, can be used to infer how many Ku are bound to the DNA. Other measurements include: lengths to determine the approximate number of bp Ku interacts with, or position on the DNA to determine if Ku is regularly travelling a specific distance onto DNA, or if Ku freely slides up and down the entire DNA length.

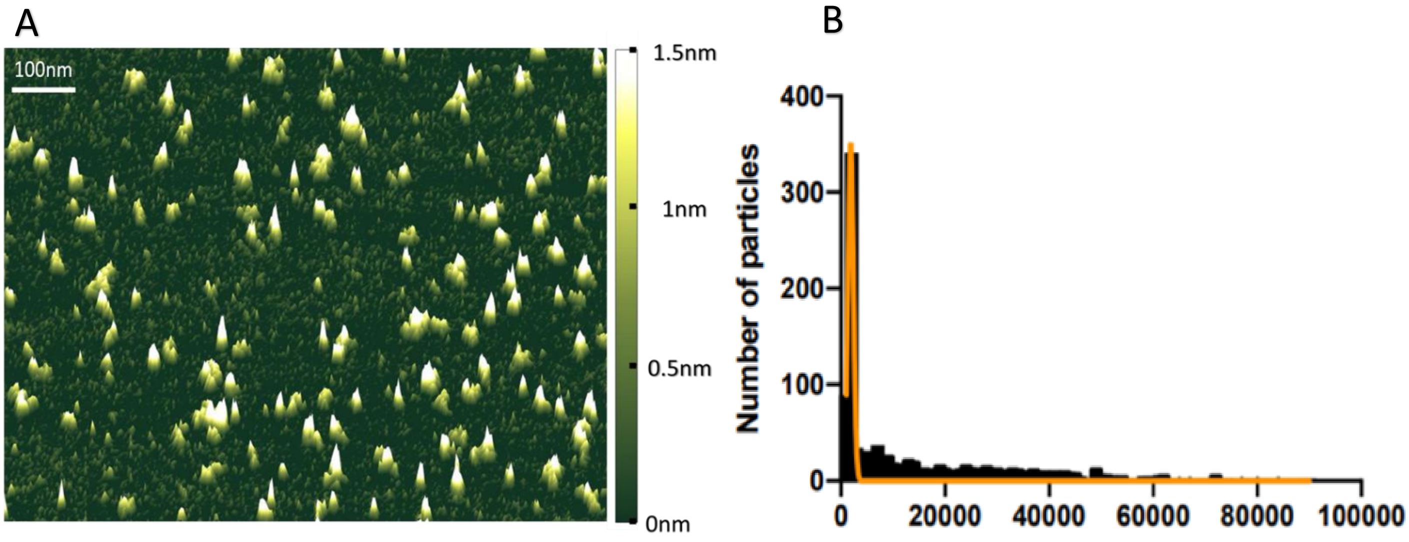


Figure 2.10. *Particle Analysis of Mycobacterium tuberculosis Ku.* (A) Representative 3D image of Ku at a concentration of 25nM in 40mM HEPES pH 7, 200mM NaCl, 5mM MgCl, and 1mM DTT. Image was taken using a ScanAsyst Air probe on a Bruker Bioscope Catalyst atomic force microscope at 256 samples/line, at a frequency of 1Hz and processed using ImageMetrics. (B) Frequency distribution of Ku particle volume, non-linear Gaussian fit, $R^2 = 0.9$. n=1237 number of particles. Analysis carried out using ImageMetrics software. Graphical analysis by GraphPad Prism.

Following up on the WT Ku data will include expanding this technique to the different truncation mutants to clarify the role of the C-terminus. This will provide details as to whether features such as the position on DNA, are affected by the C-terminus. These future results will corroborate the DNA binding data. Since Ku's affinity for DNA is higher without the C-terminus, the Ku truncations will likely travel shorter distances onto the DNA as the interaction is tighter. These results would provide valuable support for the prediction that the C-terminus destabilizes Ku, to facilitate the ease of its displacement by other proteins, because it would directly demonstrate that Ku ΔExt C-term is less mobile on DNA. Demonstrating this feature would provide more support for targeting the C-terminus of Ku for inhibitor design.

In vivo characterization of Ku

To investigate the *in vivo* implications of the C-terminus destabilizing the Ku-DNA interaction, as well as to investigate the potential of NHEJ as a drug target, NHEJ deletion mutants in *P. aeruginosa* are under construction. Using a protocol adapted from Hmelo et al. (2015)[134] I began creating *P. aeruginosa* deletion strains of *ku, ligD,* or *recB*. The goal is to observe the effects of the deletion on cell survival and DNA repair after antibiotic treatment, as well as to be able to complement the Δ*ku* and Δ*ku*Δ*recB* strain with plasmids expressing Ku mutants that impacted DNA binding. Initial cloning of the deletion mutants involved amplifying regions immediately upstream and immediately downstream of the gene, then using splicing by overhang extension (SOE) PCR to stitch the fragments together for cloning into the exchange vector. Amplified fragments for each gene were obtained, as well as a stitched PCR product the size of the two individual fragments combined (Fig. 2.11). The initial PCR products provide a starting point for the cloning of the NHEJ and HR deletion strains.

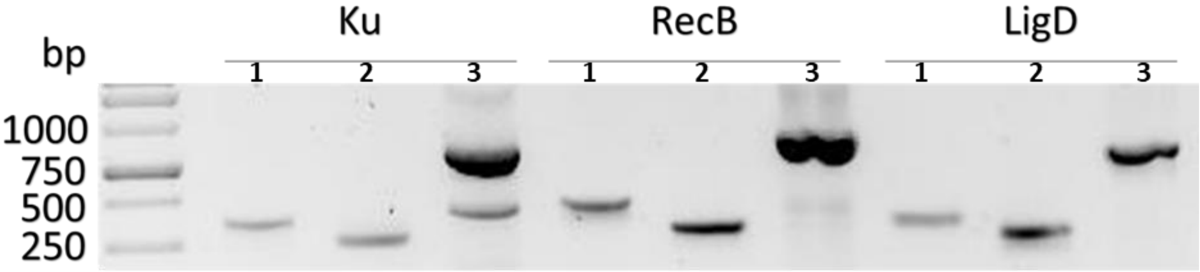


Figure 2.11. *Amplification of Homologous DNA fragments for allelic exchange.* Regions flanking the 3 genes of interest were amplified from the genomic DNA of *P.aeruginosa* PAO1. These regions were stitched together through splicing by overhang (SOE) PCR. Lane 1 represents the PCR amplified region upstream of the corresponding gene, lane 2 is the downstream fragment, and lane 3 is the SOE product. Samples were run on a 10% agarose gel and stained with GelRed. The darkest band for each SOE product represents the correctly stitched together product as each is seen at approximately the sum of the corresponding fragments.

Summary

While studying the DNA binding properties of bacterial Ku, this work unexpectedly found that the C-terminal domain, which has previously been implicated in DNA binding[27,54,55], decreases the affinity of the Ku-DNA interaction. I propose a model to explain the biological significance of this, where the proteins of other cellular processes, which require access to DNA, are able to displace Ku because of its lowered DNA binding affinity, regulated by the Ku C-terminus. In an effort to understand how this domain accomplishes this, crystallography work was started and crystals of 2 separate truncations, 1 in complex with DNA, have been achieved, however achieving diffraction quality crystals is still in progress. SAXS analysis provided an *ab initio* envelope for both WT Ku and Ku ΔExt C-term where a region potentially occupied by the C-terminus isn't observed in the Ku ΔExt C-term envelope potentially hinting at some level of domain organization. The analysis also produced logical numbers for MW and maximum particle distance for these two constructs suggesting the data is good, validating the modeled envelope. Biophysical analysis through AFM has determined the average volume of Ku is 1870nm^3, which will be crucial in expanding the biophysical characterization to the Ku-DNA interaction. These results will facilitate further investigation into the role of the C-terminus in the Ku-DNA interaction, by comparing the biophysical properties of WT Ku-DNA to ΔC-term Ku-DNA. Based on the DNA binding results, I hypothesize that AFM results will demonstrate that ΔC-term Ku will display differences in its position on DNA. Being more restricted by the tighter binding, ΔC-term Ku should be trapped closer to the DNA ends. This would carry significance by supporting the proposed model that a low affinity is important for Ku so as to not block other cellular processes accessing DNA because this would demonstrate that the mobility of Ku on DNA is directly tied to its affinity for DNA. Thus, stabilizing Ku-DNA complexes may prove to be toxic to the cell in much the same way quinolones work by stabilizing topoisomerase-DNA complexes. This would suggest that stabilized Ku-DNA complexes would block crucial processes such as replication and transcription, a promising route to explore in drug development. This work could be validated through the completion of the *P. aeruginosa* deletion strain library. Using this library, future work can test the viability of NHEJ inhibition through Ku, as well as test the viability of Ku inhibition via the C-terminus and investigate whether a higher affinity Ku is toxic to the cell. Finally, this library can extend the results to investigate synergy between HR and NHEJ inhibition, since both pathways are responsible for DNA DSB repair, and in doing so, potentially reveal a promising pathway for antimicrobial drug development.

Chapter 3: Solubilizing ImuABC

Project overview:

Project rationale:

The ImuABC complex is a highly mutagenic polymerase complex of three proteins, involved in the bacterial DNA damage response[108,135–138]. Work done *in vivo* across different bacterial species has demonstrated that this complex increases survival in DNA damaging stress conditions, such as UV exposure and chemically induced alkylation DNA damage, while also causing an increased number of mutations in strains with active ImuABC. These mutations are directly and indirectly linked to the development of antimicrobial resistance[108,135–138]. Therefore, this complex warrants further investigation into its biochemical, structural, and physical properties, as a potential target for antimicrobial drug development. To date, one significant problem has stood in the way of characterizing this complex: protein solubility[139,140]. All three components are difficult to express and purify in a soluble wildtype form, making any *in vitro* analysis impossible. The work done for this chapter was with the intent of expressing soluble ImuA, B and/or C *in vitro*.

Background:

In contrast to the proteins discussed in the previous chapters, the ImuABC complex is involved in DNA damage tolerance, not repair[108,135–138]. This means rather than fix damage, it is involved in circumnavigating DNA damage. More specifically, this complex is part of the translesion synthesis (TLS) pathway and it fulfills a functionally conserved role to polV in *E. coli,* the best studied bacterial TLS pathway[141,142]. These TLS pathways are part of the DNA damage SOS-response in bacteria. The SOS-response is initiated upon the accumulation of ssDNA, often a direct result of a stalled replication fork, which leads to RecA binding and polymerization into a nucleoprotein filament. The polymerization of RecA promotes the self-cleavage of the LexA repressor protein, initiating the expression of multiple DNA damage repair and tolerance proteins[143]. The initially induced genes are primarily repair genes, such as those involved in BER and NER. When the damage persists, TLS genes are induced, allowing the cell to survive periods of high DNA damage that overcome its repair capacity[143].

Translesion synthesis is a process in which specialized polymerases are recruited to stalled replication forks at "lesions", which is a term that encompasses any type of damage that the replicative polymerase cannot replicate past[141,142]. These specialized polymerases are able to relax the DNA backbone to insert a nucleotide across from the DNA damage. Because of this, the polymerases are error-prone and thus often associated with mutagenic behaviour[141,142]. This itself is often a beneficial trait, as it allows for an inducible state of mutagenesis, which directly leads to the development of new bacterial phenotypes. These TLS systems are induced during DNA-damaging stress conditions such as antibiotic or ROS exposure. This is a means for cells to adapt and overcome the incurring stress[108,137,141]. Between providing both immediate survival benefits and lasting ones, mutagenic mechanisms are crucial for life, and thus we see similar mechanisms in all forms of life[142]. PolV and ImuABC have a highly conserved function but no apparent structural relation; they are completely different protein complexes fulfilling the same role. This suggests that this function evolved separately. This is supported by the finding that the two systems are mutually exclusive[144], further highlighting the importance of TLS.

The ImuABC complex is particularly interesting because of its unique architecture. The *imuB* and *imuC* genes each display sequence homology to different polymerases. *imuB* shows homology to other TLS polymerases, but lacks key catalytic residues, while *imuC* resembles a subunit of the replicative polymerases and is the catalytically active component of this complex[138,140,145]. In contrast, *imuA* only shows minor sequence homology to the *recA* gene, the aforementioned initiator of the SOS-response[138,139]. Though they all show some genetic similarity to other known proteins, these homologous proteins are not known to interact, which distinguishes the ImuABC complex as an exceptionally unique piece of cellular machinery.

In addition to its unique nature, the inducible mutagenesis or "inducible evolution" phenotype linked to the ImuABC complex provides bacteria a means of developing resistance to antibiotics[108,135–138]. Moreover, it has been directly linked to the development of antibiotic resistance to, and in the tolerance of, multiple antibiotics[108,135,138]. The final straw is the presence of ImuABC in multiple problem pathogens, notably like Ku, *Mycobacterium tuberculosis,* which was responsible for 1.5 million deaths in 2018[126]. These three factors all suggest this complex may be a good drug target. ImuABC is a unique piece of cellular machinery, which means that inhibitors of this complex would be less likely to have undesirable, off target effects resultant from interactions with similar proteins because there are no known similar complexes, thus, there are targetable interaction interfaces unlike any other in the cell. As discussed in chapter 1, the induced mutator phenotype is an ideal trait to target in the fight against AMR because it is a means of eliminating the bacteria's primary method of overcoming our arsenal of antibiotics. This becomes increasingly important when considering that the ImuABC system is present in bacterial strains such as *M. tuberculosis* and that in these problem pathogens the ImuABC complex improves their survivability in DNA-damaging conditions[108,135–138]. Further still, ImuABC has been linked directly to an increased virulence of *M. tuberculosis* in mice models[2]. Together, these features suggest that inhibition of this complex would have a direct impact specifically on some of the worst bacterial infections. These reasons form a compelling argument for the further characterization of ImuABC. To date, characterization of ImuABC has been plagued by protein solubility and the fact that all 3 components are highly insoluble. The insoluble nature of ImuABC has significantly hampered our understanding of these proteins, their individual roles, and how their domains facilitate function.

Experimental procedure:

To tackle the issue of solubility, I took several approaches. The primary method was to add a variety of solubility tags when expressing the wildtype protein. *M. tuberculosis* genes, codon optimized for an *E. coli* expression system, were cloned into pMCSG- 7, 9, and 10, plasmids, which encode an N-terminal, TEV-cleavable His_6, His_6-MBP, and His_6-GST tags respectively, along with a 4th custom plasmid containing a His_6-lysozyme tag. The second method used co-expression inclusive of the *imuA, B and C* genes, together with pET-Duet1, pET-29a, and pMCSG-14. The pET-Duet1 and pMCSG-14 plasmids both have multiple cloning sites facilitating co-expression from a single plasmid while the pMCSG-14 and pET-29a plasmids have alternative antibiotic resistance markers (chloramphenicol and kanamycin respectively) facilitating co-expression through double/triple transformant selection. Constructs were expressed in varying combinations of *imuA'*, *imuB*, and *imuC* (A'B, A'C, BC, and A'BC) in an attempt to increase solubility. *imuA'* is the designation for the version of *imuA* present in *M. tuberculosis*. Bacteria with ImuABC either have ImuA or ImuA' which are believed to be distant homologs that have undergone a large amount of differentiation as their sequence have very few similarities[138,144]. Together, I trialed a range of expression conditions with these two methods of protein expression, including 37ºC and 16ºC IPTG induction, as well as with varying IPTG concentrations from 50µM to 1mM; however, this did not affect the end results. I also tested a 25ºC autoinduction using specialized media. Further still, I tested multiple different *E. coli* expression cell lines.

Results:

The first method to improve solubility involved cloning each of the constructs into pMCSG- 7, 9, and 10, as well as a custom plasmid, pSA063, a custom plasmid which adds an N-terminal, TEV-cleavable, His_6-lysozyme tag. Table 3.1 summarizes the successful constructs. Cloning was confirmed first by colony PCR using agarose gel electrophoresis, and then by Sanger sequencing.

Table 3.1: *Cloning summary of TLS genes into protein expression plasmids.* Y, yes; N, no.

Cloning Status				
Gene	**His-tag**	**MBP-His-tag**	**GST-His-tag**	**Lysozyme-His-tag**
imuA'	**N**	**N**	**Y**	**Y**
imuB	**Y**	**N**	**N**	**Y**
imuC	**Y**	**Y**	**Y**	**Y**

Following cloning, expression constructs were transformed into: *E. coli* BL21, BL21*, BL21 codon+, and SoluBL21 cell lines. Most of these transformants were screened in 3 separate expression conditions: 37ºC for 3 hours post IPTG-induction; 16ºC for 20 hours post IPTG-induction; and 25ºC for 72 hours in autoinduction media. The results are summarized in Table 3.2; none of the conditions produced soluble proteins. Expression was determined using samples taken before, during, and after induction. The cells were lysed, and the cell contents were visualized on a 10% SDS-PAGE gel (Fig 3.1). Figure 3.1 shows the only expression construct that produced successful expression: ImuA' in pMCSG-10, as demonstrated by the appearance of a band at ~48kDa, corresponding to the predicted MW of the GST-His_6-ImuA of 47kDa. The other constructs failed to produce bands at the MW which correlated to their affinity tag plus the TLS protein (ImuA=26kDa+tag, ImuB=56.7kDa+tag and ImuC=118.7kDa+tag). Solubility of successful expression was confirmed through fractionating cell lysate via centrifugation at 40,000xg to separate soluble cell components from insoluble, and determining if expressed proteins were in the insoluble or soluble fractions by SDS-PAGE (Fig. 3.2). Figure 3.2 demonstrates that GST-His_6-ImuA was found exclusively in the insoluble fraction.

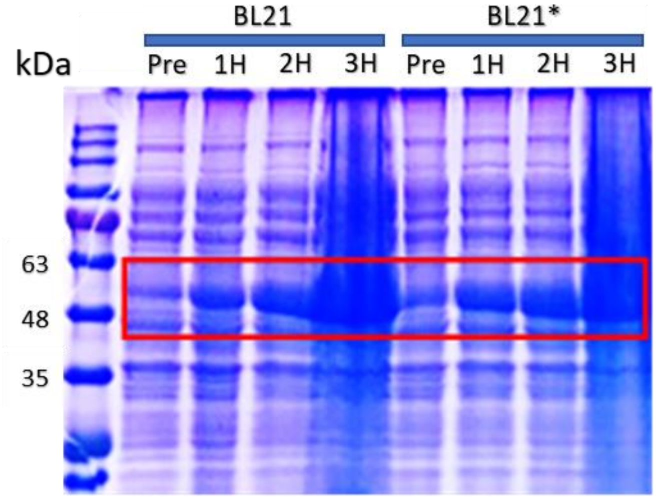


Figure 3.1. *Expression testing of Mycobacterium tuberculosis GST-tagged imuA'.* Expression of GST-His- tagged *imuA'* was induced in BL21 and BL21* *E. coli* expression strains with 50mM IPTG for 3 hrs at 37° C after reaching an OD_{600}=0.8. Samples were taken pre-induction (Pre), and at 1 (1H), 2 (2H), and 3 hours (3H) after induction. The red box highlights a band that correlates to the expected molecular weight of ImuA'.

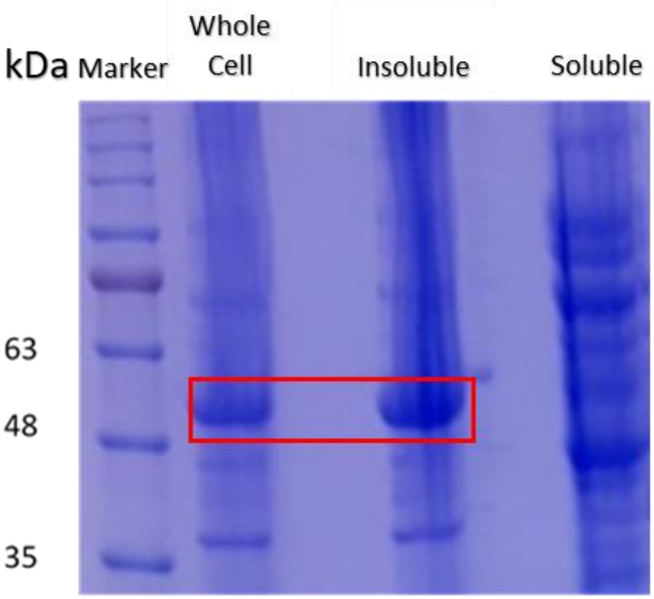


Figure 3.2. *Solubility testing of M. tuberculosis GST- tagged imuA'.* Soluble samples were extracted from expression testing samples. Highlighted with a red box are the bands associated with the ImuA' construct.

Table 3.2. *Expression conditions tested for ImuA', ImuB and ImuC*. The cell lines used were all *Escherichia coli* expression strains. Induction was done using 50mM IPTG for the 37ºC and 16ºC inductions in LB media, while the AIM (autoinduction media) method contained 115mM lactose for a gradual induction. The notation is as follows: S=soluble expression observed, I=insoluble expression observed, N=no expression observed, and X=untested.

Construct	Cell line	37ºC 3 hours	16ºC 20 hours	AIM media 25ºC 72 hours	Co-expression with at least one other construct
GST-His-imuA'	BL21	I	N	I	I
	BL21*	I	N	I	I
	BL21 Codon+	I	N	X	X
	SoluBL21	I	N	I	I
Lysozyme-His-imuA'	BL21	I	N	I	I
	BL21*	I	N	I	I
	BL21 Codon+	X	N	X	X
	SoluBL21	I	N	I	I
Lysozyme-His-imuB	BL21	N	N	N	N
	BL21*	N	N	X	X
	BL21 Codon+	N	N	X	X
	SoluBL21	N	N	N	N
Lysozyme-His-imuC	BL21	N	N	N	N
	BL21*	N	X	X	X
	BL21 Codon+	N	X	X	X
	SoluBL21	N	N	N	N
GST-His-imuC	BL21	N	N	N	N
	BL21*	N	X	X	X
	BL21 Codon+	N	X	X	X
	SoluBL21	N	N	N	N
MBP-His-imuC	BL21	N	N	N	N
	BL21*	N	X	X	X
	BL21 Codon+	N	X	X	X
	SoluBL21	N	N	N	N

For all of the successfully cloned constructs, either no expression or insoluble expression was observed, as summarized in Table 3.2. The limited observed expression success using solubility tags led to the start of co-expression trials. One approach used was co-transforming individual plasmids with different antibiotic markers. In the initial work, all plasmids contained ampicillin resistance so new constructs using pMCSG-14 and pET-29a were established because these plasmids had chloramphenicol and kanamycin resistance respectively. The second approach was co-expressing two genes from a single expression plasmid with multiple cloning sites, pET-Duet1b (this was also possible in pMCSG-14 however cloning difficulties prevented this). Table 3.3 summarizes the constructs produced in these 3 plasmids, as well as whether they were successfully sequenced.

Table 3.3: *Cloning summary of TLS genes into plasmids for co-expression or co-transformation*. S, confirmed via sequencing; I, inconclusive sequencing following positive PCR amplification gel-electrophoresis identification; N, no.

Cloning Status					
Gene	**pET-Duet1 MCS2 (Empty MCS1)**	**pET-Duet1 MCS1 with *imuB* in MCS2**	**pET-Duet1 MCS1 with *imuC* in MCS2**	**pMCSG-14 MCS1**	**pET-29a**
imuA'	**N**	**I**	**I**	**N**	**I**
imuB	**S**	**N**	**I**	**I**	**N**
imuC	**S**	**N**	**N**	**N**	**N**

Co-expression trials followed a similar course of testing, however none of the tested combinations (Table 3.4) were able to improve the results, no bands appeared at a MW that corresponded to those of any of the 3 proteins, plus their tags.

Table 3.4. *Co-expression trial summary*. Blue squares indicate that the constructs in that intersect were not tested in a co-expression system together while white squares indicate that the plasmids from that intersect were tested in a co-expression system. None of the tested co-expression systems successfully expressed soluble protein.

	pMCSG-10 *imuA'*	pSA063 *imuA'*	pSA063 *imuB*	pSA063 *imuC*	pMCSG-10 *imuC*	pMCSG-7 *imuC*	pET-Duet1 *imuB*	pET-Duet1 *imuC*	pET-Duet1 *imuB* + *imuA'*	pET-Duet1 *imuC* + *imuA'*	pET-Duet1 *imuC* + *imuB*	pMCSG-14 *imuB*	pET-29a *imuA'*
pMCSG-10 *imuA'*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	🟦
pSA063 *imuA'*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	🟦
pSA063 *imuB*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜
pSA063 *imuC*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	⬜
pMCSG-10 *imuC*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	⬜
pMCSG-7 *imuC*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	⬜
pET-Duet1 *imuB*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜
pET-Duet1 *imuC*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
pET-Duet1 *imuB* + *imuA'*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
pET-Duet1 *imuC* + *imuA'*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜	🟦
pET-Duet1 *imuC* + *imuB*	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	⬜
pMCSG-14 *imuB*	⬜	⬜	🟦	⬜	⬜	⬜	🟦	🟦	🟦	⬜	🟦	🟦	⬜
pET-29a *imuA'*	🟦	🟦	⬜	⬜	⬜	⬜	⬜	🟦	🟦	🟦	⬜	⬜	🟦

The poor solubility observed can potentially be attributed to a high amount of predicted disorder in various regions of these proteins. Figure 3.3 demonstrates a high level of predicted disorder for ImuB, particularly towards the C-terminus, using D^2P^2 online pooled prediction software[124]. High levels of disorder like this suggest that truncating these proteins and removing predicted disordered regions, may improve protein solubility because disordered proteins have a tendency to aggregate[146].

Figure 3.3. *Disordered region prediction for ImuB*. The bar represents the entirety of the *imuB* gene; the colour gradient represents order to disorder prediction across D^2P^2's pooled prediction software. White is unanimously predicted to be ordered while green represents unanimously predicted disorder. In between are the blues, lighter blue was predicted to be disordered by fewer algorithms than darker blue regions. Disorder prediction analysis by D^2P^2 online software[124].

Currently published theses addressing this challenge[139,140] have focused on *M. tuberculosis* genes and thus another potential direction for subverting this issue, is to pursue homologs from other organisms. However, anecdotal evidence from other experts at the annual Buffalo DNA repair symposium, suggested this was actively being pursued with limited success by other groups. It was these conversations that ultimately led to the shift in focus to the Ku project that was discussed in earlier chapters.

There does remain one glaring gap in efforts both anecdotally and from published sources: native expression. Most efforts have focused on expression in *E. coli*, an organism lacking these genes and thus a foreign environment. Additional future directions for this study should be focused on expression from a bacterium that naturally produce these proteins and have a developed genetic toolbox amenable to easy *in vivo* genetic manipulation, such as the clinically relevant *P. aeruginosa.* Protocols described earlier for developing *ku* and *ligD* knockouts could be used to facilitate expression and purification of the ImuA, B and C proteins from *P. aeruginosa*[134]. The genetics of *P. aeruginosa* would facilitate the addition of purification tags to native genes or overexpression of the genes from a plasmid within the host organism. This technique may be the key to purifying soluble TLS proteins, a necessary precursor to biochemical and structural analyses.

Chapter 4: Investigating protein-DNA complexes with atomic force microcopy

Project overview:

Project rationale:

Our lab was invited to write a paper for the Journal of Visualized Experiments on the technique of using AFM to study protein DNA interactions. Given my focus on using AFM to study Ku, I was the primary researcher on this project. However, to preserve our data on Ku for a research publication, we chose another protein to illustrate the benefits of AFM in studying protein-DNA interactions. A key component of our approach was to highlight the utility of ImageMetrics, an image processing software designed by Zimeng Li from the lab of Dr. Dorothy Erie, UNC Chapel Hill. Image Metrics is a powerful tool for high- throughput analysis of AFM images. To carry out this study I chose to use RecA, a well characterized DNA binding protein which is readily available and has had its DNA binding properties examined by AFM previously[147]. This system would facilitate a focus on the techniques, protocols, and tools, rather than on the protein and its interaction with DNA.

Background:

RecA is a key protein in the bacterial DNA damage response associated with multiple DNA damage tolerance and repair pathways. In chapter 1 , RecA is identified as a key component of HR and as a highly conserved protein across all bacteria, while in chapter 3, the role of RecA in the SOS-response is briefly discussed, where it plays a role in activating the DNA damage cascade[143,148]. In both cases, multiple RecA monomers bind ssDNA, polymerizing to form a nucleoprotein filament. In DSB repair, this filament facilitates homologous strand invasion, whereas when initiating the SOS-response, RecA promotes self-cleavage of the LexA repressor protein, thereby initiating the expression of SOS-genes[143,148].

Both of these roles involve ssDNA binding, however RecA is also capable of binding dsDNA, and the dsDNA binding is ATP-dependent. DNA binding occurs at a faster rate and forms a tighter complex in the presence of ATP[149–151]. Importantly, Shivashankar *et al.* (1999)[150] demonstrate that the nucleoprotein filament is stabilized by the addition of a non-hydrolysable ATP suggesting that the ATP-hydrolase function of RecA is important for the dissociation from DNA. This property makes this system ideal for observation by AFM as it facilitates the formation of stable DNA-protein complexes.

AFM is a robust tool for studying DNA protein interactions. Previous studies on RecA have demonstrated the utility of the technique by identifying that RecA elongates the DNA and determining the nucleation rate of the nucleoprotein filament[147]. These measurements are possible because of the high-resolution images which AFM produces on the nanometer scale, which permit the distinction between protein and DNA, and facilitate dimensional measurements such as height, width, and length of the particles. These measurements translate into physical traits, which include: particle volume, as analyzed for Ku in chapter 2, which can be converted to MW, informing on the number of monomers involved; the fiber length, which is useful for first identifying whether the protein forms a nucleoprotein filament, but also may be used to determine the nucleation rate; and peak distance, which identifies the footprint of a protein bound to DNA. These types of analyses have been used for multiple different DNA repair proteins, including RecA, and Ctp1, a eukaryotic HR protein, where volume analysis translated to MW supported a tetrameric assembly, while fiber length and peak spacing analysis was used to identify the formation of filaments and the approximate footprint of each Ctp1 tetramer on DNA[147,152].

Crucial to the depth of analysis by AFM is the image processing software used. The capabilities of the software heavily determine the possible traits which may be analyzed, as well as the ease by which analysis is done. The ImageMetrics image processing software designed by Zimeng Li from the lab of Dorthie Erie, UNC Chapel Hill, is a particularly powerful software program offering a wide range of potential analyses, such as those listed above, as well as traits like eccentricity, or major and minor access lengths. More importantly, this software offers high-throughput processing as it is able to identify and analyze all of the particles within one image simultaneously for all desired traits and provide the output in an easy to export excel style format with numbers demonstrated on the image correlating values to their respective particles. Further still, there are easy-to-use tools designed to maximize the image quality, such as the flattening tool, which in turn improves the data quality. The flattening tool does this by leveling out an image. There are also tools that help ensure that only the particles, which the user desires, are being analyzed, such as a masking scroll bar. The scroll bar essentially sets a height threshold for which a particle must reach to be analyzed. By providing a diverse array of tools for AFM analysis, this software supports AFM as an insightful way to visualize and study protein-DNA interactions, such as RecA-DNA. For a JoVE article then, the RecA-DNA system is ideal to demonstrate the power of AFM and ImageMetrics software provides the perfect tool to do so with.

Experimental procedure:

For this work I adapted a protocol from previous RecA AFM research[147]. RecA (New England BioLabs Ltd.) was buffer exchanged into a reaction buffer (10 mM HEPES pH 7.0, 100 mM NaCl, 7 mM $MgCl_2$) via centrifugal filter units (Vivaspin 500, 3 kDa MWCO Polyethersulfone, GE Healthcare). The dsDNA substrate was pUC-19 plasmid, digested with HindIII (New England Biolabs Ltd.) and purified using EZ-10 Spin Coloumn PCR Products Purification Kit (Bio Basic Inc.). RecA and DNA were combined with Adenosine-5'-(β,γ-imido) triphosphate lithium salt hydrate (Sigma Life Science), a non-hydrolysable ATP and incubated for 30 minutes at room temperature to facilitate the formation of stable nucleoprotein filaments. After incubation, the samples were diluted with a deposition buffer (20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 5 mM KCl, 2 mM $ZnCl_2$), and plated on freshly cleaved mica. The sample was incubated on mica for 2 minutes at room temperature. Excess protein was rinsed away using HPLC grade water (Fisher Scientific), and the mica was dried first by blotting the sides of the disk (without touching the mica surface) with Whatman Grade 1 Qualitative Filter Paper Standard Grade (GE Healthcare), then using a stream of compressed nitrogen gas. Samples were imaged using a Bruker Bioscope Catalyst atomic force microscope with a SCANASYST-AIR probe. Images were processed using ImageMetrics software (Zimeng Li, UNC Chapel Hill). Image processing of all images included flattening and plane subtraction. To select the desired particles a combination of the masking scroll bar, together with both the draw region of interest (ROI) and draw inverse ROI tools were used. My initial results end here. Further work is still needed for the JoVE publication and will require optimization of the protocol to achieve consistent depositions that clearly image nucleoprotein filaments, as well as the necessary controls including imaging of only RecA or DNA.

Results:

Initial AFM imaging identified both DNA alone (Fig. 4.1A) and RecA alone or in complex with DNA (Fig. 4.1B). DNA may be observed in these images as a thread of low height (average height: 2.57nm, n=33) while RecA appears as higher peaks either binding DNA, or as individual particles. Preliminary analysis of RecA volume analogous to that done on Ku in chapter 2 suggests an average volume of 6350nm^3 (n=48). Data collection was truncated due to COVID-19. Additional particles are still needed for a statistically significant analysis.

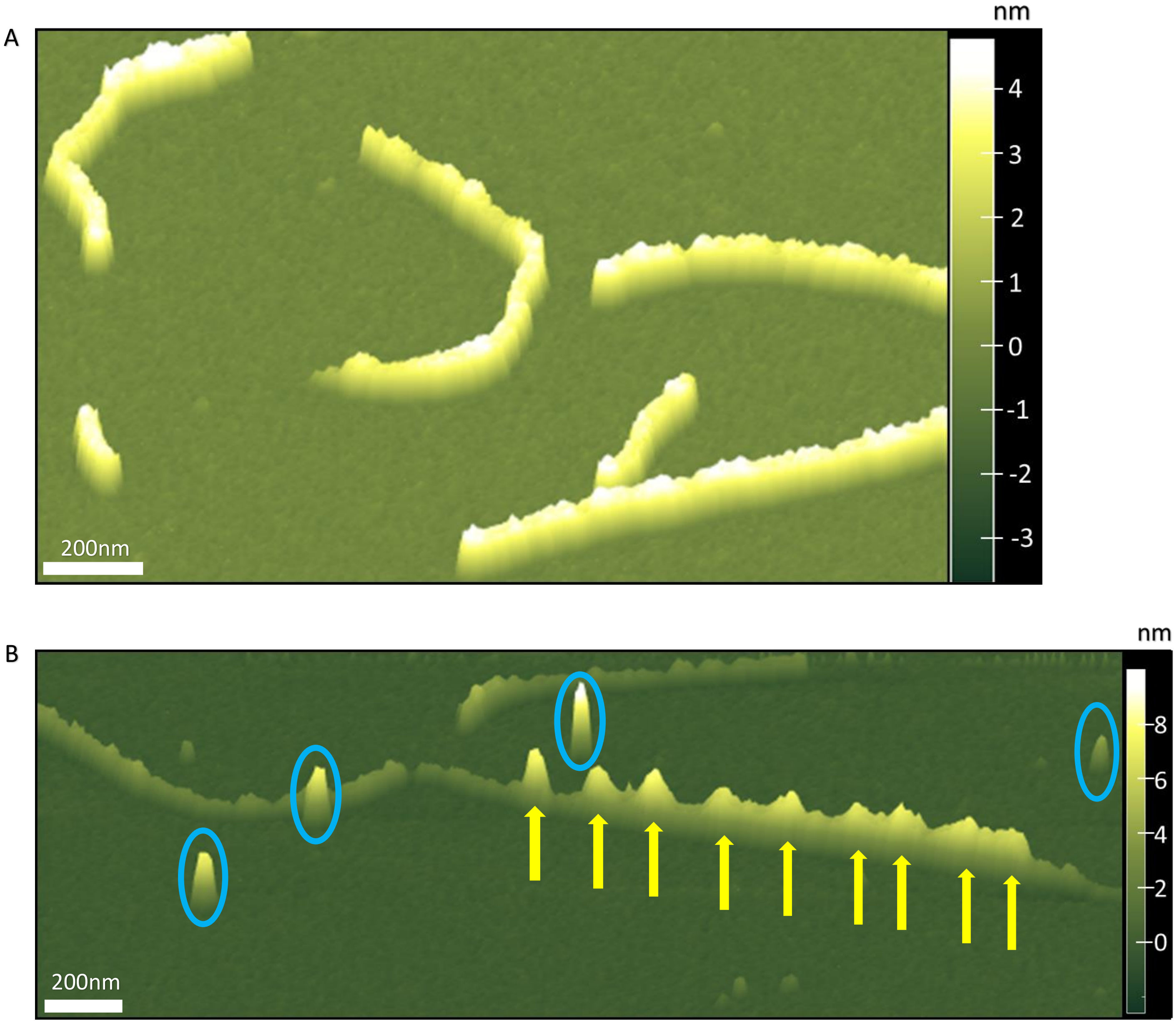


Figure 4.1. *AFM images of DNA and RecA.* A) Representative image of linearized pUC-19 plasmid DNA taken using a Bruker Bioscope Catalyst atomic force microscope with a ScanAsyst Air probe, 256 samples/line, 0.501Hz, processed using ImageMetrics software. B) Representative image of RecA and RecA-DNA complexes using linearized pUC-19 plasmid DNA taken using a Bruker Bioscope Catalyst atomic force microscope with a ScanAsyst Air probe, 256 samples/line, 0.501Hz, processed using ImageMetrics software. The blue circle represents unbound RecA, yellow arrows indicate RecA-DNA complexes.

Unfortunately, very few RecA-DNA complexes were observed, suggesting optimizations to the protocol are needed, such as increasing the incubation time before deposition on the mica. This is further supported by the images of RecA-DNA complexes, where individual RecA-DNA binding events are observed rather than a continuous nucleoprotein filament. This observation disagrees with previous evidence, which suggests that for each piece of DNA there is only one nucleation event (one initial binding) that extends into a filament eventually covering the DNA[147], whereas here in figure 4.2, there are clearly multiple bound RecA molecules suggesting this is demonstrating multiple nucleation events. If this were a filament, the RecA binding events should be much closer together according to the previous work that found a RecA periodicity of 10nm in a nucleoprotein filament, whereas figure 4.2B demonstrates that in this instance, there is a periodicity of ~0.2µm, a 20-fold difference. This further suggests a need for increased incubation time to see filamentation (the extension of individual binding events into a continuous filament).

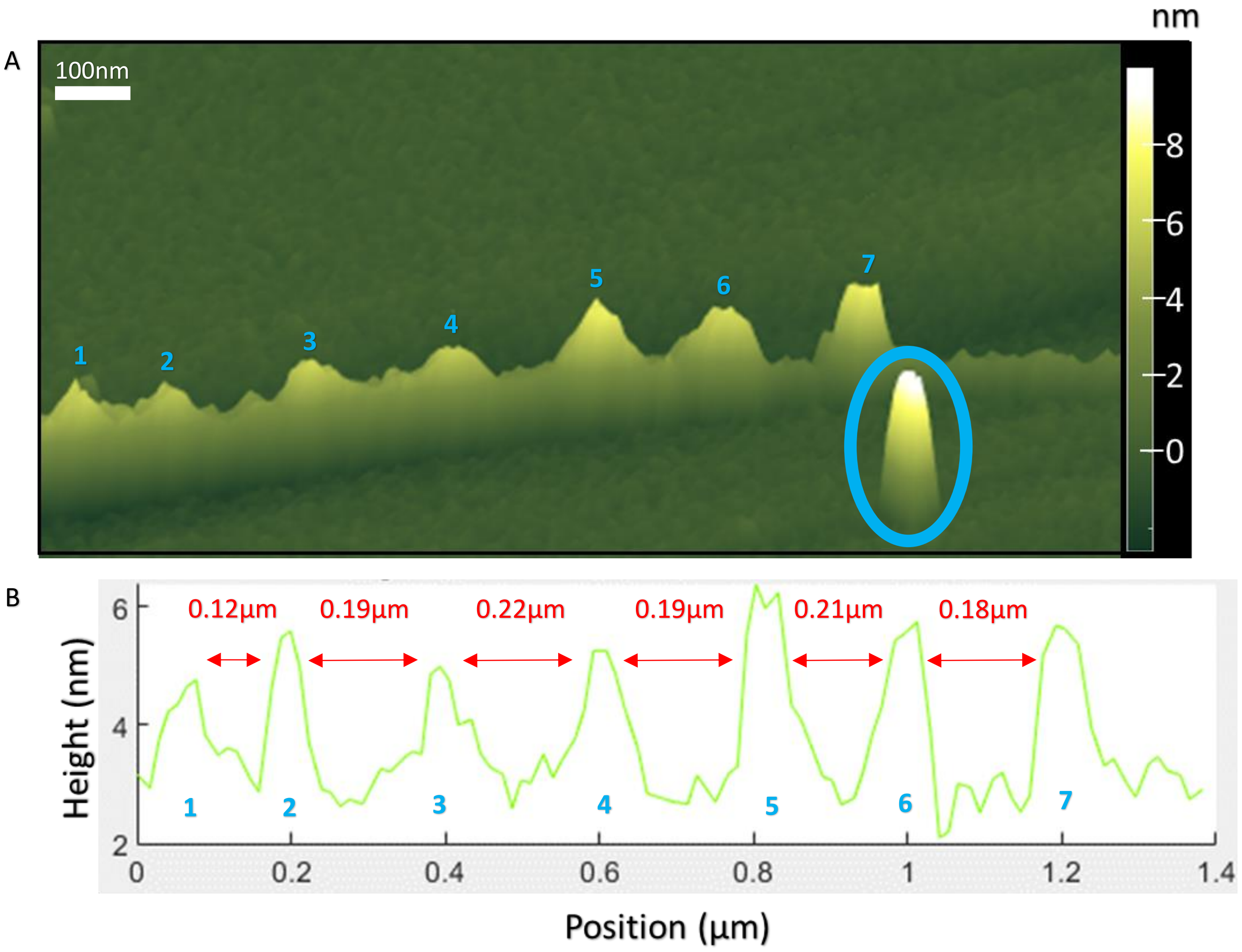


Figure 4.2. *RecA bound to linear dsDNA.* A) Representative 3D image of RecA bound to linearized pUC-19 plasmid DNA taken using a Bruker Bioscope Catalyst atomic force microscope with a ScanAsyst Air probe, 256 samples/line, 0.501Hz, processed using ImageMetrics software. The blue circle represents unbound RecA. Numbered peaks correlate to the same peaks in the trace in B) representing RecA bound to DNA. B) Cross-sectional height mapping of the peaks in A). The red numbers identify the distance between each peak's maxima, demonstrating a regularity of RecA-DNA binding of approximately 0.2µm between RecA molecules. Image generated in ImageMetrics v1.47 (Zimeng Li, UNC Chapel Hill)

These results represent the initial stages of data for a JoVE article. However, this work was interrupted by COVID-19, preventing protocol optimizations to address the limited observed binding events and further investigate potential discrepancies with previously published work. The data analysis is also limited as AFM data analysis relies on the statistical power of multiple observations of any given event. However, these preliminary results demonstrate the power of AFM in observing protein-DNA interactions and only minor optimizations should be required in order for a full analysis of RecA DNA-binding. An increased RecA-DNA incubation time should allow for full filamentation to occur, along with proper controls, that will provide the data required for statistically relevant analysis.

Chapter 5: Conclusion

Conclusions:

The work done on both RecA-DNA and ImuABC have brought these projects to points where they are primed for the next researcher to take over. Although the solubility of ImuABC proved a particularly challenging issue to solve, I believe the work I have done will provide a substantial amount of direction. By exploring a few possible solutions to the issue in a great depth, I've demonstrated a new approach is needed, and proposed what I believe to be a promising solution using native expression. The work done to demonstrate the power of AFM for studying protein-DNA interactions has established a working protocol for the AFM available at McMaster to image clean RecA-DNA complexes, it is simply a few minor optimizations and extended data collection away from ready for the JoVE article.

The bacterial Ku was where the most was accomplished, though. DNA binding assays allowed me to establish an unexpected relationship between DNA and the C-terminal domain of Ku. DNA binding assays demonstrated that the presence of the C-terminal domain reduced Ku's affinity for DNA substantially. Some residues appear to have a larger impact on this destabilization with D250A and R262A mutations demonstrating an even higher affinity than full C-terminal truncations, suggesting these residues are likely positioned in unfavorable orientations to limit DNA binding. I predict that this feature benefits bacteria by increasing Ku's freedom to thread freely when bound to DNA, which reduces the likelihood that Ku blocks other proteins accessing DNA. This model would benefit tremendously from a structure of Ku, particularly in complex with DNA, to best understand the interaction and facilitate the design of an inhibitor. I have crystallized C-terminally truncated Ku (both ΔExt C-term and ΔC-term, as well as ΔC-term with DNA), which is on its way to producing a structure. Unfortunately, these structures would lack some of the information regarding the C-terminal domain. However, the core domain isn't quite completely homologous to the eukaryotic core domain, thus even the core domain structure may provide valuable insight. The lack of crystallization to any degree in WT Ku supports predictions that the C-terminus of Ku is disordered, which correlates with the C-terminus having multiple roles and may make an inclusive structure impossible without techniques, such as crosslinking, to reduce Ku mobility in the crystal. I have established the basis for a deeper biophysical analysis of Ku using AFM by imaging Ku only and calculating the average Ku volume of 1870nm^3. This will serve as a comparison for Ku-DNA values, facilitating a more robust analysis. To simulate Ku inhibition and investigate NHEJ's relation to HR, I began creating a strain library of *P. aeruginosa* mutants deficient in either *ku*, *ligD*, or *recB* or a combination of these. This library will facilitate both expanding the *in vitro* DNA binding observations to *in vivo* phenotypes through complementation, as well as provide a means of studying the effects of NHEJ inhibition under multiple different conditions (such as stationary phase versus exponential growth or quinolone exposure). Ultimately, this work has laid the groundwork for future characterization of the Ku-DNA interaction. Most importantly though, the aforementioned DNA-binding results have provided direction to future antibiotic development studies by suggesting that a stable Ku-DNA complex would be bad for the bacterium. Aligning with what we know about quinolones turning topoisomerases into toxins, the DNA-binding results can be expanded upon by finishing the deletion library I started and simulating tight Ku binding using truncated Ku. I predict this will have an adverse effect on bacteria, potentially even more significantly in stationary phase bacteria when NHEJ is more active. I believed targeting NHEJ was promising before elucidating a mechanism that may turn Ku into the next topoisomerase due to a strong link to mutagenesis and thus the development of AMR traits. Having revealed a potential relationship that would allow Ku to be used as a weapon against bacteria the results of this research almost necessitate the beginning of Ku focused drug design research to fight against AMR.

References:

1. Hanai, R., Yazu, M. & Hieda, K. On the experimental distinction between ssbs and dsbs in circular DNA. *International Journal of Radiation Biology* **73**, 475–479 (1998).
2. Chatzipapas, K. P. *et al.* Quantification of <scp>DNA</scp> double-strand breaks using Geant4-<scp>DNA</scp>. *Medical Physics* **46**, mp.13290 (2018).
3. Mehta, A. & Haber, J. E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harbor Perspectives in Biology* **6**, (2014).
4. de Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169–185 (2004).
5. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nature Reviews Microbiology* vol. 10 449–458 (2012).
6. Ward, J. F. The complexity of DNA damage: Relevance to biological consequences. *International Journal of Radiation Biology* **66**, 427–432 (1994).
7. Chen, J. & Stubbe, J. A. Bleomycins: Towards better therapeutics. *Nature Reviews Cancer* vol. 5 102–112 (2005).
8. Dupuy, P., Sauviac, L. & Bruand, C. Stress-inducible NHEJ in bacteria: function in DNA repair and acquisition of heterologous DNA. *Nucleic Acids Research* **47**, 1335–1349 (2018).
9. González-Torres, P., Rodríguez-Mateos, F., Antón, J. & Gabaldón, T. Impact of homologous recombination on the evolution of prokaryotic core genomes. *mBio* **10**, (2019).
10. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).
11. Roth, D. B. V(D)J Recombination: Mechanism, Errors, and Fidelity. in *Mobile DNA III* vol. 2 313–324 (American Society of Microbiology, 2015).
12. Lieber, M. R. Mechanisms of human lymphoid chromosomal translocations. *Nature Reviews Cancer* vol. 16 387–398 (2016).
13. Malkova, A. & Haber, J. E. Mutations Arising During Repair of Chromosome Breaks. (2012) doi:10.1146/annurev-genet-110711-155547.
14. Jekimovs, C. *et al.* Chemotherapeutic compounds targeting the DNA double-strand break repair pathways: The good, the bad, and the promising. *Frontiers in Oncology* vol. 4 APR (2014).
15. Gutierrez, A., Stokes, J. M. & Matic, I. Our evolving understanding of the mechanism of quinolones. *Antibiotics* vol. 7 32 (2018).

16. Hong, Y., Zeng, J., Wang, X., Drlica, K. & Zhao, X. Post-stress bacterial cell death mediated by reactive oxygen species. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 10064–10071 (2019).

17. Amarh, V. & Arthur, P. K. DNA double-strand break formation and repair as targets for novel antibiotic combination chemotherapy. *Future Science OA* vol. 5 (2019).

18. Gong, C. *et al.* Mechanism of nonhomologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku, ligase D and ligase C. *Nature Structural and Molecular Biology* **12**, 304–312 (2005).

19. Pardo, B., Gómez-Gonzµlez, B., Aguilera, A. & Vespucio, A. DNA double-strand break repair: how to fix a broken relationship. *Cellular and Molecular Life Sciences* **66**, 1039–1056 (2009).

20. Ayora, S. *et al.* Double-strand break repair in bacteria: a view from Bacillus subtilis. *FEMS Microbiology Reviews* **35**, 1055–1081 (2011).

21. Hiom, K. Homologous recombination: How RecA finds the perfect partner. *Current Biology* vol. 22 R275–R278 (2012).

22. Forget, A. L. & Kowalczykowski, S. C. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. *Nature* **482**, 423–427 (2012).

23. Brandsma, I. & Gent, D. C. Pathway choice in DNA double strand break repair: Observations of a balancing act. *Genome Integrity* vol. 3 9 (2012).

24. Moeller, R. *et al.* Roles of the Major, Small, Acid-Soluble Spore Proteins and Spore-Specific and Universal DNA Repair Mechanisms in Resistance of Bacillus subtilis Spores to Ionizing Radiation from X Rays and High-Energy Charged-Particle Bombardment Downloaded from. *JOURNAL OF BACTERIOLOGY* **190**, 1134–1140 (2008).

25. Moeller, R. *et al.* Role of DNA repair by nonhomologous-end joining in Bacillus subtilis spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *Journal of Bacteriology* **189**, 3306–3311 (2007).

26. Aravind, L. & Koonin, E. v. 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 Research* **11**, 1365–1374 (2001).

27. Mcgovern, S. *et al.* C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. *Nucleic Acids Research* **44**, 4785–4806 (2016).

28. Sharda, M., Badrinarayanan, A., Sai, A. & Seshasayee, N. Evolutionary analysis of bacterial Non-Homologous End Joining Repair. *bioRxiv* 869602 (2019) doi:10.1101/869602.

29. Pannunzio, N. R., Watanabe, G. & Lieber, M. R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry* vol. 293 10512–10523 (2018).

30. Bertrand, C., Thibessard, A., Bruand, C., Lecointe, F. & Leblond, P. Bacterial NHEJ: a never ending story. *Molecular Microbiology* **111**, 1139–1151 (2019).

31. Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair. *Nature* **412**, 607–614 (2001).

32. Fell, V. L. & Schild-Poulter, C. The Ku heterodimer: Function in DNA repair and beyond. *Mutation Research - Reviews in Mutation Research* vol. 763 15–29 (2015).

33. Mimitou, E. P. & Symington, L. S. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *The EMBO Journal* **29**, 3358–3369 (2010).

34. Britton, S., Coates, J. & Jackson, S. P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *Journal of Cell Biology* **202**, 579–595 (2013).

35. Hammel, M. *et al.* Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *Journal of Biological Chemistry* **285**, 1414–1423 (2010).

36. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nature Cell Biology* **9**, 675–682 (2007).

37. Roberts, S. A. *et al.* Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* **464**, 1214–1217 (2010).

38. Goodarzi, A. A. *et al.* DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO Journal* **25**, 3880–3889 (2006).

39. Chang, H. H. Y., Watanabe, G. & Lieber, M. R. Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependentartemis resection atbluntdnaends. *Journal of Biological Chemistry* **290**, 24036–24050 (2015).

40. Kurosawa, A. *et al.* The Requirement of Artemis in Double-Strand Break Repair Depends on the Type of DNA Damage. *DNA and Cell Biology* **27**, 55–61 (2008).

41. Grundy, G. J. *et al.* APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO Journal* **32**, 112–125 (2013).

42. Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. *Cell* **134**, 981–994 (2008).

43. McElhinny, S. A. N. *et al.* A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Molecular Cell* **19**, 357–366 (2005).

44. Ma, Y. *et al.* A biochemically defined system for mammalian nonhomologous DNA end joining. *Molecular Cell* **16**, 701–713 (2004).

45. Bebenek, K., Pedersen, L. C. & Kunkel, T. A. Structure-function studies of DNA polymerase λ. *Biochemistry* **53**, 2781–2792 (2014).

46. Moon, A. F. *et al.* Sustained active site rigidity during synthesis by human DNA polymerase μ. *Nature Structural and Molecular Biology* **21**, 253–260 (2014).

47. Pryor, J. M. *et al.* Essential role for polymerase specialization in cellular nonhomologous end joining. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4537–E4545 (2015).

48. Mahaney, B. L., Hammel, M., Meek, K., Tainer, J. A. & Lees-Miller, S. P. XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochem Cell Bio.* **91**, 31–41 (2013).

49. Ochi, T. *et al.* PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347**, 185–188 (2015).

50. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301–313 (2006).

51. Crut, A., Nair, P. A., Koster, D. A., Shuman, S. & Dekker, N. H. Dynamics of phosphodiester synthesis by DNA ligase. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6894–6899 (2008).

52. Mcwilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* **41**, W597–W600 (2013).

53. de Ory, A., Zafra, O. & de Vega, M. Efficient processing of abasic sites by bacterial nonhomologous end-joining Ku proteins. *Nucleic Acids Research* **42**, 13082–13095 (2014).

54. Kushwaha, A. K. & Grove, A. Mycobacterium smegmatis Ku binds DNA without free ends. *Biochemical Journal* **456**, 275–282 (2013).

55. Zhu, H. & Shuman, S. Gap filling activities of Pseudomonas DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *Journal of Biological Chemistry* **285**, 4815–4825 (2010).

56. Pitcher, R. S., Tonkin, L. M., Green, A. J. & Doherty, A. J. Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. *Journal of Molecular Biology* **351**, 531–544 (2005).

57. Wright, D., DeBeaux, A., Shi, R., Doherty, A. J. & Harrison, L. Characterization of the roles of the catalytic domains of Mycobacterium tuberculosis ligase D in Ku-dependent error-prone DNA end joining. *Mutagenesis* **25**, 473 (2010).

58. Yakovleva, L. & Shuman, S. Nucleotide misincorporation, 3'-mismatch extension, and responses to abasic sites and DNA adducts by the polymerase component of bacterial DNA ligase D. *Journal of Biological Chemistry* **281**, 25026–25040 (2006).

59. Pitcher, R. S. *et al.* Structure and Function of a Mycobacterial NHEJ DNA Repair Polymerase. *Journal of Molecular Biology* **366**, 391–405 (2007).

60. Pitcher, R. S., Brissett, N. C. & Doherty, A. J. Nonhomologous End-Joining in Bacteria: A Microbial Perspective. *Annual Review of Microbiology* **61**, 259–282 (2007).

61. Gong, C., Martins, A., Bongiorno, P., Glickman, M. & Shuman, S. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *Journal of Biological Chemistry* **279**, 20594–20606 (2004).

62. Ukkivi, K. & Kivisaar, M. Involvement of transcription-coupled repair factor Mfd and DNA helicase UvrD in mutational processes in Pseudomonas putida. *DNA Repair* **72**, 18–27 (2018).

63. Sinha, K. M., Stephanou, N. C., Gao, F., Glickman, M. S. & Shuman, S. Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *Journal of Biological Chemistry* **282**, 15114–15125 (2007).

64. Li, Z., Wen, J., Lin, Y., Wang, S. & Xue, P. A Sir2-Like Protein Participates in Mycobacterial NHEJ. *PLoS ONE* **6**, 20045 (2011).

65. Zhou, Y. *et al.* Discovery and characterization of Ku acetylation in Mycobacterium smegmatis. *FEMS Microbiology Letters* **362**, (2015).

66. *TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY: FINAL REPORT AND RECOMMENDATIONS THE REVIEW ON ANTIMICROBIAL RESISTANCE CHAIRED BY JIM O’NEILL*. (2016).

67. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*. (2016).

68. Taylor, J. *et al.* Estimating the economic costs of antimicrobial resistance: Model and Results. (2014).

69. Kinch, M. S., Patridge, E., Plummer, M. & Hoyer, D. An analysis of FDA-approved drugs for infectious disease: Antibacterial agents. *Drug Discovery Today* vol. 19 1283–1287 (2014).

70. Deak, D., Outterson, K., Powers, J. H. & Kesselheim, A. S. Progress in the fight against multidrug-resistant bacteria? A review of U.S. food and drug administration-approved antibiotics, 2010-2015. *Annals of Internal Medicine* vol. 165 363–372 (2016).

71. Boucher, H. W. *et al.* Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clinical Infectious Diseases* **48**, 1–12 (2009).

72. Baharoglu, Z. & Mazel, D. Vibrio cholerae triggers SOS and mutagenesis in response to a wide range of antibiotics: A route towards multiresistance. *Antimicrobial Agents and Chemotherapy* **55**, 2438–2441 (2011).

73. Tyers, M. & Wright, G. D. Drug combinations: a strategy to extend the life of antibiotics in the 21st century. *Nature Reviews Microbiology* **17**, 141–155 (2019).

74. Silver, L. L. Multi-targeting by monotherapeutic antibacterials. *Nature Reviews Drug Discovery* vol. 6 41–55 (2007).

75. Lim, C. S. Q. *et al.* Identification of a potent small-molecule inhibitor of bacterial DNA repair that potentiates quinolone antibiotic activity in methicillin-resistant Staphylococcus aureus. *Bioorganic and Medicinal Chemistry* **27**, (2019).

76. Wilkinson, M. *et al.* Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. *eLife* **5**, (2016).

77. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: From targets to networks. *Nature Reviews Microbiology* vol. 8 423–435 (2010).

78. Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T. & Tomizawa, J. I. Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4772–4776 (1977).

79. Sugino, A., Peebles, C. L., Kreuzer, K. N. & Cozzarelli, N. R. Mechanism of action of nalidixic acid: Purification of Escherichia coli nalA gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4767–4771 (1977).

80. Mizuuchi, K., Fisher, L. M., O'Dea, M. H. & Gellert, M. DNA gyrase action involves the introduction of transient double-strand breaks into DNA. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1847–1851 (1980).

81. Marians, K. J. & Hiasa, H. Mechanism of quinolone action. A drug-induced structural perturbation of the DNA precedes strand cleavage by topoisomerase IV. *Journal of Biological Chemistry* **272**, 9401–9409 (1997).

82. Chen, C. R., Malik, M., Snyder, M. & Drlica, K. DNA gyrase and topoisomerase IV on the bacterial chromosome: Quinolone-induced DNA cleavage. *Journal of Molecular Biology* **258**, 627–637 (1996).

83. Alt, S., Mitchenall, L. A., Maxwell, A. & Heide, L. Inhibition of DNA gyrase and DNA topoisomerase IV of Staphylococcus aureus and Escherichia coli by aminocoumarin antibiotics. *Journal of Antimicrobial Chemotherapy* **66**, 2061–2069 (2011).

84. Schroder, W., Goerke, C. & Wolz, C. Opposing effects of aminocoumarins and fluoroquinolones on the SOS response and adaptability in Staphylococcus aureus. *Journal of Antimicrobial Chemotherapy* **68**, 529–538 (2013).

85. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810 (2007).

86. Hemnani, T. & Parihar, M. S. *REACTIVE OXYGEN SPECIES AND OXIDATIVE DNA DAMAGE*. *Indian J Physiol Pharmacol* vol. 42 (1998).

87. Cabiscol Català, E., Tamarit Sumalla, J. & Ros Salvador, J. *Oxidative stress in bacteria and protein damage by reactive oxygen species*. https://repositori.udl.cat/handle/10459.1/56751 (2000).

88. Becerra, M. C. & Albesa, I. Oxidative stress induced by ciprofloxacin in Staphylococcus aureus. *Biochemical and Biophysical Research Communications* **297**, 1003–1007 (2002).

89. Waddell, S. J. *et al.* The use of microarray analysis to determine the gene expression profiles of Mycobacterium tuberculosis in response to anti-bacterial compounds. *Tuberculosis* **84**, 263–274 (2004).

90. Belenky, P. *et al.* Bactericidal Antibiotics Induce Toxic Metabolic Perturbations that Lead to Cellular Damage. *Cell Reports* **13**, 968–980 (2015).

91. Vilchèze, C. *et al.* Enhanced respiration prevents drug tolerance and drug resistance in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4495–4500 (2017).

92. Keren, I., Wu, Y., Inocencio, J., Mulcahy, L. R. & Lewis, K. Killing by bactericidal antibiotics does not depend on reactive oxygen species. *Science* **339**, 1213–1216 (2013).

93. Rowe, S. E. *et al.* Reactive oxygen species induce antibiotic tolerance during systemic Staphylococcus aureus infection. *Nature Microbiology* vol. 5 282–290 (2020).

94. Amundsen, S. K., Fero, J., Salama, N. R. & Smith, G. R. Dual nuclease and helicase activities helicobacter pylori AddAB are required for DNA repair, recombination, and mouse infectivity. *Journal of Biological Chemistry* **284**, 16759–16766 (2009).

95. Henderson, M. L. & Kreuzer, K. N. Functions that protect Escherichia coli from tightly bound DNA-protein complexes created by mutant EcoRII methyltransferase. *PLoS ONE* **10**, (2015).

96. Role of double strand DNA break repair for quinolone sensitivity in Escherichia coli: therapeutic implications. *Rev Esp Quimioter* **28**, 139–144 (2015).

97. Dharmaraja, A. T. Role of Reactive Oxygen Species (ROS) in Therapeutics and Drug Resistance in Cancer and Bacteria. (2017) doi:10.1021/acs.jmedchem.6b01243.

98. O'Rourke, E. J. *et al.* Pathogen DNA as target for host-generated oxidative stress: Role for repair of bacterial DNA damage in Helicobacter pylori colonization. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2789–2794 (2003).

99. Foti, J. J., Devadoss, B., Winkler, J. A., Collins, J. J. & Walker, G. C. Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* **336**, 315–319 (2012).

100. Burby, P. E. & Simmons, L. A. A bacterial DNA repair pathway specific to a natural antibiotic. *Molecular Microbiology* **111**, 338–353 (2019).

101. Sartorelli, A. C., Tomasz, M. & Rockwell, S. Studies on the mechanism of the cytotoxic action of the mitomycin antibiotics in hypoxic and oxygenated EMT6 cells. *Advances in Enzyme Regulation* **33**, 3–17 (1993).

102. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* vol. 13 42–51 (2015).

103. Beaber, J. W., Hochhut, B. & Waldor, M. K. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72–74 (2004).

104. Cirz, R. T. *et al.* Inhibition of Mutation and Combating the Evolution of Antibiotic Resistance. doi:10.1371/journal.pbio.0030176.

105. Alves, I. R. *et al.* Effect of SOS-induced levels of imuABC on spontaneous and damage-induced mutagenesis in Caulobacter crescentus. *DNA Repair* **59**, 20–26 (2017).

106. Popa, O., Hazkani-Covo, E., Landan, G., Martin, W. & Dagan, T. Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Research* **21**, 599–609 (2011).

107. Kivisaar, M. Mechanisms of stationary-phase mutagenesis in bacteria: Mutational processes in pseudomonads. *FEMS Microbiology Letters* vol. 312 1–14 (2010).

108. Boshoff, H. I. M., Reed, M. B., Barry, C. E. & Mizrahi, V. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* **113**, 183–193 (2003).

109. Cromie, G. A., Connelly, J. C. & Leach, D. R. F. Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans. *Molecular Cell* vol. 8 1163–1174 (2001).

110. Ponder, R. G., Fonville, N. C. & Rosenberg, S. M. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Molecular Cell* **19**, 791–804 (2005).

111. Kohanski, M. A., DePristo, M. A. & Collins, J. J. Sublethal Antibiotic Treatment Leads to Multidrug Resistance via Radical-Induced Mutagenesis. *Molecular Cell* **37**, 311–320 (2010).

112. Thi, T. D. *et al.* Effect of recA inactivation on mutagenesis of Escherichia coli exposed to sublethal concentrations of antimicrobials. *Journal of Antimicrobial Chemotherapy* **66**, 531–538 (2011).

113. Mesak, L. R., Miao, V. & Davies, J. Effects of subinhibitory concentrations of antibiotics on SOS and DNA repair gene expression in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* **52**, 3394–3397 (2008).

114. Baharoglu, Z., Krin, E. & Mazel, D. RpoS Plays a Central Role in the SOS Induction by Sub-Lethal Aminoglycoside Concentrations in Vibrio cholerae. *PLoS Genetics* **9**, (2013).

115. Gutierrez, A. *et al.* β-lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nature Communications* **4**, 1610 (2013).

116. Ynés Valencia, E., Esposito, F., Spira, B., Blázquez, J. & Galhardo, R. S. Ciprofloxacin-Mediated Mutagenesis Is Suppressed by Subinhibitory Concentrations of Amikacin in Pseudomonas aeruginosa. (2017) doi:10.1128/AAC.02107-16.

117. Paris, Ü. *et al.* NHEJ enzymes LigD and Ku participate in stationary-phase mutagenesis in Pseudomonas putida. *DNA Repair* **31**, 11–18 (2015).

118. Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A. & Leblond, P. Genome plasticity is governed by double strand break DNA repair in Streptomyces. *Scientific Reports* **8**, (2018).

119. Nagel, M., Reuter, T., Jansen, A., Szekat, C. & Bierbaum, G. Influence of ciprofloxacin and vancomycin on mutation rate and transposition of IS256 in Staphylococcus aureus. *International Journal of Medical Microbiology* **301**, 229–236 (2011).

120. Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B. & Claverys, J. P. Antibiotic stress induces genetic transformability in the human pathogen streptoccus pneumoniae. *Science* **313**, 89–92 (2006).

121. Slager, J., Kjos, M., Attaiech, L. & Veening, J. W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157**, 395–406 (2014).

122. Overballe-Petersen, S. *et al.* Bacterial natural transformation by highly fragmented and damaged DNA. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19860–19865 (2013).

123. Gengenbacher, M. & Kaufmann, S. H. E. Mycobacterium tuberculosis: Success through dormancy. *FEMS Microbiology Reviews* vol. 36 514–532 (2012).

124. Oates, M. E. *et al.* D 2 P 2 : database of disordered protein predictions. *Nucleic Acids Research* **41**, D508–D516 (2013).

125. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* vol. 16 18–29 (2015).

126. WHO | Global tuberculosis report 2019. *WHO* (2020).

127. Oldfield, C. J. *et al.* Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1834**, 487–498 (2013).

128. Su, X. D. *et al.* Protein crystallography from the perspective of technology developments. *Crystallography Reviews* vol. 21 122–153 (2015).

129. Weichenberger, C. X., Afonine, P. v., Kantardjieff, K. & Rupp, B. The solvent component of macromolecular crystals. *Acta Crystallographica Section D: Biological Crystallography* **71**, 1023–1038 (2015).

130. Heras, B. & Martin, J. L. Biological Crystallography Post-crystallization treatments for improving diffraction quality of protein crystals. *topical reviews Acta Cryst* **61**, 1173–1180 (2005).

131. Cousido-Siah, A. *et al.* Crystal packing modifies ligand binding affinity: The case of aldose reductase. *Proteins: Structure, Function and Bioinformatics* **80**, 2552–2561 (2012).

132. Dyer, K. N. *et al.* High-throughput SAXS for the characterization of biomolecules in solution: A practical approach. *Methods in Molecular Biology* **1091**, 245–258 (2014).

133. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* **50**, 1212–1225 (2017).

134. Hmelo, L. R. *et al.* Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. *Nature Protocols* **10**, 1820–1841 (2015).

135. Galhardo, R. S., Rocha, R. P., Marques, M. v & Menck, C. F. M. An SOS-regulated operon involved in damage-inducible mutagenesis in Caulobacter crescentus. *Nucleic Acids Research* **33**, 2603–2614 (2005).

136. Jatsenko, T., Sidorenko, J., Saumaa, S. & Kivisaar, M. DNA Polymerases ImuC and DinB Are Involved in DNA Alkylation Damage Tolerance in Pseudomonas aeruginosa and Pseudomonas putida. *PLoS ONE* **12**, (2017).

137. Luján, A. M. *et al.* ImuB and ImuC contribute to UV-induced mutagenesis as part of the SOS regulon in Pseudomonas aeruginosa. *Environmental and Molecular Mutagenesis* **60**, 594–601 (2019).

138. Warner, D. F. *et al.* Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13093–13098 (2010).

139. Anye, V. *Structural Analysis of Induced Mutagenesis A' Protein from Mycobacterium tuberculosis and of a Thermophillic GH9 Cellulase*. (2014).

140. Boonzaier, J. *Structural Analysis of Induced Mutagenesis Protein B from Mycobacterium tuberculosis*. http://etd.uwc.ac.zahttp//etd.uwc.ac.za (2016).

141. Ippoliti, P., DeLateur, N., Jones, K. & Beuning, P. Multiple Strategies for Translesion Synthesis in Bacteria. *Cells* **1**, 799–831 (2012).

142. Vaisman, A., McDonald, J. P. & Woodgate, R. Translesion DNA Synthesis. *EcoSal Plus* **5**, (2012).

143. Patel, M., Jiang, Q., Woodgate, R., Cox, M. M. & Goodman, M. F. A new model for SOS-induced mutagenesis: How RecA protein activates DNA polymerase v. *Critical Reviews in Biochemistry and Molecular Biology* vol. 45 171–184 (2010).

144. Erill, I., Campoy, S., Mazon, G. & Barbé, J. Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Research* **34**, 66–77 (2006).

145. Jadaun, A., Sudhakar D, R., Subbarao, N. & Dixit, A. In Silico Screening for Novel Inhibitors of DNA Polymerase III Alpha Subunit of Mycobacterium tuberculosis (MtbDnaE2, H 37 R v). *PLoS ONE* **10**, (2015).

146. Churion, K. A. & Bondos, S. E. Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification. *Methods in Molecular Biology* **896**, 415–427 (2012).

147. Sattin, B. D. & Goh, M. C. Direct observation of the assembly of RecA/DNA complexes by atomic force microscopy. *Biophysical Journal* **87**, 3430–3436 (2004).

148. Michel, B. After 30 Years of Study, the Bacterial SOS Response Still Surprises Us. *PLoS Biology* **3**, e255 (2005).

149. West, S. C., Cassuto, E., Mursalim, J. & Howard-Flanders, P. Recognition of duplex DNA containing single-stranded regions by recA protein. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 2569–2573 (1980).

150. Shivashankar, G. v., Feingold, M., Krichevsky, O. & Libchaber, A. RecA polymerization on double-stranded DNA by using single-molecule manipulation: The role of ATP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7916–7921 (1999).

151. Lusetti, S. L. & Cox, M. M. The Bacterial RecA Protein and the Recombinational DNA Repair of Stalled Replication Forks. *Annual Review of Biochemistry* **71**, 71–100 (2002).

152. Andres, S. N., Li, Z. M., Erie, D. A. & Williams, R. S. Ctp1 protein–DNA filaments promote DNA bridging and DNA double-strand break repair. *Journal of Biological Chemistry* **294**, 3312–3320 (2019).

Appendix:

Chapter 2 Materials and Methods

Cloning

A codon optimized for *Escherichia coli* Ku gene was ordered as a geneblock (IDT Inc.). Primers were designed to amplify and add ligation independent cloning (LIC) sites to either end. To begin LIC, empty pMCSG-7 vector was digested with SSPI for 6 hours at 37ºC then purified using a PCR cleanup kit (BioBasic Inc.). Then 2 concurrent reactions are carried out, one for the gene fragment with LIC ends, and one for the digested plasmid. For the gene fragment a 20µL of reaction was setup with 0.2pmol of the gene, 2µL of 10X T4 DNA polymerase buffer, 1µL of 100mM DTT, 2µL of 25mM dCTP, 0.4µL (1U) of T4 DNA polymerase (NEB Ltd.), and ddH20 to 20µL. For the digested plasmid a 20µL of reaction was setup with 100ng of the plasmid, 2µL of 10X T4 DNA polymerase buffer, 1µL of 100mM DTT, 2µL of 25mM dGTP, 0.4µL (1U) of T4 DNA polymerase (NEB Ltd.), and ddH20 to 20µL. Both reactions are progressed at room temperature for 30 minutes and then heat inactivated at 70ºC for 20 minutes after which they were combined in entirety and put back at 70ºC for 2 minutes. After the 2 minutes, the mixture was rested at room temperature for 2 minutes to allow annealing, and transformed in Top10 *Escherichia coli*, using the whole 40µL with 50µL of competent cells. Cells were grown on LB agar plates containing ampicillin overnight at 37ºC, which selected for successful transformants using pMCSG-7 antibiotic resistance tag. Successful colonies were picked using a pipette tip which was dipped into 10mL of LB media with ampicillin to grow overnight on a shaking incubator at 37ºC. Plasmid was purified from the overnight culture using Plasmid DNA extraction miniprep kits (BioBasic Inc.) then used in a PCR reaction with the initial primers to confirm the presence of the cloned fragment. The reaction mixture after PCR was run on a 1% agarose gel at 100V in TBE buffer for 40 minutes to confirm the presence of a band at the appropriate size for the Ku gene validating successful cloning. To produce the various constructs, primers were designed for site directed mutagenesis to produce versions of the gene which were cloned using the same protocol.

Protein Purifications

Competent BL21 Codon plus *Escherichia coli* cells were transformed with the plasmid containing the desired construct and then plated on LB agar plates to grow at 37ºC overnight. A colony is picked from the plate using a pipette tip then the whole tip is dropped into 600mL of AIM media (recipe available below) in a Fernbach flask and incubated in a shaking incubator at 27ºC for 72 hours. Cells were collected via centrifugation at 4ºC, 9000xg for 10 minutes. Cell pellets were resuspended in 35mL of lysis buffer (recipe below) and allowed to lyse for 20 minutes on ice. The solution was sonicated at an output wattage of approximately 20 for 6 cycles of 20 seconds on, 20 seconds off. The lysate was fractionated by centrifugation at 4ºC, 40000xg for 30 minutes. The soluble fraction was incubated with 5 mL of Ni-NTA resin per 35mL of sample on ice while gently shaking for 30 minutes. The mixture was loaded onto a gravity-flow column, where the beads settled at the bottom. Three successive washes were done, first 15 column volumes (CV) of lysis buffer, then 5CV of Ni-NTA buffer (50mM tris pH 8, 0.4M NaCl, 10% glycerol, 1mM BME) + 20mM imidazole, finally 5CV of Ni-NTA buffer plus 40mM imidazole. The protein was eluted using 25mL Ni-NTA buffer plus 400mM imidazole. The last two purification steps use an AKTA Explorer FPLC, first anion exchange on a 5mL HiTrap Q-sepharose column (GE Healthcare), then size exclusion chromatography (SEC) on a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). The anion exchange column was equilibrated with Q_A buffer (20mM tris pH 8, 2mM BME) and the sample was diluted 5-fold in Q_A before being loaded onto the column. The protein was eluted gradually through the introduction of a progressive gradient of Q_B (20mM tris pH 8, 1M NaCl, 2mM BME). This consistently produced a singular A_{280} peak so all fractions which produced an A_{280} reading were pooled and then concentrated using an Amicon Ultra-15 Centrifugal Filter unit (Millipore Sigma), spun at 4ºC, 4000RPM, for 10 minute intervals until either protein began to precipitate out of solution or until there was approximately 5mL of solution remaining. The SEC column was equilibrated using Ku storage buffer (20mM tris pH 8, 0.4M NaCl, 10% glycerol) then 5mL of sample was loaded via manual injection per run. Once again, fractions with any A_{280} were pooled and concentrated using an Amicon Ultra-15 Centrifugal Filter unit (Millipore Sigma) until protein begins to crash out of solution. For storage, the protein was aliquot into 50µL, 100µL, and 1mL, and flash frozen using liquid nitrogen to be stored at -80ºC.

Crystal screening and optimization

Broad screening

Multiple broad screens were used, including the JCSG II, PEGS II and Classics I screen (Qiagen). Each condition was added to the appropriate well of a 96-well deep well block for use with the Phenix crystallization robot (Art Robbins Instruments). Ku constructs removed from deep freeze storage and diluted to a high medium and low concentrations in PCR tubes with storage buffer. The concentrations varied depending on the construct due to a difference in solubility. Initially wild-type Ku was screened at 7, 5, and 2, mg/mL as 7mg/mL appeared to be the maximal solubility, ΔExt C-term Ku was screened at 8, 5, and 2, mg/mL for a similar reason, and ΔC-term Ku was screened at 10, 5, and 2, mg/mL, this was later changed to 16, 12, and 8, mg/mL as this construct appeared far more soluble and better results were observed at the higher concentrations. The total volume of each concentration was made to be 25*(number of screens)µL + 50µL. Similarly for ΔExt C-term Ku plus DNA screens high, medium, and low concentrations were made, however the DNA substrate was limiting to the concentrations, the desired ratio of Ku:DNA was 1:0.6 due to Ku functioning as a dimer, so the tested ratios were 166.7µM (5mg/mL): 100µM, 133.3µM (4mg/mL):80µM, and 100µM (3mg/mL):60µM. The Phenix crystallization robot then plated 250nL of screen and 250nL of protein (or protein+DNA) into each of 3 sitting drop locations in each of 96 wells of an intelli 3-well plate, the wells were filled with 50 µL of screen. These plates were sealed with tape and either stored at 25 or 4ºC.

Optimization

Visual inspection of the broad screen plates identified promising conditions with some degree of crystallization that were chosen for optimization. The first step of optimization was identifying the conditions of a hit by examining the associated table to the screen, all hits contained 3 components: A salt, a buffer, and a precipitant. 200µL optimization screens were created for the hit by adjusting the buffer pH, the salt concentration, and the precipitant concentration, screening at levels below and above the initial hit condition. When this new screen is made, crystal plates are set in an identical fashion to the broad screen instead substituting out the broad screen for the custom screen. The second type of optimization used the Additive Screen (Hampton Research), a purchased set of 96 unique conditions shown to aid crystallization. The identical condition to the hit was replicated in an Intell-3-well crystallization plate (Art Robbins Instruments) with the addition of 1 new component per well from the expansion screen to a final volume of 200µL. These screens were also plated in the same way as broad screens, substituting out the broad screen for the custom one. The final type of optimization involved switching the condition to from a sitting drop to hanging drop crystal tray, where ratios of protein:crystallization screen were altered and total volume of the drop was increased to produce bigger crystals. These trays were set by hand with 500µL of precipitant added to the well and drop sizes ranging from 2-6µL.

Screening for Protein Crystals with Crystal Violet

Promising crystal screen hits were stained using 0.1% crystal violet. 5% of the initial drop volume of 0.1% crystal violet was added to the drop (example: 0.2µL crystal violet in a 4µL drop) and incubated at the temperature the tray had been at for 1 hour. Crystals were then viewed under a stereo microscope to confirm the dye had concentrated in the crystals, validating they are protein.

Crystal Screening

Single crystals were picked from crystal trays using a nylon crystal loops (Hampton Research). The crystal was immediately submerged in liquid nitrogen, then mounted on the goniometer of the xray diffractometer (Rigaku MicroMax-007 HF) in a stream of nitrogen gas cooled to a temperature of 100K. The crystal position was adjusted to center the beam path on the crystal. HKL3000 software was used to control the diffractometer. 2 different angles were screened for each crystal (0^o and 90^o), the detector distance was set to 150mm, and the exposure time was set to 120s.

SAXS sample prep

Protein was removed from storage in 1mL aliquots, 500µL of which was re-purified using the SEC column equilibrated with storage buffer lacking glycerol. The 2 fractions encompassed by the maxima of the A_{280} were pooled and the concentrations were adjusted to 1, 3, and 5, mg/mL for the low medium and high concentrations. For buffer 1 and 2, the buffer flow through from a fraction far from those where Ku was eluted was used. For the Ku plus DNA samples, the remaining 500µL aliquot was used to make 1:1 mixtures of protein and DNA, which then was loaded onto the SEC column and prepared the same as protein only samples. For each of the 5 samples for each construct 20µL of sample was shipped to the SIBYLS beamline (ALS, Berkeley, California) for SAXS data collection.

DNA binding Assays

A 10% Native-PAGE gel was set up and pre-run using 0.5x TBE buffer at 200V for 30 minutes. While the gel was running, protein was removed from storage and diluted to 200µM using EMSA buffer (0.05M tris pH 8, 0.05M NaCl, 1mM DTT, 5mM $MgCl_2$, 4% glycerol, 0.01mg/ml BSA) then 2-fold serially diluted down to 10nM. The final volume of each reaction was 40µL using 4µL of 10xEMSA buffer, 4µL of each concentration of protein (resulting in a final 10-fold dilution bringing concentrations down to 20µM to 1nM), 2µL of 200nM FAM-labelled DNA (final concentration 10nM), and 30µL ddH_2O. The reactions were then incubated at 30ºC for 20 minutes. The samples were then loaded onto the gel in 0.5x TBE, 15µL per well, and run at 180V for 1 hour. The gel was imaged using the Typhoon imager (GE Healthcare).

DNA binding assays data analysis

Intensity of the unbound DNA was quantified using ImageJ (https://imagej.nih.gov/) and used to calculate the fraction of DNA bound by Ku. Plots of DNA bound against protein concentration were plotted and used to calculate the dissociation constant (GraphPad Prism).

Allelic exchange work

Primers were designed to amplify approximately 500bp regions flanking either end of *ku* and *recB* in *P. aeruginosa* PA01. Primers farthest from the gene of interest (P1 and P4) were designed to introduce restriction sites, HindIII upstream (P1) and XbaI downstream (P4) to facilitate cloning at a later stage. The middle primers each contained overlapping regions to the gene of interest in a manner that the first 3 and last 3 codons of the gene are preserved in frame, on top of this overlap, primers 2 and 3 were designed to have an overlapping complimentary sequence with one another to facilitate SOE PCR, this region was designed to be in frame with 3 codons preserved on either end of the gene. The flanking regions were amplified from purified *Pseudomonas aeruginosa* genomic DNA using PCR. The reaction mixture for each fragment was run on a 1% agarose gel at 100V for 40 minutes, appropriately sized fragments were gel extracted using a gel extraction kit (BioBasic Inc.). Purified fragments were used for SOE PCR, approximately 25ng of either fragment along with primers 1 and 4 were combined with traditional PCR components for a PCR reaction. These reaction mixtures were run on a 1% agarose gel at 100V for 40 minutes, appropriately sized fragments (sum of two initial fragments) were gel extracted using a gel extraction kit (BioBasic Inc.).

Chapter 3 Materials and Mehods

Cloning

Two methods for cloning were used, LIC cloning mimicking that used for *ku* described in chapter 2.1, and traditional restriction digest cloning. For the latter, primers were designed to introduce restriction digest sites on either end of the *imu* gene of interest via PCR. Gene fragments and plasmids were digested overnight with the appropriate restriction enzymes at 37ºC. Fragments were purified using aPCR cleanup kit (BioBasic Inc.) then mixed in a 1:3 ration (plasmid:gene) and ligated using T4 DNA ligase (NEB Ltd.). Ligation reaction mixtures were then transformed in the same way as the mixture of the 2 LIC components and the remaining steps for cloning were carried out the same.

Expression testing

Competent cells from a combination of 5 strains of *Escherichia coli* (BL21, BL21*, BL21 Codon+, Rosetta, and SoluBL21) were transformed with plasmid containing the *imu* gene constructs and grown in 10mL of LB media plus antibiotic at 37ºC. From each overnight 1mL was used to inoculate two 50mL cultures in LB plus antibiotic, these were grown at 37ºC on a shaking incubator until OD=1 at which point 500µL samples were removed and they were induced with 50µL of 1M IPTG. Each strain had 1 50mL culture then moved back to 37ºC while the other to a 16ºC incubator, samples were taken from both at the one- and two-hour marks. The 37ºC culture was removed from the incubator after 3 hours at which point a fourth sample was taken, the 16ºC culture was removed from the incubator after 20 hours at which point a fourth sample was taken. All of these 500µL samples are spun down at 6000RPM for 10 minutes to pellet the cells. These pellets are resuspended and lysed in 50µL of 2x SDS-load dye. Each of the 4 samples for each condition were loaded (10µL) onto and run side by side on a 10% SDS-PAGE gel in Tris-glycine buffer at 200V for 40 minutes to identify conditions in which an appropriate size band for the construct appeared after induction. To test AIM expression, the protein purification process from chapter 2.2 was used.

Chapter 4 Materials and Methods

Sample Preparation for Atomic Force Microscopy

RecA sample preparation

Pure RecA was purchased from NEB biolabs. 10µL of RecA was transferred into reaction buffer (10 mM Tris, pH = 7, 7mM $MgCl_2$, 100 mM NaCl) in a final volume 100µL. This 0.2mg/mL RecA was mixed in an 8:6:1 ratio with AMPPNP (Sigma-Aldrich) diluted to 1mg/mL and 15nM linearized pUC-19 plasmid (digested with HindIII overnight and PCR-cleanup kit purified (BioBasic Inc.)). This mixture was incubated at room temperature for 30 minutes after which 3µL was transferred into 19.5µL of deposition buffer (20 mM TRIS, pH=7.5, 5 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 2 mM $ZnCl_2$). This was then ready to be spotted onto mica.

Ku sample preparation

Wild-type Ku was removed from frozen stocks, thawed, and diluted in storage buffer (see protein purification in chapter 2) to a final volume of 2mL. This 2mL sample was loaded onto an AKTA Explorer FPLC and run on an SEC column equilibrated with Ku AFM buffer (40mM HEPES pH 7, 200mM NaCl 5mM MgCl 1mM DTT). Fraction volumes were set to 0.5mL, the 2 fractions encompassing the most intense part of the A280 peak were pooled, the concentration was calculated using a DeNovix spectrophotometer. This was adjusted to 100nM using Ku AFM buffer. For Ku only images, this was further diluted to 25nM at which point it was ready to be spotted onto mica. For Ku with DNA images, the 100nM Ku was mixed in a 1 to 1 ratio with 2nM linearized pUC-19 plasmid (digested with HindIII overnight and PCR-cleanup kit purified (BioBasic Inc.)) and incubated at room temperature for 20 minutes at which time it was ready to be spotted onto mica.

Preparing the AFM sample

15mm, V1 Mica disks (Ted Pella Inc.) were attached onto magnetic metal disks with hot glue. The mica was then cleaved using Scotch tape. Immediately after cleaving, 20µL of sample was spotted onto the mica and left for 5 minutes to facilitate binding. Excess sample was removed by rinsing the mica with 1mL of HPLC grade water. The bulk of excess water was removed using filter paper, and fully dried with a gentle stream of nitrogen gas. These were then ready for AFM imaging.

Data Collection

Prepared samples were mounted on magnetic glass slides and placed in the sample holder of the atomic force microscope (Bruker BioScope Catalyst). A ScanAsyst tip (Bruker) was then mounted on the cantilever of the microscope which was then mounted on the microscope. The laser was then activated by opening the NanoScope Analysis software (Bruker) and launching a large amplitude ScanAsyst air experiment. The laser was then aligned onto the tip using the Bruker easy align tool and the adjustment knobs on the microscope were used to center the beam onto the detector. The microscope was then placed over the sample stage and using the manual controls, the tip is lowered until just above the sample. To begin scanning, the engage function was selected which lowers the tip the last of the distance to the sample stage. To scan for features, the scan radius was set to 10µm, the samples per line was set to 128, and the scan rate was set to 0.5Hz. To find features the x-y offsets were changed in increments of 15µm. When a feature was found, the scan radius was changed to 1µm for protein only and 2µm for protein plus DNA and it was centered on the feature. The samples per line was increased to 256 and the scan rate to 1Hz to improve the image quality. Other variables were occasionally adjusted to improve image quality, however these varied sample to sample and needed to be tuned to the individual sample to produce the best-looking image. When the necessary adjustments were made the image was saved using the capture function. Then the process was restarted at the scanning for features step to collect more images from the sample.

Data Analysis

Image analysis

Images were imported into ImageMetrics software (Zimeng Li, University of North Carolina, Chapel Hill) where masking was reduced to 0, and sample images were treated to planefitting and flattening as needed, to reduce image noise. To further improve image quality the particle appearance offset was reduced to 0 while the scale was adjusted to between 4 and 6 x10^{-9}m. At this point, the mask was adjusted so that the larger features believed to be sample are completely covered while background noise is minimally masked. Using the analyze tab the metrics for analysis were selected. For this study, the volume of the Ku protein was analyzed. After the metrics are selected, the analyze function was used to quickly analyze all masked particles of the image, the results button will open the output assigning each particle a number and an associated value.

Data analysis

The values from all images were exported into GraphPad Prism, and used to create a frequency of distribution plot. This was fit to a normal curve, the maxima of which represented an average volume of the particles and thus the volume of Ku.