Investigating replication fork blocks, replication-transcription conflicts and replication restart dynamics in *Escherichia coli*

A Thesis Submitted for the Degree of Doctor of Philosophy

By

Stelinda Peros

Department of Biosciences,

Brunel University London

2025

To my family for illuminating every step of my life! 

Contents

Abstract 1

Acknowledgments 3

Funding Statement 5

Publications, Conferences & Awards 6

Declaration 8

Chapter 1 10

1.1 DNA features across all domains of life 11

1.1.1 Milestones in DNA Discovery and Research 11

1.1.2 DNA features in humans 12

1.1.3 DNA replication errors as the cause of human genetic diseases 13

1.1.4 DNA features in bacteria 18

1.1.5 DNA replication errors as a source of antimicrobial resistance 20

1.2 The bacterium *Escherichia coli* as a model organism 27

1.2.1 *Escherichia coli* as a model organism to understand the principles of DNA replication 27

1.2.2 The characteristics of *Escherichia coli* 29

1.2.3 *Escherichia coli* DNA replication initiation 31

1.2.4 Replication extension in Escherichia coli 37

1.2.5 Replication termination in *Escherichia coli* 41

1.3 DNA replication mechanism in humans 46

1.3.1 Human DNA replication initiation 46

1.3.2 Human DNA replication elongation 49

1.3.3 Human DNA replication termination 51

1.4 Obstacles and Collisions between the replisome and other molecules 53

1.4.1 Collisions between the DNA replication and RNA transcription machineries 54

1.5 Replication Restart Proteins 57

1.5.1. Why the Replication Restart Proteins are essential 57

1.5.2 The properties of *priA* and *priA300* 59
1.5.3. The properties of PriB, PriC and DnaT 62
1.6 Aims & Objectives 65
Chapter 2 66
2.1 Strains and Plasmids 67
2.2 Media, Buffers and Antibiotics 78
2.2.1 LB medium 78
2.2.2 Mu medium 78
2.2.3 M9 Minimal salt medium 78
2.2.4 Antibiotics and supplements 79
2.3 Transductions 80
2.3.1 Production of P1 lysates 80
2.3.2 P1 Transductions 80
2.3.3 Streak Tests 81
2.4 Bacterial transformations 82
2.4.1 Preparation of electrocompetent cells 82
2.4.2 Cell transformation via electroporation 82
2.5 Spot dilutions assay 83
2.6 Fluorescent microscopy 83
2.7 Computational Image Analysis and Cell Quantification 84
2.7.1 Image Preprocessing and Deep-Learning Segmentation 85
2.7.2 Post-Processing and Data Integrity Filters 85
2.7.3 Feature Extraction and Statistical Framework 86
2.8 Extraction of bacterial gDNA 87
Chapter 3 89
3.1 Introduction and research aims 90
3.2 The experimental set up 92
3.2.1 The inducible *lacO*-LacI system 93
3.2.2 The control strains 98

3.3 Cells lacking replication restart proteins struggle to grow in the presence of a replication fork block 99

3.3.1 Cells deficient of *priB* experience growth defects in the presence of the *lacO*$_{34}$- LacI replication fork block 99

3.3.2 Cells deficient of *priC* experience growth defects in the presence of the *lacO*$_{240}$- LacI replication fork block 101

3.3.3 Cells deficient of the helicase activity of *priA* experience growth defects in the presence of the *lacO*$_{34}$- LacI replication fork block 103

3.3.4 The absence of either replication restart protein leads to cellular division defects 105

3.4 Discussion 111

Chapter 4 117

4.1 Introduction and research aims 118

4.2 The experimental set-up 119

4.3 Results 122

4.3.1 The absence of either replication restart protein leads to normal DNA replication and cell division in the absence of replication-transcription conflicts 122

4.3.2 In the absence of *priB* or the helicase activity of *priA* the cells exhibit replication and cell division stress during replication-transcription conflicts 124

4.3.3 The presence of *rpoB*35* results in the resolution of replication-transcription conflicts 127

4.3.4 In the double origin ori*C*$^{+}$ ori*X*$^{+}$ cells there are not any replication-transcription conflicts 131

4.4 Discussion 137

Chapter 5 141

5.1 Introduction and research aims 142

5.2 The experimental set-up 144

5.3 Results 144

5.3.1 In the presence of saccharin wild type *E. coli* cells undergo DNA replication defects and cellular division perturbations 144

5.3.2 In the presence of saccharin *E. coli* cells deficient of either *priB* or *priC* undergo DNA replication defects, cellular division perturbations and cell death 146

5.4 Discussion 147

Chapter 6 150

References **156**

Appendix **201**

Appendix A: Python Scripts for Data Processing (Chapter 3) **202**

A.1: Cell Segmentation Pipeline (CpnResNeXt) **202**

A.2: Data Processing and Statistical Analysis **210**

Appendix B: R Statistical Analysis and Visualization (Chapter 4) **341**

Abstract

DNA replication is essential for genome stability, but it is constantly jeopardized by various obstacles such as nucleoprotein complexes and transcription–replication conflicts. If not properly resolved, these impediments lead to replication fork collapse, genomic instability, and even cell death. This thesis investigates how *Escherichia coli* preserves its replication integrity using three experimental systems: site-specific protein–DNA blockades, engineered replication–transcription conflicts, and chemical stress induced by saccharin exposure. These studies were supported by the development of an automated bioimage informatics pipeline, utilizing deep-learning segmentation to enable high-throughput quantitative analysis of cellular morphology and SOS-induced stress phenotypes in live-cell time-lapse microscopy.

Using a novel inducible fork-block model, I demonstrate that the PriA–PriB–DnaT pathway is the primary restart mechanism at nucleoprotein obstacles, with PriA helicase activity being essential for efficient replication restart. Surprisingly, when large tandem repeats were placed on the opposite replichore, PriC rather than PriB played the dominant role, raising the possibility that restart pathway usage is influenced by obstacle size or chromosomal context.

Replication–transcription conflicts, generated by introducing an ectopic origin of replication (oriZ), similarly required PriA helicase and PriB for efficient fork restart. In their absence, cells displayed severe filamentation, heterogeneous stress phenotypes, and elevated Cas1–Cas2 foci. To further define the nature of these collisions, I utilized an alternative origin (*oriX*); the comparison between head-on and co-directional orientations confirmed that

cellular pathology was specifically conflict- dependent. Genetic suppression with an RNA polymerase–destabilizing mutation confirmed that these defects stem directly from transcriptional collisions rather than indirect effects.

Finally, I show that saccharin, a widely used artificial sweetener, induces replication stress in *E. coli,* with PriB-deficient cells exhibiting pronounced defects and loss of viability. These findings highlight how dietary compounds may disrupt gut microbial physiology.

Collectively, this work establishes PriA helicase as a central player in replication restart and a promising antibacterial target. Since stalled fork rescue is also critical in cancer cells, these results also provide conceptual bridges between bacterial DNA replication and oncogene-induced replication stress, opening avenues for both antimicrobial and cancer therapeutic development.

Acknowledgments

First and foremost, I would like to say a massive thank you to my PhD supervisor, Dr Christian J. Rudolph for his mentorship, guidance and support throughout this PhD research. It has been an honour to work in your lab and learn many things from you. Thank you very much for explaining science in such an interesting and inspirational way and for believing in my abilities!

I would also like to thank my secondary PhD supervisor, Dr Terry Roberts for his great help with my AFHEA application and for allowing me to shadow the excellent work carried out in his lab on breast cancer! I must also thank my review committee, Dr Concetta Bubici, Dr Sibylle Ermler and Dr Thomas Hoefken for the engaging and very enlightening discussions we had together during the reviews throughout my PhD research and for giving me the opportunity to deliver a lecture in the courses they lead!

Also, I must thank Miss Lesley Warren, the Head of the STEM Ambassador team, who has been one of the nicest people I have met and to whom I am very grateful for all the help she offered to me!

Many thanks to Brunel university Modern Language Centre and especially to Miss Pascale McFadden, to Miss Silvia Fischer and to Miss Mayumi Oyagi from Modern Languages who taught me French, German and Japanese respectively! I really had a fantastic time learning French and German from you and thank you so much for that!

Many thanks to Dr Richard Bingham for being an excellent educator and an exemplary person and for his immense support since my undergraduate studies. I feel very lucky that you taught me during my undergraduate studies and I will always be very grateful to you!

I must also thank my friends: Emine Efendi, Sarah Aldulaimi and Reham AlFashtaki for being very kind, approachable and for all the nice and funny conversations we had together and for allowing me to shadow them while they were working in the lab and learn from them techniques used in cancer research!

Last but certainly not least I thank my family for always supporting me in every possible way and for always believing in me, helping me in every possible way and encouraging me to achieve my dreams even when I did not believe in myself! Specifically, my mother, my best friend and inspiration, who has always loved me immensely from the very first days of my existence, my father who already somehow envisioned when I was only 6 years old that one day I would get a PhD, my brother for all the great and funny moments we have had together and for being the best brother I could ask for, and certainly my grandparents for all their valuable life-advice, support and love! I feel very lucky to have you all!

Funding Statement

This PhD research was carried out from 1st September 2022 until 31st August 2025 and was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) through the project "*How bacteria replicate their DNA in spite of barriers, one molecule at a time*" (Project Reference: BB/W000393/1)

Publications, Conferences & Awards

Data from this PhD work have so far resulted in the following peer reviewed publication:

- de Dios R., Gadar K., Proctor C.R., Maslova E., Han J., Soliman M.A.N., Krawiel D., Dunbar E.L., Singh B., Peros S., Killelea T., Warnke A.-L., Haugland M.M., Bolt E.L., Lentz C.S., Rudolph C.J. and R.R. McCarthy (2025). **Saccharin disrupts bacterial cell envelope stability and interferes with DNA replication dynamics.** EMBO Mol Med. DOI: 10.1038/s44321-025-00219-1

Data from this PhD research were presented in the following Conferences:

- **Microbiology Society Annual Conference 2025, Liverpool** Organiser: **Microbiology Society**, 31/03 – 03/04/2025 Poster presentation

- **10th CHMLS Conference**, Brunel University of London Organiser: **Brunel University of London**, 22/01/2025 Oral presentation

- **DNA replication meeting**, Churchill College, University of Cambridge Organiser: **Biochemical Society**, 10/09/2024- 12/09/2024 Poster presentation

- **9th CHMLS Conference**, Brunel University of London Organiser: **Brunel University of London**, 22/01/2025 Oral presentation

This PhD research was recognised with the following award:

- **Silver Award**: 2nd best Oral PhD presentation at the 10th CHMLS Postgraduate Research conference, Brunel university of London (January 2025)

Declaration

The research presented in this PhD thesis is my own unless otherwise specified and has not been submitted for any other degree.

Stelinda Peros

Even during the hardest days keep dreaming… and pipetting; Science holds wonders far greater than anything you could imagine! 💫

- Stelinda Peros

Chapter 1

Introduction

1.1 DNA features across all domains of life

1.1.1 Milestones in DNA Discovery and Research

DNA is one the most important molecules for all the organisms and its replication is fundamental for the continuation of life. The seminal studies by Johann Friedrich Miescher which led to the serendipitous discovery of the DNA molecule in 1869 were the foundation for all the subsequent discoveries in the field of DNA research (Dahm, 2005, 2008; Hall and Sankaran, 2021).

The team of Oswald Avery, Colin MacLeod, and Maclyn McCarty from the Rockefeller Institute for Medical Research carried out novel experiments to prove that DNA is the genetic material (Avery, Macleod and McCarty, 1944). Following from their interesting experiments, Alfred Hershey and Martha Chase at Carnegie Institution of Washington established DNA as the molecule of heredity (Hershey and Chase, 1952).

In 1953, the X-ray diffraction image known as Photo 51 contributed to the elucidation of DNA's double-helical structure and revolutionized the knowledge of how DNA is encoded and transmitted (Franklin and Gosling, 1953; Watson and Crick, 1953). These discoveries are among the greatest scientific breakthroughs of the 20th century because novel studies and discoveries regarding the genome of all the organisms have taken place since then.

These foundational discoveries led in 1990 to the launch of the Human Genome Project (HGP), which was a global scientific collaboration that aimed to sequence the entire human genome. The first draft of the human genome sequence was published in 2001 by both the International Human Genome

Sequencing Consortium and Celera Genomics (Lander *et al.*, 2001; Venter *et al.*, 2001). In 2003 the project was declared complete, providing the first almost complete map of the around 3 billion nucleotide base pairs of human DNA (International Human Genome Sequencing Consortium, 2004). The HGP revolutionised modern biology because it enabled large-scale studies of genetic variation and disease-associated mutations and it laid the foundation for personalised medicine and genomics. In 2023, the successful completion of the Telomere-to-Telomere (T2T) project was announced which produced the first truly complete, gapless sequence of the entire human genome, resolving all previously inaccessible repetitive regions and achieving 100% coverage of all human chromosomes (Nurk *et al.*, 2022).

1.1.2 DNA features in humans

The results from biophysical studies indicate that each human diploid cell contains DNA material of around 2m (Piovesan *et al.*, 2019). Based on the fact that humans have in total around 3 x 10^{13} cells, this means that every human is made up of an astronomical length of DNA (Bianconi *et al.*, 2013). In humans there is DNA in two organelles, in the nuclei and in the mitochondria (Nass and Nass, 1963). The human nuclear genetic material consists of more than 3 billion nucleotide pairs and is organised into approximately 20,000 genes (Clamp *et al.*, 2007). The mitochondrial genome, which as the genetic data suggest is maternally passed down to the progenies, consists of approximately 17,000 nucleotides (Hutchison *et al.*, 1974; Anderson *et al.*, 1981; Cann, Stoneking and Wilson, 1987; Stoneking and Soodyall, 1996).

1.1.3 DNA replication errors as the cause of human genetic diseases

DNA replication must be carried out with both impressive velocity and sufficient accuracy to ensure that all the chromosomes contain the same nucleotide sequence and that no deleterious mutations are introduced into the progeny cells. This is because even a single nucleotide change among the 3 billion nucleotides is sometimes enough to result in life-threatening human diseases, such as sickle cell anaemia, Marfan syndrome, achondroplasia and phenylketonuria (Blau, 2016; Ornitz and Legeai-Mallet, 2017; Inusa *et al.*, 2019; Du *et al.*, 2021). The research findings by Drake in 1969 and by Loeb and Kunkel in 1982 established that the accumulation of rare errors during DNA replication can contribute to the initiation of human diseases (Drake, 1969; Loeb and Kunkel, 1982). Thus, these early studies were fundamental because they proved the importance of replication fidelity for the maintenance of genomic stability.

Impairments in the DNA repair pathways and the mechanisms that ensure the accurate DNA replication and integrity result in genomic instability (Negrini, Gorgoulis and Halazonetis, 2010). Genomic instability refers to the increased susceptibility of the genome to acquire mutations that can trigger disease initiation and progression (Yao and Dai, 2014). Genomic instability represents a key feature of cancer, the second most frequent reason of death worldwide (Hosea *et al.*, 2024; Wu, Xia and Lin, 2024). The early work by Hartwell & Weinert cemented the idea that genomic instability is an important driver of cancer initiation and progression and established that damage in the surveillance systems, known as DNA damage checkpoints, allow the replication errors both to remain undetected and to persist (Hartwell and Weinert, 1989).

Cancer is a complex multistage disease that can arise in any human organ and can affect individuals of all ages (Brown *et al.*, 2023). Although cancer used to be more prevalent among older populations, alarmingly both cancer incidence and death rates among younger individuals are increasing (Miller *et al.*, 2020; Li *et al.*, 2023; Laskar *et al.*, 2024). In 2020 alone, cancer was responsible for approximately 10 million deaths worldwide (Sung *et al.*, 2021). Even more concerning are recent statistical projections which predict that both cancer incidence and mortality due to cancer will continue to rise, with more than 30 million cancer cases expected in total every year by 2050 (Chen *et al.*, 2023).

Cancer is a genetic progressive disease which arises as the result of gradual accumulation of mutations. Pioneering early cancer genetics studies revealed that mutations in proto-oncogenes such as *RAS*, *MYC*, and *HER2*, as well as in tumour suppressor genes such as *RB1* and *TP53* (Friend et al., 1986; Baker et al., 1990), disrupt replication control, cell-cycle progression, and DNA damage responses, thereby leading to and accelerating malignant transformation (Stehelin *et al.*, 1976; Friend *et al.*, 1986; Baker *et al.*, 1989; Kelland, 2005). These discoveries established cancer as a multistage disease driven by replication-associated mutations. When proto-oncogenes acquire activating mutations or become overexpressed, they are converted into oncogenes that promote excessive and uncontrolled cell proliferation. Cancer cells often show abnormal shapes under the microscope and can replicate indefinitely due to multiple genetic changes (Hanahan and Weinberg, 2000).

Tumour suppressor genes, often referred to as the “guardian genes” of genomic integrity, are essential for maintaining cellular homeostasis through rectifying DNA damage, regulating cell cycle progression, and inducing apoptosis in severely damaged cells (Sherr, 2004; Vousden and Prives, 2009). Among the most frequently mutated tumour suppressor genes is *TP53*,

BRCA1 and *BRCA2* often implicated in breast and ovarian cancer and RB1 often mutated in osteosarcoma (Miki *et al.*, 1994; Wooster *et al.*, 1995; Levine and Oren, 2009; Olivier, Hollstein and Hainaut, 2010; Kleinerman, Schonfeld and Tucker, 2012). Mutations that lead to reduced function of these genes or render them non-functional are detrimental because they facilitate genomic instability, evasion of apoptosis, and limitless genetic and cellular proliferation—all of which are hallmarks of cancer development, as shown in figure 1 (Hanahan, 2022). This limitless replication results in the formation of progressively more cancerous cells which accumulate even more mutations during every DNA and cell replication (Loeb, Loeb and Anderson, 2003; Brown *et al.*, 2023). At the same time, the cancerous cells avoid all the cellular signals to halt replication and employ resistance mechanisms against the immune response mechanisms which they also attack (Hanahan and Weinberg, 2011). The resulting cluster of the malignant cells is known as tumour (Au *et al.*, 2017).

To sustain their rapid proliferation, the tumour's cancerous cells deplete the body's oxygen and nutrient supplies by inducing the formation of new blood vessels, a process known as angiogenesis (Nishida *et al.*, 2006). Even worse, the malignant cells, can intravasate into the bloodstream and migrate to the other organs through the process of metastasis (Fares *et al.*, 2020; de Visser and Joyce, 2023; Gerstberger, Jiang and Ganesh, 2023). Metastasis is progressive which indicates that it can take many years, sometimes even decades, from the time when the single cancerous cell appears until the time the patient experiences the symptoms (Hu and Curtis, 2020).

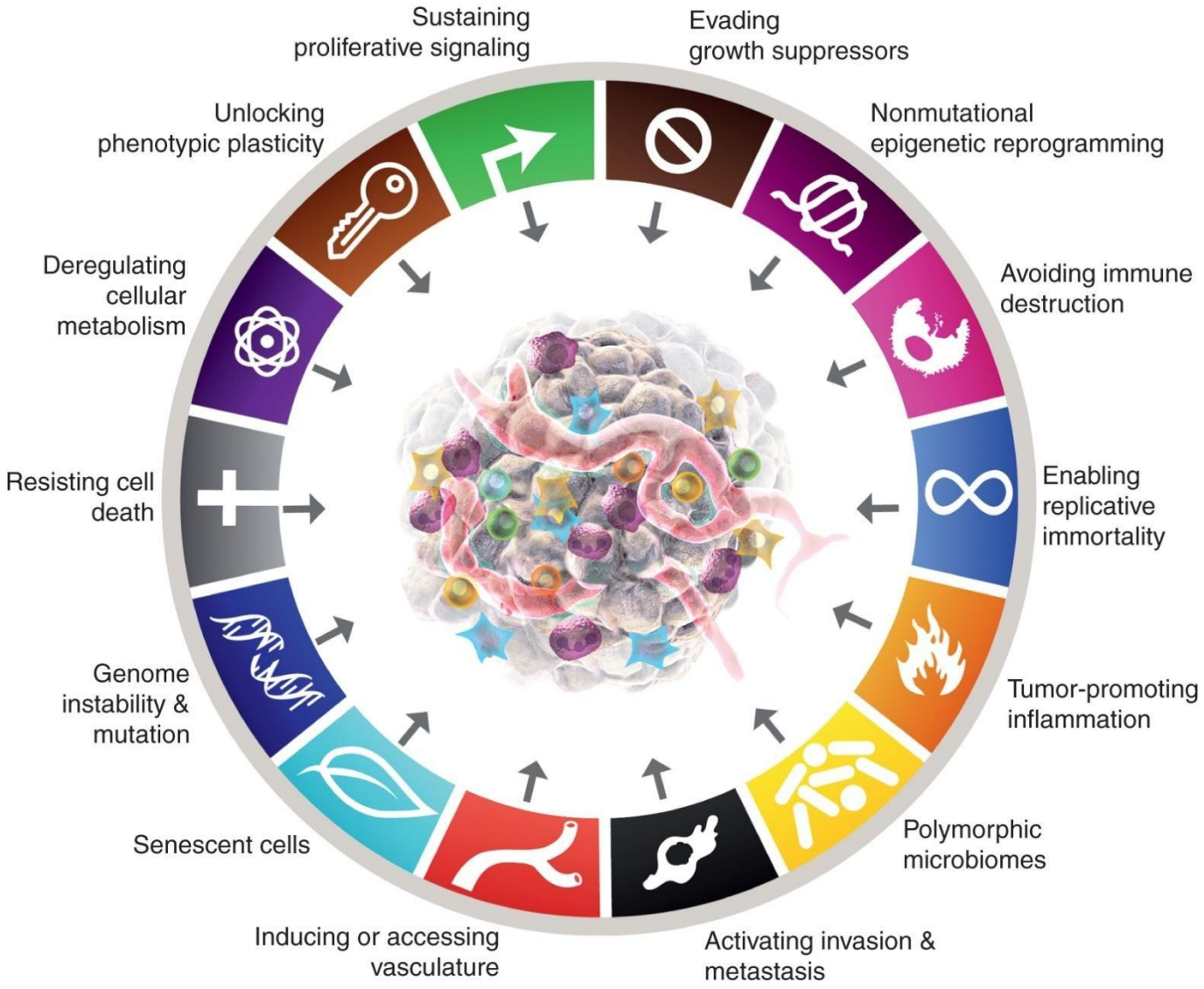

Figure 1. The established hallmarks of every cancer type. The circular schematic presents the 12 established hallmarks of every cancer type which can be divided in three categories: core hallmarks, enabling characteristics, and tumour microenvironmental influences. The category of core hallmarks comprises the following features: constant cellular proliferation, evasion of tumour suppression genes, resistance to cellular death, replicative immortality, abuse of blood supplies, tissue invasion, metastasis to other tissues and organs, rewiring of energy production pathways and attack of immune surveillance systems. The enabling characteristics include the following: genomic instability and mutation accumulation, epigenetic modifications and variety of the microbial communities. Finally, the category of the tumour microenvironmental influences includes senescent cells, which although are cells which are unable to replicate inn reality produce and release secrete factors which support the tumour environment. Figure from (Hanahan, 2022).

There are four stages of cancer from 1-4 depending on how severe the disease is (Brierley, Gospodarowicz and O'Sullivan, 2016). Stage 4 cancers are the most dangerous because they indicate that the tumour has metastasized to other organs, making them more difficult to treat (McPhail *et al.*, 2015; Waxler, 2016; Koo *et al.*, 2020). This highlights that it is crucial to diagnose cancer in its early stages, ideally during stage 1. Otherwise, the cancerous cells will outnumber the healthy cells, destroy the human organs and ultimately lead to death (Boire *et al.*, 2024; Mani *et al.*, 2024).

In the past scientists believed that cancer would result only because of the exposure to environmental carcinogens (such as UV, smoking and drinking) and/or heritability. The inheritance of certain mutations can considerably increase the risk of suffering from cancer during lifetime because these mutations increase the likelihood of cancer onset (Lichtenstein *et al.*, 2000). Indeed, research has shown that individuals who carry hereditary mutations in the *BRCA1* and *BRCA2* genes have considerably increased risk of developing breast and ovarian cancer during their lifetime (Milne and Antoniou, 2011). However, a scientific paper from Tomasetti, Li and Vogelstein established that in reality the majority of the mutations that lead to cancer are the result of errors during DNA replication (Tomasetti, Li and Vogelstein, 2017). This novel result placed DNA replication research at the forefront of cancer biology research and established early detection of cancer as equally essential as its prevention.

Cancer is a dangerous disease that causes both emotional and physical pain to the patients as well as great financial burden to the global economy (Niedzwiedz *et al.*, 2019; Chen *et al.*, 2023). The research findings during the last 50 years have considerably helped in the fight against cancer (Surh, 2021). However, still both the number of cases and mortalities appear to be

increasing. This means that still there are many unknown aspects regarding DNA replication, the occurrence and correction of replication errors, as well as the reasons why some replication errors escape repair and lead to disease mutations. All these indicate that mean that accurately delineating the mechanisms of DNA replication and understanding how replication errors arise and how they are normally dealt with may be the foundation for the diagnosis and treatment of human diseases and even the key to prevent their onset.

1.1.4 DNA features in bacteria

Bacteria, archaea and eukaryotes are the established domains of life which were distinguished primarily based on by differences in their ribosomal RNA sequences. (Woese, Kandler and Wheelis, 1990). Bacteria are prokaryotic organisms meaning that they lack a nucleus (Rosselló-Mora and Amann, 2001). Although they lack nucleus and mitochondria, they do contain genetic material which is also present in the form of DNA (McCutcheon, 2016). There are more than 30,000 officially discovered and classified bacterial species and it is estimated that there are more than a billion yet undiscovered bacterial species in the world (Dykhuizen, 2005; Parte *et al.*, 2020).Also, interestingly, during homeostatic conditions the human and bacterial cells exist and cooperate in complete harmony (Sender, Fuchs and Milo, 2016).

The majority of bacteria contain a single chromosome which is usually circular (Wang, Llopis and Rudner, 2013). However, there are some bacteria the genome of which is linear, with the most notable examples being the bacteria that belong to the *Borrelia* and *Steptomyces* species (Volff and Altenbuchner, 2000). Also, the genome of *Agrobacterium tumefaciens*, the

bacterium that infects plants, is composed of two chromosomes, one circular and one linear (Goodner *et al.*, 1999). Also, bacteria contain genetic material of various lengths. More specifically, *Mycoplasma genitalium* is a self-replicating bacterium the genome of which is less than 600 kilobases and so far, this is the smallest recorded genome size of a self –replicating bacterium (Fraser *et al.*, 1995). In contrast, the genome of the bacterium myxobacterium *Sorangium cellulosum* is made up of more than 13×10^6 base pairs and it is considered the largest documented and sequenced bacterial genome so far (Schneiker *et al.*, 2007).

Bacteria also can contain extrachromosomal supercoiled DNA sequences known as plasmids which are characterised by their ability to replicate independently of the chromosome (Lederberg, 1952; Shintani, Sanchez and Kimbara, 2015). Although plasmids are not crucial for the bacterial survival, they are important because their genes can produce proteins that make the bacteria resistant against antibiotics (Rozwandowicz *et al.*, 2018). Plasmids are able to transfer between the bacteria, either of the same or different species (Redondo-Salvo *et al.*, 2020). This plasmid-dependent DNA transfer is known as conjugation (Zatyka and Thomas, 1998). Scientists take advantage of the conjugation mechanism, and they utilise them by inserting into them genetic sequences of interest (Smillie *et al.*, 2010).

1.1.5 DNA replication errors as a source of antimicrobial resistance

The discovery of the antibiotic penicillin by Alexander Fleming advanced medicine and considerably reduced mortality cases attributed to bacterial infections (Chain *et al.*, 1940; FLEMING, 1944; Aminov, 2010). This serendipitous discovery paved the way for the discovery and development of the next antibiotics, such as streptomycin, used for the treatment of tuberculosis, and chloramphenicol, used to treat a range of diseases such as meningitis and cholera (Schatz, Bugle and Waksman, 1944; Ehrlich *et al.*, 1947; Nagao, 1953; Dinos *et al.*, 2016). Antibiotics usually kill the disease-causing bacterial cells without affecting the human cells and the non-pathogenic bacteria by targeting essential cellular processes such as DNA replication, protein synthesis and cell wall synthesis (Kohanski, Dwyer and Collins, 2010).

However, antimicrobial resistance (AMR), the emergence and spread of diseases-causing bacterial strains that are resistant to antimicrobials, is considered an urgent global health danger affecting every country. As also illustrated in figure 2, the phenomenon of AMR has exponentially exacerbated and as of 2025, among the bacterial pathogens leading to AMR related deaths worldwide include *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa* (Murray *et al.*, 2022; Salam *et al.*, 2023; Ahmed *et al.*, 2024).

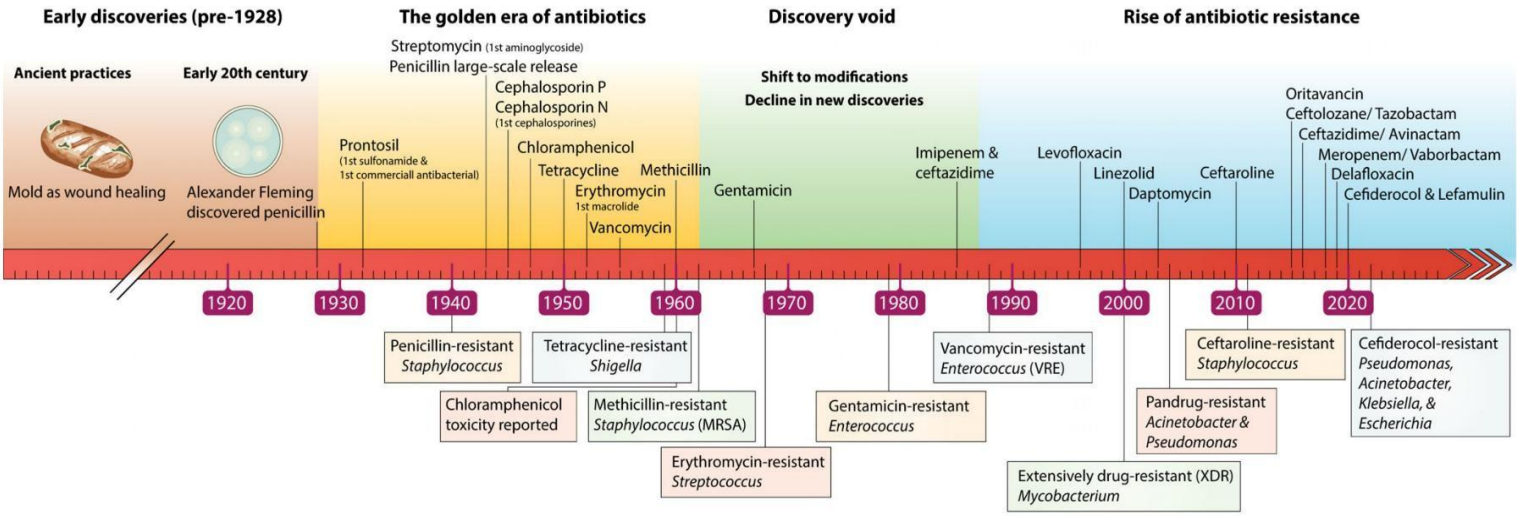

Figure 2. Timeline of key events from antibiotic discovery and development to antibiotic resistance emergence and spread. This figure visually spans key events from the 1920s to 2025 and is organized into four phases. In the first phase the discovery of the antibiotic penicillin in 1929 was instrumental for the second phase from 1940 until 1960, also known as the "golden era", during which major antibiotics—such as aminoglycosides, tetracyclines, and macrolides—were discovered and developed. In parallel, gradually antimicrobial resistant (ARMR) strains emerged, such as methicillin-resistant Staphylococcus aureus (MRSA), carbapenem-resistant Enterobacteriaceae (CRE), and multidrug-resistant Mycobacterium tuberculosis. From 1960 until around 1990 there were not many discoveries or development of new antibiotics, but new AMR strains continued emerging and spreading. The final phase which spans from 1990s till 2025 highlights the increasing global threat challenge posed by the multidrug-resistant and the extensively drug-resistant bacteria. Figure from (Tahmasebi *et al.*, 2025).

World Health Organisation (WHO) reports are alarming because they estimate that by 2050 AMR will lead to 10 million deaths annually, and according to World Bank the global financial losses can reach up to $100 trillion based on the current global trends (de Kraker, Stewardson and Harbarth, 2016; Price, 2016; Ho *et al.*, 2025). Among the key drivers of AMR emergence and spread is the natural ability of bacteria to evolve in response to adverse environmental conditions with a remarkable speed (Prestinaci, Pezzotti and Pantosti, 2015). Specifically, AMR can emerge through the accumulation of spontaneous mutations in the genome of the pathogenic bacteria (Davies, 1997). During each bacterial DNA replication cycle, spontaneous errors—such as point mutations, insertions, or deletions—can occur, some of which will lead to genes which confer antibiotic resistance. For example, during every DNA replication cycle in *E. coli*, the base substitution rate is approximately 10^{-10}. This indicates that every nucleotide has a 10^{-10} probability of being added incorrectly by the DNA polymerase (Drake, 1991; Schaaper, 1993).

Furthermore, taking into account that *E. coli* has an optimal doubling time of around 20 minutes, this means in a day (24 hours) a single bacterial lineage can undergo up to 72 generations. As a result, starting from just a single bacterial cell in the end of day 33.1 mutations will be incorporated. And while this number might appear low, in an average population of 10^9 bacteria, ultimately this translates to approximately 3.3×10^{10} mutations in just a day. Many of these mutations are certainly neutral or even deleterious for the bacterial population. However, there are also rare but beneficial mutations like those leading to antibiotic resistance (Rocha, 2018). In accordance with Darwinian selection in the presence of the selective pressure, in this case the antibacterial, the bacterial cells with the

advantageous gene will multiply more rapidly than the rest, eventually outnumbering them (Uddin *et al.*, 2021).

Among specific mutations that have been directly linked to resistance against antimicrobials, are mutations in *rpoB* conferring resistance to rifampicin, as well as in *gyrA* and *parC* conferring resistance to fluoroquinolones (Campbell *et al.*, 2001; Hooper and Jacoby, 2016). In addition, horizontal gene transfer allows the pathogenic bacteria to acquire resistance genes from other bacteria either of the same or different species, through the mechanisms of conjugation, transformation, and transduction (Munita and Arias, 2016).

The phenomenon of AMR unfortunately is also further exacerbated by human activities (Mancuso *et al.*, 2021). Specifically, the overuse and even misuse of antibiotics in medicine, and all domains of life (livestock, agriculture), as well as the inconsiderate disposal of antimicrobial agents into the aquaculture and generally in the environment, create significantly stronger selective pressures that ultimately increase exponentially both the emergence and spread of AMR strains (Manyi-Loh *et al.*, 2018). Even worse, this selective pressure has now led to the emergence and spread of "superbugs", which are pathogenic bacteria that are resistance to more than antimicrobials, referred to multi-drug-resistant (MDR) bacteria, and even to all antimicrobials, known as extensively drug-resistant (XDR) pathogens (Aslam *et al.*, 2018). Notable examples include methicillin resistant *Staphylococcus aureus* (MRSA) and carbapenem-resistant *Klebsiella pneumoniae* (Salam *et al.*, 2023). AMR bacteria employ various novel mechanisms to render an existing antimicrobial ineffective which are depicted in figure 3.

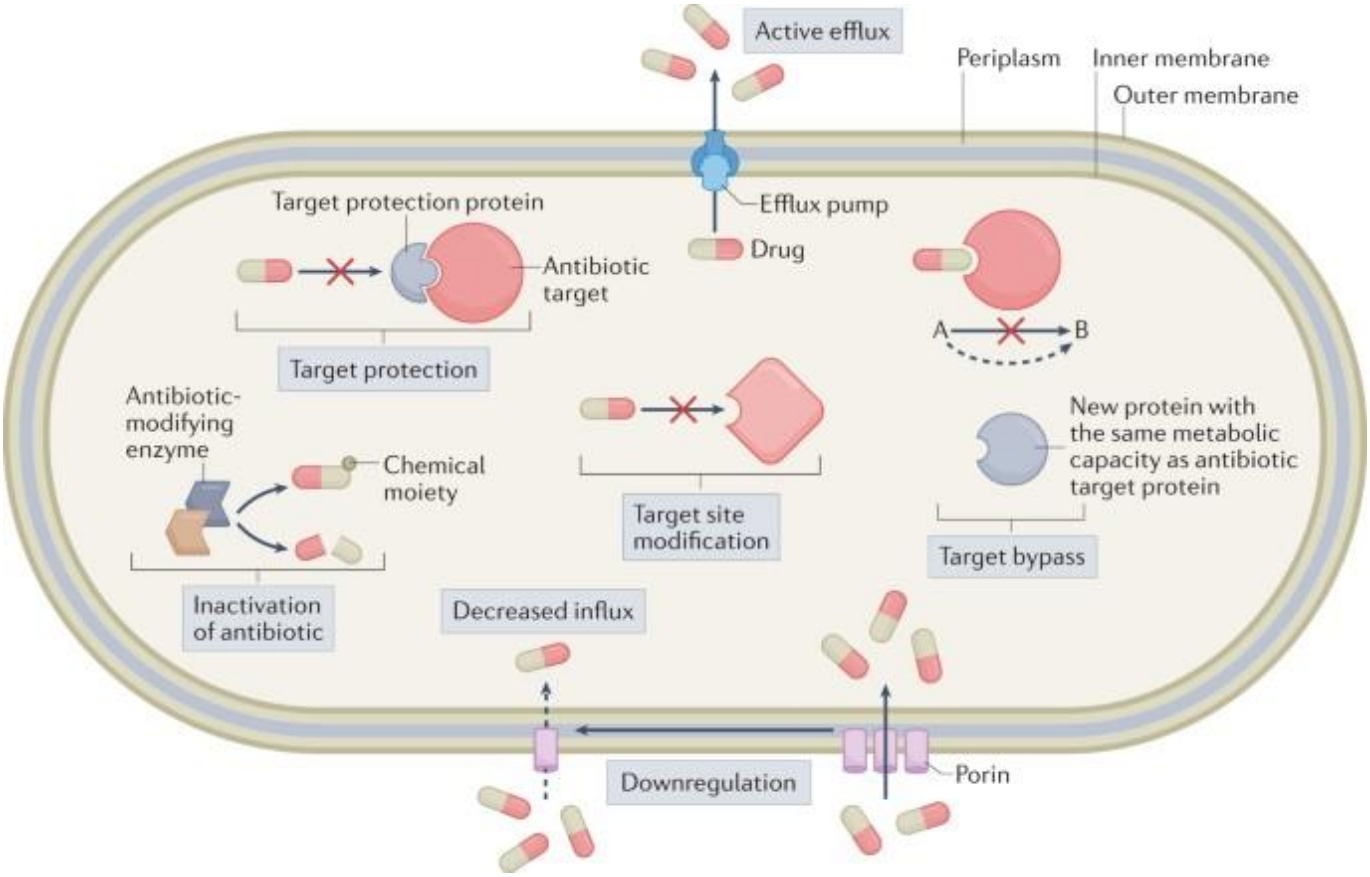

Figure 3. The mechanisms antibacterial resistant pathogens utilise. Antibacterial resistant bacteria can specifically change the target protein of the antibiotic so that the antibiotic cannot identify it. They can synthesize novel proteins that destroy the antibacterial and produce genes that render the bacterial cell wall difficult to be penetrated by the antibiotic. Finally, they can create efflux pumps which remove the antibiotic outside of the cell before it gets activated. Figure from (Darby *et al.*, 2023).

All these indicate that DNA replication is an important driving force both for the expression of mutations and for horizontal gene transfer which ultimately lead to antimicrobial resistance. Thus, discovering the unidentified mechanisms around bacterial DNA replication is of paramount importance. This is because by deepening our understanding of bacterial replication mechanisms, new antibacterials can be produced which for example can target previously unidentified targets and destabilize the plasmids carrying resistance determinants. Ultimately, prioritizing DNA replication research offers one of the most promising pathways to slow the evolution of resistance and restore the effectiveness of existing antimicrobials.

Even though the emergence and spread of AMR is an alarmingly increasing threat worldwide, the development of new antibacterials is decreasing in an unprecedented way (Boucher *et al.*, 2009; Llor and Bjerrum, 2014). One of the key factors is the decline of financing in antibiotic research by the leading pharmaceutical companies because in reality, pharmaceutical companies do not regard antibiotic research as a lucrative business (Towse *et al.*, 2017). This is because research on antibiotic development is prolonged and costly and ultimately leads to minimal financial return (Projan, 2003). Specifically, the profit of developing a novel antibiotic can reach up to $50 million, which although might appear substantial, in reality is scarce in comparison to the profits for other drugs, such musculoskeletal disorder drugs, which can reach up to $1 billion (Falagas, Fragoulis and Karydis, 2006; Clancy and Nguyen, 2020). On top of that, antibiotics are among the types of drugs that are not prescribed for lifetime, which further limits sales and hence profit (Nathan and Cars, 2014; Ventola, 2015).

Emerging technologies like artificial intelligence (AI) and phage therapy are promising complementary strategies that can be utilised in synergy with DNA replication research (Pires *et al.*, 2017; Arnold, McLellan and Stokes, 2025). AI can revolutionize the discovery and development of antibacterials by analysing vast datasets and predicting the antibacterial resistance patterns (De La Fuente-Nunez, Cesaro and Hancock, 2023; Liu *et al.*, 2024; Cesaro *et al.*, 2025). For instance, researchers have utilized AI and designed innovative antibiotics that can treat dangerous bacterial pathogens such as MRSA and gonorrhoea (Krishnan *et al.*, 2025). Additionally, phage therapy can efficiently utilise bacteriophages and kill pathogenic bacteria (Lin, Koskella and Lin, 2017). These viruses can particularly take advantage of the bacterial replication machinery to self-replicate eventually eradicating the bacteria (Shariati, Noei and Chegini, 2023). Advances in understanding the phage-bacteria interactions and in in phage engineering can enhance the probability of discovering or developing a novel antibacterial (Strathdee *et al.*, 2023).

1.2 The bacterium *Escherichia coli* as a model organism

1.2.1 *Escherichia coli* as a model organism to understand the principles of DNA replication

Since, its discovery, *Escherichia coli* has been a model organism in molecular biology (Ruiz and Silhavy, 2022). It is inexpensive to cultivate, requires minimal growth conditions, and its genome has been fully sequenced, thus all its genes are already known (Blattner *et al.*, 1997; Francis and Page, 2010). Additionally, its rapid doubling time of around 20 minutes under ideal conditions leads to fast and reproducible experiments (Cooper and Helmstetter, 1968).

Importantly, DNA replication is a vital process, and it was found that many fundamental principles of DNA replication are highly conserved between bacteria and humans, such as origin firing, helicase loading and replisome assembly as well as fork progression (Dutta *et al.*, 2011; Lang and Merrikh, 2018). Interestingly, novel studies utilising as a model organism *E. coli* provided important insights into the mechanisms of eukaryotic DNA replication (Bell and Kaguni, 2013).

Beyond these, utilisation of *E. coli* offers advanced genetic tractability. Specifically, single-molecule and live-cell imaging techniques, using fluorescently tagged replisome components such as the sliding clamp DnaN, allow direct visualization of the DNA replication mechanisms, such as replication forks movements, stalling and rescue (Reyes-Lamothe, Sherratt and Leake, 2010; Howell, Daniel and Brown, 2017). Furthermore, the fact that *E. coli* is a haploid organism, as opposed to humans which are diploid, is advantageous for the study of mutations (Verma, Qian and Adhya, 2019). This

is because any introduced mutations in its single circular chromosome will result to phenotypic changes, because there is not any possibility for these mutations to be masked by the second allele, as is the case with humans (Romero Romero *et al.*, 2022). This has applicability both to antimicrobial resistance research, because replication errors can lead to mutations, and to cancer biology research, because as already established replication stress underlies genome instability, one of the hallmarks of cancer (Woodford and Ellington, 2007; Huang and Zhou, 2021).

Notably, *E. coli* is also a powerful system to study replication–transcription conflicts, which is a threat across all domains of life, and one of the main interests of this PhD research (Dutta *et al.*, 2011; Lang and Merrikh, 2018). Additionally, R-loop formation during replication–transcription conflicts is well characterized in *E. coli*, where RNase H enzymes resolve RNA–DNA hybrids to prevent fork collapse (Castillo-Guzman and Chédin, 2021). These mechanisms are increasingly recognized as drivers of replication stress in cancer cells. *E. coli* also helped researchers investigate fork reversal which is an important survival strategy in cancer cells highlighting evolutionary conservation (Atkinson and McGlynn, 2009). All of these characteristics of *E. coli* render it both an efficient and powerful model organism for discovering the mechanisms of DNA replication and genome instability as well as the cellular responses to replication stress.

1.2.2 The characteristics of *Escherichia coli*

Escherichia coli is a bacterium that is normally present in the intestines of humans and many animals and contributes to the normal food digestion and production of various vitamins, such as vitamin K (Tenaillon *et al.*, 2010; Blount, 2015a). It was first discovered in 1885, and its name derives from the person who first identified it Theodor von Escherich (Escherich, 1988; Shulman, Friedmann and Sims, 2007).

In the 1950s the studies of John Cairns revealed that the genome of *E. coli* is circular (Figure 4) (Cairns, 1963). Also, according to *E. coli* Genome Project and subsequent studies the genetic material of *Escherichia coli* is circular and composed of around 4×10^6 base pairs which produce in total more than 4000 genes (Blattner *et al.*, 1997; Serres *et al.*, 2001; Verma, Qian and Adhya, 2019). According to microscopy data a physiologically normal *E. coli* cell is rod-shaped, with a length of about 2.0μm long and a diameter of 0.5μm (Shiomi, Mori and Niki, 2009).

The studies by the Nobel laureates Tatum and Lederberg using *Escherichia coli* were valuable because they established it as one of the most widely used model organisms (Lederberg and Tatum, 1946; Blount, 2015b). Among the most notable discoveries utilising *Escherichia coli* as the model organism were first the establishment that the DNA, and not the proteins, is indeed the genetic material by the Hershey-Chase Experiment in 1952 and then the discovery that the DNA is copied in a semi-conservative way by the Meselson-Stahl experiment in 1958 (Hershey and Chase, 1952; Meselson and Stahl, 1958).

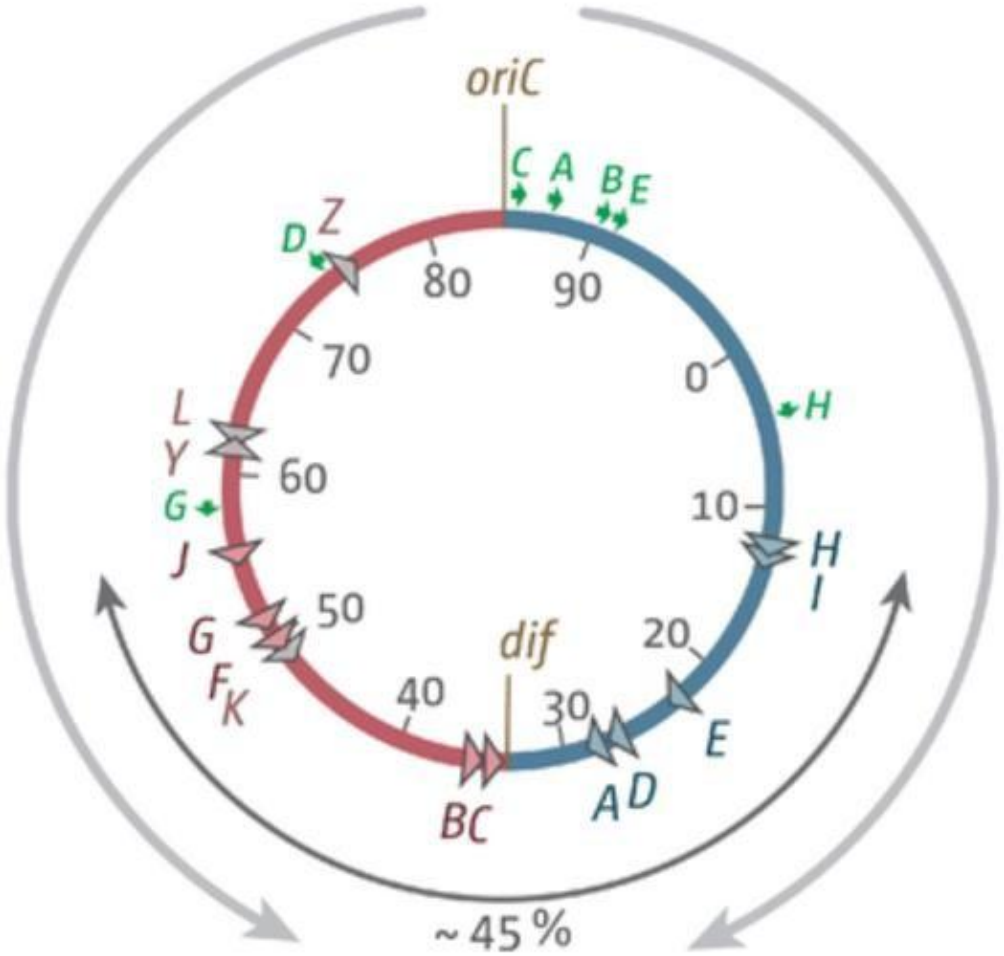

Figure 4. Organization of the *E. coli* chromosome. Schematic representation of the circular E. coli chromosome (~4.6 Mb) showing the chromosomal organisation relative to DNA replication. The origin of replication (oriC) is positioned at the top of the map and the replication terminus region (dif) at the bottom. Bidirectional DNA replication proceeds from oriC via two replication forks: the clockwise-moving fork (right replichore, blue) and the counter-clockwise-moving fork (left replichore, red). Numbers inside the circle (0–90) indicate minutes of the standard genetic recombination map. Grey arrowheads indicate the orientation of highly transcribed rRNA operons and other major transcription units, illustrating regions prone to replication–transcription encounters. Letters (A–J) denote chromosomal loci used in this study, mapped to their relative genomic positions. Green arrows mark the positions and orientations of replication termination (ter) sites, with arrow size reflecting relative activity or usage. Adapted from (Dimude *et al.*, 2016).

There are more than 700 *E. coli* serotypes, the majority of which do not pose health dangers (Al Qabili *et al.*, 2022). However, there are some *E. coli* serotypes that can cause dangerous human complications, with one of them being the O157:H7 strain (Gyles, 2007). O157:H7 is an enterohemorrhagic *E. coli* strain carried by ruminants and transmitted to humans via undercooked meat and contaminated water (Kaper, Nataro and Mobley, 2004; Lim, Yoon and Hovde, 2010). O157:H7 infects the humans by releasing into the human body a hazardous toxin known as Shiga and the release of this toxin leads to symptoms like stomach cramps and vomiting (Riley *et al.*, 1983; Wells *et al.*, 1983; Paton and Paton, 1998). This toxin can also attack the red blood cells leading to a severe human condition known as heamolytic-uremic syndrome, destroy the kidneys and even lead to death (Wadolkowski *et al.*, 1990; O'Loughlin and Robins-Browne, 2001; Joseph *et al.*, 2020).

1.2.3 *Escherichia coli* DNA replication initiation

In *Escherichia coli* there are many proteins and enzymes that are specifically synthesised to interact efficiently. Their harmonious cooperation is crucial to ensure that DNA replication is carried out with remarkable precision. In contrast to the human chromosomes, where DNA replication initiates from a lot of locations along the chromosome, in the circular chromosome of *Escherichia coli* it starts from one specific site (Kaguni and Kornberg, 1984; Mackiewicz *et al.*, 2004). This specific location which is also known as the origin of replication and is abbreviated as ori*C* is composed of around 250 base pairs (Tabata *et al.*, 1983).

The first protein which recognises and binds to ori*C* is protein DnaA. DnaA is composed of 473 amino acids, has a molecular weight of around 52kDa and

has four domains, illustrated in figure 5 (Sutton and Kaguni, 1997; Katayama *et al.*, 2010; Kaguni, 2011). DnaA is an important protein, and this is highlighted by its conservation across the bacterial species and the structural similarity it presents with archaeal and eukaryotic initiator proteins (Messer *et al.*, 1999; Erzberger, Pirruccello and Berger, 2002; Giraldo, 2003).

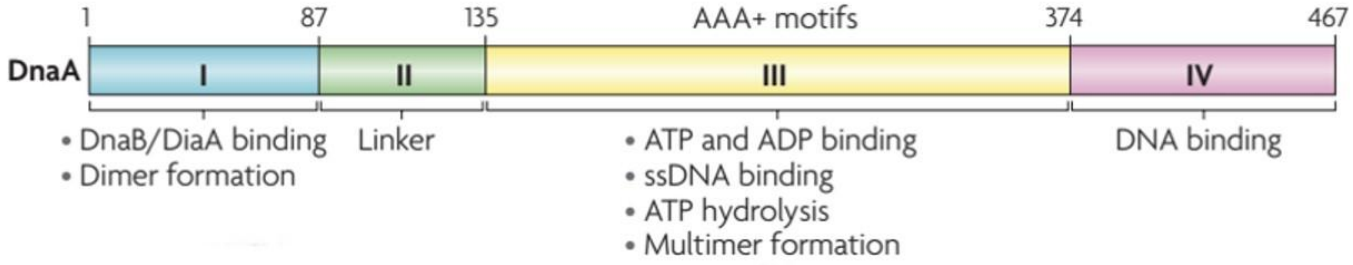

Figure 5. The domains of the bacterial initiator protein DnaA. *E. coli* initiator protein DnaA is composed of 467 amino acids which code for its four domains. Domain 1 which spans the first 87 amino acids is responsible for efficiently interacting with DnaB helicase. Domain 2, which is made up by the amino acids 87–134 acts primarily as a linker. Domain 3 is the product of the amino acids 135–373 contributes to the helicase activity of DnaA and is involved in the interaction between DnaA and ATP/ADP, a second interaction site with DnaB. Domain 4 is made by the last 94 amino acids, and it is where the DNA binding domain is. Figure adapted from (Katayama *et al.*, 2010).

The recognition and initial binding of the DnaA protein to the *oriC* is attributed to the presence of five DnaA boxes, which are known as: R1, R2, R3, R4 and R5 and are depicted in figure 6 (Fujikawa *et al.*, 2003). Up to 20 copies of the DnaA bind to the R sites creating a strong enough interaction that leads to bending of the DNA, thus creating stress that ultimately destabilizes the region of DUE, a region which is located to the left of the R-boxes and is composed of only A and T bases. Importantly, the two histone-like proteins, IHF and FIS, are located between the DnaA boxes and efficiently introduce bends enhancing DnaA assembly (Duderstadt and Berger, 2008). The presence

of the weaker AT base pairs leads to the unwinding of the DUE region, which as a result generates the initial open complex necessary for loading the next enzyme, which is DnaB helicase, and initiating DNA replication (Fuller, Funnell and Kornberg, 1984).

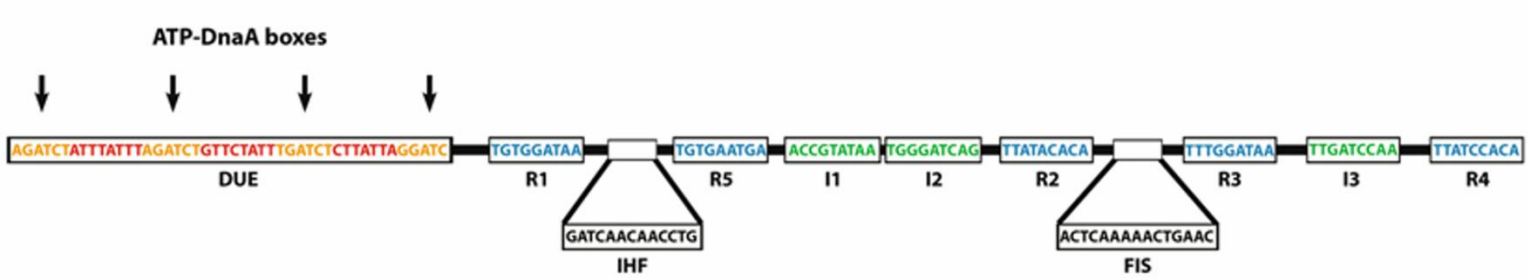

Figure 6. The sequences of DnaA boxes, DUE, I1-I3, IHF and FIS in the *E. coli* oriC. The enzyme DnaA attaches to the five DnaA boxes which are in the following order: R1, R5, R2, R3, and R4. I1–I3 are DNA sequences the interaction of which with DnaA is weaker than the interaction with DnaA boxes. The cooperative action of the histone-like proteins IHF and FIS triggers conformational changes that lead the melting of DUE. Figure from (Duderstadt and Berger, 2008).

Escherichia coli DnaB is an enzyme that is part of the distinct domain of SF4 helicases and is composed of six monomers. As soon as it binds to the DNA, it first encircles and then translocates along the lagging strand in the direction of 5′→3′ (Behrmann *et al.*, 2021). Following this, it interacts with DnaG primase. Specifically, its N-terminal domain binds with the primase DnaG, and its C-terminal domain, efficiently interacts with ATP as well as the helicase loader DnaC (Soultanas, 2005; Makowska-Grzyska and Kaguni, 2010; Li and Araki, 2013). However, importantly DnaB is an enzyme that is unable to bind on by its own on the DNA sequence and it is the interaction between DnaB and the enzyme DnaC, known as the helicase loader, that enables the binding of DnaB

on the sequence (Arias-Palomo *et al.*, 2013). DnaB undergoes significant conformational changes, which are also illustrated in figure 7. Specifically, in the beginning before even being able to bind to the DNA, DnaB exists in a constricted state. Then, the interaction with DnaC helicase loader results in the opening of DnaB and this state favours the binding of DnaG primase (Tougu and Marians, 1996; Strycharska *et al.*, 2013).

The helicase loader DnaC can bind DnaB without the cooperation with ATP, but in vivo *ATP* is essential for the loading of DnaB onto the origin (Davey *et al.*, 2002). Additionally, the results indicate that ATP can enhance the affinity of DnaC for single-stranded DNA, and thus the affinity of DnaB (Biswas-Fiss, 2006). As shown in figure 7 six DnaC monomers associate with one DnaB hexamer (C6B6) to form the active helicase-loading complex (Chodavarapu *et al.*, 2016; Blaine, Simmons and Stallings, 2023). Following DnaB docking at the unwound DNA, DnaC is released upon interaction with DnaG, allowing DnaB to begin unwinding (Hayashi *et al.*, 2020; Sakiyama *et al.*, 2022).

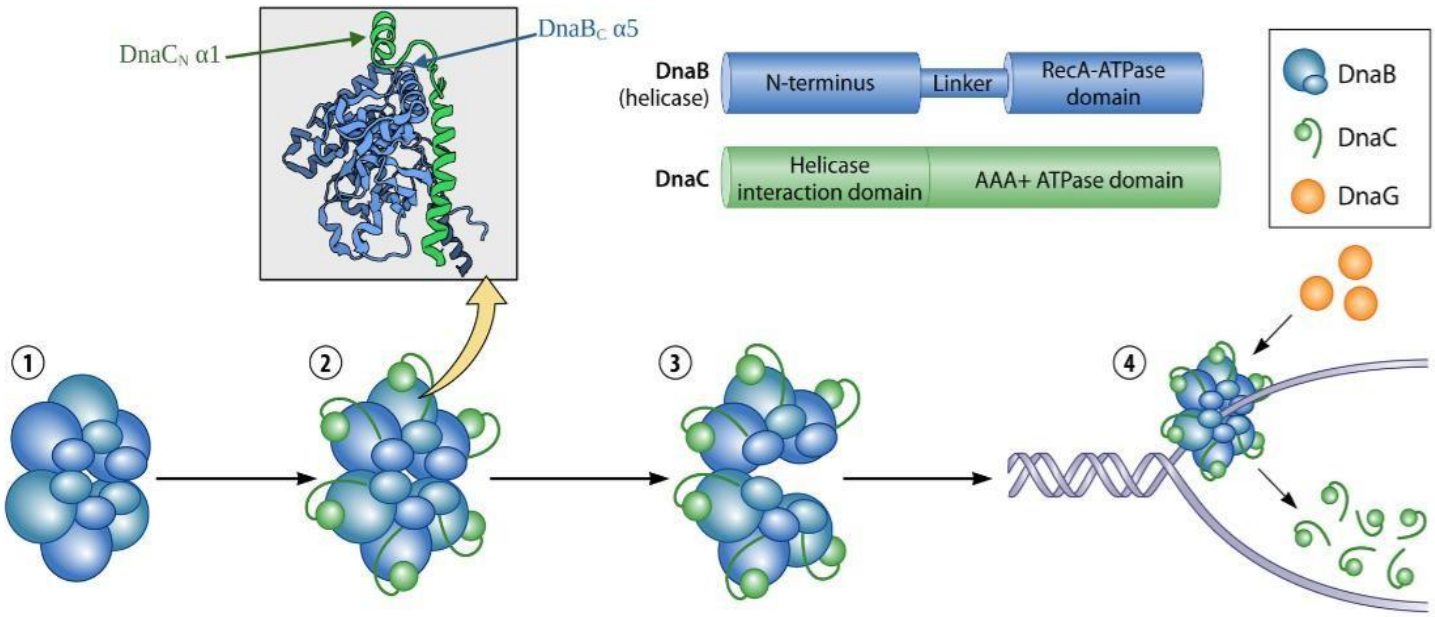

Figure 7. The interaction between *E. coli* DnaB helicase and DnaC helicase loader. DnaB (depicted in blue) is composed of the same 6 monomers which bind with one another and lead to the formation of a tight ring structure. DnaB is composed of the N terminus, a RecA- ATPase domain and a linker region between them. DnaC (depicted in green) is composed of two domains: the helicase interaction domain and the AAA+ ATPase domain. Each DnaC monomer binds and interacts with a single DnaB monomer, thus the reaction is made up of 6 DnaC monomers which bind with 6 DnaB monomers. This interaction changes the conformation of the DnaB resulting in its opening. The DnaB-DnaC complex binds to the replication fork. As soon as DnaB gets loaded, DnaC completely dissociates. Then DnaG primase (depicted in orange) binds on the replication complex to produce the RNA primers. The crystal structure which illustrates the interaction between a DnaC monomer (depicted in green) and a DnaB monomer (depicted in blue) has PDB accession number 6KZA. Figure from (Blaine, Simmons and Stallings, 2023).

The coordinated binding of DnaA, DnaB, and DnaC results in the onset of replication. Once it gets activated and bound to the origin of replication by the interactions with DnaC and DnaG, DnaB is able to unwind the double stranded DNA efficiently generating two single strands of DNA which is illustrated in figure 8. The action of another set of proteins, known as single-stranded DNA-binding proteins (SSB) which stabilizes unwound strands so that during DNA replication the two single DNA strands cannot rebind together (Figure 8). Also, DNA gyrases and topoisomerases are essential to relieve the torsional stress generated during helix separation (Akama *et al.*, 2025).

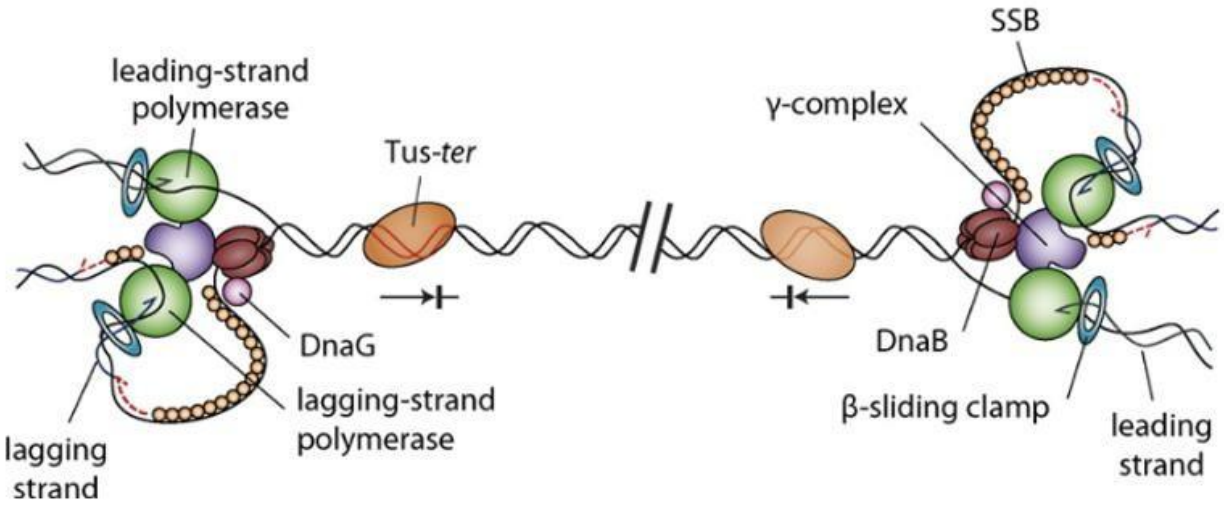

Figure 8: Illustration of the proteins that comprise the replisome on the leading and lagging strand. The DnaB helicase moves in one direction separating the two strands and the single stranded DNA binding protein abbreviated here as SSB acts as a protection for the DNA single strands. The primase DnaG and the DNA polymerase act in orchestration to add the single primer of the leading strand and the multiple primers at the lagging along with the new nucleotide bases. The replication forks encounter the Tus–ter complex, which mediates replication termination. Reproduced from (Rudolph *et al.*, 2007).

1.2.4 Replication extension in Escherichia coli

The process of DNA replication in the single chromosome of the organism *Escherichia coli* is carried out in both directions as is also illustrated in figure 9 (Prescott and Kuempel, 1972). This is achieved by the action of the two identical replisomes that are formed on the origin of replication and then independently from one another continue moving on opposite directions and with the same velocity which can reach up to 1000 nucleotides per second (Reyes-Lamothe, Wang and Sherratt, 2008).

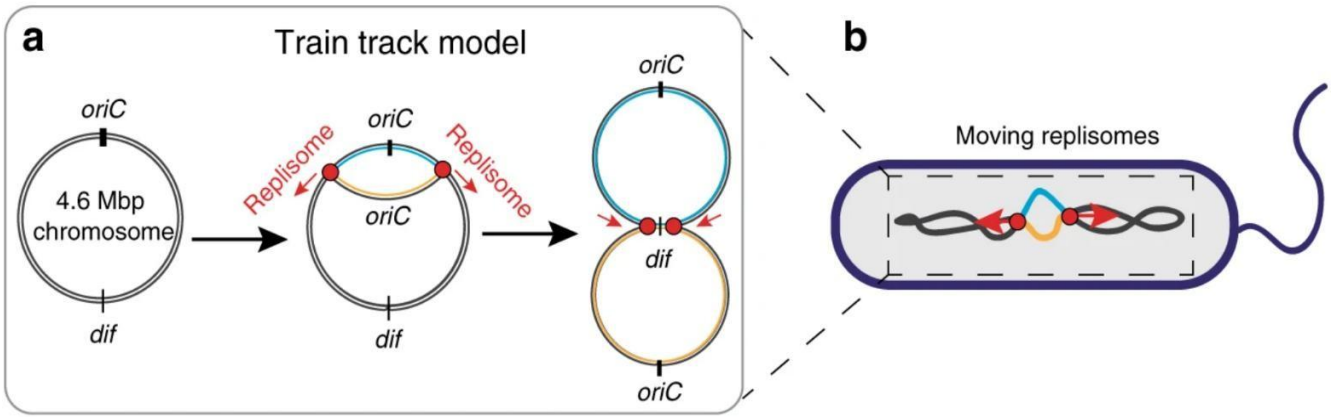

Figure 9: Schematic depiction of the bidirectional replication of the Escherichia coli and the direction of the two replication forks that are formed in the replication bubble. The two replication forks start at the origin of replication and move in opposite directions until they reach the termination of replication which is also the location where the two newly synthesised DNA molecules, each composing of a parental and a daughter strand, are separated. Reproduced from (Japaridze *et al.*, 2020).

The protein which is responsible for the addition of the bases in the two daughter strands is DNA polymerase III. However, DNA polymerase III is unable to initiate DNA replication alone so before DNA polymerase III can act and initiate the addition of the new nucleotides on the newly formed daughters strands though, short RNA sequences, of no more than 20 base pairs in length need to be added first so that DNA polymerase III can extend those sequences (Bell, 2019). These short RNA sequences are called primers and the enzyme that is mainly responsible for the production and addition of these RNA sequences on the daughter strand is the DnaG primase (Kuchta and Stengel, 2010).

Although DNA replication takes place in both strands simultaneously, the two strands are not replicated in exactly the same way, and this is mainly attributed to two facts. The first is that the DNA molecule is composed of two antiparallel strands, one with 5′ to 3′ direction and the other with 3′ to 5′ direction (Stillman, 2015). The second reason is that DNA polymerase III is an enzyme and thus it can only act from the 5′- end to the 3′-end and not the other way round. As a result, on one of the daughter strands, which is known as the leading strand, the DNA replication takes place from one end to the other continuously without any interruptions (Langston and O'Donnell, 2006).

However, the other strand, which is called the lagging strand is replicated in smaller fragments. This means that in contrast to the leading strand where only one RNA primer is needed to be added by DnaG primase, in the lagging strand multiple primers must be produced and added on the lagging strand so that the process can continue as also shown in figure 10 (Hernandez, Lee and Richardson, 2016). Notably, the removal of the RNA primers is a fast and continuous process that take place before the termination of DNA replication.

DNA polymerase I is the enzyme that removes the RNA primers and swaps them with the equivalent DNA sequences (Garcia-Diaz and Bebenek, 2007).

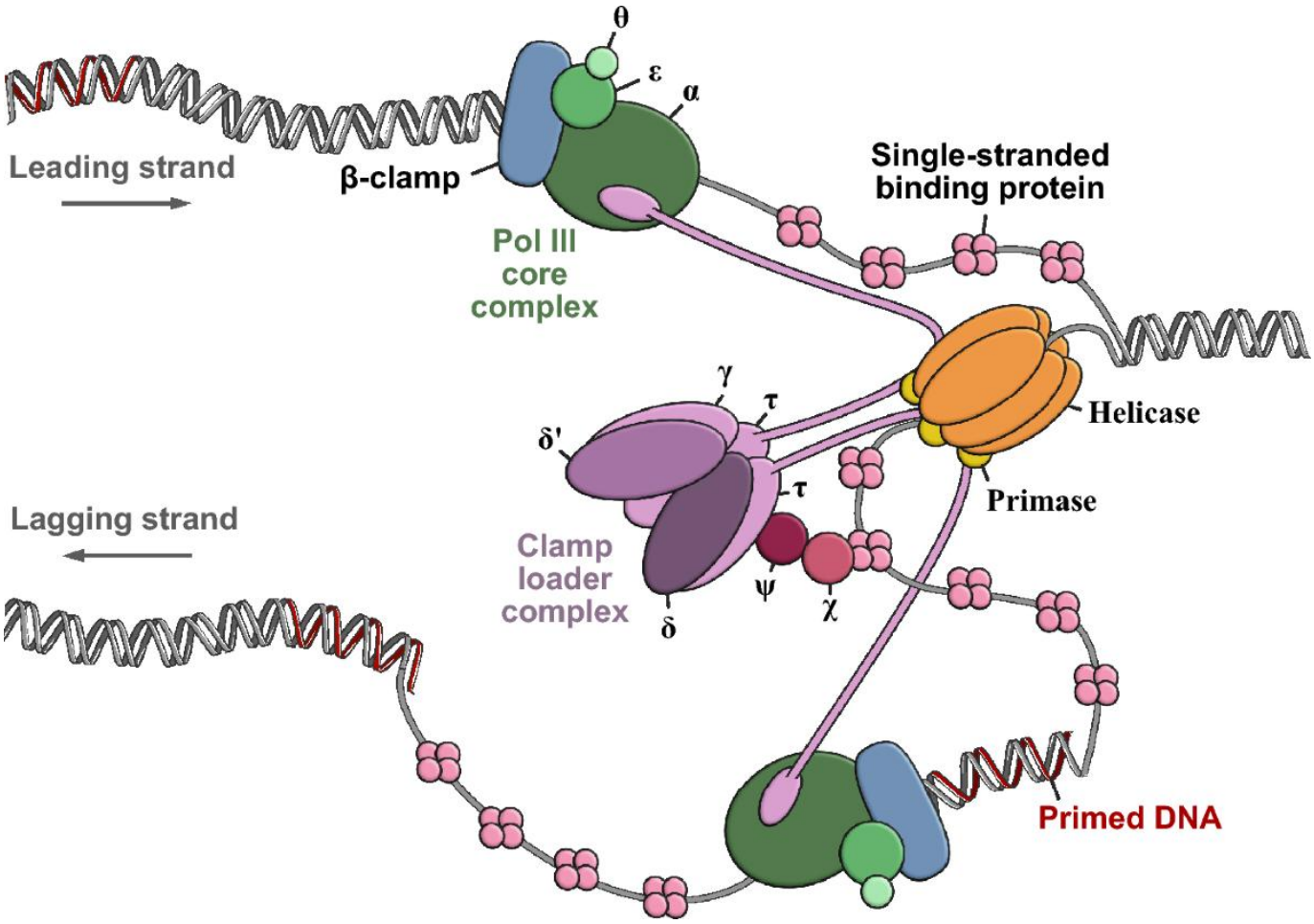

Figure 10: Structural arrangement of the bacterial replisome at the replication fork during bidirectional chromosome replication. The enzymes act in the following order: First DnaB helicase (in orange) disrupts the hydrogen bonds that keep the parental strands connected with one another. Then, DnaG primase (in yellow) synthesizes the single RNA primer (shown in red) in the leading strand and the multiple RNA primers (also shown in red) in the lagging strand. The clamp loader (composed of τ and γ subunits in light purple; δ, δ′, χ, and ψ subunits in darker purple) positions the β sliding clamp (in blue) around DNA polymerase III core complex (composed of subunits α shown in dark green and subunits ε θ shown in lighter green). Then DNA polymerase III follows DnaB helicase and adds the new nucleotides both in the leading and the lagging strand. In parallel, SSB (Single-strand binding) protein (in pink) wraps the single strands so that they do not reform hydrogen bonds with one another. Lagging-strand synthesis proceeds discontinuously via Okazaki fragments. Figure from (Kurth and O'Donnell, 2009). Figure from (Simonsen *et al.*, 2024).

Even though DNA replication is a fundamental process for all the organisms it would be expected that this process would be carried in complete harmony and without any impediments. However, it is important to underline that in reality DNA replication both in *Escherichia coli* and in humans does not take place perfectly without any mistakes and the wrong bases can be added on the newly synthesised DNA molecules. More specifically, it has been shown experimentally that the subunit a of DNA polymerase III adds incorrect bases with the frequency of 1 out of every 10^5 nucleotides (Bloom *et al.*, 1997). Although this error rate appears to be too low, in fact knowing that the genome of *E. coli* is 4×10^6 base pairs that would mean that in every cell cycle 40 nucleotides would be incorporated incorrectly into the *E. coli* genome.

However, the *E. coli* cells have developed protecting and editing mechanisms that identify and remove the incorrectly inserted bases from the daughter strands and replace them with the correct ones. More specifically, *E. coli* utilizes the ε subunit of DNA polymerase III, which is its proof-reading subunit, and it increases the replication by a factor of 100 (Scheuermann *et al.*, 1983; Scheuermann and Echols, 1984). This means that from 1 wrong base in every 10^5, 1 wrong base will be incorporated in every 10^7. Then, methyl- directed mismatch repair increase the accuracy of DNA replication to a factor of another 100, so the resulting error of DNA replication will be 1 base out of every 10^9 (Acharya *et al.*, 2003; Hsieh and Yamane, 2008). Thus, these two mechanisms of *E. coli* decrease the error rate of DNA replication by a factor of 10^4, from 10^{-5} to 10^{-9} (Fijalkowska, Schaaper and Jonczyk, 2012; Niccum *et al.*, 2018).

1.2.5 Replication termination in *Escherichia coli*

DNA replication termination in *Escherichia coli* takes place in a single chromosomal region, often referred to as termination zone, which is located approximately opposite the single origin of replication (Neylon *et al.*, 2005; Rudolph *et al.*, 2019). Within this region there are 10 termination (Ter) sites, 5 on the left replichore and 5 on the right replichore (Figure 11)(Bastia and Zaman, 2014). Each Ter site is made up of 23 nucleotide base pair which interestingly, although is not symmetric and not exactly the same for every Ter site, has a strictly conserved G-C6 base pair and also a conserved 13-bp core region (Goodall *et al.*, 2021). This difference between the Ter sites is depicted by the differences they present at how efficiently they can block the replication forks (Duggin and Bell, 2009; Moolman *et al.*, 2016).

Also, the 10 Ter sites are divided among the two replichores, meaning that 5 Ter sites are located on the left replichore and are responsible to trap the right replication fork and 5 Ter sites are located on the right replichore and are responsible to trap the left replication fork. This efficient safety mechanism ensures first that all the parts of the *E.coli* chromosome will be copied only once, preventing thus any genetic imbalances, and also that the two replication forks will only converge once and only on this area (Mulugu *et al.*, 2001; Dimude *et al.*, 2016).

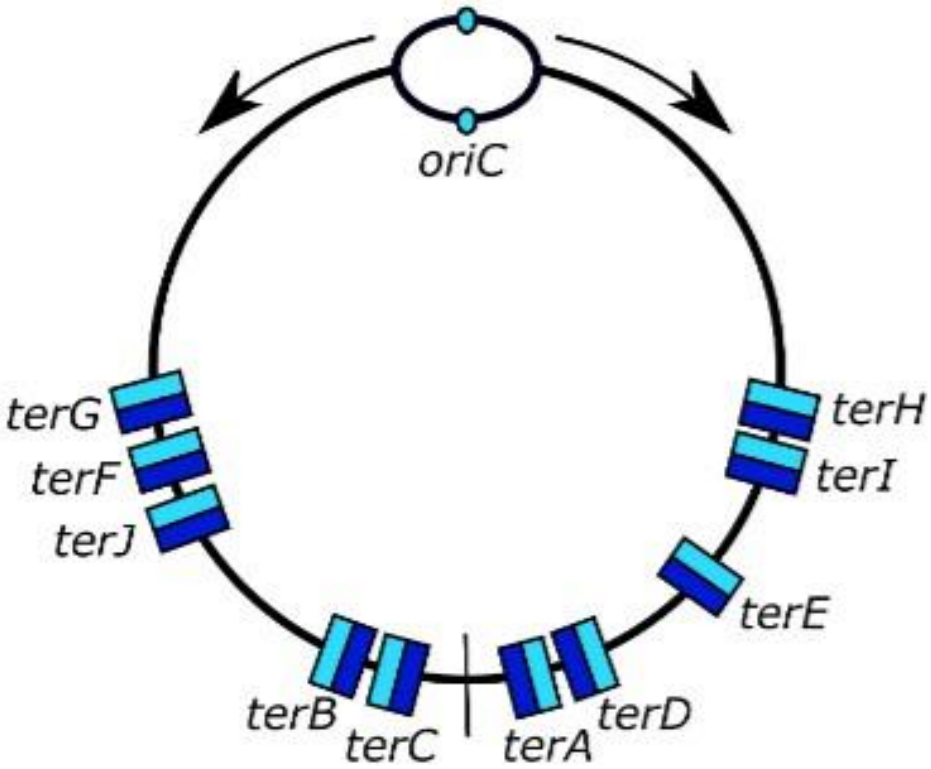

Figure 11. The termination sites of *E. coli*. *E. coli* chromosome is circular and to be replicated two replisomes need to bind to the ori*C* and travel with the same speed on opposite directions until they reach the termination region. On every replichore there are 5 Ter (termination) sites bound the Tus protein. The light blue corresponds to the permissive state whereas the dark blue corresponds to the non-permissive state. The replication fork which initiates from the left will meet the following ter sites: *terG*, *terF*, terF, terB, terC and because they meet these sites on their permissive sate they can pass through them. However, as soon it meets terA, the first Ter site of the right replichore they will get stuck because it will first face its non-permissive state. Respectively, the replication fork originating from the right replichore will be able to pass through its permissive sites (terH, terI, terE, terDa and terA) until it encounters the non -permissive state of terC which belongs on the left replichore and will block its progression. Figure adapted from (Jameson, Rudolph and Hawkins, 2021).

Additionally, each Ter site is bound by a protein known as the Tus protein (abbreviation for terminus utilization substance), which has a molecular weight of 36 kDa, is encoded by the gene *tus* and is the same for all the Ter sites (Hill and Marians, 1990). The interaction between the Ter site and Tus protein leads to a polar block to the replication fork (Khatri *et al.*, 1989). Each Tus protein comprises two different sides, a permissive and a non- permissive side. Interestingly, this indicates that the left replication fork can enter the left termination zone but cannot pass through the right termination zone and vice versa (Neylon *et al.*, 2005).

The specific mechanism of the conversion between the permissive and nonpermissive state is interesting. Specifically, as depicted in figure 12, it was found that in the non-permissive state the base cytosine at the nucleotide position 6 of the Ter sequence, which as mentioned earlier is conserved among all the Ter sites, is removed from the Ter nucleotide sequence and interacts with amino acids in the Tus protein (Kaplan, 2006). Their tight binding between the nucleotide C and the Tus amino acids creates a block which is too strong to be overcome by DnaB helicase and thus by the replication fork, so the replication stops there. As is shown by the two crystal structures in figure 13, in vivo the Tus-ter complex undergoes a significant conformational change when it transitions from its permissive to its non-permissive state.

The experimental findings suggest that the replication forks convergence, can lead to unreplicated sections or sometimes over-replicated regions that need to be processed to prevent mutagenesis and thus genomic instability (Jameson, Rudolph and Hawkins, 2021). Among the enzymes involved, the helicase RecG contributes to efficiently resolve the under- or over- replicated structures (McGlynn and Lloyd, 2002; Rudolph, Upton and Lloyd, 2009; Rudolph *et al.*, 2010, 2013; Midgley-Smith, Dimude and Rudolph, 2019).

After *E. coli* replication termination, the two resulting circular daughter chromosomes are interlinked. Their accurate separation requires the cooperative action of enzymes, such as DNA gyrase, topoisomerase IV and FtsK (Kato *et al.*, 1990; Espeli, Lee and Marians, 2003; Sayyed *et al.*, 2016; Helgesen, Sætre and Skarstad, 2021). This coordinated and efficient termination mechanism prevents unrestrained replication fork progression and facilitates the accurate chromosomal segregation into each daughter cell (Wang *et al.*, 2011).

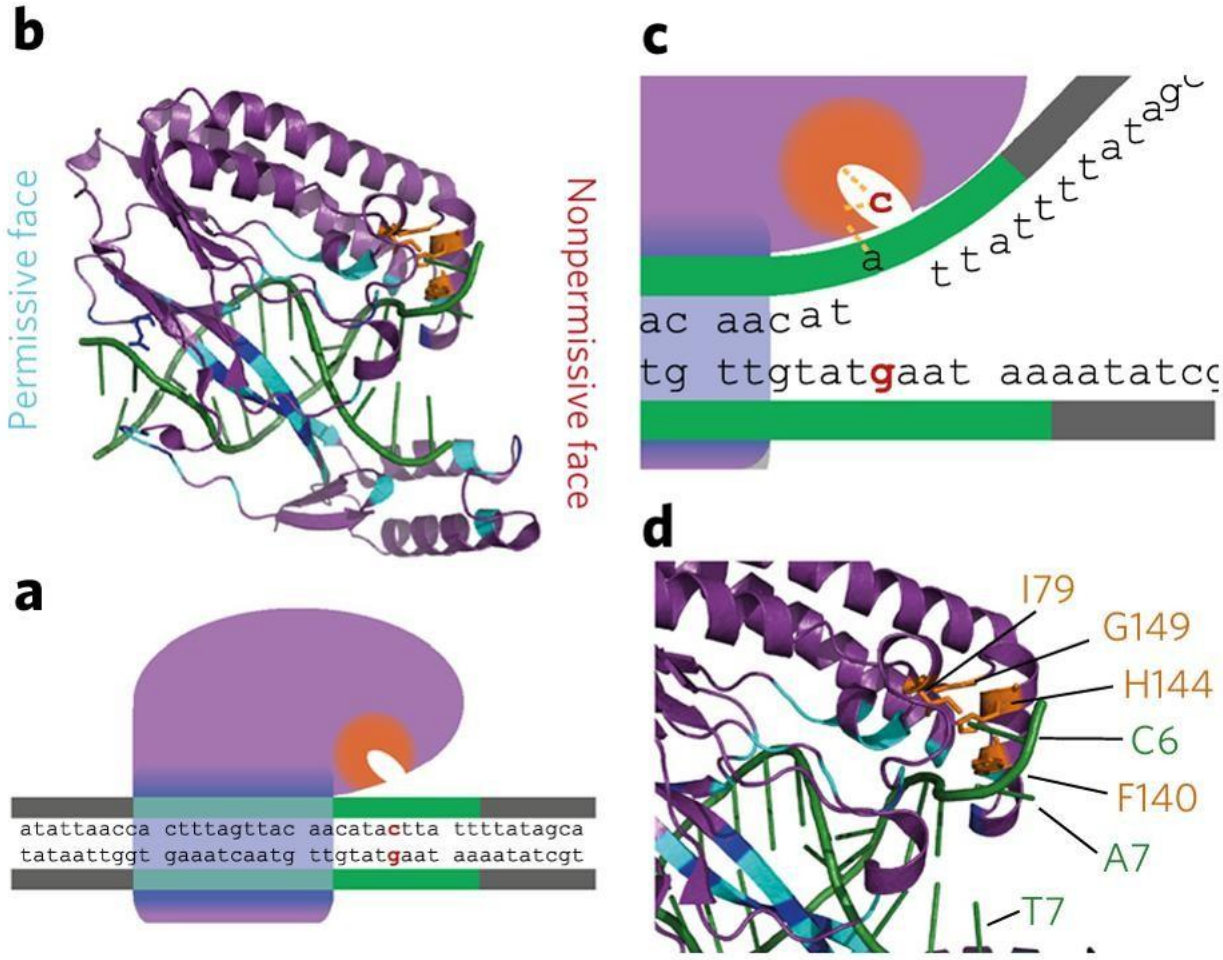

Figure 12. The difference between the permissive vs the non-permissive state of the Tus-ter regions. The left side illustrates the permissive state, and the right side illustrates the non-permissive state. Image a depicts the binding of the Ter site nucleotides with the Tus protein (shown in purple) and the empty pocket in the permissive state. Image b

illustrates the crystal structure of the permissive state of the Tus-ter complex when bound to the incoming replication fork. Figure c shows the interaction of the base Cytosine at nucleotide position 6 of the Ter with the enzymatic pocket of the Tus protein. Figure d depicts a close-up of the interaction between the base Cytosine at nucleotide position 6 of the Ter with the amino acids of the Tus protein in its non-permissive state. Figure adapted from (Berghuis *et al.*, 2015).

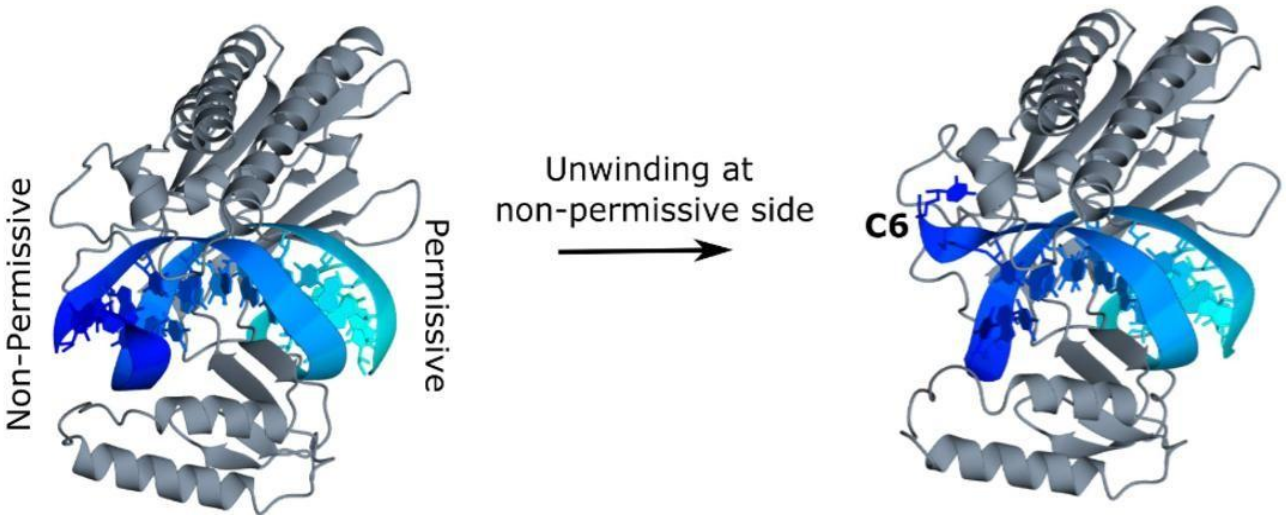

Figure 13. The crystal structure of the permissive vs the non-permissive state of the Tus-Ter during *E. coli* replication termination. The PDB accession number of the complex is 2I06. On the left crystal structure, the amino acids with dark blue comprise the non-permissive state and the amino acids with light blue comprise the permissive state of the Tus protein. On the right crystal structure, the interaction between the cytosine at nucleotide position 6 of ter and the amino acids of the non-permissive state (in dark blue) of the Ter protein leads to a state that blocks DnaB helicase. Figure adapted from (Jameson, Rudolph and Hawkins, 2021).

1.3 DNA replication mechanism in humans

1.3.1 Human DNA replication initiation

In the domain of eukaryotes DNA replication initiates from many origins of replication which is contrary to the prokaryotic DNA replication which initiates from just a single origin of replication (Hu and Stillman, 2023). The considerably bigger genome of the humans compared to the prokaryotes is one of the reasons for the existence of the multiple replication origins in the human genome (O'Donnell, Langston and Stillman, 2013).

Human DNA replication is a more complex procedure compared to the bacterial DNA replication and is a process that is highly dependent on the human cell cycle (Masai and Arai, 2000). The cell cycle is broadly divided in two phases: Interphase and Mitosis. Interphase is further divided in three subphases which are G1 (Gap1 phase), S (Synthesis Phase) and G2 (Gap2 phase) (Wang, 2021). It is established that human DNA replication takes place during the S phase (Limas and Cook, 2019). The ultimate goal of DNA replication is cellular replication which takes place during the Mitosis phase (Bournaka *et al.*, 2024).

More specifically, the human replication origins are established during the G1 phase, which precedes the S phase (Schwob, 2004). This means that in the G1 phase the pre-replicative complex (pre-RC), illustrated in figure 14, gets composed and binds to the DNA sequence (Sun and Kong, 2010). The organization of the following proteins and enzymes compose the eukaryotic pre-replicative complex: Origin Recognition Complex (ORC), Cdc6 (Cell Division Cycle 6), Cdt1 (Chromatin Licensing and DNA Replication Factor 1),

and the MCM2–7 helicase complex (DePamphilis, 1998). However, importantly the eukaryotic pre-replicative complex does not get activated during G1 phase but only during S-phase and this is because the enzymes that compose the pre-replicative complex have to be phosphorylated (Shen and Prasanth, 2012). The only enzymes that can phosphorylate the prereplication complex enzymes are two: cyclin-dependent kinases (CDKs) and Dbf4-dependent kinase (DDK) (Sclafani and Holzen, 2007). The successful phosphorylation of the pre-replication complex enzymes leads to the recruitment of the next two proteins: Cdc45 (Cell Division Cycle 45) and the GINS (Go-Ichi-Ni-San) complex (Fu *et al.*, 2022). Cdc45 and GINS are important because they recognise and bind to MCM2–7 forming thus the CMG (Cdc45 + GINS + MCM2–7) helicase, which is the helicase structure that can efficiently unwind the two parental DNA strands (Ilves *et al.*, 2010).

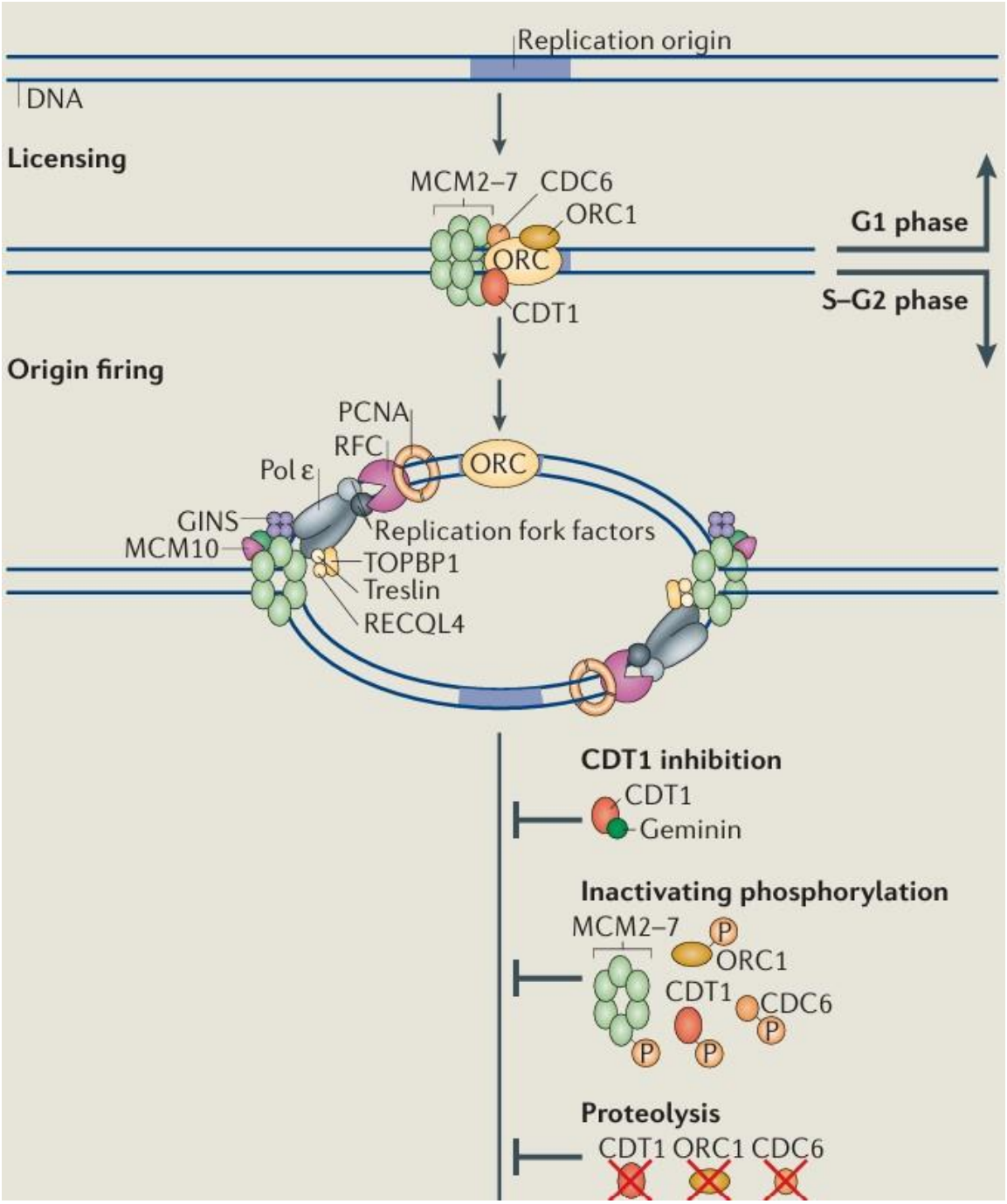

Figure 14. Origin licensing, activation, and inhibition of re-licensing during the eukaryotic cell cycle. Replication origins are first "licensed" during late mitosis and G1 phase through sequential recruitment of the origin recognition complex (ORC), CDC6, and CDT1. This leads to the loading of MCM2–7 helicase forming the pre-replicative complex (pre-RC). During S phase, cell-cycle-regulated kinases load the following onto the pre-RC: MCM10, GINS, TOPBP1, Treslin, RECQL4, PCNA, RFC, and DNA polymerase ε. This results in the activation of the helicase enabling thus the initiation of bidirectional replication. To prevent re-initiation, Geminin binds and inhibits the action of CDT1. Also, all the replication molecules that comprise the replication licensing factors become phosphorylated, hence

inactivated. Finally, ORC1, CDC6, CDT1 are targeted for ubiquitination and then degraded by the proteasome, ensuring thus their irreversible removal after origin firing. Figure adapted from (Fragkos *et al.*, 2015).

1.3.2 Human DNA replication elongation

As is the case with *E. coli* replication, the human DNA replication is impossible without the initial addition of primers. This is achieved by the orchestration of PrimPol and DNA polymerase α enzymes (Rechkoblit *et al.*, 2016; Díaz-Talavera *et al.*, 2022). However, an interesting difference between the *E. coli* and human DNA replication is that different DNA polymerases replicate the leading and lagging strand. Specifically, DNA polymerase ε is responsible for the addition of nucleotides on the continuous leading strand whereas DNA polymerase δ is responsible for the addition of nucleotides on the discontinuous the lagging strand also leading to the formation of Okazaki fragments (Kunkel and Bebenek, 2000; Zheng and Shen, 2011).

In addition, as is also the case with *E. coli* replication, the two types of DNA polymerase bind to the sliding clamp known as proliferating cell nuclear antigen PCNA. Then, the clamp loader known as replication factor C RFC attaches PCNA on the DNA polymerases (Jeruzalmi, O'Donnell and Kuriyan, 2001; Yao and O'Donnell, 2012). The equivalent of the *E. coli* single strand binding proteins is RPA which also is responsible to ensure the existence of the parental strands in a single stranded form (Dickey, Altschuler and Wuttke, 2013). During DNA replication supercoiling can also accumulate so there are two types of topoisomerases that can create either a cut on one of the DNA strands or a cut on both DNA strands and they are known as TOP1 and TOP2 respectively (Pommier *et al.*, 2016; Atkin, Raimer and Wang, 2019; Menger *et al.*, 2022).

Similarly with bacterial DNA replication, during human DNA replication, the removal of RNA primers is important. Specifically, RNase H and FEN1 are the enzymes which are responsible for the removal of the RNA primers from the daughter DNA strands. As the RNA primers are removed from the daughter strands and substituted with the equivalent DNA nucleotides, DNA ligase connects the newly added DNA nucleotides with the rest nucleotides by creating the phosphodiester bonds (Al-Behadili *et al.*, 2018). Replication termination in humans is a complicated procedure with a similar general principle as the bacterial DNA replication meaning that it takes place when two replication forks converge. However, in humans there are not any Ter-tus sites that trap the replication forks (Dewar and Walter, 2017). Instead, the disassembly of the human CMG (Cdc45-Mcm2-7-GINS) helicase complex leads to replication termination (DePamphilis *et al.*, 2006; Sclafani and Holzen, 2007).

Also, topoisomerases are responsible for alleviating DNA topological stress from the initiation of DNA replication until its termination (Pommier *et al.*, 2022). Specifically, the helicase unwinding of the DNA and the movement of the replication forks along the DNA, naturally leads to supercoiling of the DNA ahead of them with tightly interlinked DNA strands known as catenanes forming as well (Postow *et al.*, 2001). There are two types of topoisomerases Topoisomerase I which targets only single strands and makes temporary cuts only on single strands and Topoisomerase II which make temporary cuts on double strands (Wendorff *et al.*, 2012; Takahashi *et al.*, 2022).

1.3.3 Human DNA replication termination

There is a distinct difference between bacterial DNA replication termination and human DNA replication termination. Human DNA is linear, and not circular as the bacterial DNA, and DNA polymerases are able to add new nucleotides only from 5' to 3' direction. In addition, they require an RNA primer to start (Ishikawa and Naito, 1999). This means that after the removal of the RNA primers these sequences will not be able to be replicated. If these unreplicated sequences were responsible for the coding of genes, then this would mean that potentially important genes would not be able to be expressed which would have negative effects for the humans.

However, in reality the ends of the human and of many other eukaryotic organisms are made up of DNA sequences which do not code for any gene. More specifically repetitive sequences of TTAGGG make up the ends of the chromosomes and are covered by an enzyme known as shelterin (Moyzis *et al.*, 1988). So, this means that although every time the DNA replicates some terminal nucleotides remain unreplicated in reality the telomeres become shorter and shorter without affecting the gene coding nucleotides. It is estimated that a cell can replicate around 50 times before the telomeres get diminished and after that the cells normally enter the senescence state (Hayflick and Moorhead, 1961; Srinivas, Rachakonda and Kumar, 2020). The discovery of telomerase, an enzyme that can efficiently lengthen the shortened telomeres, was an important discovery because it provided new insight into the mechanisms of aging and disease onset (Szostak and Blackburn, 1982; Greider and Blackburn, 1985; Blackburn, 2001). Normally telomerases remain activated only in the human germ and stem cells. However, cancer cells manage to reactivate the action of telomerase for all the human cells resulting thus in immortal DNA replication. Since, infinite DNA replication is one of the key

features of cancerous cells, telomerase remains one of the key enzymes that are researched with the aim to develop novel anti-cancer drugs.

1.4 Obstacles and Collisions between the replisome and other molecules

Based on the essentiality of DNA replication it would be expected that this fundamental process would be carried out continuously and harmoniously. However, the results from genetic studies demonstrate that in vivo both the prokaryotic and eukaryotic DNA replicative machineries constantly face a lot of obstacles along their way as well as different types of clashes with many other molecules, such as proteins, DNA molecules as well as RNA molecules (Gupta *et al.*, 2013). For example, damaged DNA areas, which can result from UV radiation, secondary DNA structures, such as G-quadruplexes (G4s), and especially collisions between replication and transcription are among the factors that constantly threaten the cellular homeostatic maintenance (Mirkin and Mirkin, 2007; Kciuk *et al.*, 2020; Parekh *et al.*, 2023). The most frequent and significant clashes are the ones that take place between DNA polymerase and RNA polymerase. These conflicts are inevitable because both DNA replication and RNA transcription utilise DNA as the template and there are many times when both DNA replication and RNA transcription must take place simultaneously (Zeman and Cimprich, 2014; Hizume and Araki, 2019). All these obstacles, clashes and conflicts are dangerous for the organisms because they jeopardize the normal process of DNA replication (Shay and Keith, 2008; Eglenen-Polat *et al.*, 2024).

1.4.1 Collisions between the DNA replication and RNA transcription machineries

Both DNA replication and RNA transcription are two of the most significant processes that need to take place so that the final protein product is produced. The clashes between DNA replication and RNA transcription machineries are nevertheless inevitable and this is because both the DNA replisome complexes and the RNA transcriptome complexes have to share the same molecule (Rudolph *et al.*, 2007; Merrikh *et al.*, 2012; Gaillard and Aguilera, 2016; Syeda *et al.*, 2020; Stoy *et al.*, 2023).

The results from biophysical studies demonstrate that the DNA- RNA conflicts are considered as among the most critical threats for the genetic stability of the organisms (Hamperl and Cimprich, 2016; Lalonde *et al.*, 2021). There are two types of collisions between the DNA and RNA machineries, as illustrated in figure 15 as well, the ones that take place codirectionally and the ones that are head-on (Lin and Pasero, 2017). Both conflicts are dangerous but as the results indicate, the head-on collisions lead to more harmful consequences for the organisms (Mirkin and Mirkin, 2005a).

One of the main results of the DNA- RNA conflicts is the formation of distinct DNA: RNA structures, which are called R- loops, also illustrated in figure 16 (Hamperl and Cimprich, 2014; Sollier *et al.*, 2014). The production of R- loops is a great hazard for the cells and the scientific findings present that there is a direct correlation between the formation of R- loops and aging as well as a range of human diseases, such as cancer, Friedrich ataxia and fragile X syndrome (Lee, Kwak and Kim, 2025).

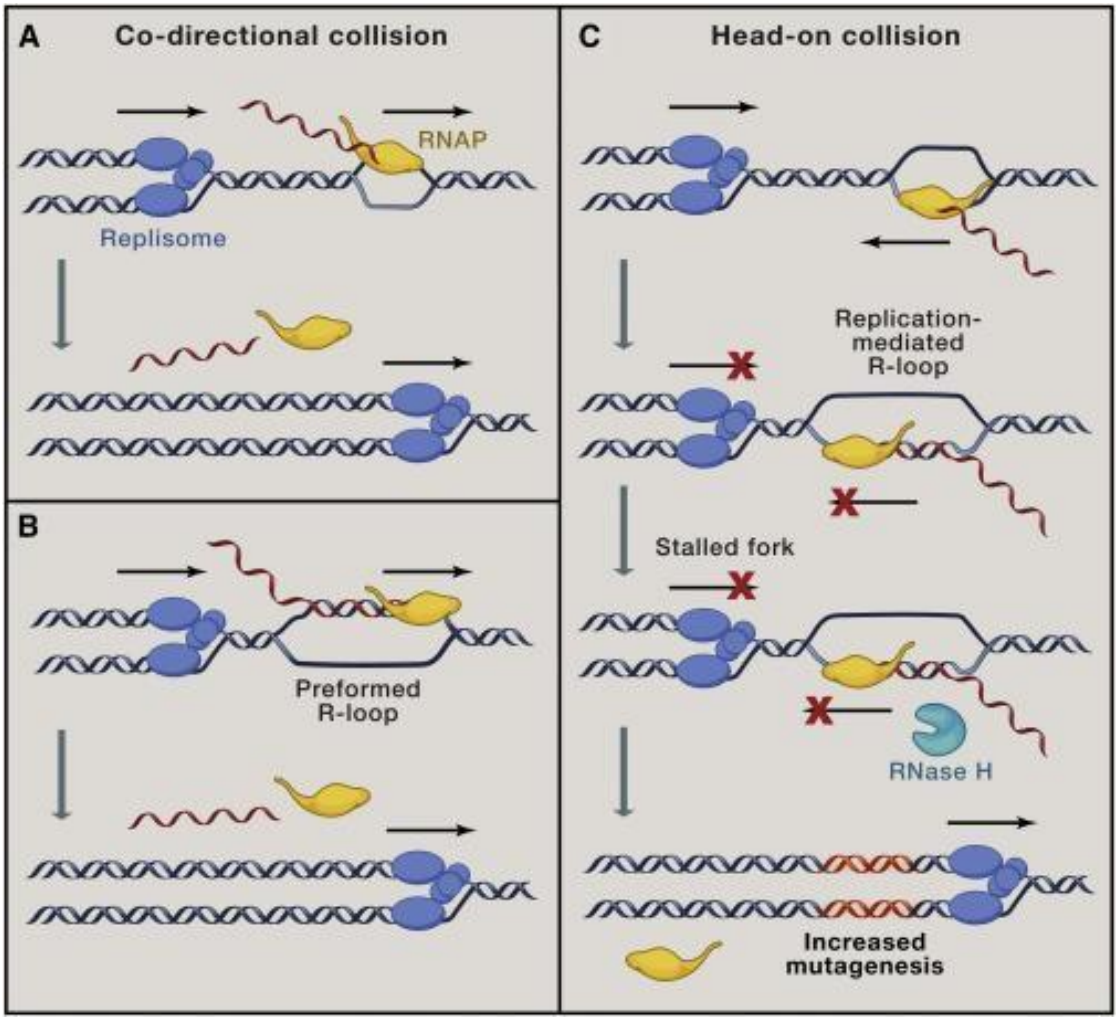

Figure 15: Graphical depiction of the two types of conflicts between the DNA replisome and the RNA transcription mechanisms along with the resulting effects for each situation. The head-on collisions are more serious than the co-directional collisions because they can lead to increased mutagenesis and the dissociation of the RNase H. Figure reproduced from (Lin and Pasero, 2017).

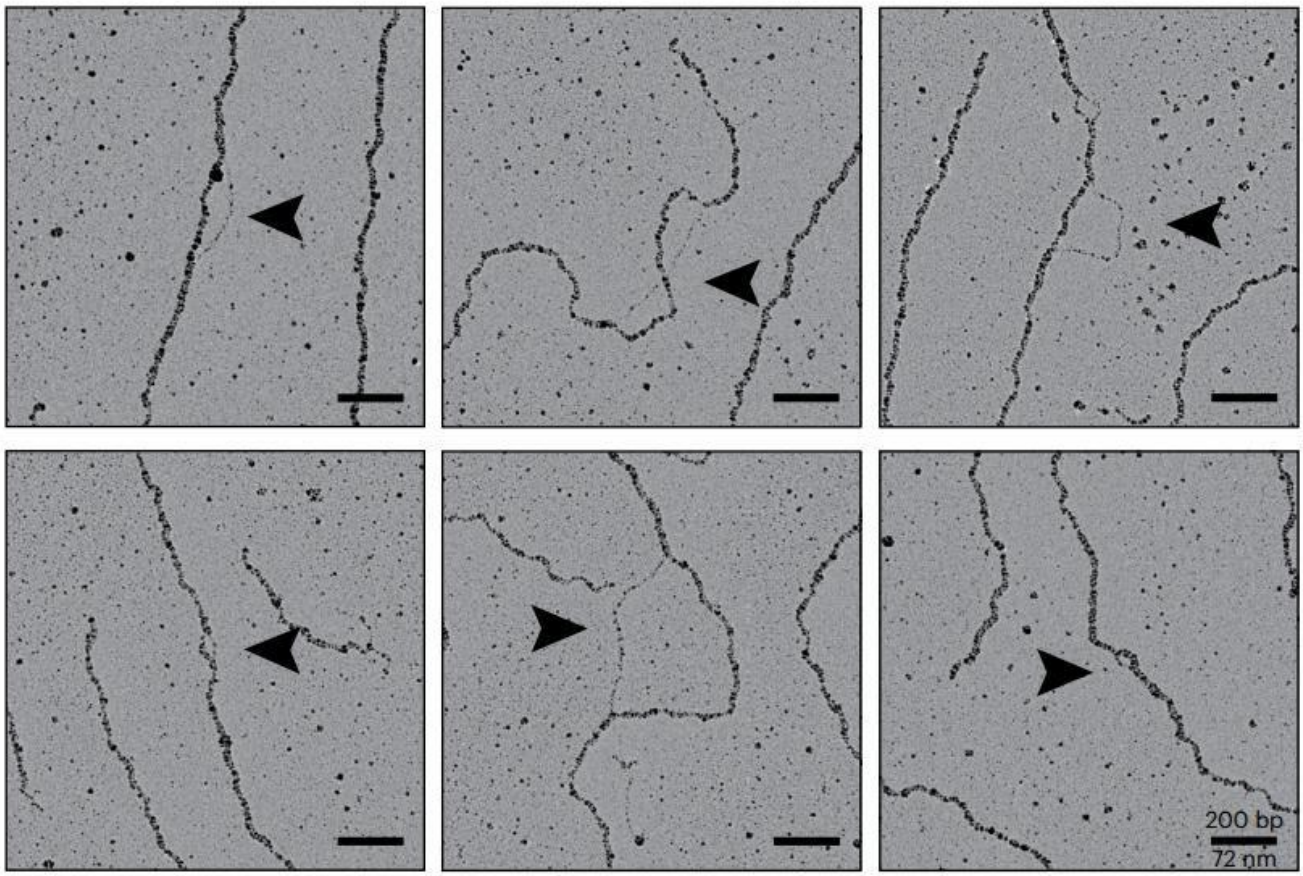

Figure 16. R-loops under the microscope. R-loops consist of an RNA–DNA hybrid and a displaced non-template DNA strand. In electron microscopy, only the DNA strands are usually visible, and here the displaced DNA strand appears as a single-stranded loop forming a bubble. Figure from (Stoy *et al.*, 2023).

1.5 Replication Restart Proteins

1.5.1. Why the Replication Restart Proteins are essential

DNA-RNA clashes occur so frequently that published data have demonstrated that the replisome detachment from the DNA molecule occurs at least one time during every cell replication cycle (Cox *et al.*, 2000). The hindered or disassembled replication forks cannot be left unrepaired though because if they are not rescued then this can be lethal for the cells and as a consequence for the organisms as well.

One of the major problems is that once the replisome stalls or even in the worst case dissociates, the DnaA protein which is a key component for the replication initiation is not anymore able to attach to the stalled sequences to rescue. This is because, as previously mentioned, DnaA needs to first recognise specific nucleotide sequences, namely DnaA boxes, that are only present on the origin of replication *oriC* and cannot be found anywhere else in the genome. As a consequence, DnaB—the replicative helicase in *E. coli*—cannot be reloaded at genomic sites outside oriC, as DnaA-dependent loading is restricted to the origin of replication. Furthermore, it is important to underline that the termination site of the *E. coli* chromosome which is opposite of the origin of replication is surrounded by 10 *ter*-Tus complexes, with half of them being on its right side and half of them being on its left side, and they block the incoming replication forks so that they do not pass on the other chromosomal half and over replicate the chromosome. This means that if one of the two replisomes stalls or dissociates, the other replication fork located at the other half of the chromosome cannot rescue the replication because it will get trapped by the *ter*-Tus complexes. This indicates that if

replication is not restarted the chromosome will be under replicated and thus the cells will eventually die.

Escherichia coli cells however have developed a set of very important evolutionary methods and unique proteins which are known as DNA replication restart proteins (Michel and Sandler, 2017a). These distinct DNA replication restart proteins are involved in identifying these stalled or even detached replication forks. This special identification is achieved based on the DNA structure and irrespectively of the presence of any specific sequences of nucleotides in the DNA structures, as is the case with the DnaA protein like previously mentioned. One of the aims of their action is to counteract the inability of the DnaA to bind on the stalled or disconnected replication forks. DnaB helicase can then rebind and act again on the DNA strand for the DNA replication to continue and terminate normally (Windgassen *et al.*, 2018). The replication restart proteins were first discovered in the bacteriophage ΦX174 (Schekman *et al.*, 1975). In *E. coli* there are several replication restart pathways which are composed of the following proteins: PriA, PriB, PriC and DnaT (Schekman *et al.*, 1975; Michel and Sandler, 2017b).

1.5.2 The properties of *priA* and *priA300*

PriA protein has a molecular weight of 82kDa, is composed of 733 amino acids and is made up of 6 domains which are illustrated in figure 17 (Nurse *et al.*, 1990; Leroux, Jani and Sandler, 2017). The importance of PriA is similar with that of the DnaA which is suggested by the fact that it presents homology in all the bacteria (Sandler and Marians, 2000).

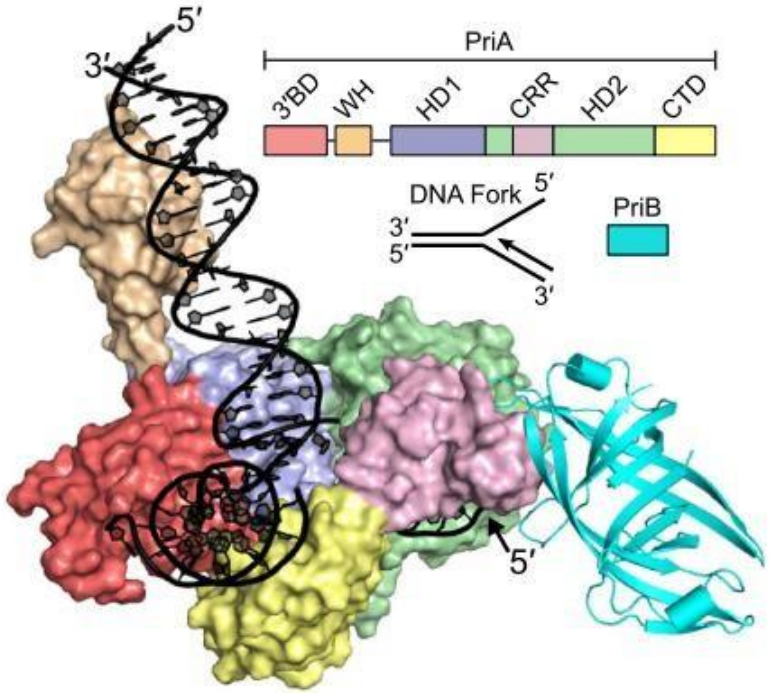

Figure 17. Cryo-EM and schematic depiction of the structure of *priA*, its 6 structural domains, the replication fork as well as PriB. PriA is composed of the following 6 domains which are illustrated with different colours: 3′BD (3′-Binding Domain), WH (Winged-Helix), HD1 and 2(Helicase Domains 1 and 2), CRR (Cysteine Rich Region) and CTD (C-terminal domain). Figure from (Duckworth *et al.*, 2023).

Importantly, the results from genetic studies demonstrate that the 3' binding domain of *priA* presents structural and functional homology to the HIRAN domain of HLTF, which is a eukaryotic protein involved in DNA damage tolerance (Achar *et al.*, 2015; Kile *et al.*, 2015; Chavez, Greer and Eichman, 2018). Moreover, it has been shown by in vivo studies that if *priA* is not present in the prokaryotic cells, then even if the other three replication restart proteins are present, the cells will not be able to divide normally and so they will not be viable (Sandler and Marians, 2000).

This fact can be accredited to the multiple roles of *priA*, meaning that except of being a restart protein it also has helicase, ATPase and translocase functions, which are unique roles that none of the other replication restart proteins have been found out to possess (Gabbai and Marians, 2010). PriA protein is valuable because by presenting a helicase function this means that it is the only replication restart protein that can create single DNA strands by separating the two existing parental strands. This proves to be particularly essential because the binding and the action of the DnaB helicase will be possible only in the presence of single strands (Cadman and McGlynn, 2004). A recent study has shown that once PriA protein identifies a stalled or disassembled replication it undergoes a significant structural remodelling, shown in figure 18, with its cysteine rich region domain rotating approximately ~85° and this conformational change results in the binding of both the replication fork and of PriB (Duckworth *et al.*, 2023).

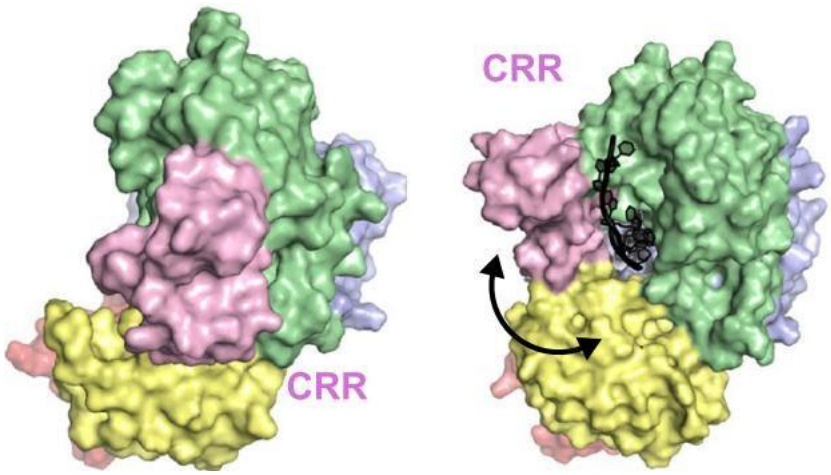

Figure 18. Illustration of the conformational change of PriA and its CRR domain during replication restart. On the left there is PriA before the replication fork binding and on the right after the replication fork binding. The arrow shows the ~85° rotation that the cysteine-rich region domain (CRR), which is shown in pink, undergoes to enable the binding of the replication fork and PriB. Figure from (Duckworth *et al.*, 2023).

Since, PriA needs to be present in the bacterial sets, a point mutation of the *priA* gene is used for in vivo studies which is the *priA300* (Gabbai and Marians, 2010). The difference between *priA* and *priA300* is a single nucleotide change at position 300 that in the final protein product results into the single amino acid change K230R. PriA300 retains the replication restart function n of the PriA protein but in reality, does not possess the helicase activity of PriA (Zavitz and Marians, 1992; Sandler *et al.*, 2001). The PriA helicase mutants are viable whereas the PriA mutants where the replication restart activity has been inactivated are very sick and very rarely survive. This indicates that the replication restart function of PriA is more important than the helicase activity because even if the helicase activity is lost from the cells the replication restart activity of PriA300 will assist in the continuation of DNA replication.

1.5.3. The properties of PriB, PriC and DnaT

PriB and PriC are also significant proteins, although cells that are deficient either of the PriB or the PriC grow without accumulating a plethora of mutations (Windgassen *et al.*, 2018). However, if both PriB and PriC are missing the cells are sicker than when PriA restart activity is absent. PriB, which was initially knowns as 'protein n', has a molecular weight of around 11 kDa, is made up of 208 amino acids and, as shown in figure 19, is composed of two identical polypeptide chains (Liu *et al.*, 2004; Fujiyama *et al.*, 2019).

PriB is mainly responsible for stabilizing the biochemical interactions that are developed between the other replication restart proteins and especially between PriA and DnaT (Cadman *et al.*, 2005). Although it presents some structural similarity with the single stranded binding protein it has been found out that it does not present functional similarity (Shioi *et al.*, 2005; Huang *et al.*, 2006).

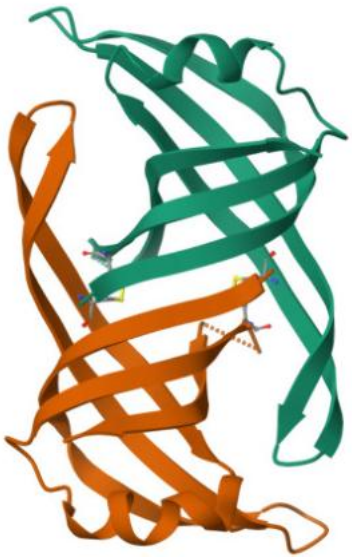

Figure 19. Crystal structure of the replication restart protein PriB. PriB is composed of two identical monomers, and each polypeptide chain is composed of 104 amino acids. The pdb accession code for the crystal structure of PriB is 5WQV. Figure from (Fujiyama *et al.*, 2019).

PriC has a molecular weight of 24kDa and is composed of five alpha helices (Figure 20) (Wessel *et al.*, 2016). It is mainly responsible in the final step for the actual recruitment of the DnaC and as a result of the DnaB helicase as well, which is the ultimate goal of the action of the management of all the replication restart proteins (Heller and Marians, 2005).

Figure 20. Crystal structure of the replication restart protein PriC. PriC is composed of five alpha helices. The pdb accession code for the crystal structure of PriC is 2NCJ. Figure from (Wessel *et al.*, 2016).

DnaT is responsible for the normal reloading of the helicase loader DnaC and DnaB for the DNA replication continuation. This means that the replication restart protein DnaT is needed for the normal functioning of all replication restart pathways, and it has been found out that in the mutant ΔDnaT cells many spontaneous mutations emerge like in the case of ΔPriA cells and that is why both ΔPriA and ΔDnaT cells are extremely difficult to be handled experimentally (Huang, Lin and Huang, 2013). The great importance of the replication restart protein DnaT is also evident based on the experimental results because the *Escherichia coli* strains that lack the DnaT cannot survive, as is the case with the replication restart protein PriA. Based on experimental data, it has been established that there are three replication

restart pathways in *E.coli*: PriA- PriB- DnaT, PriA- PriC- DnaT and PriC- DnaT (Michel and Sandler, 2017b; Sandler *et al.*, 2021)

1.6 Aims & Objectives

Since its discovery and establishment as the genetic material, DNA as well as the principles of DNA replication have been researched extensively, and important knowledge has been gained from the studies. However, there are still a lot of questions regarding the replication- transcription conflicts. So, the main aim of my research is to thoroughly understand the replication-transcription conflicts.

Recent experimental data from the Rudolph lab demonstrate interesting effects regarding the role of the replication restart proteins in the resolution of replication- transcription conflicts. As a result, another aim of this research is to determine how replication restart proteins and their associated pathways promote the resolution of replication–transcription conflicts and maintain replication fork stability.

By doing so, the detailed visualization of the blocks, clashes and collisions that the replisome faces combined with the detailed examination of all the proteins involved in the replication restart pathway will be feasible. The knowledge will help us decipher all the molecular proteins and reactions that are involved in vivo. Moreover, it will help us understand the exact reasons of the disassembly of some replication forks.

Chapter 2

Materials & Methods

2.1 Strains and Plasmids

Table 2.1. List of strains constructed and used during this PhD research.

STRAIN NUMBER	GENOTYPE	REFERENCE
SP001	oriZ::cm	Wang et al., 2011
SP002	srgA1	JD1092
SP003	*priA300 ΔlacIZYA oriZ-cat* pAM374	JD1314 × P1 RCe544 to Cm^r Ap^r
SP004	*priA300 ΔlacIZYA* pAM374	JD1314
SP005	*ΔlacIZYA srgA1 oriZ-cat*	JD1092 X P1 RCe504 to Cm^r
SP006	MG1655 *lacO*34	JD1746
SP007	MG1655 *lacO*22	JD1747
SP009	*MG1655*	RCe698
SP012	*W3110*	Lab stock
SP013	*priA*300	JD1382
SP014	*ΔpriB::dhfr*	RCe815
SP015	*ΔpriC::kan*	RCe814
SP017	*priA*300 *argEC*::apra-*lacO*22	SP013 x P1.SP007 to Ap^r (1)
SP018	*priA*300 *argEC*::apra-*lacO*22	SP013 x P1.SP007 to Ap^r (2)
SP019	Δ*priB* ::dhfr *argEC*::apra-*lacO*22	SP014 x P1.SP007 to Ap^r (1)
SP020	Δ*priB* ::dhfr *argEC*::apra-*lacO*22	SP014 x P1.SP007 to Ap^r (2)
SP021	Δ*priC*::kan *argEC*::apra-*lacO*22	SP015 x P1.SP007 to Ap^r (1)
SP022	Δ*priC*::kan *argEC*::apra-*lacO*22	SP015 x P1.SP007 to Ap^r (2)
SP023	*priA*300 *argEC*::apra-*lacO*34	SP013 x P1.SP006 to Ap^r (1)

SP024	*priA*300 *argEC*::apra-*lacO*34	SP013 x P1.SP006 to Ap^r (2)
SP025	Δ*priB* ::dhfr *argEC*::apra-*lacO*34	SP014 x P1.SP006 to Ap^r (1)
SP026	Δ*priB* ::dhfr *argEC*::apra-*lacO*34	SP014 x P1.SP006 to Ap^r (2)
SP027	Δ*priC*::kan *argEC*::apra-*lacO*34	SP015 x P1.SP006 to Ap^r (1)
SP028	Δ*priC*::kan *argEC*::apra-*lacO*34	SP015 x P1.SP006 to Ap^r (2)
SP029	*argEC*::apra-*lacO*22 lacI	SP007 xf lacI to Amp + IPTG^r
SP030	*priA*300 *argEC*::apra-*lacO*22 lacI	SP017 xf lacI to Amp + IPTG^r
SP031	Δ*priB* ::dhfr *argEC*::apra-*lacO*22 lacI	SP019 xf lacI to Amp + IPTG^r
SP032	Δ*priC*::kan *argEC*::apra-*lacO*22 lacI	SP022 xf lacI to Amp + IPTG^r
SP033	*argEC*::apra-*lacO*34 lacI	SP006 xf lacI to Amp + IPTG^r
SP034	*priA*300 *argEC*::apra-*lacO*34 lacI	SP023 xf lacI to Amp + IPTG^r
SP035	Δ*priB* ::dhfr *argEC*::apra-*lacO*34 lacI	SP024 xf lacI to Amp + IPTG^r
SP036	Δ*priC*::kan *argEC*::apra-*lacO*34 lacI	SP028 xf lacI to Amp + IPTG^r
SP037	*argEC*::apra-*lacO*240	RCe304
SP038	*priA*300 *argEC*::apra-*lacO*240	SP038 x P1.SP013 to Ap^r (1)
SP039	*priA*300 *argEC*::apra-*lacO*240	SP038 x P1.SP013 Ap^r (2)
SP040	*priA*300 *argEC*::apra-*lacO*240	SP038 x P1.SP013 to Ap^r (3)
SP041	Δ*priB* ::dhfr *argEC*::apra-lacO240	SP038 x P1.SP013 to Ap^r (4)
SP042	Δ*priB* ::dhfr *argEC*::apra-*lacO*240	SP038 x P1.SP014 to Ap^r (2)
SP043	Δ*priB* ::dhfr *argEC*::apra-lacO240	SP038 x P1.SP014 to Ap^r (3)
SP044	Δ*priB* ::dhfr *argEC*::apra-*lacO*240	SP038 x P1.SP014 to Ap^r (4)
SP045	Δ*priC*::kan *argEC*::apra-*lacO*240	SP038 x P1.SP015 to Ap^r (1)
SP046	*priA*300 *argEC*::apra-*lacO*240 lacI	SP039 xf lacI to Amp + IPTG^r

SP048	Δ*priB* ::dhfr *argEC*::apra-lacO240 lacI	SP042 xf lacI to Amp + IPTG^r
SP049	Δ*priC*::kan *argEC*::apra-*lacO*240 lacI	SP046 xf lacI to Amp + IPTG^r
SP050	*argEC*::apra-*lacO*240 lacI	SP038 xf lacI to Amp + IPTG^r
SP051	Δ*priB argEC*::apra-*lacO*22 lacI	SP021 xf lacI to Amp + IPTG^r
SP052	Δ*priB argEC*::apra-*lacO*34 lacI	SP027 xf lacI to Amp + IPTG^r
SP053	Δ*priB argEC*::apra-*lacO*240 lacI	SP042 xf lacI to Amp + IPTG^r
SP055	Δ*priC argEC*::apra-*lacO*22	SP007 x P1.SP015 to Ap^r (1)
SP056	Δ*priC argEC*::apra-*lacO*22	SP007 x P1.SP015 to Ap^r (2)
SP057	Δ*priC argEC*::apra-*lacO*22	SP007 x P1.SP015 to Ap^r (3)
SP058	TB28 *priA*300	Lab stock
SP059	*argEC*::apra-*lacO*34	SP009 x P1.SP006 to Ap^r (1)
SP060	*argEC*::apra-*lacO*34	SP009 x P1.SP006 to Ap^r (2)
SP061	Δ*priC argEC*::apra-*lacO*34	SP015 x P1.SP006 to Ap^r (1)
SP062	Δ*priC argEC*::apra-*lacO*34	SP015 x P1.SP006 to Ap^r (2)
SP063	TB28	Bernhard & de Boer, 2004
SP064	TB28 *argEC*::apra-*lacO*22	SP063 x P1.SP007 to Ap^r (1)
SP065	TB28 *argEC*::apra-*lacO*22	SP063 x P1.SP007 to Ap^r (2)
SP066	TB28 *argEC*::apra-*lacO*22	SP063 x P1.SP007 to Ap^r (3)
SP067	TB28 *argEC*::apra-*lacO*34	SP063 x P1.SP008 to Ap^r (1)
SP068	TB28 *argEC*::apra-*lacO*34	SP063 x P1.SP008 to Ap^r (2)
SP069	TB28 *argEC*::apra-*lacO*34	SP063 x P1.SP008 to Ap^r (3)
SP070	TB28 *priA*300 *argEC*::apra-*lacO*22	SP058 x P1.SP007 to Ap^r (1)
SP071	TB28 *priA*300 *argEC*::apra-*lacO*22	SP058 x P1.SP007 to Ap^r (2)

SP072	TB28 *priA*300 *argEC*::apra-*lacO*22	SP058 x P1.SP007 to Ap^r (3)
SP073	TB28 *priA*300 *argEC*::apra-*lacO*34	SP058 x P1.SP008 to Ap^r (1)
SP074	TB28 *priA*300 *argEC*::apra-*lacO*34	SP058 x P1.SP008 to Ap^r (2)
SP075	TB28 *priA*300 *argEC*::apra-*lacO*34	SP058 x P1.SP008 to Ap^r (3)
SP077	*priA*300 pTK135	SP013 xf SP076 to Amp^r
SP078	Δ*priB* pTK135	SP104 xf SP076 to Amp^r
SP079	Δ*priC* pTK135	SP015 xf SP076 to Amp^r
SP080	TB28 Δ*priB*	SP063 x P1.SP014 to Ap^r (1)
SP081	TB28 Δ*priB*	SP063 x P1.SP014 to Ap^r (2)
SP082	TB28 Δ*priC*	SP063 x P1.SP015 to Ap^r (1)
SP083	TB28 Δ*priC*	SP063 x P1.SP015 to Ap^r (2)
SP084	Δ*priB* TB28 *lacO*22	SP064 x P1.SP014 to Ap^r (1)
SP085	Δ*priB* TB28 *lacO*22	SP064 x P1.SP014 to Ap^r (2)
SP086	Δ*priC* TB28 *lacO*22	SP064 x P1.SP015 to Ap^r (1)
SP087	Δ*priC* TB28 *lacO*22	SP064 x P1.SP015 to Ap^r (2)
SP088	TB28 lacI	SP063 xf lacI to Amp + IPTG^r
SP089	TB28 *lacO*22 lacI	SP064 xf lacI to Amp + IPTG^r
SP090	*priA*300 TB28 *lacO*22 lacI	SP070 xf lacI to Amp + IPTG^r
SP091	Δ*priB* TB28 *lacO*22 lacI	SP084 xf lacI to Amp + IPTG^r
SP092	Δ*priC* TB28 *lacO*22 lacI	SP086 xf lacI to Amp + IPTG^r
SP093	Δ*priC* TB28 *argEC*::apra-*lacO*34	SP067 x P1.SP014 to Ap^r (1)
SP094	Δ*priC* TB28 *argEC*::apra-*lacO*34	SP067 x P1.SP014 to Ap^r(2)
SP095	Δ*priC* TB28 *argEC*::apra-*lacO*22	SP067 x P1.SP015 to Ap^r (1)

SP096	Δ*priC* TB28 *argEC*::apra-*lacO22*	SP067 x P1.SP015 to Ap^r (2)
SP097	TB28 *argEC*::apra-*lacO*34 lacI	SP067 xf lacI to Amp + IPTG^r
SP098	TB28 *priA*300 *argEC*::apra-*lacO*34 lacI	SP058 xf lacI Amp + IPTG^r
SP099	TB28 Δ*priB argEC*::apra-*lacO*34 lacI	SP093 xf lacI Amp + IPTG^r
SP100	Δ*priC* TB28 *argEC*::apra-*lacO*34 lacI	SP096 xf lacI Amp + IPTG^r
SP101	MG1655	RCe698
SP102	MG1655 oriC$^+$ oriZ$^+$::cm	RCe779
SP103	Δ*priB* oriC$^+$ oriZ$^+$	SP014 x P1.SP102 to Cm^r
SP104	Δ*priC* oriC$^+$ oriZ$^+$	SP015 x P1.SP102 to Cm^r
SP105	*priA*300 oriC$^+$ oriZ$^+$	SP013 x P1.SP102 to Cm^r
SP106	Δ*priC*::dhfr	JD1858
SP107	MG1655 oriC$^+$ oriZ$^+$ pTK135	SP102 xf pTK135 to Amp^r
SP108	*priA*300 oriC$^+$ oriZ$^+$ pTK135	SP105 xf pTK135 to Amp^r
SP109	Δ*priB* oriC$^+$ oriZ$^+$ pTK135	SP103 xf pTK135 to Amp^r
SP110	Δ*priC* oriC$^+$ oriZ$^+$ pTK135	SP104 xf pTK135 to Amp^r
SP111	TB28 *lacO*22 Δ*priC*::dhfr	SP064 x P1.SP106 to Tm^r (1)
SP112	TB28 *lacO*22 Δ*priC*::dhfr	SP064 x P1.SP106 to Tm^r (2)
SP113	TB28 Δ*priC*::dhfr *argEC*::apra-*lacO*34	SP067 x P1.SP106 to Tm^r (1)
SP114	TB28 Δ*priC*::dhfr *argEC*::apra-*lacO*34	SP067 x P1.SP106 to Tm^r (2)
SP115	TB28 *argEC*::apra-*lacO*22 Δ*priC*::dhfr lacI	SP111 xf lacI to Amp +IPTG^r
SP116	TB28 *argEC*::apra-*lacO*22 Δ*priC*::dhfr lacI	SP113 xf lacI to Amp +IPTG^r
SP117	Kan-YPet-DnaN	RCe723
SP118	MG1655 *priA*300 rpo*	JD1020

SP119	TB28 *priA*300	JD1303
SP120	MG1655 *priA*300	JD1382
SP121	TB28 oriC$^+$ oriZ$^+$ dinG argE::Tn10	JD1399
SP122	TB28 *lacO*240	SP063 x P1.SP038 to Ap^r (1)
SP123	TB28 *priA*300 *argEC*::apra-*lacO*240	SP063 x P1.SP038 to Ap^r (2)
SP124	TB28 *priA*300 *argEC*::apra-*lacO*240	SP058 x P1.SP038 to Ap^r (1)
SP125	TB28 *lacO*240	SP058 x P1.SP038 to Ap^r (2)
SP126	oriC$^+$ oriZ$^+$ rpo*35	RCe566
SP127	*priA*300 oriC$^+$ oriZ$^+$ kan-YPet-DnaN	SP105 x P1.SP117 to Kan^r
SP128	*priA*300 oriC$^+$ oriZ$^+$ rpo*	SP118 x P1.SP102 to Cm^r
SP129	oriC$^+$ oriZ$^+$ dnaN-YPet	RCe749
SP130	MG1655 dnaN-YPet	JD1473
SP131	TB28 *argEC*::apra-*lacO*240 ΔpriB	SP122 x P1.SP014 to Tm^r
SP132	TB28 priC::dhfr *argEC*::apra-*lacO*240	SP122 x P1.SP106 to Tm^r
SP133	TB28 *argEC*::apra-*lacO*240 lacI	SP122 xf lacI to Amp + IPTG^r
SP134	TB28 *priA*300 *argEC*::apra-*lacO*34 lacI	SP124 xf lacI to Amp + IPTG^r
SP135	TB28 Δ*priB argEC*::apra-*lacO*34 lacI	SP131 xf lacI to Amp + IPTG^r
SP136	TB28 Δ*priC argEC*::apra-*lacO*34 lacI	SP132 xf lacI
SP137	*recA*	DG063
SP138	*recB*	JD1269
SP139	*recF*	JD1254
SP140	Δ*priB* Δ*sbcCD*	JD1850
SP141	Δ*priC* Δ*sbcCD*	JD1859

SP142	*priA*300 Δ*sbcCD*	JD1852
SP143	Δ*priB priA*300	JD1847
SP144	Δ*priC priA*300	JD1842
SP145	*srgA1*	JD1092
SP146	oriC⁺ oriZ⁺ *srgA1*	JD1849
SP147	oriC⁺ oriZ⁺ Δ*priB*	JD1672
SP148	oriC⁺ oriZ⁺ Δ*priC*	JD1673
SP149	oriC⁺ oriZ⁺ *priA*300	JD1305
SP150	oriC⁺ oriZ⁺ Δ*sbcCD*	JD1428
SP151	oriC⁺ oriZ⁺ Δ*priB* Δ*sbcCD*	JD1844
SP152	oriC⁺ oriZ⁺ Δ*priC* Δ*sbcCD*	JD1860
SP153	oriC⁺ oriZ⁺ *priA*300 Δ*sbcCD*	JD1854
SP154	oriC⁺ oriZ⁺ Δ*priB priA*300	JD1848
SP155	oriC⁺ oriZ⁺ Δ*priC priA*300	JD1851
SP156	oriC⁺ oriZ⁺ rpo*35	RCe566
SP157	*priA*300 dnaN-YPet	SP013 x P1.SP117 to Kanʳ (1)
SP158	*priA*300 dnaN-YPet	SP013 x P1.SP117 to Kanʳ (2)
SP159	oriC⁺ oriZ⁺ *priA*300 dnaN-YPet	SP105 x P1.SP117 to Kanʳ (1)
SP160	oriC⁺ oriZ⁺ *priA*300 dnaN-YPet	SP105 x P1.SP117 to Kanʳ (2)
SP161	MG1655 TB28 oriC⁺ oriZ⁺ ΔdinG argE::Tn 10	SP009 x P1.SP121 (1)
SP162	MG1655 TB28 oriC⁺ oriZ⁺ ΔdinG argE::Tn 10	SP009 x P1.SP121 (2)

SP163	oriC⁺ oriZ⁺ rpo*	SP161 x P1.SP126 (1)
SP164	oriC⁺ oriZ⁺ rpo*	SP161 x P1.SP126 (3)
SP165	*priA*300 oriC⁺ oriZ⁺ ΔdinG argE::Tn 10	SP105 x P1.SP121 (1)
SP166	*priA*300 oriC⁺ oriZ⁺ ΔdinG argE::Tn 10	SP105 x P1.SP121 (3)
SP167	TB28 lacI	SP063 xf lacI to Amp + IPTG^r
SP168	TB28 *argEC*::apra-*lacO*22 lacI	SP064 xf lacI to Amp + IPTG^r
SP169	TB28 *argEC*::apra-*lacO*34 lacI	SP067 xf lacI to Amp + IPTG^r
SP170	TB28 *argEC*::apra-*lacO*240 lacI	SP122 xf lacI to Amp + IPTG^r
SP171	TB28 *priA*300 lacI	SP058 xf lacI to Amp + IPTG^r
SP172	TB28 *priA*300 *argEC*::apra-*lacO*22 lacI	SP070 xf lacI to Amp + IPTG^r
SP173	TB28 *priA*300 *argEC*::apra-*lacO*34 lacI	SP058 xf lacI to Amp + IPTG^r
SP174	TB28 *priA*300 *argEC*::apra-*lacO*34 lacI	SP124 xf lacI to Amp + IPTG^r
SP175	TB28 Δ*priB* lacI	SP080 xf lacI to Amp + IPTG^r
SP176	TB28 Δ*priB argEC*::apra-*lacO*22 lacI	SP084 xf lacI to Amp + IPTG^r
SP177	TB28 Δ*priB argEC*::apra-*lacO*34 lacI	SP093 xf lacI to Amp + IPTG^r
SP178	TB28 Δ*priB argEC*::apra-*lacO*240 lacI	SP131 xf lacI to Amp + IPTG^r
SP179	*priA*300 TB28 oriZ+ ΔdinG argE::Tn10	SP013 x P1.SP121 (1)
SP180	*priA*300 TB28 oriZ+ ΔdinG argE::Tn10	SP013 x P1.SP121 (2)
SP181	ΔsbcCD	RCe562
SP182	*priA*300 rpo*	SP179 x P1.SP126 (1)
SP183	*priA*300 rpo*	SP179 x P1.SP126 (2)
SP184	*priA*300 oriC+ oriZ+ rpo*	SP165 x P1.SP126
SP185	AB5075 (1)	(Jacobs *et al.*, 2014)

SP186	AB5075 (2)	(Jacobs *et al.*, 2014)
SP187	oriC^{+} oriZ^{+} rpo* pTK135	SP126 xf pTK135 to Amp^{r}
SP188	rpo* pTK135	SP163 xf pTK135 to Amp^{r}
SP189	*priA*300 rpo* pTK135	SP182 xf pTK135 to Amp^{r}
SP190	*priA*300 oriC^{+} oriZ^{+} rpo* pTK135	SP184 xf pTK135 to Amp^{r}
SP191	TB28 ΔpriC lacI	SP082 xf lacI to Amp + IPTG^{r}
SP192	TB28 *argEC*::apra-*lacO*34pBad24	SP067 xf pBad24 to Amp^{r}
SP193	*priA*300 TB28 *argEC*::apra-*lacO*34 pBad24	SP073 xf pBad24 to Amp^{r}
SP194	Δ*priB* TB28 *argEC*::apra-*lacO*34 pBad24	SP093 xf pBad24 to Amp^{r}
SP195	Δ*priC* TB28 *argEC*::apra-*lacO*34 pBad24	SP095 xf pBad24 to Amp^{r}
SP196	*priA*300	SP013
SP197	*priA*300 kan-YPet-DnaN	SP013 x P1.RCe273 to Kan^{r} (1)
SP198	*priA*300 kan-YPet-DnaN	SP013 x P1.RCe273 to Kan^{r} (2)
SP199	rpo* kan-YPet-DnaN	SP163 x P1.RCe273 Kan^{r} (1)
SP200	rpo* kan-YPet-DnaN	SP163 x P1.RCe273 to Kan^{r} (1)
SP201	*priA*300 rpo* kan-YPet-DnaN	SP182 x P1.RCe273 to Kan^{r} (1)
SP202	*priA*300 rpo* kan-YPet-DnaN	SP182 x P1.RCe273 to Kan^{r} (1)
SP203	oriC^{+} oriZ^{+} *priA*300 kan-YPet-DnaN	SP105 x P1.RCe273 to Kan^{r} (1)
SP204	oriC^{+} oriZ^{+} *priA*300 kan-YPet-DnaN	SP105 x P1.RCe273 to Kan^{r} (2)
SP205	oriC^{+} oriZ^{+} rpo* kan-YPet-DnaN	SP126 x P1.RCe273 to Kan^{r} (1)
SP206	oriC^{+} oriZ^{+} rpo* kan-YPet-DnaN	SP126 x P1.RCe273 to Kan^{r} (2)
SP207	oriC^{+} oriZ^{+} *priA*300 rpo* kan-YPet-DnaN	SP184 x P1.RCe273 to Kan^{r}
SP208	oriC^{+} oriZ^{+} *priA*300 rpo* kan-YPet-DnaN	SP184 x P1.RCe273 to Kan^{r}

SP209	MG1655 oriX	DG014
SP210	TB28 oriX:cat	RCe751
SP211	MG1655 oriX (S)	SP009 x P1.SP209 to Cm^r
SP212	MG1655 oriX (M)	SP009 x P1.SP209 to Cm^r
SP213	MG1655 oriX (L)	SP009 x P1.SP209 to Cm^r
SP214	*priA*300 oriX (S)	SP013 x P1.SP209 to Cm^r
SP215	*priA*300 oriX (M)	SP013 x P1.SP209 to Cm^r
SP216	*priA*300 oriX (L)	SP013 x P1.SP209 to Cm^r
SP217	Δ*priB* oriX (S)	SP014 x P1.SP209 to Cm^r
SP218	Δ*priB* oriX (M)	SP014 x P1.SP209 to Cm^r
SP219	Δ*priB* oriX (L)	SP014 x P1.SP209 to Cm^r
SP220	Δ*priC* oriX (S)	SP015 x P1.SP209 to Cm^r
SP221	Δ*priC* oriX (M)	SP015 x P1.SP209 to Cm^r
SP222	Δ*priC* oriX (L)	SP015 x P1.SP209 to Cm^r
SP223	TB28 *argEC*::apra-*lacO*22 lacI	SP064 xf lacI to Amp + IPTG^r
SP224	*priA*300 TB28 *argEC*::apra-*lacO*22 lacI	SP070 xf lacI to Amp + IPTG^r
SP225	Δ*priB* TB28 *argEC*::apra-*lacO*22 lacI	SP084 xf lacI to Amp + IPTG^r
SP226	Δ*priC* TB28 *argEC*::apra-*lacO*22 lacI	SP086 xf lacI to Amp + IPTG^r
SP227	TB28 *argEC*::apra-*lacO*34 lacI	SP067 xf lacI to Amp + IPTG^r
SP228	*priA*300 TB28 *argEC*::apra-*lacO*34 lacI	SP073 xf lacI to Amp + IPTG^r
SP229	Δ*priB* TB28 *argEC*::apra-*lacO*34 lacI	SP093 xf lacI to Amp + IPTG^r
SP230	Δ*priC* TB28 *argEC*::apra-*lacO*34 lacI	SP095 xf lacI to Amp + IPTG^r
SP231	MG1655 oriX pTK135 (M)	SP212 xf pTK135 to Amp + IPTG^r

SP232	*priA*300 oriX pTK135 (M)	SP215 xf pTK135 to Amp + IPTG^r
SP233	Δ*priB* oriX pTK135 (S)	SP217 xf pTK135 to Amp + IPTG^r
SP234	Δ*priB* oriX pTK135 (M)	SP218 xf pTK135 to Amp + IPTG^r
SP235	Δ*priC* oriX pTK135 (S)	SP220 xf pTK135 to Amp + IPTG^r
SP236	Δ*priC* oriX pTK135 (M)	SP221xf pTK135 to Amp + IPTG^r
SP237	Δ*priC* oriX pTK135 (L)	SP222 xf pTK135 to Amp + IPTG^r
SP238	oriC$^+$ oriZ$^+$	RCe926
SP239	oriC$^+$ oriZ$^+$ *priA*300	RCe928
SP240	oriC$^+$ oriZ$^+$ Δ*priB* DnaN-YPet	SP103 x P1.SP117 to Kan^r (1)
SP241	oriC$^+$ oriZ$^+$ Δ*priB* DnaN-YPet	SP103 x P1.SP117 to Kan^r (2)

Table 2.2. List of plasmids used during this PhD research.

PLASMID	DESCRIPTION	REFERENCE
pRC7	Unstable mini-F plasmid, *lacIZYA*+	(Bernhardt and de Boer, 2004a)
pPM306	*lacI* expression vector	Lab Stock
pKD46	Lambda Red recombinase *araBAD*p	(Datsenko and Wanner, 2000)
pTK135	Cas1-Cas2 fluorescent reporter; used as a biomarker for DNA replication/spacer acquisition	Killelea *et al.* 2023

2.2 Media, Buffers and Antibiotics

All liquid media and agar used for the experiments were made and autoclaved the same day at 121°C and 103 kPa for 15 min.

2.2.1 LB medium

LB Broth was made by mixing the following: 1% bacto tryptone (BD Biosciences), 0.5% yeast extract (BD Biosciences), 0.05% sodium chloride (Fisher), 0.002 M sodium hydroxide (Fisher), pH~7.0. To make LB agar, the LB broth was evenly distributed into 200 ml aliquots, and 1.5% agar (Sigma Aldrich) was added (3 g for every 200ml of LB broth).

2.2.2 Mu medium

Mu Broth was made by mixing the following: 1% bacto tryptone (from BD Biosciences), 0.5% yeast extract (BD Biosciences), 1% sodium chloride (from Fisher), 0.002 M sodium hydroxide (Fisher), and the pH was adjusted to 7.0. To make Mu agar, the Mu broth was evenly distributed into 200 ml aliquots, and 1% agar (Sigma Aldrich) was added (2 g for every 200ml of Mu broth).

2.2.3 M9 Minimal salt medium

DifcoTM M9 Minimal Salts base (BD Biosciences) was used in the preparation of M9 minimal media. The powder makes a 5 × M9 minimal salts solution, which was distributed into 50 ml aliquots and autoclaved.

One 50 ml aliquot was diluted 5-fold, re-distributed into 50 ml aliquots and autoclaved. M9 agar: A 50 ml 5 × M9 minimal salts aliquot was added to 200 ml molten 1.8% water agar (3.6 g agar). Final concentrations of 2 mM magnesium sulphate (MgSO4), 0.1mM calcium chloride (CaCl2) (Sigma Aldrich), 0.4% glucose and 0.05% casamino acids were aseptically added to the M9 agar, ensuring that the bottle remained at 50°C to prevent it setting prematurely. M9 minimal medium: The 1 × M9 minimal salts was used for serial dilutions of bacterial cultures and bacteriophage P1 lysate cultures. Magnesium sulphate (2 mM) and calcium chloride (0.1 mM) were added to 50 ml of sterile 1× M9 salts.

2.2.4 Antibiotics and supplements

Table 2.3 List of antibiotics and supplements used along with their stock and final concentrations

Supplement	Stock concentration	Final concentration
Ampicillin	5 mg/ml	50 µl/ml
Apramycin	2 mg/ml	20 µl/ml
Kanamycin	4 mg/ml	40 µl/ml
Chloramphenicol	1 mg/ml	10 µl/ml
Tetracycline	10 mg/ml	10 µl/ml
Trimethoprim	50 mg/ml	10 µl/ml
Arabinose	20 %	0.20 %
Glucose	20 %	0.20 %

2.3 Transductions

2.3.1 Production of P1 lysates

For all the P1 lysates and transductions the starting lysate was the P1 phage from the wild type MG1655 from *Escherichia coli*. For their preparation, an overnight culture of the donor strain was grown and then 100 µl of it was used and inoculated using 11 ml Mu so that the absorbance at 600nm was 0.05. The culture was incubated at 37 °C until the absorbance at 600nm was 0.3. After that, 200 µl of calcium chloride of concentration 0.5M was added and the culture was incubated at 37 °C for 10 min. Then, 30 µl of the wild type P1 phage was added and incubated at 37 °C and after 4 hours lysis was visible. 500 µl of chloroform was added to the culture and the culture was vortexed. The produced P1 lysate was transferred into a fresh tube and was spun for 10 min, 5000 g and 500 µl of chloroform were added and then stored at 4 °C.

2.3.2 P1 Transductions

For the P1 transductions a fresh culture of the *E. coli* recipient strain was grown at 37°C overnight in LB. Then, 50 µl of sterile calcium chloride of concentration 0.5 M was added and the culture was mixed. After that, 200 µl of the desired P1 phage was added with 200 µl of the culture and mixed. The mix was incubated for 20 mins at 37 °C. Then, 1ml of LB broth and 2 µl of Sodium Citrate were added and the mixture was incubated for 1.5 hours at

37 °C at 100x g . Finally, the mixture was spun down at 6000x g for 1min. 200 µl of Sodium Citrate were added and the mixture was plated onto the selective antibiotic plates.

2.3.3 Streak Tests

Four single colonies from the selective antibiotic plate were picked and purified by streaking to single cells on LB plates and incubated overnight at 37°C. The next day three single colonies were used to grow overnight cultures together with the recipient strain and incubated at 37°C. 10 µl of the three transductants and the recipient strain were plated on the following antibiotic plates: LB, Chloramphenicol, Ampicillin, Kanamycin, Apramycin, Tetracycline and Trimethoprim. Also, the culture was plated on one LB plate and another LB plate which was exposed to 60 J/m^2 UV for 60 s and was named UV plate. All the plates were incubated at 37°C for at least 24 hours and then the growth pattern of the transductants was compared to that of the recipient.

If the growth pattern of the transductants was the expected one, two of the transductants were frozen. For freezing a bacterial strain, an overnight culture was grown and the next day 900 µl of it were mixed with 900 µl of 80% glycerol solution and the mixture was stored in a cryotube (Nunc) in two freezers, one at -20°C and one at -80°C as a backup storage.

2.4 Bacterial transformations

2.4.1 Preparation of electrocompetent cells

For the preparation of electrocompetent bacterial cells, an overnight culture of the recipient strain was used to inoculate 11 ml LB to an optical density of 0.05 and the culture was grown to an optical density of 0.6 in a shaking water bath. Once grown to the correct density, the cells and all reagents were kept on ice for the whole procedure. The cells were washed four times in M9 to reduce the ionic strength of the culture then pelleted and resuspended in 10 ml ice cold 10% glycerol solution. This was repeated three more times in M9 with the following volumes of glycerol: 5 ml, 1 ml and 0.5 ml. After the last wash step, the cell pellet was resuspended in the final drop that remained after the final 0.5 ml of glycerol solution.

2.4.2 Cell transformation via electroporation

To transform the bacterial cultures using electroporation, first the plasmid of interest was added to the cells, and the mixture was incubated on ice for 10 min. After that, it was transferred to a chilled electroporation cuvette. Immediately after electroporation at 1.75 kV for at least 4 ms using the Eppendorf Eporator, the cells were mixed with 1.6 ml SOC and incubated at 37°C for two hours in a tube rotator. After that, 100 µl of the transformation culture were plated onto the relevant antibiotic plates for selection of transformants and streaked to single colonies. The plates were then incubated at 37°C for 24 hours. The appearance of single colonies on the plates indicated that the electroporation was successful.

2.5 Spot dilutions assay

To carry out the spot dilutions assay, the bacterial strain of interest was streaked on the appropriate plate (which normally is LB but if a plasmid is present in a strain, then it is plated onto medium with ampicillin) and a fresh overnight culture of it was grown. The next day, 100 µl of the bacterial overnight cultures were inoculated with 11ml of Mu and the appropriate antibiotics if needed to the optical density of 0.05.

After that, the 11ml of the culture was grown until the optical density reached 0.4. Immediately, the cells were put on ice for a couple of minutes, and the cells were spun in a pre-cooled centrifuge for 5 mins at 5000 rpm and the 11ml of Mu only were used to resuspend the pellet. After that, 100 µl were used to perform serial dilutions in 1x M9 medium from 10^{-1} to 10^{-5}. Finally, 10 µl of the dilutions were plated on the appropriate selection plate or plates which were then incubated at 37°C overnight.

2.6 Fluorescent microscopy

For the fluorescent microscopy experiment, 11 ml of Mu were used to inoculate 100 µl of fresh to an optical density of 0.05. The 11 ml culture was then incubated in a shaking water bath at 37°C until the optical density was 0.3. While the culture was incubated, the agar platforms were prepared on microscope slides. A Thermo Scientific adhesive gene frame with dimensions 1.5 × 1.6 cm was applied to the microscope slide. Then, 100 µl of 1% agarose solution in M9 0.2 % glucose minimal medium were pipetted in to the well, with extra care so that no air bubbles were generated that would interfere

with the results, and at the same time was covered with a second slide to create an even surface. After 20 mins at room temperature, the slides were carefully separated and allowed to dry for 12 min at room temperature, or 5 min at 42 and then used immediately. 1 µl of the bacterial culture was transferred to the agar platform of the microscope slide. The microscope slide was then carefully rotated five times so that the bacterial culture drop spread out evenly. After the drop of the bacterial culture had dried, a cover slip was used to cover it. Finally, the Nikon Ti-U inverted microscope equipped with an ORCA Flash 4.0 camera from Hamamatsu was used to visualize the bacterial culture.

2.7 Computational Image Analysis and Cell Quantification

Following the acquisition of raw microscopy data, a high-throughput computational pipeline was employed to quantify *E. coli* morphological characteristics. This pipeline was developed in Python by Dr. Lewis Frame (University of York) to ensure an objective, reproducible, and statistically rigorous analysis of cellular dimensions. The Python-based preprocessing and segmentation scripts utilized for this workflow are documented in Appendix A. To address the specific requirements of focus quantification and non-parametric population analysis in Chapter 4, a secondary analytical pipeline was independently developed by the candidate using the R programming language, as documented in Appendix B.

2.7.1 Image Preprocessing and Deep-Learning Segmentation

To achieve precise cell identification, automated segmentation was performed using a ResNeXt-based Convolutional Neural Network (CNN), specifically the CpnResNeXt101UNet architecture. Before the segmentation phase, raw images underwent a standardization protocol where they were normalized for intensity and resized to a uniform resolution of 1024×1024 pixels. This preprocessing ensured that the model could perform pixel-wise classification with high sensitivity, effectively delineating individual cell boundaries even in high-density fields where cells were in close physical proximity. The model generated high-fidelity binary masks for every identified object, which served as the foundation for all subsequent measurements.

2.7.2 Post-Processing and Data Integrity Filters

To maintain the integrity of the morphological dataset, a series of automated post-processing filters were implemented using the skimage and scipy libraries. To avoid sampling bias and measurement errors, objects touching the edges of the imaging frame were automatically excluded. Furthermore, rigorous biological criteria were applied: objects were only included in the final analysis if they exceeded a minimum area threshold of 100 pixels and maintained an eccentricity value greater than 0.8. These parameters were critical for distinguishing healthy, single *E. coli* cells from non-cellular artifacts, background noise, or unresolved cell clusters that could otherwise skew the population distribution.

2.7.3 Feature Extraction and Statistical Framework

Morphological features, primarily cell length and width, were extracted from the refined masks. To translate these computational units into biological metrics, all pixel-based measurements were converted to micrometers using a calibrated conversion factor of 0.065 µm per pixel. Data management and statistical calculations were conducted using the pandas, numpy, and scipy.stats libraries.

For Chapter 3 data management and statistical calculations for cell dimensions were conducted using the *pandas*, *numpy*, and *scipy.stats* Python libraries. The resulting cellular populations were visualized through distribution plots to represent mean and median values across a minimum of three independent biological replicates. To determine the statistical significance of observed morphological shifts between experimental conditions and wild-type controls, independent two-sample Welch's t-tests were performed. This approach was specifically selected to provide a robust analysis that accounts for potentially unequal variances between mutant and control populations, ensuring the scientific validity of the findings presented in Results Chapter 3.

The resulting cellular populations were visualized through distribution plots, including histograms and violin plots, to represent the mean and median values across a minimum of three independent biological replicates. To determine the statistical significance of observed morphological shifts between experimental conditions and wild-type controls, independent two-sample Welch's t-tests were performed. This approach was specifically selected to provide a robust analysis that accounts for potentially unequal variances between mutant and control

populations, ensuring the scientific validity of the findings presented in Results Chapter 1.

For the analysis of Cas1-Cas2 foci accumulation in Chapter, a distinct statistical framework was employed to account for the non-parametric nature of discrete count data. These analyses and high-resolution visualizations were performed independently by the candidate using R (v4.x) within the RStudio environment. Data normality was assessed using the Shapiro-Wilk test; given the non-normal distribution of foci counts, the Wilcoxon rank-sum test (Mann-Whitney U) was utilized for all pairwise comparisons against the oriC^+ oriZ^+ control.

Populations were visualized using violin plots to represent data density, supplemented with jittered raw data points to display the full distribution of the N > 2,000 cells analyzed. All R source code and specific package dependencies (ggplot2, ggpubr, rstatix) are documented in Appendix B.

2.8 Extraction of bacterial gDNA

For the extraction of bacterial genomic DNA up to 2 x 10^9 bacterial cells were harvested in a 2 ml centrifuge tube by centrifugation for 10 mins at 7000 g at room temperature. The supernatant was then removed and discarded taking care not to disturb the pellet. The bacterial pellet was then resuspended in 180 µl Buffer ATL and the sample was transferred to a 2ml sample tube. 20 µl of lysozyme with a concentration 100 mg/ml was added and mixed by tapping the tube. The sample was placed in a thermomixer and incubated at 37 °C with shaking at 1500x g for 30 mins. After that, 20 µl of proteinase K

were added and mixed by tapping the tube. The mixture was then incubated at 56 °C with shaking at 1500x g for 2hrs.

After 2 hrs, 4 µl of RNase A were added to the sample, mixed by pulse vortexing and incubated for 10 mins at room temperature. Then 20 µl MagAttract Suspension G and 280 µl Buffer MB were added to the sample and mixed by pulse vertexing. The sample was transferred to the tube holder of the MagAttract Magnetic Rack and incubated at room temperature at 1000x g for 10 min. The tube holder was then placed on the magnetic base for around 1 min until bead separation had been completed and the supernatant was removed. 750 µl Buffer MW1 were added to the sample, and the tube holder was placed onto the mixer. The sample was incubated at room temperature for 3 mins at 1000x g and the tube was placed on the magnetic rack until bead separation had been completed. The same procedure was repeated by the addition of 750 µl Buffer PE and distilled water. The final step of the procedure to extract the bacterial genomic DNA was the addition of TE. The extracted gDNA was stored at 4 °C.

Chapter 3

Dynamics of Replication Fork Blocks and their Resolution

3.1 Introduction and research aims

As established, maintenance of DNA integrity and DNA replication are essential for cell survival and continuation of life. However, the fundamental process of DNA replication is continuously susceptible to various blocks and conflicts with other molecules which, if not resolved timely and efficiently, will lead to death. These include DNA lesions arising from UV radiation, the formation of secondary DNA structures such as G-quadruplexes (G4s), and particularly collisions between the replication and transcription machineries, all of which pose persistent threats to cellular homeostasis (Mirkin and Mirkin, 2007; Kciuk *et al.*, 2020; Parekh *et al.*, 2023). However, knowledge regarding the specific mechanisms by which the cells successfully overcome the various replication fork blocks appears to be limited. Also, although it is established that there are three replication restart pathways: PriA-PriB-DnaT, PriA-PriC-DnaT and PriC-DnaT (Michel and Sandler, 2017b), it remains unclear which of the three restart pathways is more actively involved when DNA replication forks encounter DNA and/or protein obstacles of various lengths.

So, the aims of the first research chapter were to investigate in vivo how replisomes interact with replication blocks of various lengths and chromosomal positions and to accurately determine which of the three replication restart pathways is predominantly utilized in the resolution of stalled replication forks as a result of nucleoprotein blocks.

To achieve these aims with high quantitative precision, all morphological assessments and cellular dimensions reported in this chapter were analyzed using a custom-developed Python computational pipeline (detailed in Section 2.7 and Appendix A). By utilizing a ResNeXt-based Convolutional Neural Network for automated cell segmentation, we ensured the objective identification of cell boundaries across thousands of individual cells (n > 2000

per condition). This deep-learning approach allowed for the extraction of accurate length and width measurements following a calibrated conversion of 0.065 μm per pixel. Statistical significance between the various replication restart mutants and wild-type controls was determined using independent two-sample Welch's t-tests, providing a robust mathematical foundation for the phenotypic differences observed when replication forks are challenged by nucleoprotein blocks.

3.2 The experimental set up

To answer these questions the following experimental set up was employed. *E. coli* MG1655 derivative strains containing 22, 34, and 240 tandem repeats of the *lacO* operator were utilized. These strains were originally constructed and characterized by the laboratory of Professor Peter McGlynn (University of Nottingham and University of York) to create site-specific protein-DNA barriers (Figure 21) (Payne *et al.*, 2006a; McGlynn and Guy, 2008a). Then *priB, priC* and *priA300* were deleted one at a time from the cells. Finally, the cells were transformed with a plasmid that allowed expression of the lac repressor LacI, which is controlled via the p_{araBAD} promoter (Payne *et al.*, 2006b).

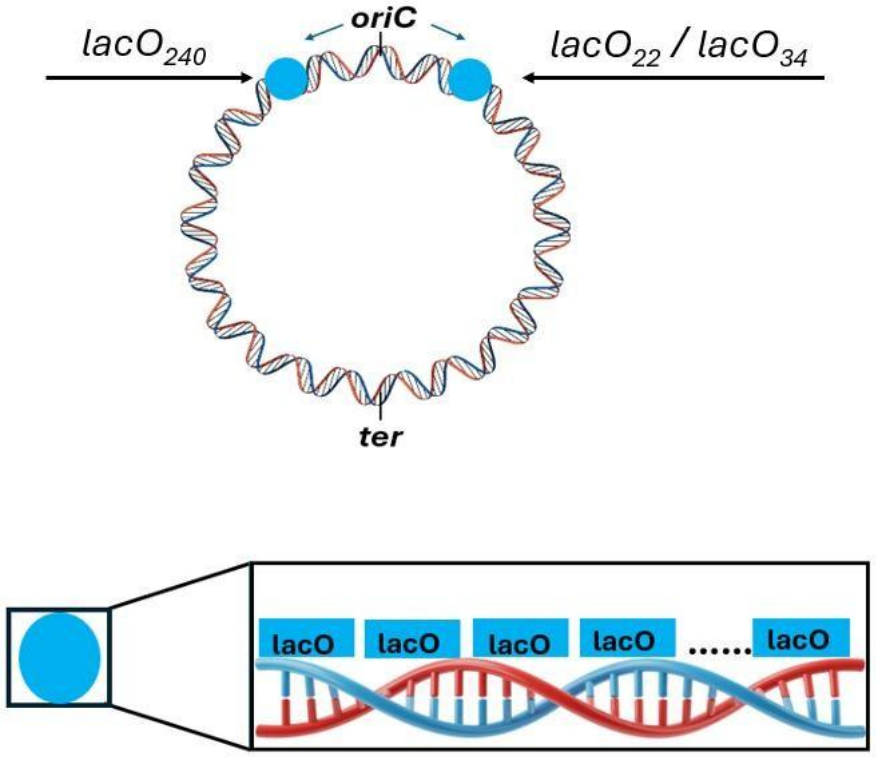

Figure 21. The locations of inserting the tandem *lacO* repeats in the *E. coli* chromosome. The 22 and 34 tandem repeats of the lac operator sequences are integrated in the right replichore while the 240 tandem repeats are integrated in the left replichore. The strain that contains the 22 tandem repeats of the lac operator is the PM222 and the

strain that contains the 34 tandem repeats of the lac operator is the BP41(Payne et al., 2006). The strain that contains 240 lac operator sites was already present in the lab. Image created with BioRender.

It is important to highlight and explain the reasons that *DnaT* and *priA* were not deleted from the cells. *DnaT* was not deleted from the cells because published experimental results report that the deletion of *DnaT*, although not fatal for the cells, causes severe growth defects in the cells. Also, the presence of *DnaT* in all replication restart pathways is an additional strong indicator of its essential presence. As for *priA*, it was not deleted from the cells either because, even though it is not present in all replication restart pathways like *DnaT*, *priA* mutants also undergo extreme cellular defects as is the case with *DnaT* mutants.

However, cells with a *priA300* allele were also tested. Specifically, although the cells with a *priA300* allele retain the replication restart function of *priA*, they are deficient of the helicase PriA[K230R] protein but are viable and to date literature does not report that they undergo growth defects (Zavitz and Marians, 1992; Sandler, Samra and Clark, 1996). Also, importantly it has been reported that PriA helicase might also be involved in the PriA-PriC mediated pathway (Sandler *et al.*, 2001; Windgassen *et al.*, 2018). This means that the use of *priA300*, *ΔpriB* and *ΔpriC* cells should enable the establishment of the hierarchy of the three replication restart pathways needed for restarting DNA synthesis at the model block.

3.2.1 The inducible *lacO*-LacI system

After all the cells were constructed and transformed with the plasmid that expresses LacI protein, the cultures were grown to mid-exponential phase in the presence of IPTG and then plated on 3 types of plates which all contained

Ampicillin to ensure the presence and therefore expression of the LacI plasmid. The first plate contained arabinose, the second contained IPTG and the third contained both arabinose and IPTG. The choice of these media is attributed to the characteristics of this specific plasmid that expresses the LacI protein and to the principles of the interaction between LacI and *lacO*. More specifically, since the gene that expresses this specific LacI protein has an arabinose-induced promoter this means that the presence of arabinose will lead to the expression of LacI plasmid, hence LacI protein will be produced and will bind to the tandem *lacO* repeats and the *lacO*-LacI repeats will be a replication fork block (figure 22A). On the other hand, the presence of IPTG will cause a conformational change of the LacI protein preventing its binding on the tandem *lacO* repeats and so there will not be a block for the replication fork (figure 22B). Finally, regarding the medium containing both arabinose and IPTG, although the presence of arabinose will lead to the expression and production of LacI, the presence of IPTG will prevent binding of the produced LacI protein on the tandem *lacO* repeats, hence there will not a block for replication fork progression (figure 22C).

This means that this is an inducible system that can be turned on, by the addition of arabinose and turned off by the addition of IPTG, also schematically illustrated in figure 22. So, this inducible system leads to the presence of two states, one where there is not a replication fork block, and one where there is a replication fork block. Thus, this experimental setup provides a novel system to investigate how each replication restart protein contributes to cell viability and growth.

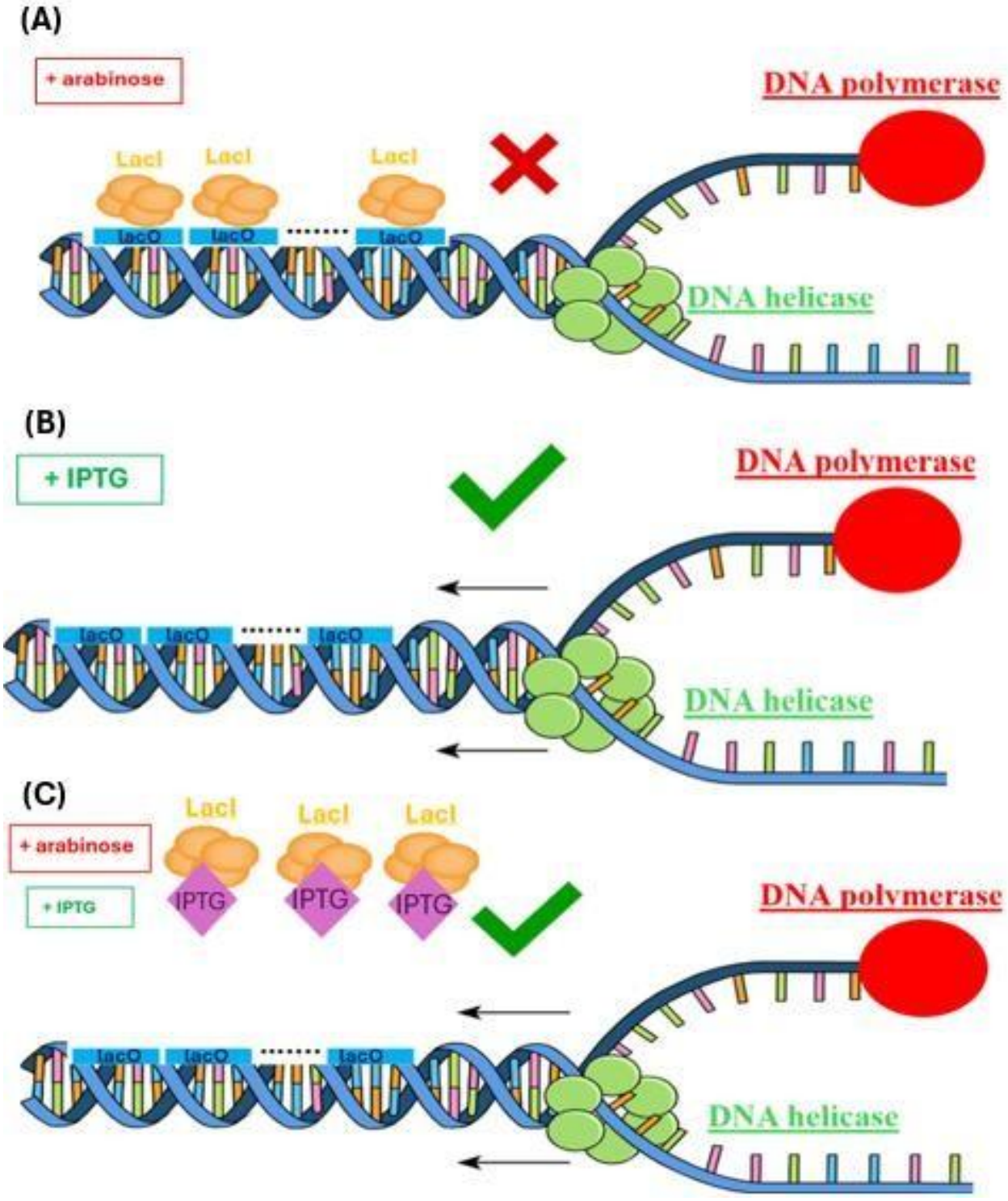

Figure 22. Regulation of replication fork progression in the araBAD–placI system. (A) Arabinose-induced fork block. Addition of arabinose in the medium induces expression of the LacI repressor. LacI then binds to tandem *lacO* repeats integrated into the *E.coli* chromosome, creating a physical barrier that stalls replication fork progression. (B) IPTGmediated fork progression. In the presence of IPTG only, LacI is produced but unable to bind to lacO repeats due to conformational inhibition. The replication fork therefore

progresses normally without blocks. (C) Arabinose and IPTG together. Arabinose strongly induces LacI expression; however, IPTG binding prevents LacI from interacting with *lacO* repeats. As a result, despite elevated LacI levels, replication fork progression is not blocked. Image created with BioRender.

The experiment initiated by utilising the wild-type *E. coli* strain. However, although the results indicated clear growth difficulties, the results also showed large degree of inconsistency between experiments. More specifically, in almost every experiment one of the strains lacking a restart protein grew better under expressing conditions than in the previous experiments, where for instance bacterial colonies of the strain lacking *priC* appeared larger under LacI -inducing conditions than if LacI was not induced which is not possible (Figure 23). Similar effects were also observed for the other mutants as well in all the other experiments. This result is a strong indication of the emergence of rapidly arising spontaneous suppressor mutations that lead to faster growth in the presence of *lacO*-LacI model blocks. However, even when the cells were freshly transformed with the LacI expression plasmid, again some variability was observed.

This signifies that suppressor mutations existed already in the *lacO* array strains even before the introduction of the LacI expression plasmid. It is also important to highlight that during generating strains which carry the *lacO* arrays, all construction steps were carried out in the presence of IPTG to prevent any LacI binding. This is attributed to the fact that the chromosomal *lacI* gene is expressed in MG1655 at low levels. Even though there were low levels of LacI in the presence of IPTG again there was some interference, enabling thus again the emergence of spontaneous suppressor mutants which outgrew the normal wild type cells (Figure 23).

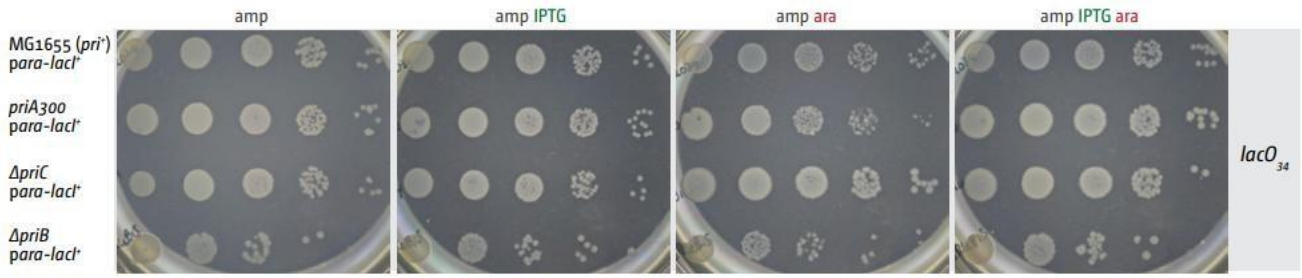

Figure 23. Emergence of growth suppressor mutations in MG1655 strains. Representative spot dilution assays showing inconsistent growth phenotypes in the MG1655 background. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. When utilising MG655 strains for the model fork block experiments in almost every experiment one of the strains lacking restart proteins grew better under expressing conditions than in previous experiments, where colonies of the *ΔpriC* strain appeared larger under *lacI* inducing conditions than if *lacI* was not induced. In other experiments similar effects were observed for the other mutants as well. This behaviour is indicative of rapidly arising spontaneous suppressor mutations that enable faster growth in the presence of *lacO*-LacI model blocks. To ensure reproducibility of the results the spot dilutions assay was repeated three times.

To avoid the emergence of suppressor mutations, all the strains were reconstructed in TB28 strain. TB28 is a MG1655 derivative in which the *lacIZYA* genes are deleted (Bernhardt and de Boer, 2004b). The utilisation of TB28 enabled the successful reconstruction of strains with the *lacO* arrays that behaved in a considerably more stable way. Importantly, spontaneous suppressor mutations were not observed in the strains carrying the *lacO* arrays but only when the *lacI* expression plasmid was transformed in them. To carry out the experiments, cells of the strains of interest freshly were always freshly transformed with the LacI expression plasmid, rather than freezing the stocks with the plasmid present to reduce the probability of suppressor mutations as much as possible.

3.2.2 The control strains

As controls for this experimental set up two bacterial strains were constructed. The first type included cells which had lac operator tandem repeats incorporated in their chromosome but did not express the LacI plasmid. The second type included cells which did not have lac operator tandem repeats incorporated in their chromosome but expressed the LacI plasmid. This means that the *lacO*-LacI replication fork was not expressed in either of the two control strains. Then *priB*, *priC* and the helicase activity of *priA* were deleted one by one from the strains.

The spot dilutions assay results enabled the direct comparison between the control strains will all the replication restart proteins present and when either of them was deleted. All the strains grew normally. This indicates that when the replication fork block is not present in the chromosome even after deleting either of the replication restart genes, the cells can grow normally and there are not any growth defects in any of the cells (Figure 24).

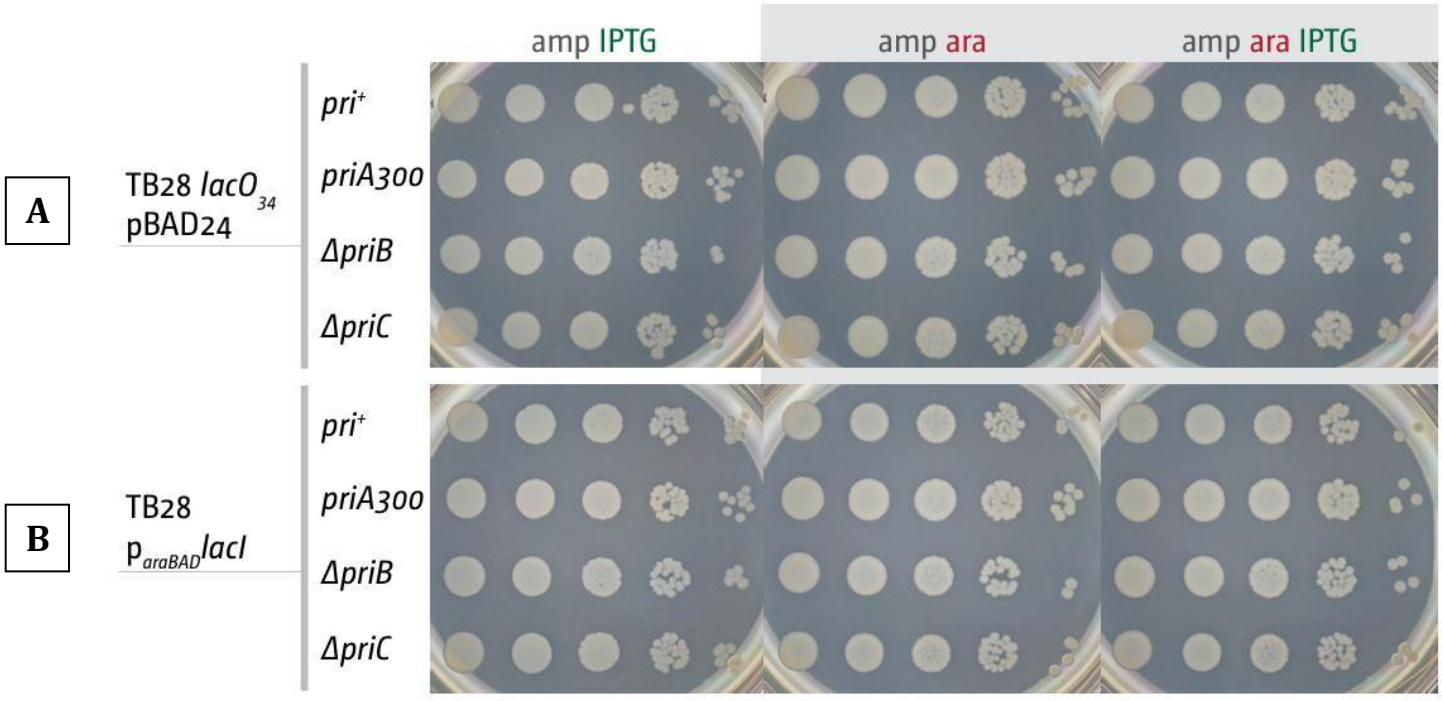

Figure 24. Cells grow normally when replication fork blocks are not present even in the absence of replication restart proteins. Representative spot dilution assays confirming that the *lacO* array and the p*ara*BAD-*lacI* plasmid do not inherently reduce fitness. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. (A) Strains containing the *lacO34* array but lacking the LacI expression plasmid. (B) Strains containing the *paraBAD-lacI* plasmid but lacking the *lacO* target array. The absence of growth defects across all replication restart mutant backgrounds (Δ*priA300*, Δ*priB*, Δ*priC*) confirms that the lethality observed in subsequent experiments is specifically due to the formation of the *lacO*- LacI nucleoprotein complex, rather than the genetic components themselves.

3.3 Cells lacking replication restart proteins struggle to grow in the presence of a replication fork block

3.3.1 Cells deficient of *priB* experience growth defects in the presence of the $lacO_{34}$- LacI replication fork block

As shown in Figure 25, wild type cells with the full contingent of DNA replication and repair proteins appeared to have little difficulties in the presence of any of the replication fork blocks $lacO_{22}$, $lacO_{34}$ and $lacO_{240}$ which is in line with the previously published experimental findings (Payne *et al.*, 2006b). *ΔpriB* strains were tested first in the presence of all replication fork blocks ($lacO_{22}$, $lacO_{34}$ and $lacO_{240}$) and in all the 3 types of media (arabinose, IPTG, arabinose + IPTG). For the cells lacking PriB the results showed considerably slower growth only in medium with arabinose only, which is the medium that induces the expression of LacI hence when the *lacO*-LacI is a replication fork block. The $lacO_{34}$ array showed a considerably stronger growth defect than the $lacO_{22}$ and $lacO_{240}$ arrays. Notably, this growth defect is not

present neither in the medium that contains IPTG only nor in the medium that contains the combination of arabinose+ IPTG (Figures 25 & 28).

Interestingly, the experiments also revealed the occurrence of many faster growing colonies, which are attributed to the occurrence of spontaneous suppressor mutations, providing thus an explanation regarding the reason why the attempts to establish this system in MG1655 strain, in which some LacI repressor is always present, caused all these difficulties (Figure 23).

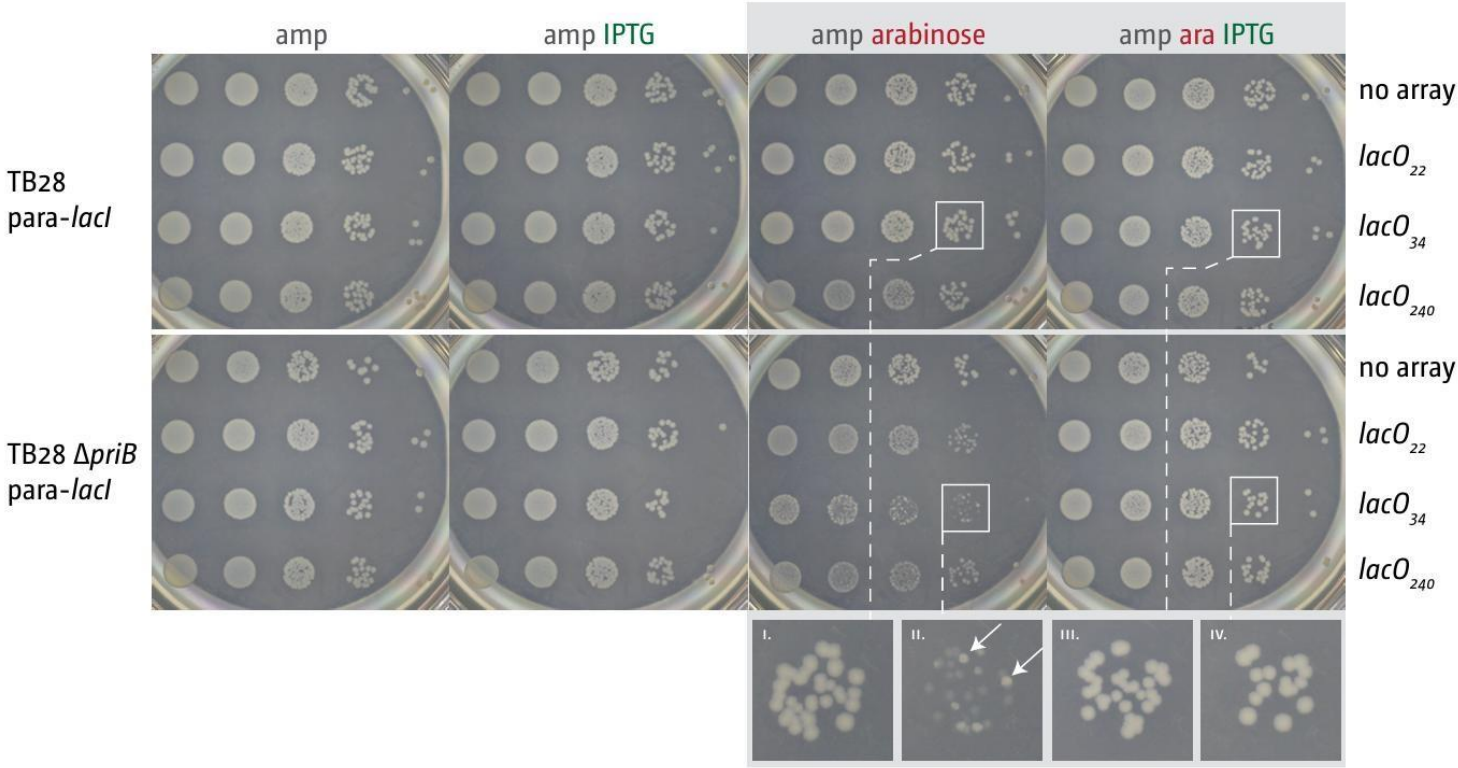

Figure 25. Cells deficient of *priB* experience growth defects in the presence *lacO*$_{34}$LacI. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. The top four plates correspond to the control strain TB28 para-*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The bottom four plates correspond to the strain control strain TB28 Δ*priB* para-*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The amp arabinose plate corresponds to the presence of the replication fork blocks in the system whereas the amp arabinose IPTG plate

corresponds to the absence of the replication fork blocks. The deletion of *priB* leads growths stress and to the emergence of spontaneous suppressor mutations when 34 tandem *lacO* repeat are added and expressed in the system.

3.3.2 Cells deficient of *priC* experience growth defects in the presence of the *lacO*$_{240}$- LacI replication fork block

Then, the effects of cells lacking PriC were tested. The results showed that in contrast to cells lacking PriB, the effects in cells lacking PriC are significantly more subtle. Specifically, no effect was observed at the *lacO*$_{22}$ array, and a similar effect was also observed at the *lacO*$_{34}$ array (Figures 26 & 28). However, interestingly enough a strong growth defect was observed in cells lacking PriC in the presence of the *lacO*$_{240}$ array. Also, in the presence of the *lacO*$_{240}$ array the emergence of fast-growing strains with suppressor mutations was observed (Figure 26). Notably, the emergence of fast-growing strains with suppressor mutations was not fully suppressed even in the presence of IPTG (Figure 26).

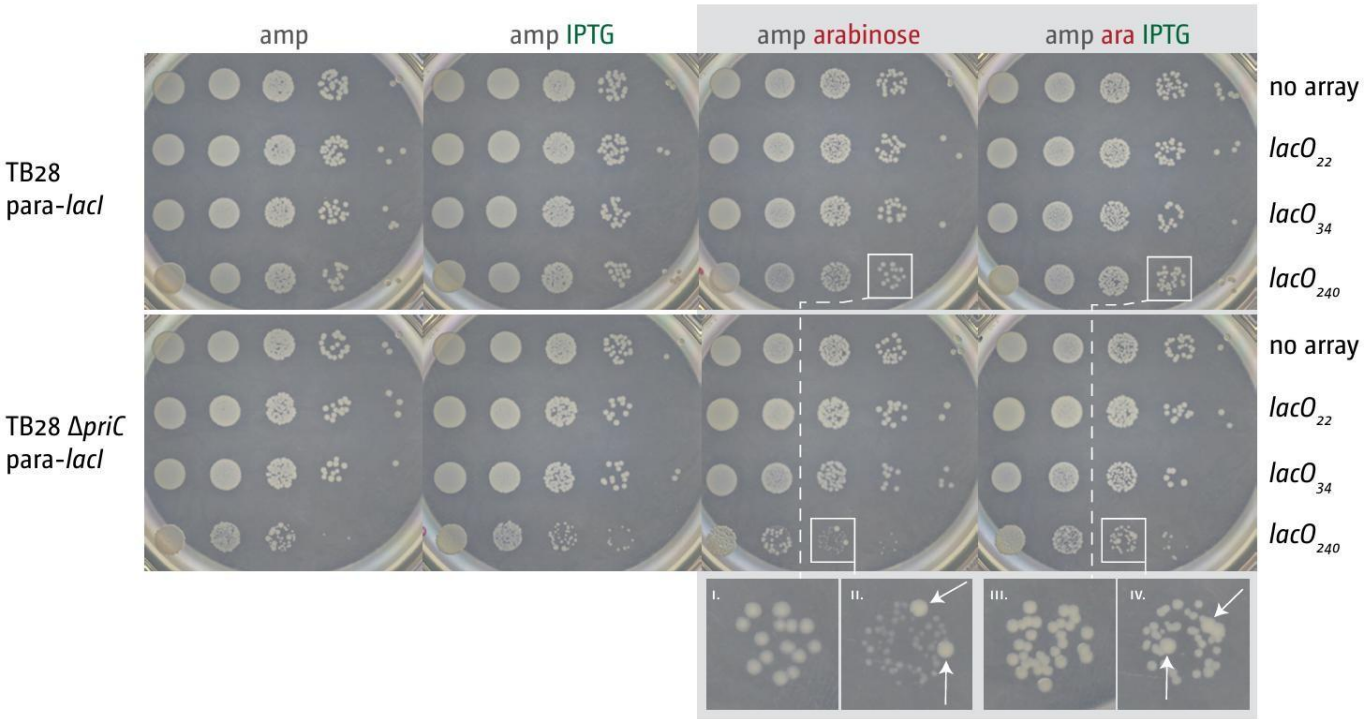

Figure 26. Cells deficient of *priC* experience growth defects in the presence *lacO*$_{240}$LacI. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. The top four plates correspond to the control strain TB28 para-*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The bottom four plates correspond to the strain control strain TB28 Δ*priC* para-*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The amp arabinose plate corresponds to the presence of the replication fork blocks in the system whereas the amp arabinose IPTG plate corresponds to the absence of the replication fork blocks. The deletion of *priC* leads growths stress and to the emergence of spontaneous suppressor mutations when 240 tandem *lacO* repeats are added and expressed in the system.

3.3.3 Cells deficient of the helicase activity of *priA* experience growth defects in the presence of the *lacO*$_{34}$- LacI replication fork block

PriA is an important replication restart protein that is present in 2 of the three replication restart pathways. Also, its helicase activity plays an important role in overcoming model replication blocks. Remarkably, it was suggested before that PriA helicase might also play a significant role as part of the PriA-PriC restart pathway (Sandler *et al.*, 2001; Windgassen *et al.*, 2018). Nevertheless, based on the results for the cells lacking *priC*, the initial hypothesis was that the effect that will be observed for the *priA300* cells should also be similarly minor.

Surprisingly though the opposite effect was observed: growth of *priA300* cells was reduced both in the presence of *lacO*$_{22}$-*LacI* and in the presence of *lacO*$_{34}$LacI. More specifically, the *lacO*$_{34}$ array had a stronger effect than the *lacO*$_{22}$ array, but importantly a distinct growth delay was also observed in the presence of the *lacO*$_{22}$ model block (Figures 27& 28).

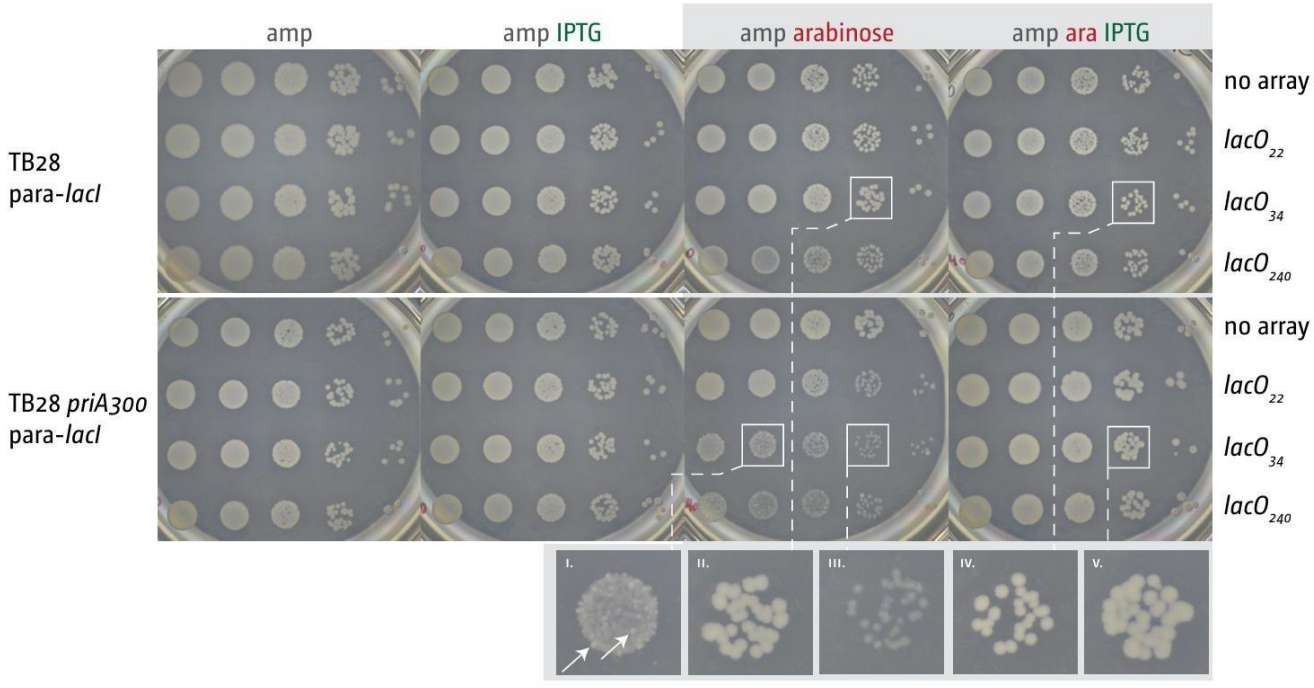

Figure 27. Cells deficient of the helicase activity of *priA* experience growth defects in the presence *lacO$_{34}$*-LacI. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. The top four plates correspond to the control strain TB28 para*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The bottom four plates correspond to the strain strain TB28 Δ*priA300* para-*lacI* with either no tandem *lacO* repeats, either 22 tandem *lacO* repeats, 34 tandem *lacO* repeats or 240 tandem *lacO* repeats. The amp arabinose plate corresponds to the presence of the replication fork blocks in the system whereas the amp arabinose IPTG plate corresponds to the absence of the replication fork blocks. The deletion of the helicase activity of *priA* leads growths stress and to the emergence of spontaneous suppressor mutations when 34 tandem *lacO* repeats are added and expressed in the system.

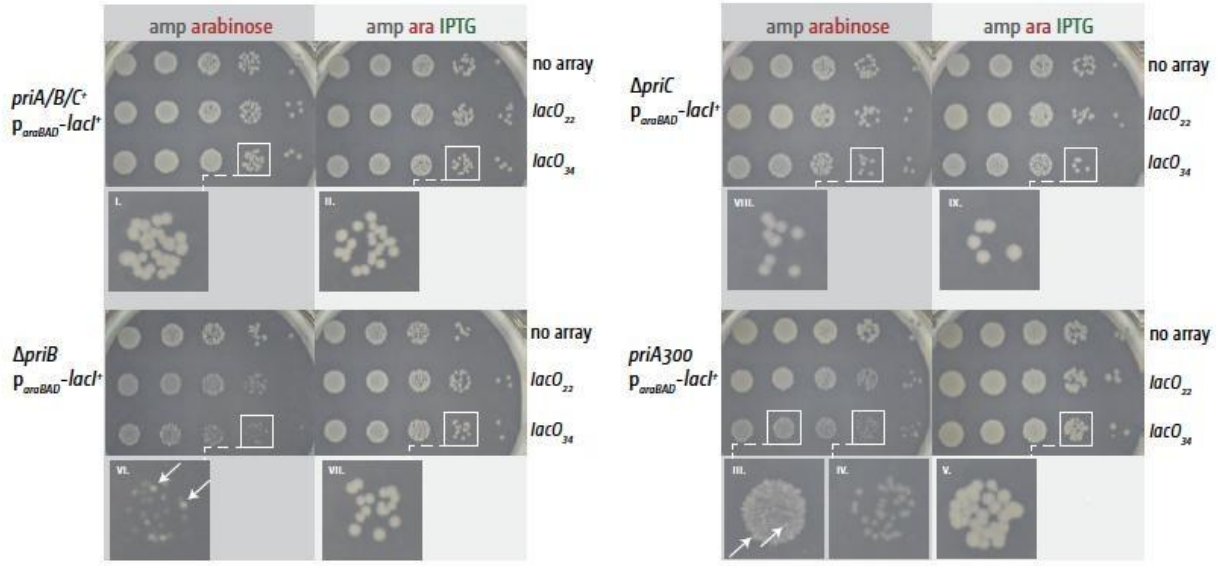

Figure 28. The expression of the *lacO34*–LacI leads to growth difficulties and the emergence of growth suppressor mutations in the cells lacking either *priB* or the helicase activity of *priA*. Cultures were serially diluted 10-fold (10^{-1} to 10^{-5}) and spotted from left to right onto LB agar. Comparison between the amp arabinose plate and the amp arabinose IPTG plate indicates that the *lacO$_{34}$*- LacI replication fork block leads to growth difficulties in cells deficient of either *priB* or the helicase activity of *priA*.

3.3.4 The absence of either replication restart protein leads to cellular division defects

This unexpected and interesting result lead to a next experiment with the aim to analyse the nature of this growth defect. Specifically, the growth of individual cells was observed and followed under conditions where the *lacO*$_{34}$-LacI block was either active or inactive (Figure 29A) The presence of arabinose only led to the presence of the *lacO*$_{34}$-LacI block, whereas the addition of IPTG led to the absence of the *lacO*$_{34}$-LacI block.

More precisely, the control strain *lacO*$_{34}$-LacI and the strain *lacO*$_{34}$-LacI *priA300* were first grown at 37 °C until mid-exponential phase. As soon as the optical density reached 0.2. 1ul of each strain was put in a microscope slide and incubated in the microscope at 37 °C for 2 hr (120 min). Each of the strains were grown under two conditions. First, they were grown in the presence of IPTG, hence there would not be any block for the replisome since the lac inhibitor will detach because of the presence of IPTG. In parallel, they were also grown in the presence of arabinose which means that the LacI would be expressed and bound to the lac operators, hence there would be a block for the replisome.

When the control wild type *lacO*$_{34}$-LacI cells were followed via time lapse microscopy for 2 hr, interestingly enough it was observed that although the vast majority of cells appeared to grow normally when the replication fork block was present, infrequent but notably regular formation of filamentous cells was also observed (Figure 29 B). The filamentation was transient (Figure 29A) and the filaments broke down into normal-sized cells that continued to grow normally. This effect was observed less in the presence of IPTG (Figure 29A).

Considering the observed growth defect of priA300 cells in the presence of the model replication fork blocks, the hypothesis was that these cells will filament more frequently than the wild type cells. The observations strongly confirmed the hypothesis because indeed when the model replication fork block was present, the *priA300* cells showed a more regular occurrence of filaments and this effect was mostly suppressed by the presence of IPTG (Figure 29B). Interestingly enough though it was observed that the filamentous cells after some cell divisions were breaking down to form normal-growing cells, as is the case for the wild type.

One inherent challenge about a time lapse microscopy experiment is that there is a restriction regarding the number of cells that can be observed at a specified time. So, to verify that filamented cells occur more frequently in *lacO*$_{34}$ *priA300* cells an advanced experimental set-up was devised. Specifically, the two strains were grown and once they reached early exponential phase they were split in two cultures, in one LacI expression was induced by the addition of arabinose, and in the other culture IPTG was also added preventing thus the LacI repressor from binding to the tandem *lacO* repeats to from a replication fork block. The cultures were then let to grow for 120 min and then images were taken, and the cell lengths were analysed. Although the number of cells that could be analysed was not standard, this experimental approach enabled the capture and analysis of around 500 cells at time point 0, and then up to 2000 cells at time point 120 min (Fig. 30). To ensure the accuracy and objectivity of this large-scale analysis, cell boundaries were identified using the automated ResNeXt-based segmentation pipeline (Section 2.7). This deep-learning approach allowed for the rapid processing of thousands of individual cells, utilizing the 0.065 μm per pixel calibration for precise length extraction. By employing this computational framework, we could reliably quantify the

shift in population distributions and identify the subtle increases in filamentation that might be missed by manual inspection.

Directly after splitting the cultures only small variations in the lengths of cells were observed (Fig. 30). Although at time point 0 min *priA300* cells turned out to be slightly shorter than wild type cells, this effect was rather mild. After 120 min with the expression of *lacI*, a reduction of the cell size in wild type cells was observed, which is also attributed to the fact that at time point 120 min the cells had reached early stationary phase, where the size of cells is reduced (Akerlund, Nordström and Bernander, 1995). Also, in wild type $lacO_{34}$ cells the comparison between the cells that were treated with arabinose only and the cells that were with IPTG only, showed that there was a difference regarding their cell length. This means that the presence of arabinose, hence the presence of the *lacO- LacI* replication fork, resulted in an increase of the average cell lengths, and interestingly the presence of IPTG reversed this effect.

Regarding *priA300* cells importantly at time point 0 min, before the expression of *lacI*, their average length was reduced compared to the wild type cells. Then, after 120 min of growth with arabinose, hence *lacI* expression, the average length was significantly increased. The addition of IPTG resulted in a reduction in the average cell size of *priA300* cells (Figures 29 & 30).

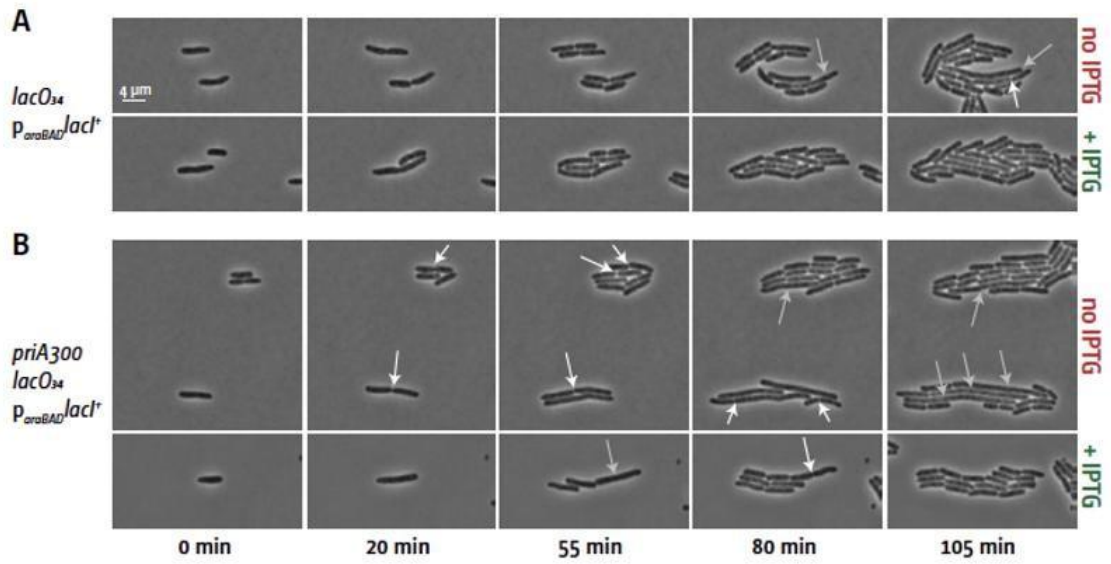

Figure 29. Time lapse- microscopy experiment of *lacO34* p*araBAD*lacI+ and *priA300 lacO34* p*araBAD*lacI+. The two strains were first grown in a shaking incubator until mid-exponential phase. Then each of the strains was split in two sub-cultures, one of them grew in the presence of arabinose and the other in the presence of IPTG. The 4 cultures were grown and their growth was followed under the microscope for 120 min. The presence of arabinose causes the first filamentous cells which are indicated with the white arrows from 55 min in the strain *priA300 lacO34* p$_{araBAD}$*lacI* and the filamentation effect continues and exacerbates until the end of the experiment. The addition of IPTG considerably reduces but does not completely eliminate the emergence of filamentous cells in the strain *priA300 lacO34* p$_{araBAD}$*lacI.*. The addition of arabinose also leads to some filamentation in the strain *lacO34* p$_{araBAD}$*lacI* but it is not so frequent and occurs in later cell divisions. The addition of IPTG in the strain *lacO34* p$_{araBAD}$*lacI* completely eliminates the emergence of filamentous cells.

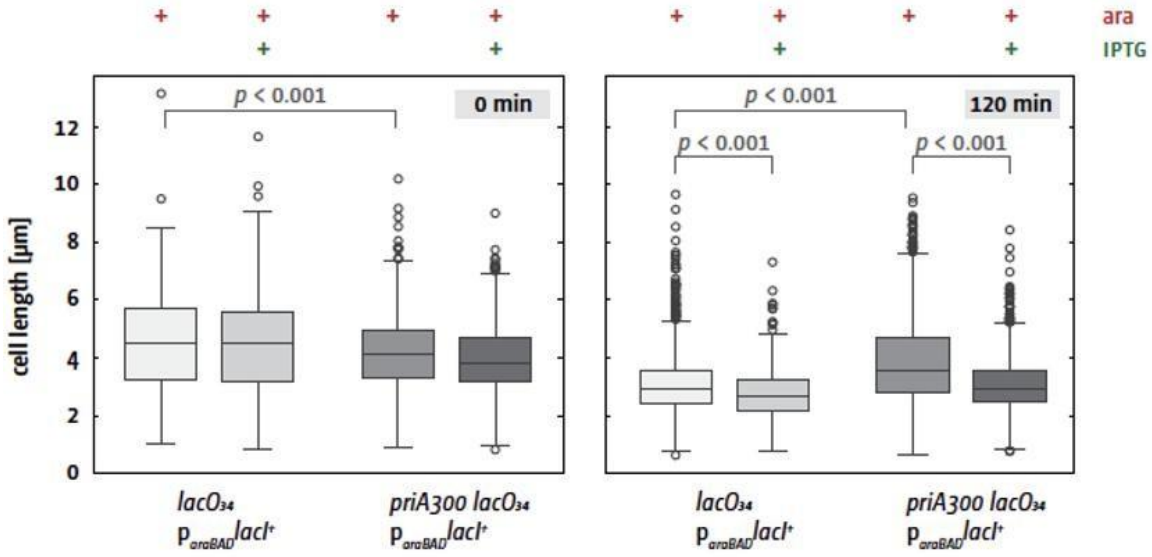

Figure 30. PriA helicase deficiency exacerbates cell elongation upon p*araBAD*lacI⁺ induction. Boxplots indicate the 25th, 50th (median), and 75th percentiles; whiskers extend to 1.5x the interquartile range (IQR), with individual dots representing extreme outliers (single filamented cells). Left Panels (0 min): At t = 0, both strains exhibit a tight, synchronized distribution of short rods (approx. 2.0–2.5 µm) regardless of induction state. Right Panels (120 min): Induction of the *lacO34*–LacI block (+ara) results in significant cell elongation in both strains, but the effect is drastically exacerbated in the *priA300* background. The *priA300* population shows a massive upward shift in median length to ~8.5 µm (compared to ~3.8 µm for WT), indicating a severe failure in cell division control. For the *priA300* mutant at 120 min, the addition of IPTG (ara +IPTG) results in a statistically significant reduction in filamentation, with the median length returning to ~2.9 µm. This confirms that the elongation is a direct physiological response to the nucleoprotein block rather than an inherent growth defect of the *priA300* mutation. Brackets and *p*-values (***p < 0.001) denote differences calculated by Welch's t-test, chosen specifically to account for the heteroscedasticity (unequal variance) caused by filamented cells in the mutant population. The R script used for data processing, distribution analysis, and Welch's t-test calculations is provided in **Appendix A**.

In addition to the defects observed in PriA-deficient cells, the quantitative analysis of Δ*priB* and Δ*priC* mutants provided further clarity on pathway specialisation. As shown in Figure 31, Δ*priB* cells exhibited a significant increase in filamentation and a wider distribution of cell lengths when challenged with *lacO34* blocks, mirroring the phenotypes seen in *priA300* backgrounds. In contrast, ΔpriC cells displayed a morphology much closer to the wild-type distribution under the same conditions. These results, quantified through the high-throughput Python pipeline and validated by Welch's t-tests, demonstrate that the PriA-PriB-DnaT pathway is the primary responder to these nucleoprotein obstacles. The inability of PriC to fully compensate for the loss of PriB in these instances suggests that the physical properties of the *lacO34* block specifically require the recruitment of the PriB-containing restart complex to ensure timely cell division.

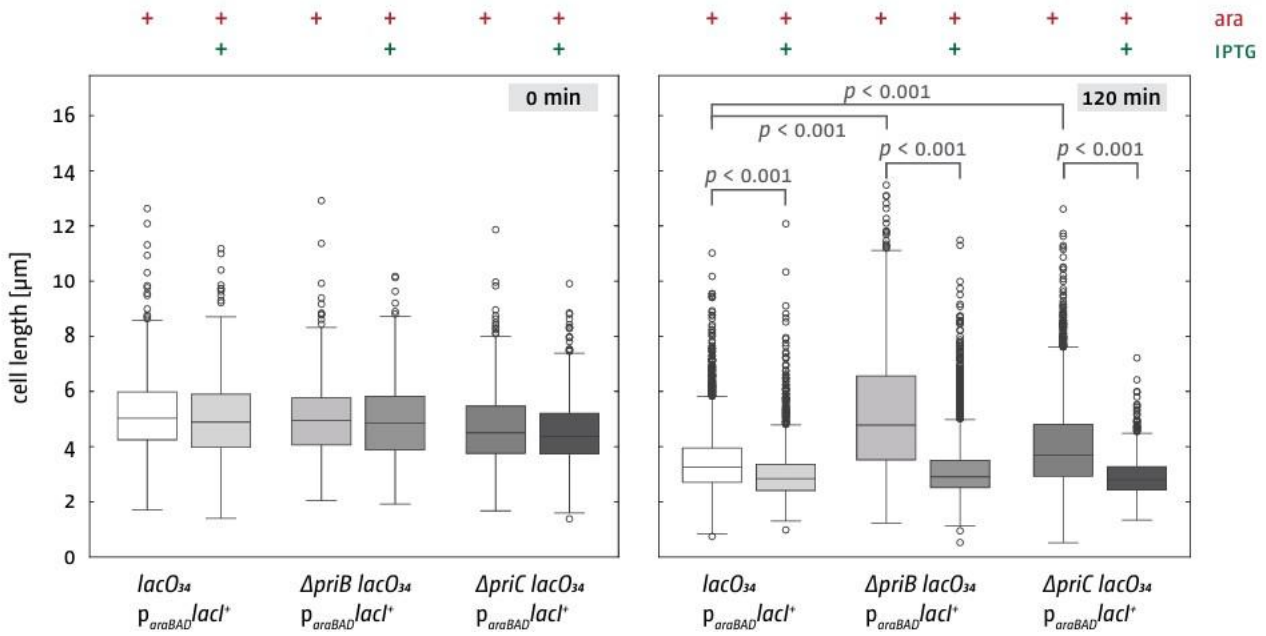

Figure 31. PriB and PriC deficiency exacerbates cell elongation upon p*araBAD*lacI*+ induction. Boxplots represent the median (horizontal line), 25th/75th percentiles (box), and 1.5x IQR (whiskers); individual outliers are plotted as points. Left Panels (0 min): At t=

0min, the wild-type (WT), Δ*priB*, and Δ*priC* populations are morphologically consistent, with median lengths of approximately 2.2 μm regardless of the induction state. Right Panels (120 min): Induction of the *lacO34*- LacI block triggers a significant increase in cell length across all strains. The Δ*priB* background exhibits the most severe phenotype, with the median length shifting to ~8 μm. In comparison, the Δ*priC* mutant displays a significant, yet more moderate, elongation with a median of ~5.5 μm. This distinction suggests that *priB* plays a more critical role than *priC* in maintaining division control during fork blockage. The addition of IPTG results in a statistically significant reduction in filamentation. The Δ*priB* median contracts to ~3 μm, while the Δ*priC* median returns to ~2.8 μm. This near-complete recovery confirms that the morphological stress is a direct consequence of the nucleoprotein impediment. Comparisons between induced (ara) and released (ara+ IPTG) conditions were evaluated using independent two-sample Welch's t-tests (***$p < 0.001$). This test was selected to maintain statistical power despite the high heteroscedasticity (unequal variance) inherent in filamented populations. Full computational details and R scripts are provided in Appendix A.

3.4 Discussion

The results and analyses presented in this chapter successfully address the questions outlined in the aims and provide novel insights regarding how replication fork blocks affect cellular physiology and viability and provide a hierarchy about the activity of the three replication restart pathways to preserve genome stability.

More specifically, by using a novel replication fork block model of various lengths in *E. coli* it was shown that during DNA replication cells deficient of either replication restart protein experience considerable growth difficulties and hence cell division defects. To ensure the statistical robustness of these findings, all morphological measurements and filamentation analyses were processed through the ResNeXt-based computational pipeline (Section 2.7). This allowed for the objective quantification of cell lengths across thousands of

individual events (n > 2000), providing a high-resolution view of the phenotypic shift from normal division to elongated states.

Interestingly, the replication forks arrest upon meeting the blocks, and even worse some of the forks can be disassembled. However, the results indicate that replication can continue after a short period, which reveals that replisomes must be re-assembled at the site of replication fork arrest. In the model organism *E. coli* this is done by specific proteins which are known as replication restart proteins and are the following: PriA, PriB, PriC and DnaT.

Although it has already been established that these replication restart proteins are involved in the following restart pathways: PriA-PriB-DnaT, PriA-PriC-DnaT and PriC-DnaT it had not been established yet which of these pathways is more active in the resolution of replication fork blocks (Sandler and Marians, 2000; Heller and Marians, 2005; Michel and Sandler, 2017b). So, by utilising a novel replication fork bl ock model we established that replication restart at nucleoprotein obstacles predominantly utilises the PriA-PriB-DnaT restart pathway (Figures 25, 27 & 28). This is attributed to the fact that cells deficient of either PriB or the helicase activity of PriA suffered the most in the presence of the replication fork block models $lacO_{22}$ and $lacO_{34}$. The significance of these differences was mathematically validated using independent two-sample Welch's t-tests, which confirmed that the increased cell lengths in PriA and PriB mutants were statistically significant ($p < 0.05$) compared to the WT control.

Also remarkably, after the significant difficulties we faced with the emergence of a plethora of suppressor growth mutations when utilising MG1655 strain (Figure 23), the reconstruction of these strains in TB28 strain also revealed that the cells deficient of either replication restart protein accumulated suppressor growth mutations but not to the extent that we observed with

MG1655 (Figures 25, 26, 27 & 28). The emergence of the suppressor growth mutations resulted in additional difficulties for us because all the strains had to be reconstructed from the beginning in TB28 and even after that they had to be freshly transformed with the LacI expressing plasmid to carry out the experiments. Otherwise, if they were frozen then again suppressor mutations appeared. However, all these technical challenges and additional experimental steps ultimately provided valuable insights into the critical and indispensable roles of replication restart proteins. Remarkably, if the absence of a single gene leads to such a rapid emergence of suppressor mutations, then this suggests that this gene expresses a vital protein which is essential for normal cellular growth (Poon, Davis and Chao, 2005; Hogan and Cardona, 2022). So, the strong selective pressure for the accumulation of compensatory mutations indicates that the loss of this protein severely damages DNA replication, thereby threatening cellular viability. As a result, the cells are forced to rapidly adapt in the absence of this essential protein to survive.

PriA is a unique protein that has both replication restart activity and a helicase activity which is absent from the other replication restart proteins (Gabbai and Marians, 2010; Leroux, Jani and Sandler, 2017). It has been demonstrated that the restart activity of PriA is important to re-recruit the replication machinery to the replication fork block location (Marians, 2000). However, it is important to underline that the exact role of its helicase activity had not been yet identified. The results from this chapter demonstrate that just the deficiency of the helicase activity of PriA is enough to cause the cells to undergo significant growth defects, which is evident by the presence of filamentous cells, which in turn affect their division when replication forks are blocked at an obstacle (Bernander and Nordström, 1990; Courcelle *et al.*, 2001) (Figures 27, 29 & 30). Interestingly enough though the results show that this effect is temporary (Figure 29). This is because the filamented cells that

initially struggle to replicate their DNA and hence struggle to undergo cell division, after a period of time are able to undergo cell division normally. This temporal recovery was quantified using the automated segmentation pipeline, which tracked the return of the population to a standard length distribution after the initial stress period. This means that they are able to undergo cell division and hence they have replicated their DNA. Also, our results show that the average viability of the cells is not significantly affected by the presence of the replication fork blocks. All these results indicate that the deletion of the helicase activity of PriA causes significant stress initially but in the end the cells are able to terminate DNA replication and divide via employing other back-up pathways.

Importantly, this research is the first to highlight the important role of the helicase activity of PriA helicase in the efficient handling of stalled replication forks as a result of nucleoprotein obstacles. Thus, our results provide the foundation for future interesting work regarding the mechanistic action of PriA helicase. Specifically, it would be very interesting to find out whether there might be a synergy as well between PriA helicase and other proteins or pathways that act on the stalled replication fork so that the proteins which belong on the recombination pathways can then act to rectify the growth defects.

Also, regarding the $lacO_{240}$ replication fork block model, it is important to highlight that the $lacO_{240}$ integration system has not been extensively utilised before as the $lacO_{22}$ and $lacO_{34}$ systems. This means that when the 240 tandem *lacO* sequences were integrated into the chromosome the results were very interesting because they show that the cells that are deficient of PriC, and not PriB or the helicase activity of PriA, will suffer the most in the presence of this specific replication fork block. This means that $lacO_{240}$ has a considerably

larger size compared to *lacO*$_{22}$ and *lacO*$_{34}$ and this observed effect was specific to the absence of PriC. All these indicate that an important connection might exist between the size of the model block, and which specific replication restart protein is the most important, hence which of the three replication restart pathways is more actively involved. Our finding is consistent with previous research in the field (Heller and Marians, 2006; Wessel *et al.*, 2013; Windgassen *et al.*, 2018). However, it is important to underline that the *lacO*$_{240}$ array that was present in our lab and utilised is integrated on the left replichore and not on the right replichore as is the case with both *lacO*$_{22}$ and *lacO*$_{34}$ arrays (Figure 21).

A very interesting future experiment thus could involve the reciprocal chromosomal relocation of the replication fork block models. This means that it would be interesting to construct new strains where the *lacO*$_{240}$ array will be integrated on the right replichore and the *lacO*$_{22}$ and *lacO*$_{34}$ arrays will be integrated on the left replichore and then evaluate how the cells deficient of either replication restart protein will behave. More specifically, it is observed that in the presence of the right *lacO*$_{240}$ array the cells deficient of PriC again suffer the most and then in the presence of the left *lacO*$_{22}$ and *lacO*$_{34}$ arrays the cells deficient of either PriB or the helicase activity of PriA again suffer the most then this would indicate that the decisive factor of which replication restart pathway will act to resolve the replication fork block is attributed to the barrier size only and not to the chromosomal location of the barriers. Finally, utilising replication fork blocks of intermediate sizes, for example 60× and 120× *lacO* repeats on the left and right replichore, might show a gradual shift from the PriA–PriB replication restart pathway action to the PriC replication restart pathway action and even provide the replication fork block size threshold at which the cell switches between the replication restart pathways. The high-throughput analytical framework established in this

thesis, combining deep-learning segmentation with robust statistical testing (Appendix A), provides the necessary toolset for these future investigations into replication restart thresholds.

Chapter 4

Mechanisms of Replication–Transcription Conflicts and their Resolution

4.1 Introduction and research aims

As already established, conflicts between DNA and RNA are unavoidable and frequent, threatening thus the genomic stability of both the eukaryotic and the prokaryotic organisms (Bayona-Feliu and Aguilera, 2025). Both DNA replication and RNA transcription are essential molecular processes which need to be carried out accurately for the production of functional proteins to ensure thus the survival of an organism (Drummond and Wilke, 2009).

However, both DNA replication and RNA transcription operate on exactly the same deoxynucleotide template (Gaillard and Aguilera, 2016; Wu, Hickson and Liu, 2020). In addition, these two processes are carried out with different speeds. There are two types of replication-transcription conflicts, co-directional and head-on with the head-on replication-transcription conflicts being more deleterious for the cells (Srivatsan *et al.*, 2010) (Figure 32).

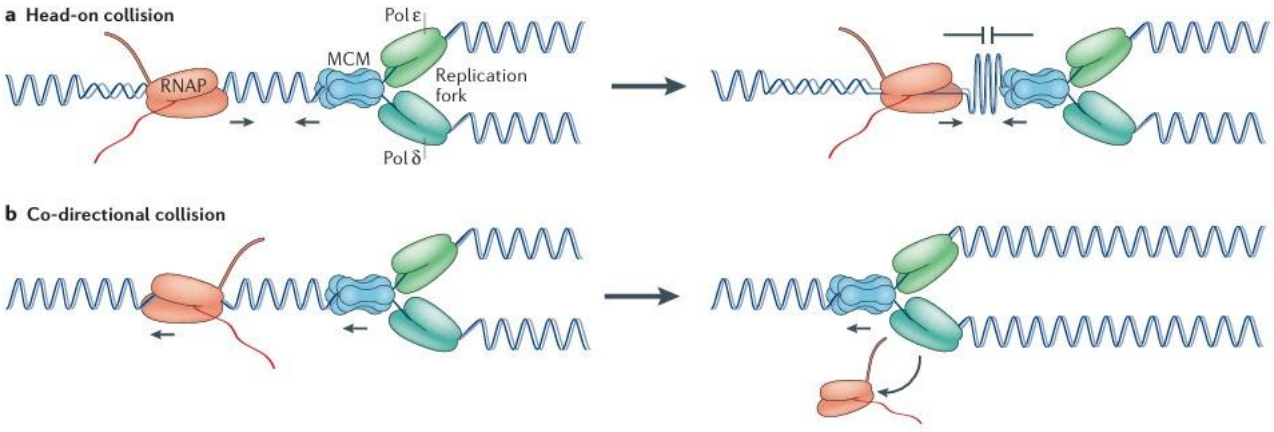

Figure 32. Effects of head-on and co-directional transcription–replication collisions in eukaryotic cells. a.) When RNA polymerase (illustrated with orange and abbreviated as RNAP) operates in an opposite direction from the eukaryotic mini chromosome maintenance complex (illustrated with blue and abbreviated as MCM) and the replication fork (illustrated with green), they collide with one another head-on. A head-on replication-transcription collision leads to stalling of the replication fork and RNA polymerase and the DNA section between them shrinks and has increased probability to

accumulate mutations. b.) When RNA polymerase operates in the same direction as the mini chromosome maintenance complex and the replication fork, they collide with one another co-directionally. A co-directional replication-transcription collision can be resolved by the removal of the RNA polymerase from the DNA template so that DNA replication can continue. Figure from (Gaillard and Aguilera, 2016).

The results from the first results chapter yielded important knowledge regarding the role of the replication restart proteins for the resolution of replication fork blocks. Specifically, we learned that replication restart proteins play an important role in the resolution of replication fork blocks and the absence of even one of them causes significant cellular growth problems.

Thus, we were interested to investigate whether the replication restart proteins also play an important role for the resolution of clashes between DNA replication and transcription and especially the head-on replication-transcription conflicts which pose more dangers for the organisms. Also, we wanted to find out which of the three replication restart pathways is more actively involved in the resolution of replication- transcription conflicts.

4.2 The experimental set-up

To study the effects of replication-transcription conflicts we incorporated a second origin of replication into the chromosome of *E. coli*. This ectopic origin of replication is known as *oriZ* but essentially has the same features as *oriC*. The insertion of this ectopic origin of replication will result in the occurrence of a head-on replication-transcription conflict. This is because, as is the case with *oriC*, two replication forks will be assembled at *oriZ*. The right replication fork initiating from *oriZ* will proceed in the normal direction of

replication. However, as is also illustrated in figure 33B, the left replication fork initiating from ori*Z* will traverse in the opposite orientation hence there will be a head-on conflict with the highly transcribed *rnh* operon (Wang *et al.*, 2011). Having a strain like this will allow us to study replication-transcription conflicts, and the Rudolph lab and colleagues have successfully utilised it in a number of publications (Ivanova *et al.*, 2015a; Dimude *et al.*, 2018; Dimude, Midgley-Smith and Rudolph, 2018; Hawkins *et al.*, 2019).

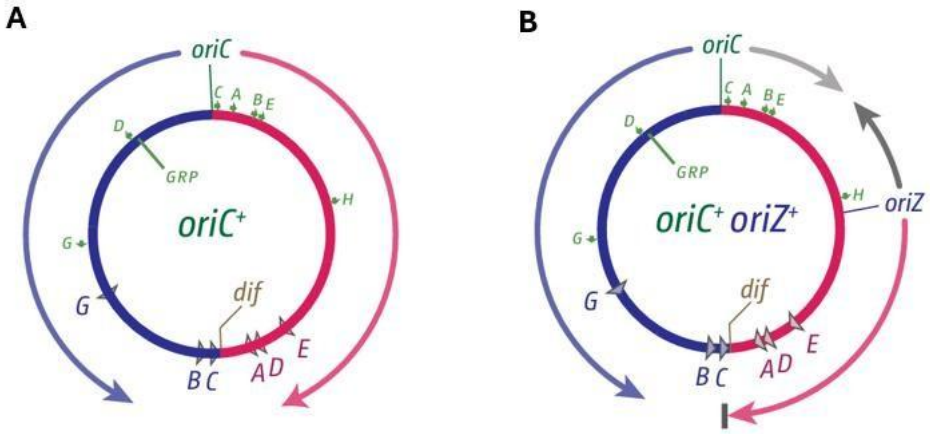

Figure 33. Differences between normal *E. coli* cells with the single origin of replication ori*C* and *E. coli* cells the additional origin of replication ori*Z* A.) Normally two replication forks will initiate from ori*C* and will replicate the DNA in opposite directions until they reach the termination zone. B.) In *E. coli* cells with the additional origin of replication ori*Z* two forks will initiate from ori*Z.* The right replication fork initiating from ori*Z,* depicted with pink arrow, will move in the normal direction. However, the left replication fork, depicted with dark grey, will move in the opposite direction and this result in a head-on collision between this replication fork and the *rnh* operon, depicted with the letter H in green.

Before I started my PhD, the Rudolph lab confirmed that the ori*Z* insertion indeed leads to replication–transcription conflicts by performing marker frequency analysis (Figure 34). These preliminary data showed that replication forks initiated from ori*Z* stall in a head-on orientation at the *rnh* operon. Because these results validated the ori*Z* system as a robust model, I adopted this strain background for all subsequent experiments described in this chapter.

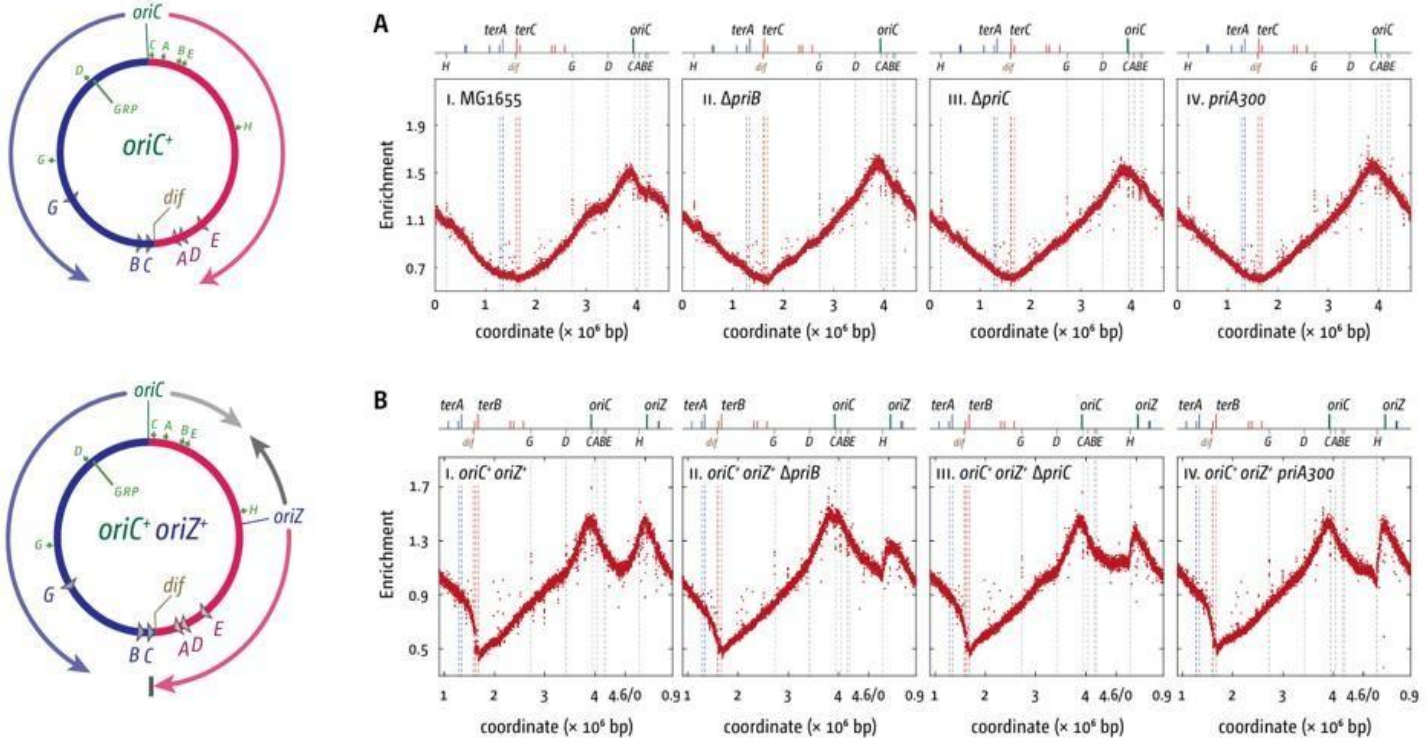

Figure 34. Marker frequency analysis of *E. coli* strains with single or dual replication origins in replication restart mutant backgrounds. *(A)* Chromosome replication profiles of ori*C* strains (single origin). Marker frequency plots were generated from deep sequencing data of MG1655 (wild type), Δ*priB*, Δ*priC*, and *priA300* strains. All profiles display a single replication peak at ori*C* with a gradual decrease toward the terminus (dif region). *(B)* Chromosome replication profiles of ori*C*$^{+}$ ori*Z*$^{+}$ strains (dual origin). Two distinct peaks at ori*C* and ori*Z* indicate efficient initiation from both origins in wild type and mutant backgrounds. Deletion of *priB* had the biggest effect on replication profiles, followed by the *priC* and priA300 mutants which also showed a reduction in peak height at both origins, suggesting impaired replication efficiency. Chromosome cartoons (left) depict replication origins and bidirectional fork movement.

4.3 Results

4.3.1 The absence of either replication restart protein leads to normal DNA replication and cell division in the absence of replication-transcription conflicts

We first started with the wild-type *E. coli* cells which have the single origin of replication ori*C*. Then *priB*, *priC* and the helicase activity of *priA* were deleted one by one from the cells. Then, to visualise the DNA replication dynamics under the microscope the plasmid pTK135 was transformed into the cells. The expression of pTK135 leads to the production of Cas1-Cas2 foci. Although the Cas1-Cas2 complex is used in the bacterial CRISPR system, it has been established as an important and accurate biomarker that is indicative of DNA replication. This means that the presence of Cas1-Cas2 foci under the microscope indicates that DNA replication is taking place, hence the number of Cas1 foci in the cells under the microscope is a good indication of the DNA replication activity (Killelea *et al.*, 2023). Cas1-Cas2 foci are as a relatively quick and accurate way to study replication dynamics in a variety of strains including strains lacking replication restart proteins.

First, the wild type MG1655 strain with the single origin ori*C* and with all the replication restart proteins was visualised under the microscope. Then, cells deficient of either *priB*, *priC* or the helicase activity of *priA* were also visualised under the microscope. The results show that under the microscope the cells with the single origin of replication ori*C* and with all the replication restart proteins present looked very similar to the cells were either *priB*, *priC* or the helicase activity of *priA* were not present. More specifically, all the cells had a similar length of around 4μm and a similar number of foci which was around 1-2 foci per cell (Figures 35 & 36).

This indicates that in the single origin *E. coli* cells the deletion of either of the restart proteins does not cause any genomic growth defects and the processes of DNA replication and cellular division do not get affected either.

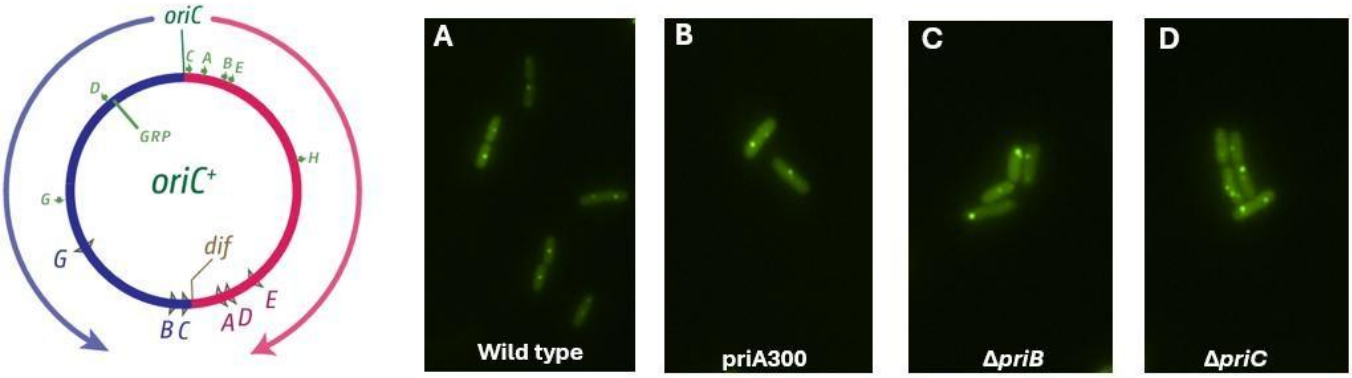

Figure 35. Single origin strains under the microscope have a healthy phenotype. Wild type *MG1655* cells depicted in A, *priA300* cells depicted in B, *ΔpriB* cells depicted in C and *ΔpriC* cells depicted in D all have a length of around and each cell has normal Cas1-Cas2 foci counts of 1-2.

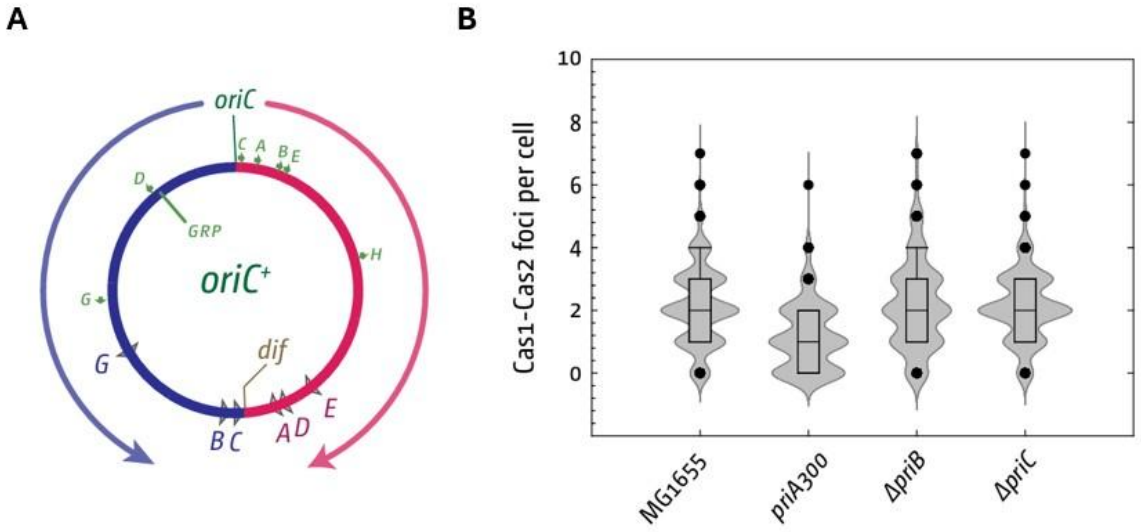

Figure 36. Cas1-Cas2 foci counts of single origin strains. Wild type *MG1655* cells, *priA300* cells, *ΔpriB* cells and *ΔpriC* cells on average all have normal Cas1-Cas2 foci counts of 1-2.

4.3.2 In the absence of *priB* or the helicase activity of *priA* the cells exhibit replication and cell division stress during replication-transcription conflicts

After that, the ectopic origin of replication ori*Z* was introduced into the wild type cells and again *priB*, *priC* and the helicase activity of *priA* were deleted one by one from the cells. Cells with the two origins of replication, ori*C* and ori*Z* and with all the replication restart proteins looked similar to the normal cells that had only the single origin of replication ori*C*. More specifically, the cells that contained both the normal origin of replication ori*C* and the ectopic origin of replication ori*Z* were around 4μm in length which means that they were healthy. Also, the average number of the Cas1-Cas2 foci in each cell was 1 or 2 which was also the case when the cells had only the single origin of replication ori*C*. Interestingly, results showed that the phenotypic characteristics in the double origin cells when *priC* was deleted were also similar to the wild type single origin and double origin cells. To formally quantify these observations, a non-parametric statistical analysis was performed using R. As shown in the violin plot in Figure 38 and table 4.1, which represents the full distribution of over 2,000 analyzed cells, the strain oriC^+ oriZ^+ Δ*priC* showed no significant difference in foci accumulation compared to the control (1.56 ± 0.06; p = 0.99 by Wilcoxon rank-sum test).

However, when the double origin cells which were deficient of either *priB* or the helicase activity of *priA* were visualised surprisingly we observed two types of cells. Specifically, although there were some cells which had normal length of around 4 μm and a normal number of Cas1- Cas2 foci, 1-2 per cell, there were also many cells which were considerably filamentous compared to the normal cells (Figure 37). The filamentation in bacterial cells has been established as a feature of cellular stress. Also remarkably, in the filamentous

cells the number of Cas1-Cas2 foci was significantly higher than what we had observed so far. Interestingly, cells deficient of the helicase activity of *priA* were more filamentous than the cells deficient of *priB* with some of the filamentous cells exceeding 20 foci (Figures 37 &38). Since, Cas1-Cas2 foci are a biomarker that is associated with DNA replication, the significant increase of Cas1-Cas2 foci indicates that there is abnormal DNA replication. The detailed statistical parameters for these populations, including precise mean values, standard error, and total cell counts (N), are summarized in Table 4.1. Statistical analysis confirmed a highly significant increase in foci for both the ori*C*⁺ ori*Z*⁺ *priA300* and the ori*C*⁺ ori*Z*⁺ Δ*priB* strain. These quantitative results suggest that *priB* and the helicase activity of *priA* are essential for resolving replication-transcription conflicts arising from the dual-origin configuration, whereas *priC* appears to be dispensable in this specific pathway.

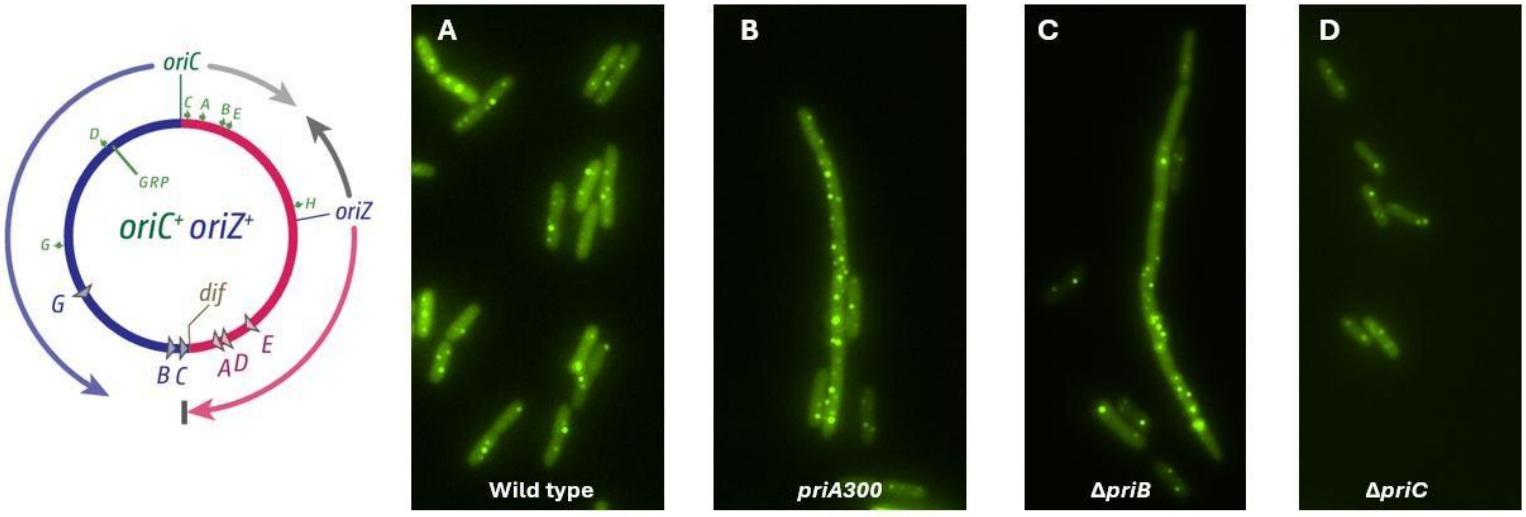

Figure 37. Double origin strains ori*C*⁺ ori*Z*⁺ under the microscope. While the control ori*C*⁺ ori*Z*⁺ and the ori*C*⁺ ori*Z*⁺ Δ*priC* strain exhibit uniform morphology, ori*C*⁺ ori*Z*⁺ *priA300* and ori*C*⁺ ori*Z*⁺ Δ*priB* strains display phenotypic heterogeneity, characterised by normal looking cells but also severely filamentous cells with stochastic accumulation of multiple Cas1- Cas2 foci.

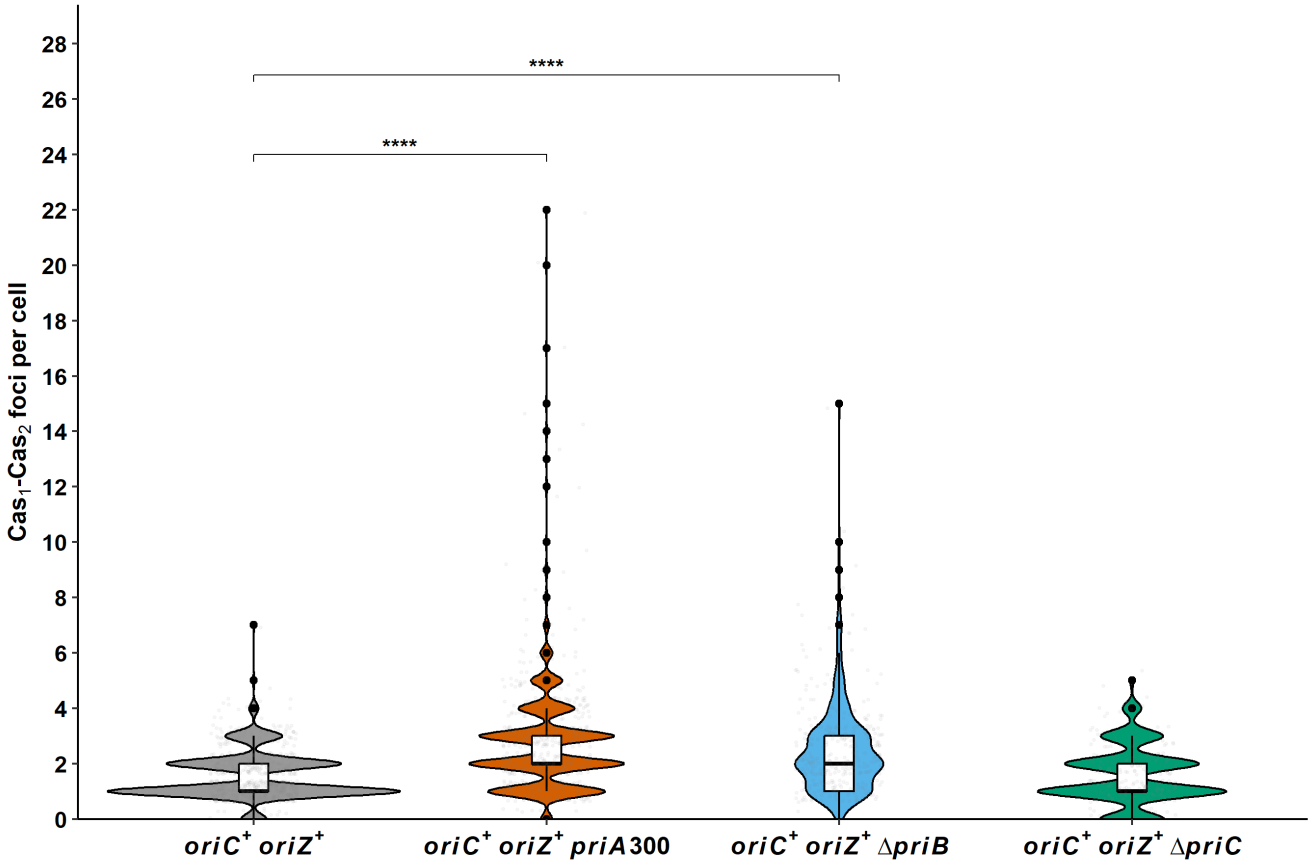

Figure 38. Quantitative distribution of Cas1-Cas2 foci counts in double origin strains ori*C*⁺ ori*Z*⁺. Violin plots represent data density, with internal boxplots indicating the median and interquartile range. Grey jittered points represent individual cellular measurements (N > 2,000 total cells). Statistical significance was determined by a two-sided Wilcoxon rank-sum test comparing each mutant strain to the ori*C*⁺ ori*Z*⁺ control. The four asterisks (****) indicate a highly significant difference ($p < 0.0001$), while 'ns' indicates no statistical significance ($p > 0.05$). Analysis and visualization were performed using R (v4.x).

Table 4.1. Statistical summary of Cas1-Cas2 foci accumulation in double origin strains ori*C*⁺ ori*Z*⁺.

Genotype	N (cells)	Mean ± SEM	Median	p-value (vs. Control)
ori*C*⁺ ori*Z*⁺	729	1.54 ± 0.03	1	—
ori*C*⁺ ori*Z*⁺ *priA300*	620	2.79 ± 0.09	2	**< 0.0001**
ori*C*⁺ ori*Z*⁺ Δ*priB*	340	2.48 ± 0.09	2	**< 0.0001**
ori*C*⁺ ori*Z*⁺ Δ*priC*	302	1.56 ± 0.06	1	**0.99 (ns)**

4.3.3 The presence of *rpoB*35* results in the resolution of replication-transcription conflicts

The assumption was that indeed in the ori*C*⁺ ori*Z*⁺ cells replication-transcription conflicts do take place and that the helicase activity of *priA* is so important that when the cells are deficient of it, they experience increased cellular stress and genomic instability. To verify this hypothesis, the next step was to make the RNA polymerase less stable. This would allow replication forks to move more easily through regions where replication-transcription conflicts occur, because RNA polymerase would be less prone to forming persistent paused or backtracked states (Dutta *et al.*, 2011; Dimude *et al.*, 2015).

To achieve this, the point mutation *rpoB**35 was introduced in oriC⁺ oriZ⁺ *priA300* cells (McGlynn, Savery and Dillingham, 2012). Normally *rpoB* is the gene that codes for the production of the beta subunit of RNA polymerase (Figure 39). However, *rpoB*35* allele leads to the amino acid change from Histidine (H) to Glutamine (Q) at amino acid residue 1244 on β subunit which results in a destabilised form of the beta subunit of RNA polymerase,

ultimately leading to an RNA polymerase that is less stably bound to DNA (Trautinger and Lloyd, 2002). As a control of this experiments the point mutation *rpoB*35* was also introduced in the *oriC*⁺ *oriZ*⁺ cells.

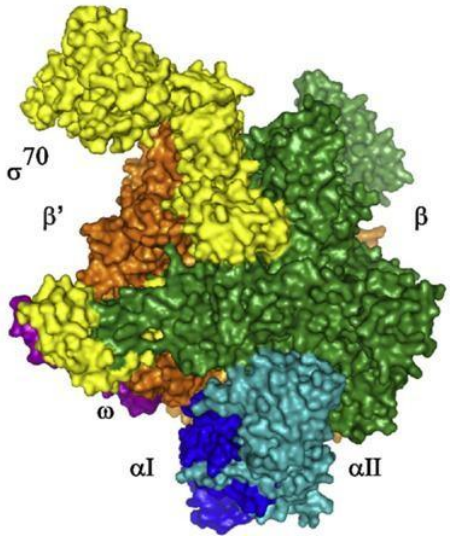

Figure 39. 3D structural view of the subunits in the *E. coli* RNA polymerase holoenzyme. The *E. coli* RNA polymerase holoenzyme is shown with its subunits: β (green), β′ (orange), αI (dark blue), αII (light blue), ω (purple), and the σ^{70} initiation factor (yellow). Figure adapted from (Mazumder and Kapanidis, 2019).

Indeed, fascinatingly enough when the ori*C*⁺ ori*Z*⁺ *rpoB**35 Δ*priA300* cells were visualised under the microscope there were not any filamentous cells at all and also the number of foci was decreased to an average of around 1-2 per cell (Figures 40 & 41). The control cells of the control strain ori*C*⁺ ori*Z*⁺ *rpoB**35 also had normal length and an average number of Cas1-Cas2 foci 12 (Figures 40 & 41). In essence, these cells now looked like the normal looking cells that had a single origin of replication.

This experiment confirmed the hypothesis that indeed in the ori*C*⁺ ori*Z* strains replication- transcription conflicts do take place. Also, more importantly this series of experiments showed that during replication-transcription conflicts the helicase role of *priA* is vital. This is because if the helicase activity of *priA* is absent from the cells, then the cells express significant cellular stress and undergo genomic instability (Figure 35 & 36)

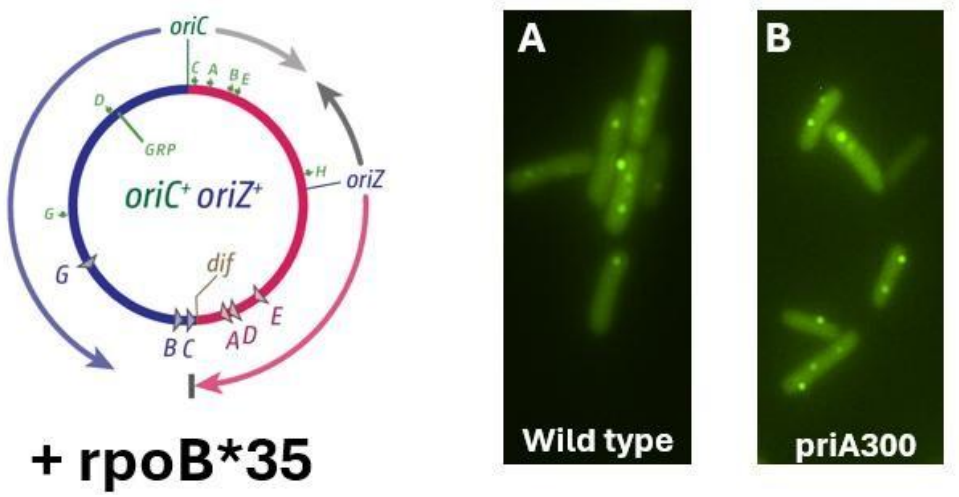

Figure 40. Strains ori*C*⁺ ori*Z*⁺ *rpoB*35* and ori*C*⁺ ori*Z*⁺ *rpoB*35 priA300* under the microscope have a healthy phenotype. Both the cells of strain and the cells of strain. Both the cells of the strain ori*C*⁺ ori*Z*⁺ *rpoB*35* and the cells of the strain ori*C*⁺ ori*Z*⁺ *rpoB*35 priA300* had an average cell length of between 3,5 –4 μm.

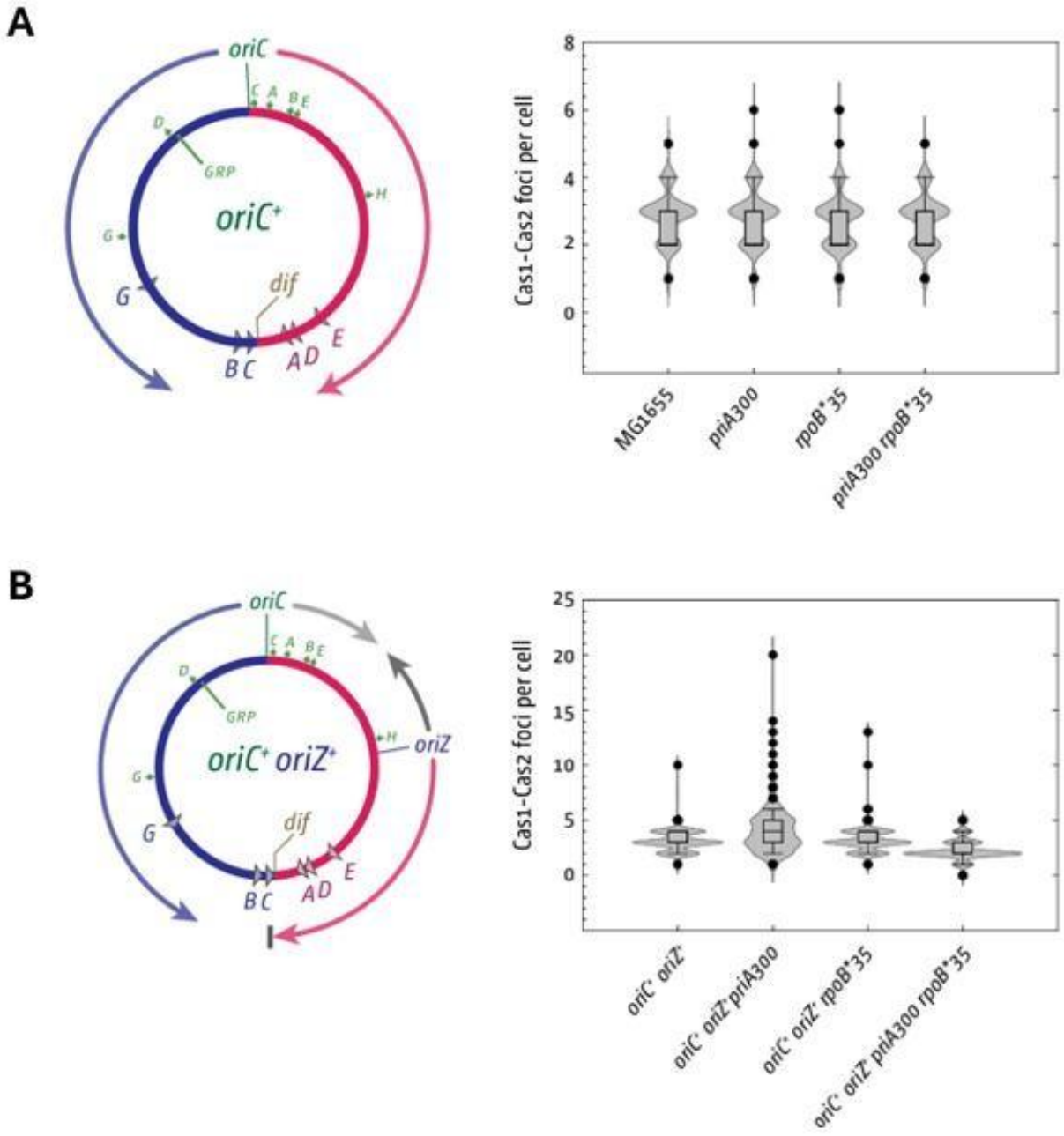

Figure 41. ori*C*⁺ ori*Z*⁺ *priA300* cells have increased Cas1-Cas2 foci counts but the presence of *rpoB*35* reverses this effect. (A) *oriC* single-origin cells. The schematic shows the *E. coli* chromosome with replication initiating at ori*C* and proceeding bidirectionally (arrows). Violin plots to the right show the number of Cas1–Cas2 foci per cell in MG1655 (wild type), *priA300*, *rpoB*35, and priA300 rpoB**35. The distribution is narrow, with medians around 2 foci per cell and a small number of outliers with a maximum of 7. (B) *oriC*⁺ *oriZ*⁺ *dual-origin cells.* The schematic shows an extra replication origin (ori*Z*), producing overlapping replication forks. Violin plots to the right show that cells deficient of the helicase activity of *priA* exhibit a broader distribution and higher number of foci, with a median of 3 foci per cell and high-value outliers that can reach up to 20. This increase is reduced in the rpoB*35* background *oriC*⁺ *oriZ*⁺ *priA300 rpoB**35,

suggesting that *rpoB*35* alleviates fork conflicts or DNA damage in cells with multiple origins.

4.3.4 In the double origin ori*C*$^{+}$ ori*X*$^{+}$ cells there are not any replication-transcription conflicts

In addition to using the *rpoB*35* mutation, which reduces transcription–replication conflicts by destabilising the beta subunit of the RNA polymerase holoenzyme, as a second experimental approach we decided to construct additional control strains carrying the second origin of replication *oriX* which is incorporated on the left replichore as show in figure 42. While the *rpoB*35* results demonstrated that the filamentation and abnormal replication dynamics in the *oriZ* strains were transcription-dependent, the hypothesis was that the *oriX* strain would provide a chromosome engineering verification approach to reconfirm the results.

This second verification experimental approach was devised by considering the fact that, as shown in figure 42, the *E. coli* chromosome is not transcriptionally symmetrical. Specifically, highly transcribed loci such as the *rnh* operon lie to the right of *oriC*, whereas the left replichore contains no comparable hot spots. This means that, unlike *oriZ*, which drives a head-on replication-transcription conflict into the highly transcribed *rnh* operon, *oriX* is positioned such that both replication forks initiating from it move codirectionally with transcription in this region of the chromosome. Specifically, as is also illustrated in figure 42C, the operons *rnd* and *rng* will be met co-directionally by the replication forks initiating from *oriC* and *oriX* respectively. In other words, although the presence of *oriX* will increase the

overall replication complexity it will not introduce a head-on collision between replication and transcription.

This experimental design thus takes advantage of the natural asymmetry of the *E. coli* chromosome, where highly transcribed operons such as *rnh* are present on the right replichore, whereas the corresponding region on the left replichore is relatively transcription- quiet. So, by comparing *oriZ* strains where replication-transcription conflicts are unavoidable with *oriX* strains where no such conflicts should occur, we should be able to verify that the observed cellular defects of filamentation and abnormal Cas1-Cas2 foci were indeed the direct consequence of replication–transcription collisions.

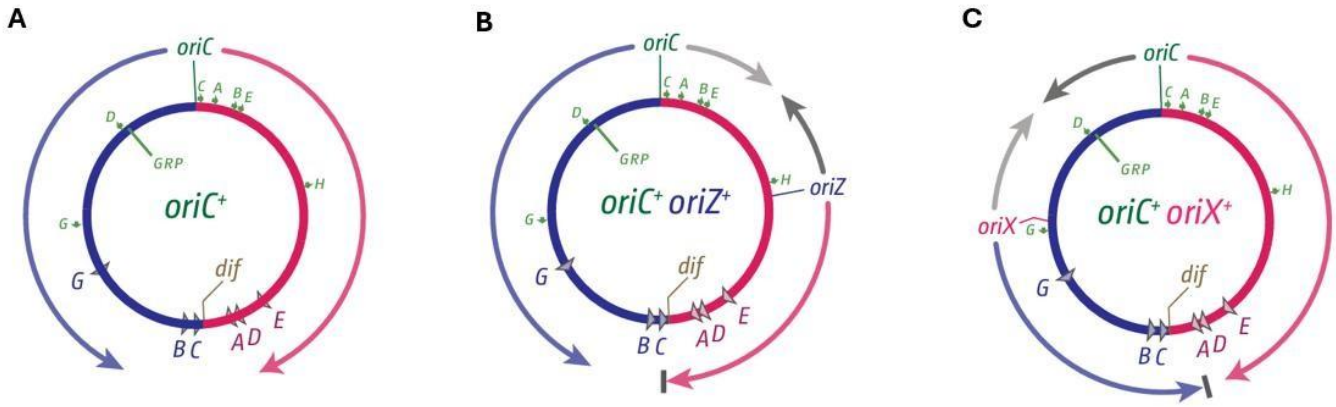

Figure 42. Chromosomal replication programs in single origin *oriC* and double origin *oriC*$^+$ *oriZ*$^+$and *oriC*$^+$ *oriX*$^+$ *E. coli* strains. A.) In wild-type cells two replication fork blocks initiate from *oriC* and move bidirectionally until they meet at the termination area. (B) In *oriC*$^+$ *oriZ*$^+$ cells, the additional ectopic origin of replication on the right replichore generates a leftward-moving fork that collides head-on with the highly transcribed genes such as *rnh*, creating head-on replication–transcription conflicts. (C) In *oriC*$^+$ *oriX*$^+$ cells, the added origin of replication on the left replichore generates symmetric replication forks without leading to head-on collisions with highly transcribed operons.

Indeed, the fluorescence microscopy results confirm our hypothesis because they revealed that all *oriX* strains showed normal cell morphology and the Cas1–Cas2 foci numbers were on average 1–2 per cell, with no evidence of filamentation (Figures 43 & 44). This behaviour mirrors ori*C* only cells and establishes that the presence of the ectopic ori*X* origin of replication does not create head-on replication–transcription conflicts, even when either of the replication restart proteins *priB* and *priC* or the helicase activity of *priA* are disrupted.

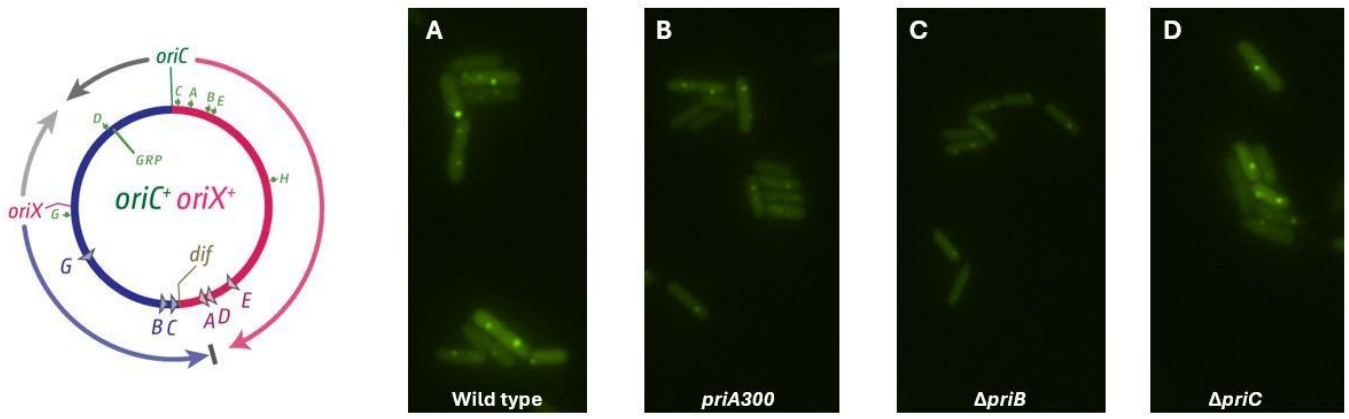

Figure 43. ori*X* strains display normal cell morphology and Cas1–Cas2 foci regardless of replication restart pathway defects. (Left) Schematic representation of the *E. coli* chromosome carrying the ectopic replication origin ori*X*, which aligns replication forks co-directionally with transcription across highly transcribed loci, thereby avoiding head-on replication–transcription conflicts. *(Right)* Fluorescence microscopy images showing Cas1–Cas2 foci in wild-type (A), *priA300* (B), Δ*priB* (C), and Δ*priC* (D) ori*X* strains. All strains exhibit normal cell morphology and 1–2 Cas1–Cas2 foci per cell, comparable to ori*C*-only control cells, indicating that ori*X* does not induce significant replication–transcription conflicts.

To complement the microscopy observations, we quantified Cas1–Cas2 foci per cell across the ori*X* strain panel (wild-type, *priA300*, Δ*priB*, Δ*priC*). As shown in Figures 44 and 45, all strains exhibited narrow distributions centred around two foci per cell, without evidence of abnormal accumulation or broadening of the distribution. These results align with the lack of filamentation seen microscopically and confirm that the addition of the addition origin of replication *oriX* does not induce head-on replication–transcription conflicts, like the addition of *oriZ* does, even when either of the replication restart pathways are impaired (Figure 45).

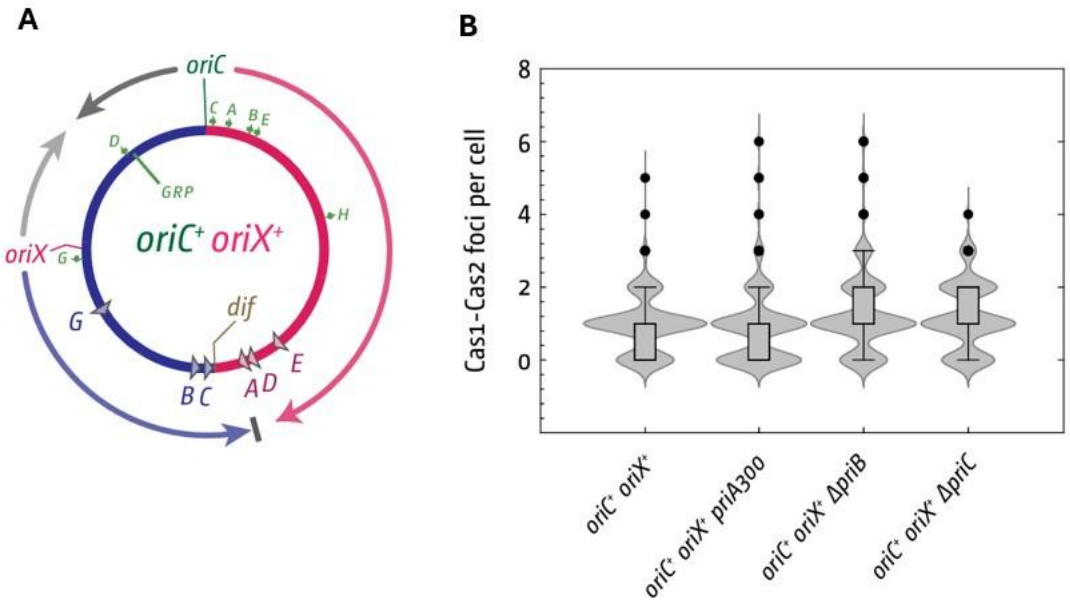

Figure 44. Quantification of Cas1–Cas2 foci confirms the absence of replication–transcription conflicts in ori*X* strains. (A) Schematic representation of the *E. coli* chromosome carrying ori*X*, where replication forks proceed co-directionally with transcription through highly expressed regions, preventing head-on conflicts. (B) Violin and box plots showing the number of Cas1–Cas2 foci per cell in ori*C*⁺ ori*X*⁺, ori*C*⁺ ori*X*⁺ *priA300*, ori*C*⁺ ori*X*⁺ Δ*priB*, and ori*C*⁺ ori*X*⁺ Δ*priC* strains. All strains display a median of around 2 foci per cell with similar distributions, which are consistent with normal replication dynamics and the absence of significant replication stress.

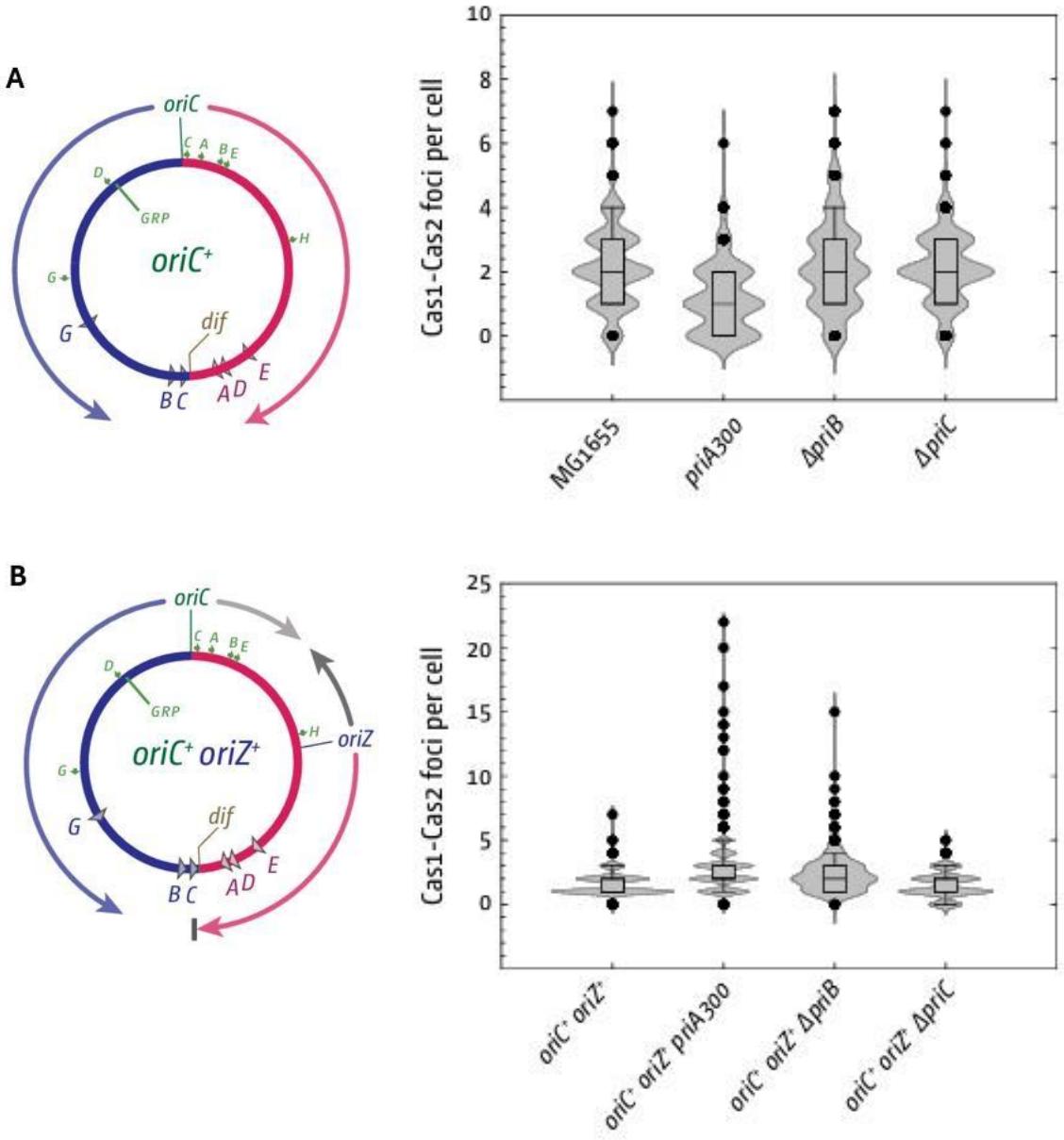
A
oriC
C A B E
D
GRP
oriC+
H
G
G
dif
B C
A D
E
Cas1-Cas2 foci per cell
10
8
6
4
2
0
MG1655
priA300
ΔpriB
ΔpriC
B
oriC
C A B E
D
GRP
oriC+ oriZ+
H
oriZ
G
G
dif
B C
A D
E
Cas1-Cas2 foci per cell
25
20
15
10
5
0
oriC- oriZ-
oriC- oriZ- priA300
oriC- oriZ- ΔpriB
oriC- oriZ- ΔpriC
C
oriC
C A B E
D
GRP
oriC+ oriX+
H
oriX
G
G
dif
B C
A D
E
Cas1-Cas2 foci per cell
8
6
4
2
0
oriC- oriX+
oriC- oriX+ priA300
oriC- oriX+ ΔpriB
oriC- oriX+ ΔpriC

Figure 45. Cas1–Cas2 focus quantification highlights replication stress in ori*Z* but not ori*C* and ori*X* strains. Violin and box plots showing the number of Cas1–Cas2 foci per cell in different replication origin contexts. (A) In the single origin *oriC* background the average Cas1-Cas2 foci counts per cell is around 2 both for the wild type and when either *priB*, *priC* or the helicase activity of *priA* were deleted. (B) In the double origin ori*C*$^{+}$ ori*Z*$^{+}$ background the deletion of either *priB* or of the helicase activity of *priA* leads to a significant increase in Cas1–Cas2 foci, indicating pronounced replication–transcription conflicts. (C) In the double origin ori*C*$^{+}$ ori*X*$^{+}$ background both the wild type strain and the strains where either of the replication restart proteins have a narrow distribution of around 2 foci per cell on average, which is similar to the ori*C* strains.

4.4 Discussion

The results and analyses presented in this chapter successfully address the questions outlined in the aims and provide novel insights regarding how replication-transcription conflicts affect cellular morphology and provide important insight regarding the activity of the three replication restart pathways. By adding the ectopic origin of replication ori*Z*, we created the conducive opportunity for head-on replication-transcription conflict to take place between the highly transcribed *rnh* operon and one of the replication forks initiating from ori*Z* (Mirkin and Mirkin, 2007; Srivatsan *et al.*, 2010; Gaillard and Aguilera, 2016) (Figure33). The results showed that in the double origin oriC^+ oriZ^+ strains the cells deficient of either *priB* or the helicase activity of *priA* experience significant growth defects which lead to filamentation and increased Cas1-Cas2 foci in the cells (Figures 37, 38), which is consistent with replication stress phenotypes reported previously (Rudolph *et al.*, 2010; Dimude *et al.*, 2015; Killelea *et al.*, 2023).

To verify that the phenotypic observations are attributable to the occurrence of head-on replication-transcription conflicts two experiments were conducted. The first experiment involved the addition of the *rpoB*35* allele in the double origin oriC^+ oriZ^+ which leads to a destabilised beta subunit of the RNA polymerase holoenzyme, thus destabilising the RNA polymerase holoenzyme (Figure 39), a strategy previously shown to alleviate replication-transcription conflicts (Trautinger and Lloyd, 2002; Dutta *et al.*, 2011; McGlynn, Savery and Dillingham, 2012). When the oriC^+ ori*Z rpoB*35* and oriC^+ ori*Z rpoB*35 priA300* cells were visualised under the microscope all the cells had a normal cell length as the ori*C* strains and on average every cell

had 1-2 Cas1- Cas2 foci which was also the same for the ori*C* strains (Figures 35, 36, 40 & 41).

In addition, we leveraged the natural asymmetry of the *E. coli* chromosome: highly transcribed operons such as *rnh* are located on the right replichore, whereas the corresponding region on the left replichore is relatively transcriptionally quiet (Figure 42) (Brewer, 1988; Rocha, 2004; Merrikh *et al.*, 2012). By introducing the second origin of replication ori*X* on the left replichore, we compared ori*Z* strains, where replication–transcription conflicts are unavoidable, with ori*X* strains, where such conflicts are minimized. This approach allowed us to validate that the observed cellular defects of filamentation and abnormal Cas1–Cas2 foci were a direct consequence of replication–transcription collisions (Figures 43, 44 & 45).

The results from this chapter establish that the PriA–PriB–DnaT pathway is the principal replication-restart pathway for resolving replication–transcription conflicts. This is attributed to the fact that during replication-transcription conflicts the absence of *priB* or inactivation of the helicase activity of *priA* severely impair DNA replication and cell division, whereas deletion of *priC* does not produce comparable defects to the cell morphology and replication dynamics. This result is consistent with previous genetic and biochemical analyses identifying PriA- PriB- DnaT as the primary replication restart pathway at stalled replication forks (Abe *et al.*, 2021; McKenzie *et al.*, 2022).

Also, the findings from our research reveal, for the first time, a critical role for the helicase activity of *priA* in resolving replication–transcription conflicts. Remarkably, cells deficient in the helicase activity of *priA* showed a degree of stress equivalent to, or exceeding, that caused by loss of *priB*. This indicates that the helicase activity of *priA* helicase function is indispensable

for maintaining replication integrity under replication- transcription collisions. This was an unexpected result, as no previous studies have highlighted the essential contribution of the helicase activity of *priA* specifically in the context of resolving replication-transcription collisions (McGlynn and Lloyd, 2002; Windgassen *et al.*, 2018; Duckworth *et al.*, 2023). Our data therefore underscore the importance of PriA helicase in preventing abnormal DNA replication and cellular stress during replication–transcription conflicts.

The data raise an important question: why do cells lacking PriA helicase activity display an increased number of Cas1–Cas2 foci? Notably, this phenotype is not uniform across the population. While some cells appear morphologically normal, others are filamentous, and it is these filamentous cells that exhibit the highest density of Cas1–Cas2 foci. Two plausible explanations may account for this phenomenon. First, blocked replication forks may be processed through homologous recombination, generating additional replication forks as a by-product (Cox *et al.*, 2000; Marians, 2004; Rudolph *et al.*, 2010). Alternatively, fork stalling may delay replication completion long enough for ori*C* to refire, leading to overlapping rounds of replication before existing forks are fully cleared (Skarstad and Katayama, 2013). Together, these findings highlight a heterogeneous stress response and point to an unexpectedly critical role for *priA* helicase activity in maintaining replication integrity under transcriptional conflict.

To answer this question some interesting experiments could be carried out. For instance, tracking via time lapse microscopy the origin of replication ori*C* using fluorescent tags might show if the origin refires again or not. Also, quantifying the levels of the initiator protein DnaA should show if the origin refires or not. Additionally, a key difference between the two explanations is that if indeed the origin of replication refires again then that the content of

DNA should increase over time since new DNA will be created. Hence, an informative experiment that could answer the question would involve DNA content flow cytometry. Specifically, if ori*C* refires then the results should show a right-shifted, multi-modal distribution with elevated mean C-values. However, if the phenomenon is attributed to homologous recombination, then the results from flow cytometry should produce a modest broadening around wild-type levels.

Filamentation provides another readout of replication stress. So, inactivating RecA in PriA helicase–deficient cells might also clarify whether the filaments arise from recombination-mediated fork processing or from refiring of the origin of replication. More specifically, if recombination is responsible, then a *priA recA* double mutant should exhibit reduced filamentation but also poor viability, reflecting thus a failure to generate restartable replication forks. However, if ori*C* refiring drives the phenotype, then filamentation should persist despite *recA* loss, as new replication forks would continue to collide without being properly repaired. Thus, the combination of origin tracking, cell morphology and DNA-content flow cytometry might provide a reliable framework to distinguish between the two mechanisms.

Chapter 5

Effects of Saccharin on Replication Restart Dynamics

5.1 Introduction and research aims

Saccharin, the 2D chemical structure of which is illustrated in figure 46, is one of the most broadly used artificial sweeteners globally. Chemist Constantin Fahlberg discovered saccharin at Johns Hopkins University in 1879 (Warner, 2008). This serendipitous discovery combined with initial results reporting its non-toxicity for humans, led to its mass production and use especially during World War I (Singh *et al.*, 2020). Interestingly research using mice as the model organism reported that it could lead to cancer, and thus resulted to its characterisation as a human carcinogen for almost 20 years by the Food and Drug Administration (Reuber, 1978; Anderson and Kirkland, 1980; Cooper, 1985; Oser, 1985). However, later research reassuring that saccharin does not affect the human physiology negatively, led to its mass reproduction from the onset of the 21st century and onwards (Renwick, 1985, 2006). As of 2025, saccharin remains among the top choices of sugar substitutes and often included in many products mainly because it has zero calories (Jurcevic Zidar *et al.*, 2025).

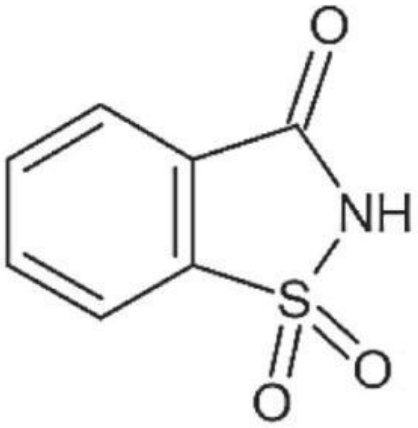

Figure 46. 2D structure of the artificial sweetener saccharin. The empirical formula of saccharin is $C_7H_5NO_3S$ and its molecular mass is 183.18 g/mol. Figure from (de Dios *et al.*, 2025).

However, there appears to be an increase of literature findings reporting that saccharin consumption can affect negatively the gut health. This is because it can cause the emergence of new bacterial species, hence an imbalance between the beneficial and the disease-causing bacteria (Payne, Chassard and Lacroix, 2012; Suez *et al.*, 2014). Also, saccharin consumption has been associated with a spark of inflammation and a range of autoimmune diseases (Cao *et al.*, 2017; Shi *et al.*, 2020). On top of that, research indicates that instead of decreasing the levels of blood sugar saccharin actually leads to the opposite effect and can contribute to cardiovascular diseases (Aeberli *et al.*, 2011; Ma *et al.*, 2022). Strikingly, research indicates that in reality saccharin is an addictive substance and its consumption can negatively affect the mental health (Lenoir *et al.*, 2007).

Given the widespread consumption of saccharin as an artificial sweetener globally and across all ages, we wanted to determine whether exposure to this compound perturbs DNA replication dynamics in *E. coli,* an important bacterium which belongs to the physiological human microbiota. Previous work has shown that saccharin can induce filamentation and stress responses in bacterial cells, but its direct effects on replication and genome maintenance still have not been investigated.

5.2 The experimental set-up

Since replication stress frequently manifests as stalled replication forks and abnormal DNA processing, we monitored replication dynamics in *E. coli* MG1655 using Cas1–LFP–Cas2 foci as accurate markers of replication associated lesions. First, we visualised normal MG1655 cells where saccharin, which was kindly provided by the lab of Professor Ronan McCarthy, was either absent or present in the system. Then, to accurately determine which replication restart pathway is most critical under these conditions, we visualized cells deficient in either PriB or PriC replication restart proteins.

5.3 Results

5.3.1 In the presence of saccharin wild type *E. coli* cells undergo DNA replication defects and cellular division perturbations

In the wild-type strain, treatment with 1.4 % and 2 % saccharin led to a marked increase in Cas1–Cas2 foci compared to mock-treated controls (Figure 47A). This accumulation of foci coincided with pronounced filamentation, indicating activation of DNA damage or replication stress responses. Quantitative analysis confirmed a significant, dose-dependent elevation in the number of foci per cell (Figure 47B).

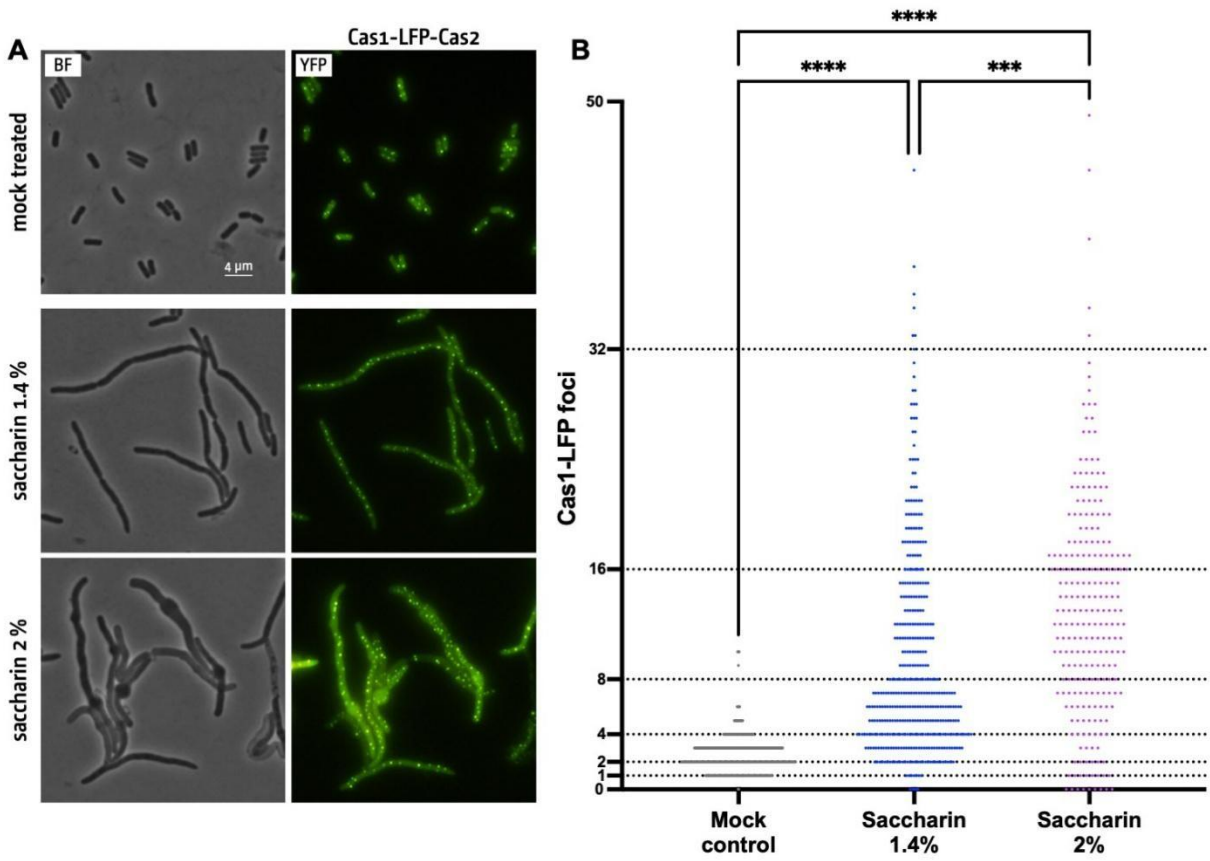

Figure 47. Replication dynamics in living cells alter dramatically following treatment with saccharin. Microscope images (A) and violin plots (B) of the wild type *E. coli* cells with and without the addition of saccharin. Before the addition of saccharin, the *E. coli* cells looked healthy with a normal number of foci. After the addition of 1.4% saccharin the cells became filamentous, and the number of foci increased dramatically reaching up to 32 per cells as the bar chart shows. The situation exacerbated with the addition of 2% saccharin. The dashed lines show the expected foci numbers based on the principles of DNA replication. In total the foci of 603 wild type *E. coli* cells, 414 cells treated with 1.4% saccharin and 241 cells treated with 2% saccharin were counted. The Kruskal–Wallis test with Dunn's correction were utilised for the analysis. Significance is shown as ***$P \leq 0.001$, ****$P \leq 0.0001$

5.3.2 In the presence of saccharin *E. coli* cells deficient of either *priB* or *priC* undergo DNA replication defects, cellular division perturbations and cell death

To test whether the replication restart machinery modulates this response, we examined Δ*priB* and Δ*priC* mutants under the same conditions. Both mutants displayed filamentous cells and fewer Cas1–Cas2 foci than wildtype cells following saccharin exposure (Figure 48). However, brightfield images revealed that Δ*priB* and Δ*priC* cells were not only severely filamentous but also some of them frequently lysed (Figure 49), indicating that these strains were unable to maintain viability under saccharin-induced stress. Importantly, deletion of *priB* produced a more dramatic phenotype than the deletion of *priC* under saccharin stress Cas1–Cas2 foci (Figure 48).

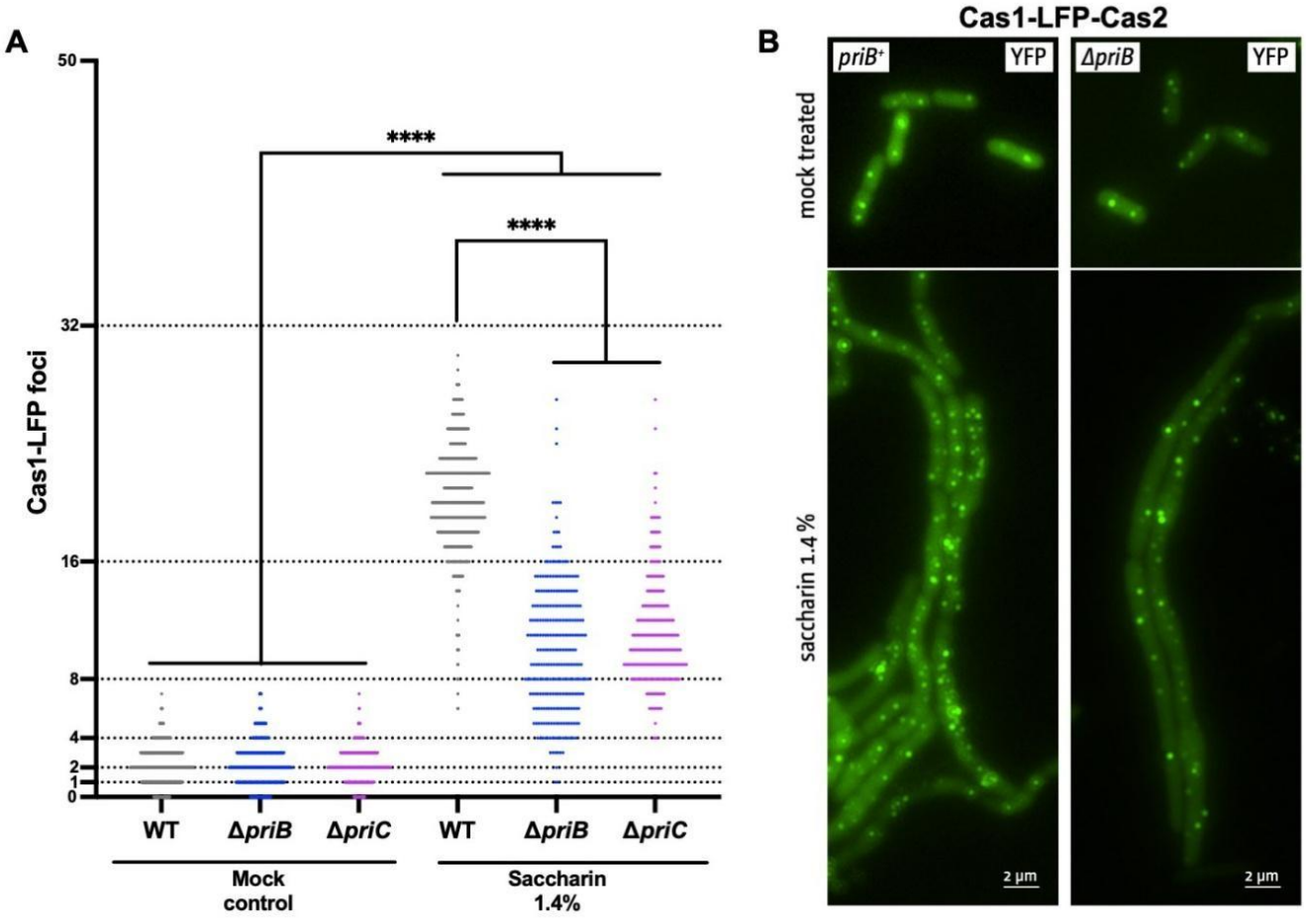

Figure 48. Cas1-Cas2 foci formation in saccharin-treated cells is significantly reduced in cells lacking the replication restart proteins PriB and PriC. Violin plots (A) and microscope images (B) of the *E. coli* cells with and without the addition of saccharin inactivating the replicative restart protein *priB*. The addition of saccharin filamented cells emerged with many foci but they were significantly fewer compared to the wild-type cells. The dashed lines show the expected foci numbers based on the principles of DNA replication. The number of cells analysed was 795 and 507 for wildtype, *ΔpriB* and *ΔpriC*, respectively, for the mock-treated samples, and 561, 274 and 297 for wild-type, *ΔpriB* and *ΔpriC*, respectively, for the saccharin-treated cells. The Kruskal–Wallis test with Dunn's correction were utilised for the analysis. Significance is shown as ****$P \leq 0.0001$

5.4 Discussion

The results from this chapter are striking because they demonstrate that saccharin, which is one of the most widely consumed artificial sweeteners globally, and especially in the last 25 years, induces striking replication stress in *E. coli* cells, impacting also significantly its replication-restart pathways. Although saccharin has been regarded as metabolically inert in bacteria, and classified as non-carcinogenic to humans, our findings reveal that exposure to this compound significantly perturbs DNA replication dynamics, cell division, and viability. This work establishes thus an important direct mechanistic link between dietary saccharin exposure and the molecular machinery that restarts stalled replication forks.

Notably, our data show that cells lacking PriB are severely compromised because they suffer the most and display pronounced filamentation and an abnormal accumulation of Cas1–Cas2 foci (Figure 48). It has already been established that filamentation is a hallmark of replication stress and activation of the SOS response, reflecting a failure to complete chromosome

duplication and segregation of DNA prior to cell division (Kuzminov, 1999; Balaban *et al.*, 2004). Although saccharin has historically been regarded as metabolically inert in bacteria, particularly following its reclassification as non-carcinogenic to humans, our findings reveal that exposure to this compound significantly perturbs DNA replication dynamics, cell division, and viability. However, the findings demonstrate that cells deficient of PriC exhibit relatively milder DNA replication defects (Figure 48). So, the results from this chapter establish that the PriA–PriB–DnaT pathway is the most important replication-restart mechanism under saccharin-induced stress. Importantly, the increased number of Cas1–Cas2 foci in filamentous cells provides important mechanistic clues. These foci likely indicate persistent DNA intermediates, possibly arising from homologous recombination–mediated fork restart or uncontrolled origin refiring events as also indicated by results chapter 2 (Killelea *et al.*, 2023).

Physiologically, these findings are remarkable because *E. coli* is an important bacterium of the human gut microbiota. The results establish that even the wild type *E.coli* cells with all the replication restart proteins undergo DNA replication and cellular defects in the presence of saccharin. Alarmingly, saccharin concentrations exceeding 1.4% are not uncommon in processed foods. All of these lead to the conclusion that if saccharin exposure also stresses other beneficial gut bacteria that rely on the PriA–PriB–DnaT pathway, this could significantly alter the composition of the gut microbiome.

In summary, these results highlight that the PriA–PriB–DnaT pathway is a crucial safeguard of genomic integrity during chemical stress, whereas the PriC pathway plays only a minor role. Saccharin, which was long assumed to be metabolically inert, in reality can dramatically alter bacterial replication dynamics, raising questions about how artificial sweetener consumption might influence the stability, evolution, and composition of the gut

microbiome. Importantly, Previous these results expand on previous studies which have shown that saccharin can alter gut microbial composition and promote microbial dysbiosis (Payne, Chassard and Lacroix, 2012; Suez *et al.*, 2014; Markus, 2024; Hetta *et al.*, 2025).

More broadly, this chapter reinforces the concept that replication restart pathways function as critical buffers against diverse forms of stress, including protein roadblocks, transcriptional collisions, and chemically induced perturbations. The repeated identification of PriA helicase and PriB as key determinants of cellular survival across multiple experimental systems highlights their central role in the bacterial DNA damage tolerance network. Considering the conservation and essential nature of the PriA-PriB-DnaT replication restart pathway, this pathway may represent promising target for antimicrobial strategies that exploit bacterial vulnerabilities under stress conditions (McGlynn and Lloyd, 2002).

Chapter 6

General Discussion

Although DNA is the building block for the survival of the organisms and thus the continuation of life, its accurate replication is constantly threatened by obstacles and collisions (Aguilera and García-Muse, 2013; Zeman and Cimprich, 2014). If these blocks and collisions are not resolved precisely and timely then the genomic and cellular stability of the organism can be irreversibly damaged. This PhD research aimed to investigate the exact mechanisms by which the model organism *Escherichia coli* efficiently manages to preserve its genomic stability when its DNA replication machinery encounters different types of obstacles and collisions.

In the first experimental approach in Results Chapter 1 by using a novel inducible replication fork block model of various lengths on the right replichore of *E. coli* I showed that during DNA replication cells deficient of either replication restart protein experience both considerable DNA replication difficulties and cell division defects. This result is consistent with previous studies which demonstrate replisome instability at protein-DNA barriers (Grompone, Ehrlich and Michel, 2004; Possoz *et al.*, 2006; McGlynn and Guy, 2008b). Interestingly, the microscopy results indicate that replication forks arrest upon meeting the nucleoprotein blocks, and even worse some of the replication forks can be disassembled. However, the results indicate that DNA replication can continue normally after a while. This indicates that replisomes must be re-assembled at the site of replication fork arrest. It has been established that in the model organism *E. coli* there are four replication restart proteins: PriA, PriB, PriC and DnaT forming the following restart pathways: PriA-PriB-DnaT, PriA-PriC-DnaT and PriC-DnaT (Sandler, 2000; Heller and Marians, 2006). However, which of the three replication restart pathways is principally involved for the resolution of

nucleoprotein blocks remained elusive. The spot dilutions and time lapse microscopy results from this PhD thesis indicate that PriA-PriB-DnaT is the replication restart pathway that is more important at nucleoprotein obstacles. In the absence of either PriB or the helicase activity of PriA, cells filament and accumulate suppressor mutations which is in agreement with published research (Michel, Sinha and Leach, 2018). Interestingly, when the same nucleoprotein obstacle composed of considerably more tandem repeats was integrated this time into the left replichore of *E. coli* the results were surprising because they indicated that PriC is the most important replication restart protein for the resolution of this block. This interesting and unexpected discovery raises the following question: Is the function of PriC activated based on the size of the nucleoprotein complexes or based on which replichore the block is? Results Chapter 1 provides interesting future experimental approaches that could answer this question.

In the second experimental approach, which is described in Results Chapter 2, I addressed replication- transcription conflicts which are recognised as a major source of genomic instability across all domains of life, and I investigated the importance of each replication restart pathway for their resolution (Hamperl and Cimprich, 2014, 2016; Dimude, Midgley-Smith and Rudolph, 2018; Lalonde *et al.*, 2021). By incorporating the second origin of replication ori*Z* on the right replichore I induced a head-on replication-transcription conflict between one of the replication forks initiating from ori*Z* and the highly transcribed *rnh* operon (Ivanova *et al.*, 2015b; Dimude *et al.*, 2018). By utilising Cas1-Cas2 foci as an efficient and accurate biomarker of replication fork presence I observed that cells deficient of either PriB or the helicase activity of PriA filamented and had significantly increased Cas1-

Cas2 foci. The deletion of PriC from the system did not cause any growth defects. Interestingly though, when either PriB or the helicase activity of PriA were deleted surprisingly the system expressed heterogeneity. More specifically, there were both healthy cells with an average number of Cas1-Cas2 foci and filamented cells with increased Cas1- Cas2 foci. To verify that the results obtained were directly the result of replication-transcription conflicts two experiments were devised. In the first approach, in the double origin ori*C*$^{+}$ ori*Z*$^{+}$ strains I incorporated the *rpoB*35* mutation which destabilises the RNA polymerase holoenzyme (Trautinger and Lloyd, 2002; Dutta *et al.*, 2011). The hypothesis was that if indeed replication-transcription conflicts resulted in the filamentation phenomenon then the integration of the destabilised RNA polymerase should alleviate this effect leading thus to the presence of normally looking cells with normal Cas1-Cas2 foci per cell. Indeed, the hypothesis was correct because in the double origin ori*C*$^{+}$ ori*Z*$^{+}$ *priA300 rpoB*35* strain all the cells had normal length and on average normal Cas1-Cas2 foci. In the second experimental approach considering the transcriptional asymmetry of the *E. coli* chromosome, I incorporated the ectopic origin of replication ori*X* on the left replichore (de Massy *et al.*, 1984). ori*X* was positioned such that both replication forks initiating from it move co-directionally with the transcription in this region of the chromosome. The results clearly showed that in the ori*C*$^{+}$ ori*X*$^{+}$ strain the deletion of either replication restart protein did not cause any DNA replication difficulties or cellular defects. So, the results from Results Chapter 2 clearly establish that PriA-PriB-DnaT is the replication restart pathway that resolves head-on replication-transcription conflicts. In the absence of either PriB or the helicase activity of PriA cells struggle to replicate their DNA and to divide during replication-transcription conflicts. Also, the results from

Results Chapter 2 re-establish that head-on replication-transcription conflicts are considerably more deleterious than co-directional conflicts, which is in agreement with published research (Mirkin and Mirkin, 2005b; Wang, Berkmen and Grossman, 2007; Lang *et al.*, 2017).

Finally, in Results Chapter 3, since *E. coli* is normally a beneficial gut bacterium and part of the human gut microbiome, I was intrigued to investigate the effects of the widely used artificial sweetener saccharin on *E. coli* DNA replication dynamics. Surprisingly, the presence of relatively low concentration of Saccharin (1.4% and 2%) is enough to cause DNA replication difficulties, cellular defects and even cellular death. Also, the deletion of PriB or PriC from the system exacerbates the defects with the more pronounced defects being visible in the PriB deficient cells. The findings of this chapter shed new light into the effects of saccharin on *E. coli* and on replication restart pathways. The results from this PhD indicate that saccharin has the potential to considerably affect *E. coli* and given the fact that saccharin is included in many products worldwide and sometimes in unknown concentrations raises the query as to how safe it is to consume and at what concentration.

This PhD work provides important information about bacterial physiology. PriA–PriB–DnaT pathway and more importantly the helicase activity of PriA appear to be the main factors involved to resolve both nucleoprotein complexes and replication-transcription conflicts. The unexpected finding of the fundamental role of the helicase activity of PriA establishes it as a novel and promising target for antibacterials. Since, the helicase activity of PriA is conserved but not exactly identical across the bacterial species small molecule inhibitors of the helicase activity of PriA could selectively kill the disease-causing bacteria which in principle should experience increase DNA

replication hence increased replication stress. So far, not any antibacterial has been developed that specifically targets the helicase activity of the antibacterial resistant bacteria. So, especially in the era of the multi-drug resistant bacteria, developing a novel drug that directly targets the replication principles of disease-causing bacteria will certainly be valuable.

Moreover, the findings from this PhD provide conceptual parallels in cancer biology research since DNA replication principles are conserved. Although bacterial and eukaryotic replication machineries are not the same, both depend on pathways that recognize and restart stalled replication forks. In human cells, cancer-associated replication stress is resolved by proteins which function analogously to bacterial fork restart factors. Just as PriA helicase is essential for bacterial viability under stress, human cancer cells rely on their restart pathways to survive oncogene-induced replication–transcription conflicts. By identifying PriA helicase as an Achilles' heel in bacteria, this research thus provides an important bridge between microbial DNA replication and the vulnerabilities of hyperproliferative cancer cells.

References

Abe, Y., Ikeda, Y., Fujiyama, S., Kini, R.M. and Ueda, T. (2021) 'A structural model of the PriB–DnaT complex in Escherichia coli replication restart', *FEBS Letters*, 595(3), pp. 341–350. Available at: https://doi.org/10.1002/1873-3468.14020.

Achar, Y.J., Balogh, D., Neculai, D., Juhasz, S., Morocz, M., Gali, H., Dhe-Paganon, S., Venclovas, Č. and Haracska, L. (2015) 'Human HLTF mediates postreplication repair by its HIRAN domain-dependent replication fork remodelling', *Nucleic Acids Research*, 43(21), pp. 10277–10291. Available at: https://doi.org/10.1093/nar/gkv896.

Acharya, S., Foster, P.L., Brooks, P. and Fishel, R. (2003) 'The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair', *Molecular Cell*, 12(1), pp. 233–246. Available at: https://doi.org/10.1016/s1097-2765(03)00219-3.

Aeberli, I., Gerber, P.A., Hochuli, M., Kohler, S., Haile, S.R., Gouni-Berthold, I., Berthold, H.K., Spinas, G.A. and Berneis, K. (2011) 'Low to moderate sugar-sweetened beverage consumption impairs glucose and lipid metabolism and promotes inflammation in healthy young men: a randomized controlled trial1234', *The American Journal of Clinical Nutrition*, 94(2), pp. 479–485. Available at: https://doi.org/10.3945/ajcn.111.013540.

Aguilera, A. and García-Muse, T. (2013) 'Causes of genome instability', *Annual Review of Genetics*, 47, pp. 1–32. Available at: https://doi.org/10.1146/annurev-genet-111212-133232.

Ahmed, S.K., Hussein, S., Qurbani, K., Ibrahim, R.H., Fareeq, A., Mahmood, K.A. and Mohamed, M.G. (2024) 'Antimicrobial resistance: Impacts, challenges, and future prospects', *Journal of Medicine, Surgery, and Public Health*, 2, p. 100081. Available at: https://doi.org/10.1016/j.glmedi.2024.100081.

Akama, Y., Yoshida, R., Ozaki, S., Kawakami, H. and Katayama, T. (2025) 'SSB promotes DnaB helicase passage through DnaA complexes at the replication origin oriC for bidirectional replication', *The Journal of Biochemistry*, 177(4), pp. 305–316. Available at: https://doi.org/10.1093/jb/mvaf003.

Akerlund, T., Nordström, K. and Bernander, R. (1995) 'Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli', *Journal of Bacteriology*, 177(23), pp. 6791–6797. Available at: https://doi.org/10.1128/jb.177.23.6791-6797.1995.

Al Qabili, D.M.A., Aboueisha, A.-K.M., Ibrahim, G.A., Youssef, A.I. and El-Mahallawy, H.S. (2022) 'Virulence and antimicrobial-resistance of shiga toxin-producing E. coli (STEC) Isolated from edible shellfish and its public health significance',

Archives of Microbiology, 204(8), p. 510. Available at: https://doi.org/10.1007/s00203-022-03114-2.

Al-Behadili, A., Uhler, J.P., Berglund, A.-K., Peter, B., Doimo, M., Reyes, A., Wanrooij, S., Zeviani, M. and Falkenberg, M. (2018) 'A two-nuclease pathway involving RNase H1 is required for primer removal at human mitochondrial OriL', *Nucleic Acids Research*, 46(18), pp. 9471–9483. Available at: https://doi.org/10.1093/nar/gky708.

Aminov, R.I. (2010) 'A brief history of the antibiotic era: lessons learned and challenges for the future', *Frontiers in Microbiology*, 1, p. 134. Available at: https://doi.org/10.3389/fmicb.2010.00134.

Anderson, R.L. and Kirkland, J.J. (1980) 'The effect of sodium saccharin in the diet on caecal microflora', *Food and Cosmetics Toxicology*, 18(4), pp. 353–355. Available at: https://doi.org/10.1016/0015-6264(80)90188-1.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) 'Sequence and organization of the human mitochondrial genome', *Nature*, 290(5806), pp. 457–465. Available at: https://doi.org/10.1038/290457a0.

Arias-Palomo, E., O'Shea, V.L., Hood, I.V. and Berger, J.M. (2013) 'The bacterial DnaC helicase loader is a DnaB ring breaker', *Cell*, 153(2), pp. 438–448. Available at: https://doi.org/10.1016/j.cell.2013.03.006.

Arnold, A., McLellan, S. and Stokes, J.M. (2025) 'How AI can help us beat AMR', *npj Antimicrobials and Resistance*, 3(1), p. 18. Available at: https://doi.org/10.1038/s44259-025-00085-4.

Aslam, B., Wang, W., Arshad, M.I., Khurshid, M., Muzammil, S., Rasool, M.H., Nisar, M.A., Alvi, R.F., Aslam, M.A., Qamar, M.U., Salamat, M.K.F. and Baloch, Z. (2018) 'Antibiotic resistance: a rundown of a global crisis', *Infection and Drug Resistance*, 11, pp. 1645–1658. Available at: https://doi.org/10.2147/IDR.S173867.

Atkin, N.D., Raimer, H.M. and Wang, Y.-H. (2019) 'Broken by the Cut: A Journey into the Role of Topoisomerase II in DNA Fragility', *Genes*, 10(10), p. 791. Available at: https://doi.org/10.3390/genes10100791.

Atkinson, J. and McGlynn, P. (2009) 'Replication fork reversal and the maintenance of genome stability', *Nucleic Acids Research*, 37(11), pp. 3475–3492. Available at: https://doi.org/10.1093/nar/gkp244.

Au, S.H., Edd, J., Haber, D.A., Maheswaran, S., Stott, S.L. and Toner, M. (2017) 'Clusters of Circulating Tumor Cells: a Biophysical and Technological Perspective', *Current opinion in biomedical engineering*, 3, pp. 13–19. Available at: https://doi.org/10.1016/j.cobme.2017.08.001.

Avery, O.T., Macleod, C.M. and McCarty, M. (1944) 'STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III', *The Journal of Experimental Medicine*, 79(2), pp. 137–158. Available at: https://doi.org/10.1084/jem.79.2.137.

Baker, S.J., Fearon, E.R., Nigro, J.M., Hamilton, S.R., Preisinger, A.C., Jessup, J.M., vanTuinen, P., Ledbetter, D.H., Barker, D.F., Nakamura, Y., White, R. and Vogelstein, B. (1989) 'Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas', *Science*, 244(4901), pp. 217–221. Available at: https://doi.org/10.1126/science.2649981.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L. and Leibler, S. (2004) 'Bacterial persistence as a phenotypic switch', *Science (New York, N.Y.)*, 305(5690), pp. 1622–1625. Available at: https://doi.org/10.1126/science.1099390.

Bastia, D. and Zaman, S. (2014) 'Mechanism and Physiological Significance of Programmed Replication Termination', *Seminars in cell & developmental biology*, 0, pp. 165–173. Available at: https://doi.org/10.1016/j.semcdb.2014.04.030.

Bayona-Feliu, A. and Aguilera, A. (2025) 'Transcription-Replication Conflicts: Unlocking New Frontiers in Cancer', *Bioessays*, 47(8), p. e70025. Available at: https://doi.org/10.1002/bies.70025.

Behrmann, M.S., Perera, H.M., Hoang, J.M., Venkat, T.A., Visser, B.J., Bates, D. and Trakselis, M.A. (2021) 'Targeted chromosomal Escherichia coli:dnaB exterior surface residues regulate DNA helicase behavior to maintain genomic stability and organismal fitness', *PLoS Genetics*, 17(11), p. e1009886. Available at: https://doi.org/10.1371/journal.pgen.1009886.

Bell, S.D. (2019) 'Initiating DNA replication: a matter of prime importance', *Biochemical Society Transactions*, 47(1), pp. 351–356. Available at: https://doi.org/10.1042/BST20180627.

Bell, S.P. and Kaguni, J.M. (2013) 'Helicase Loading at Chromosomal Origins of Replication', *Cold Spring Harbor Perspectives in Biology*, 5(6), p. a010124. Available at: https://doi.org/10.1101/cshperspect.a010124.

Berghuis, B.A., Dulin, D., Xu, Z.-Q., Van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N.E., Depken, M. and Dekker, N.H. (2015) 'Strand separation establishes a sustained lock at the Tus–Ter replication fork barrier', *Nature Chemical Biology*, 11(8), pp. 579–585. Available at: https://doi.org/10.1038/nchembio.1857.

Bernander, R. and Nordström, K. (1990) 'Chromosome replication does not trigger cell division in E. coli', *Cell*, 60(3), pp. 365–374. Available at: https://doi.org/10.1016/0092-8674(90)90588-6.

Bernhardt, T.G. and de Boer, P.A.J. (2004a) 'Screening for synthetic lethal mutants in Escherichia coli and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity', *Molecular Microbiology*, 52(5), pp. 1255–1269. Available at: https://doi.org/10.1111/j.1365-2958.2004.04063.x.

Bernhardt, T.G. and de Boer, P.A.J. (2004b) 'Screening for synthetic lethal mutants in Escherichia coli and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity', *Molecular Microbiology*, 52(5), pp. 1255–1269. Available at: https://doi.org/10.1111/j.1365-2958.2004.04063.x.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P. and Canaider, S. (2013) 'An estimation of the number of cells in the human body', *Annals of Human Biology*, 40(6), pp. 463–471. Available at: https://doi.org/10.3109/03014460.2013.807878.

Biswas-Fiss, E.E. (2006) 'Interaction of the nucleotide binding domains and regulation of the ATPase activity of the human retina specific ABC transporter, ABCR', *Biochemistry*, 45(11), pp. 3813–3823. Available at: https://doi.org/10.1021/bi052059u.

Blackburn, E.H. (2001) 'Switching and signaling at the telomere', *Cell*, 106(6), pp. 661–673. Available at: https://doi.org/10.1016/s0092-8674(01)00492-5.

Blaine, H.C., Simmons, L.A. and Stallings, C.L. (2023) 'Diverse Mechanisms of Helicase Loading during DNA Replication Initiation in Bacteria', *Journal of Bacteriology*. Edited by P.A. Champion, 205(4), pp. e00487-22. Available at: https://doi.org/10.1128/jb.00487-22.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) 'The complete genome sequence of Escherichia coli K-12', *Science (New York, N.Y.)*, 277(5331), pp. 1453–1462. Available at: https://doi.org/10.1126/science.277.5331.1453.

Blau, N. (2016) 'Genetics of Phenylketonuria: Then and Now', *Human Mutation*, 37(6), pp. 508–515. Available at: https://doi.org/10.1002/humu.22980.

Bloom, L.B., Chen, X., Fygenson, D.K., Turner, J., O'Donnell, M. and Goodman, M.F. (1997) 'Fidelity of Escherichia coli DNA polymerase III holoenzyme. The effects of beta, gamma complex processivity proteins and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies', *The Journal of Biological Chemistry*, 272(44), pp. 27919–27930. Available at: https://doi.org/10.1074/jbc.272.44.27919.

Blount, Z.D. (2015a) 'The unexhausted potential of E. coli', *eLife*, 4, p. e05826. Available at: https://doi.org/10.7554/eLife.05826.

Blount, Z.D. (2015b) 'The unexhausted potential of E. coli', *eLife*, 4, p. e05826. Available at: https://doi.org/10.7554/eLife.05826.

Boire, A., Burke, K., Cox, T.R., Guise, T., Jamal-Hanjani, M., Janowitz, T., Kaplan, R., Lee, R., Swanton, C., Vander Heiden, M.G. and Sahai, E. (2024) 'Why do patients with cancer die?', *Nature Reviews Cancer*, 24(8), pp. 578–589. Available at: https://doi.org/10.1038/s41568-024-00708-4.

Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, J.E., Gilbert, D., Rice, L.B., Scheld, M., Spellberg, B. and Bartlett, J. (2009) 'Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America', *Clinical Infectious Diseases*, 48(1), pp. 1–12. Available at: https://doi.org/10.1086/595011.

Bournaka, S., Badra-Fajardo, N., Arbi, M., Taraviras, S. and Lygerou, Z. (2024) 'The cell cycle revisited: DNA replication past S phase preserves genome integrity', *Seminars in Cancer Biology*, 99, pp. 45–55. Available at: https://doi.org/10.1016/j.semcancer.2024.02.002.

Brewer, B.J. (1988) 'When polymerases collide: replication and the transcriptional organization of the E. coli chromosome', *Cell*, 53(5), pp. 679–686. Available at: https://doi.org/10.1016/0092-8674(88)90086-4.

Brierley, J., Gospodarowicz, M. and O'Sullivan, B. (2016) 'The principles of cancer staging', *ecancermedicalscience*, 10, p. ed61. Available at: https://doi.org/10.3332/ecancer.2016.ed61.

Brown, J.S., Amend, S.R., Austin, R.H., Gatenby, R.A., Hammarlund, E.U. and Pienta, K.J. (2023) 'Updating the Definition of Cancer', *Molecular Cancer Research*, 21(11), pp. 1142–1147. Available at: https://doi.org/10.1158/1541-7786.MCR-23-0411.

Cadman, C.J., Lopper, M., Moon, P.B., Keck, J.L. and McGlynn, P. (2005) 'PriB stimulates PriA helicase via an interaction with single-stranded DNA', *The Journal of Biological Chemistry*, 280(48), pp. 39693–39700. Available at: https://doi.org/10.1074/jbc.M508521200.

Cadman, C.J. and McGlynn, P. (2004) 'PriA helicase and SSB interact physically and functionally', *Nucleic Acids Research*, 32(21), pp. 6378–6387. Available at: https://doi.org/10.1093/nar/gkh980.

Cairns, J. (1963) 'The Chromosome of Escherichia coli', *Cold Spring Harbor Symposia on Quantitative Biology*, 28, pp. 43–46. Available at: https://doi.org/10.1101/SQB.1963.028.01.011.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A. (2001) 'Structural mechanism for rifampicin inhibition of bacterial rna polymerase', *Cell*, 104(6), pp. 901–912. Available at: https://doi.org/10.1016/s0092-8674(01)00286-0.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) 'Mitochondrial DNA and human evolution', *Nature*, 325(6099), pp. 31–36. Available at: https://doi.org/10.1038/325031a0.

Cao, G., Wang, Q., Huang, W., Tong, J., Ye, D., He, Y., Liu, Z., Tang, X., Cheng, H., Wen, Q., Li, D., Chau, H.-T., Wen, Y., Zhong, H., Meng, Z., Liu, H., Wu, Z., Zhao, L., Flavell, R.A., Zhou, H., Xu, A., Yang, H. and Yin, Z. (2017) 'Long-term consumption of caffeine-free high sucrose cola beverages aggravates the pathogenesis of EAE in mice', *Cell Discovery*, 3(1), p. 17020. Available at: https://doi.org/10.1038/celldisc.2017.20.

Castillo-Guzman, D. and Chédin, F. (2021) 'Defining R-loop classes and their contributions to genome instability', *DNA repair*, 106, p. 103182. Available at: https://doi.org/10.1016/j.dnarep.2021.103182.

Cesaro, A., Hoffman, S.C., Das, P. and de la Fuente-Nunez, C. (2025) 'Challenges and applications of artificial intelligence in infectious diseases and antimicrobial resistance', *npj Antimicrobials and Resistance*, 3(1), p. 2. Available at: https://doi.org/10.1038/s44259-024-00068-x.

Chain, E., Florey, H.W., Gardner, A.D., Heatley, N.G., Jennings, M.A., Orr-Ewing, J. and Sanders, A.G. (1940) 'PENICILLIN AS A CHEMOTHERAPEUTIC AGENT', *The Lancet*, 236(6104), pp. 226–228. Available at: https://doi.org/10.1016/S0140-6736(01)08728-1.

Chavez, D.A., Greer, B.H. and Eichman, B.F. (2018) 'The HIRAN domain of helicase-like transcription factor positions the DNA translocase motor to drive efficient DNA fork regression', *The Journal of Biological Chemistry*, 293(22), pp. 8484–8494. Available at: https://doi.org/10.1074/jbc.RA118.002905.

Chen, S., Cao, Z., Prettner, K., Kuhn, M., Yang, J., Jiao, L., Wang, Z., Li, W., Geldsetzer, P., Bärnighausen, T., Bloom, D.E. and Wang, C. (2023) 'Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050', *JAMA Oncology*, 9(4), pp. 465–472. Available at: https://doi.org/10.1001/jamaoncol.2022.7826.

Chodavarapu, S., Jones, A.D., Feig, M. and Kaguni, J.M. (2016) 'DnaC traps DnaB as an open ring and remodels the domain that binds primase', *Nucleic Acids Research*, 44(1), pp. 210–220. Available at: https://doi.org/10.1093/nar/gkv961.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K. and Lander, E.S. (2007) 'Distinguishing protein-coding and noncoding genes in the human genome', *Proceedings of the National Academy of Sciences*, 104(49), pp. 19428–19433. Available at: https://doi.org/10.1073/pnas.0709013104.

Clancy, C.J. and Nguyen, M.H. (2020) 'Buying Time: The AMR Action Fund and the State of Antibiotic Development in the United States 2020', *Open Forum Infectious Diseases*, 7(11), p. ofaa464. Available at: https://doi.org/10.1093/ofid/ofaa464.

Cooper, R.M. (1985) 'Saccharin?Of Risk and Democracy', *Food, Drug, Cosmetic Law Journal*, 40(1), pp. 34–65.

Cooper, S. and Helmstetter, C.E. (1968) 'Chromosome replication and the division cycle of Escherichia coli B/r', *Journal of Molecular Biology*, 31(3), pp. 519–540. Available at: https://doi.org/10.1016/0022-2836(68)90425-7.

Courcelle, J., Khodursky, A., Peter, B., Brown, P.O. and Hanawalt, P.C. (2001) 'Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli', *Genetics*, 158(1), pp. 41–64. Available at: https://doi.org/10.1093/genetics/158.1.41.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J. and Marians, K.J. (2000) 'The importance of repairing stalled replication forks', *Nature*, 404(6773), pp. 37–41. Available at: https://doi.org/10.1038/35003501.

Dahm, R. (2005) 'Friedrich Miescher and the discovery of DNA', *Developmental Biology*, 278(2), pp. 274–288. Available at: https://doi.org/10.1016/j.ydbio.2004.11.028.

Dahm, R. (2008) 'Discovering DNA: Friedrich Miescher and the early years of nucleic acid research', *Human Genetics*, 122(6), pp. 565–581. Available at: https://doi.org/10.1007/s00439-007-0433-0.

Darby, E.M., Trampari, E., Siasat, P., Gaya, M.S., Alav, I., Webber, M.A. and Blair, J.M.A. (2023) 'Molecular mechanisms of antibiotic resistance revisited', *Nature Reviews Microbiology*, 21(5), pp. 280–295. Available at: https://doi.org/10.1038/s41579-022-00820-y.

Datsenko, K.A. and Wanner, B.L. (2000) 'One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6640–6645. Available at: https://doi.org/10.1073/pnas.120163297.

Davey, M.J., Fang, L., McInerney, P., Georgescu, R.E. and O'Donnell, M. (2002) 'The DnaC helicase loader is a dual ATP/ADP switch protein', *The EMBO Journal*, 21(12), pp. 3148–3159. Available at: https://doi.org/10.1093/emboj/cdf308.

Davies, J.E. (1997) 'Origins, acquisition and dissemination of antibiotic resistance determinants', *Ciba Foundation Symposium*, 207, pp. 15–27; discussion 27-35.

De La Fuente-Nunez, C., Cesaro, A. and Hancock, R.E.W. (2023) 'Antibiotic failure: Beyond antimicrobial resistance', *Drug Resistance Updates*, 71, p. 101012. Available at: https://doi.org/10.1016/j.drup.2023.101012.

DePamphilis, M.L. (1998) 'Initiation of DNA replication in eukaryotic chromosomes', *Journal of Cellular Biochemistry*, 72 Suppl 30–31(S30-31), pp. 8–17. Available at: https://doi.org/10.1002/(SICI)1097-4644(1998)72:30/31+%253C8::AID-JCB3%253E3.0.CO;2-R.

DePamphilis, M.L., Blow, J.J., Ghosh, S., Saha, T., Noguchi, K. and Vassilev, A. (2006) 'Regulating the licensing of DNA replication origins in metazoa', *Current Opinion in Cell Biology*, 18(3), pp. 231–239. Available at: https://doi.org/10.1016/j.ceb.2006.04.001.

Dewar, J.M. and Walter, J.C. (2017) 'Mechanisms of DNA replication termination', *Nature reviews. Molecular cell biology*, 18(8), pp. 507–516. Available at: https://doi.org/10.1038/nrm.2017.42.

Díaz-Talavera, A., Montero-Conde, C., Leandro-García, L.J. and Robledo, M. (2022) 'PrimPol: A Breakthrough among DNA Replication Enzymes and a Potential New Target for Cancer Therapy', *Biomolecules*, 12(2), p. 248. Available at: https://doi.org/10.3390/biom12020248.

Dickey, T.H., Altschuler, S.E. and Wuttke, D.S. (2013) 'Single-Stranded DNA-Binding Proteins: Multiple Domains for Multiple Functions', *Structure (London, England : 1993)*, 21(7), p. 10.1016/j.str.2013.05.013. Available at: https://doi.org/10.1016/j.str.2013.05.013.

Dimude, J.U., Midgley-Smith, S.L. and Rudolph, C.J. (2018) 'Replication-transcription conflicts trigger extensive DNA degradation in *Escherichia coli* cells lacking RecBCD', *DNA Repair*, 70, pp. 37–48. Available at: https://doi.org/10.1016/j.dnarep.2018.08.002.

Dimude, J.U., Midgley-Smith, S.L., Stein, M. and Rudolph, C.J. (2016) 'Replication Termination: Containing Fork Fusion-Mediated Pathologies in Escherichia coli', *Genes*, 7(8), p. 40. Available at: https://doi.org/10.3390/genes7080040.

Dimude, J.U., Stein, M., Andrzejewska, E.E., Khalifa, M.S., Gajdosova, A., Retkute, R., Skovgaard, O. and Rudolph, C.J. (2018) 'Origins Left, Right, and Centre: Increasing the Number of Initiation Sites in the Escherichia coli Chromosome', *Genes*, 9(8), p. 376. Available at: https://doi.org/10.3390/genes9080376.

Dimude, J.U., Stockum, A., Midgley-Smith, S.L., Upton, A.L., Foster, H.A., Khan, A., Saunders, N.J., Retkute, R. and Rudolph, C.J. (2015) 'The Consequences of Replicating in the Wrong Orientation: Bacterial Chromosome Duplication without an Active Replication Origin', *mBio*, 6(6), p. 10.1128/mbio.01294-15. Available at: https://doi.org/10.1128/mbio.01294-15.

Dinos, G.P., Athanassopoulos, C.M., Missiri, D.A., Giannopoulou, P.C., Vlachogiannis, I.A., Papadopoulos, G.E., Papaioannou, D. and Kalpaxis, D.L. (2016) 'Chloramphenicol Derivatives as Antibacterial and Anticancer Agents: Historic Problems and Current Solutions', *Antibiotics (Basel, Switzerland)*, 5(2), p. 20. Available at: https://doi.org/10.3390/antibiotics5020020.

de Dios, R., Gadar, K., Proctor, C.R., Maslova, E., Han, J., Soliman, M.A.N., Krawiel, D., Dunbar, E.L., Singh, B., Peros, S., Killelea, T., Warnke, A.-L., Haugland, M.M., Bolt, E.L., Lentz, C.S., Rudolph, C.J. and McCarthy, R.R. (2025) 'Saccharin disrupts bacterial cell envelope stability and interferes with DNA replication dynamics', *EMBO Molecular Medicine*, 17(5), pp. 993–1017. Available at: https://doi.org/10.1038/s44321-025-00219-1.

Drake, J.W. (1969) 'Comparative rates of spontaneous mutation', *Nature*, 221(5186), p. 1132. Available at: https://doi.org/10.1038/2211132a0.

Drake, J.W. (1991) 'A constant rate of spontaneous mutation in DNA-based microbes', *Proceedings of the National Academy of Sciences of the United States of*

America, 88(16), pp. 7160–7164. Available at: https://doi.org/10.1073/pnas.88.16.7160.

Drummond, D.A. and Wilke, C.O. (2009) 'The evolutionary consequences of erroneous protein synthesis', *Nature reviews. Genetics*, 10(10), pp. 715–724. Available at: https://doi.org/10.1038/nrg2662.

Du, Q., Zhang, D., Zhuang, Y., Xia, Q., Wen, T. and Jia, H. (2021) 'The Molecular Genetics of Marfan Syndrome', *International Journal of Medical Sciences*, 18(13), pp. 2752–2766. Available at: https://doi.org/10.7150/ijms.60685.

Duckworth, A.T., Ducos, P.L., McMillan, S.D., Satyshur, K.A., Blumenthal, K.H., Deorio, H.R., Larson, J.A., Sandler, S.J., Grant, T. and Keck, J.L. (2023) 'Replication fork binding triggers structural changes in the PriA helicase that govern DNA replication restart in E. coli', *Nature Communications*, 14(1), p. 2725. Available at: https://doi.org/10.1038/s41467-023-38144-x.

Duderstadt, K.E. and Berger, J.M. (2008) 'AAA+ ATPases in the Initiation of DNA Replication', *Critical Reviews in Biochemistry and Molecular Biology*, 43(3), pp. 163–187. Available at: https://doi.org/10.1080/10409230802058296.

Duggin, I.G. and Bell, S.D. (2009) 'Termination Structures in the *Escherichia coli* Chromosome Replication Fork Trap', *Journal of Molecular Biology*, 387(3), pp. 532–539. Available at: https://doi.org/10.1016/j.jmb.2009.02.027.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E. and Nudler, E. (2011) 'Linking RNA polymerase backtracking to genome instability in E. coli', *Cell*, 146(4), pp. 533–543. Available at: https://doi.org/10.1016/j.cell.2011.07.034.

Dykhuizen, D. (2005) 'Species Numbers in Bacteria', *Proceedings. California Academy of Sciences*, 56(6 Suppl 1), pp. 62–71.

Eglenen-Polat, B., Kowash, R.R., Huang, H.-C., Siteni, S., Zhu, M., Chen, K., Bender, M.E., Mender, I., Stastny, V., Drapkin, B.J., Raj, P., Minna, J.D., Xu, L., Shay, J.W. and Akbay, E.A. (2024) 'A telomere-targeting drug depletes cancer initiating cells and promotes anti-tumor immunity in small cell lung cancer', *Nature Communications*, 15(1), p. 672. Available at: https://doi.org/10.1038/s41467-024-44861-8.

Ehrlich, J., Bartz, Q.R., Smith, R.M., Joslyn, D.A. and Burkholder, P.R. (1947) 'Chloromycetin, a New Antibiotic From a Soil Actinomycete', *Science*, 106(2757), pp. 417–417. Available at: https://doi.org/10.1126/science.106.2757.417.

Erzberger, J.P., Pirruccello, M.M. and Berger, J.M. (2002) 'The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation',

The EMBO Journal, 21(18), pp. 4763–4773. Available at: https://doi.org/10.1093/emboj/cdf496.

Escherich, T. (1988) 'The intestinal bacteria of the neonate and breast-fed infant. 1884', *Reviews of Infectious Diseases*, 10(6), pp. 1220–1225. Available at: https://doi.org/10.1093/clinids/10.6.1220.

Espeli, O., Lee, C. and Marians, K.J. (2003) 'A physical and functional interaction between Escherichia coli FtsK and topoisomerase IV', *The Journal of Biological Chemistry*, 278(45), pp. 44639–44644. Available at: https://doi.org/10.1074/jbc.M308926200.

Falagas, M.E., Fragoulis, K.N. and Karydis, I. (2006) 'A Comparative Study on the Cost of New Antibiotics and Drugs of Other Therapeutic Categories', *PLOS ONE*, 1(1), p. e11. Available at: https://doi.org/10.1371/journal.pone.0000011.

Fares, J., Fares, M.Y., Khachfe, H.H., Salhab, H.A. and Fares, Y. (2020) 'Molecular principles of metastasis: a hallmark of cancer revisited', *Signal Transduction and Targeted Therapy*, 5(1), pp. 1–17. Available at: https://doi.org/10.1038/s41392-020-0134-x.

Fijalkowska, I.J., Schaaper, R.M. and Jonczyk, P. (2012) 'DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair', *FEMS microbiology reviews*, 36(6), pp. 1105–1121. Available at: https://doi.org/10.1111/j.1574-6976.2012.00338.x.

FLEMING, A. (1944) 'THE DISCOVERY OF PENICILLIN', *British Medical Bulletin*, 2(1), pp. 4–5. Available at: https://doi.org/10.1093/oxfordjournals.bmb.a071032.

Fragkos, M., Ganier, O., Coulombe, P. and Méchali, M. (2015) 'DNA replication origin activation in space and time', *Nature Reviews Molecular Cell Biology*, 16(6), pp. 360–374. Available at: https://doi.org/10.1038/nrm4002.

Francis, D.M. and Page, R. (2010) 'Strategies to Optimize Protein Expression in E. coli', *Current Protocols in Protein Science*, 61(1), pp. 5241–52429. Available at: https://doi.org/10.1002/0471140864.ps0524s61.

Franklin, R.E. and Gosling, R.G. (1953) 'Molecular Configuration in Sodium Thymonucleate', *Nature*, 171(4356), pp. 740–741. Available at: https://doi.org/10.1038/171740a0.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F.,

Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. and Venter, J.C. (1995) 'The minimal gene complement of Mycoplasma genitalium', *Science (New York, N.Y.)*, 270(5235), pp. 397–403. Available at: https://doi.org/10.1126/science.270.5235.397.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M. and Dryja, T.P. (1986) 'A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma', *Nature*, 323(6089), pp. 643–646. Available at: https://doi.org/10.1038/323643a0.

Fu, Y., Lv, Z., Kong, D., Fan, Y. and Dong, B. (2022) 'High abundance of CDC45 inhibits cell proliferation through elevation of HSPA6', *Cell Proliferation*, 55(7), p. e13257. Available at: https://doi.org/10.1111/cpr.13257.

Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T. and Yokoyama, S. (2003) 'Structural basis of replication origin recognition by the DnaA protein', *Nucleic Acids Research*, 31(8), pp. 2077–2086. Available at: https://doi.org/10.1093/nar/gkg309.

Fujiyama, S., Abe, Y., Shiroishi, M., Ikeda, Y. and Ueda, T. (2019) 'Insight into the interaction between PriB and DnaT on bacterial DNA replication restart: Significance of the residues on PriB dimer interface and highly acidic region on DnaT', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1867(4), pp. 367–375. Available at: https://doi.org/10.1016/j.bbapap.2019.01.008.

Fuller, R.S., Funnell, B.E. and Kornberg, A. (1984) 'The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites', *Cell*, 38(3), pp. 889–900. Available at: https://doi.org/10.1016/0092-8674(84)90284-8.

Gabbai, C.B. and Marians, K.J. (2010) 'Recruitment to Stalled Replication Forks of the PriA DNA Helicase and Replisome-loading Activities is Essential for Survival', *DNA repair*, 9(3), pp. 202–209. Available at: https://doi.org/10.1016/j.dnarep.2009.12.009.

Gaillard, H. and Aguilera, A. (2016) 'Transcription as a Threat to Genome Integrity', *Annual Review of Biochemistry*, 85, pp. 291–317. Available at: https://doi.org/10.1146/annurev-biochem-060815-014908.

Garcia-Diaz, M. and Bebenek, K. (2007) 'Multiple functions of DNA polymerases', *Critical Reviews in Plant Sciences*, 26(2), pp. 105–122. Available at: https://doi.org/10.1080/07352680701252817.

Gerstberger, S., Jiang, Q. and Ganesh, K. (2023) 'Metastasis', *Cell*, 186(8), pp. 1564–1579. Available at: https://doi.org/10.1016/j.cell.2023.03.003.

Giraldo, R. (2003) 'Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives', *FEMS Microbiology Reviews*, 26(5), pp. 533–554. Available at: https://doi.org/10.1111/j.1574-6976.2003.tb00629.x.

Goodall, D.J., Jameson, K.H., Hawkins, M. and Rudolph, C.J. (2021) 'A Fork Trap in the Chromosomal Termination Area Is Highly Conserved across All Escherichia coli Phylogenetic Groups', *International Journal of Molecular Sciences*, 22(15), p. 7928. Available at: https://doi.org/10.3390/ijms22157928.

Goodner, B.W., Markelz, B.P., Flanagan, M.C., Crowell, C.B., Racette, J.L., Schilling, B.A., Halfon, L.M., Mellors, J.S. and Grabowski, G. (1999) 'Combined Genetic and Physical Map of the Complex Genome of Agrobacterium tumefaciens', *Journal of Bacteriology*, 181(17), pp. 5160–5166.

Greider, C.W. and Blackburn, E.H. (1985) 'Identification of a specific telomere terminal transferase activity in Tetrahymena extracts', *Cell*, 43(2 Pt 1), pp. 405–413. Available at: https://doi.org/10.1016/0092-8674(85)90170-9.

Grompone, G., Ehrlich, D. and Michel, B. (2004) 'Cells defective for replication restart undergo replication fork reversal', *EMBO Reports*, 5(6), pp. 607–612. Available at: https://doi.org/10.1038/sj.embor.7400167.

Gupta, M.K., Guy, C.P., Yeeles, J.T.P., Atkinson, J., Bell, H., Lloyd, R.G., Marians, K.J. and McGlynn, P. (2013) 'Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 110(18), pp. 7252–7257. Available at: https://doi.org/10.1073/pnas.1303890110.

Gyles, C.L. (2007) 'Shiga toxin-producing Escherichia coli: an overview', *Journal of Animal Science*, 85(13 Suppl), pp. E45-62. Available at: https://doi.org/10.2527/jas.2006-508.

Hall, K. and Sankaran, N. (2021) 'DNA translated: Friedrich Miescher's discovery of nuclein in its original context', *The British Journal for the History of Science*, 54(1), pp. 99–107. Available at: https://doi.org/10.1017/S000708742000062X.

Hamperl, S. and Cimprich, K.A. (2014) 'The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability', *DNA repair*, 19, pp. 84–94. Available at: https://doi.org/10.1016/j.dnarep.2014.03.023.

Hamperl, S. and Cimprich, K.A. (2016) 'Conflict Resolution in the Genome: How Transcription and Replication Make It Work', *Cell*, 167(6), pp. 1455–1467. Available at: https://doi.org/10.1016/j.cell.2016.09.053.

Hanahan, D. (2022) 'Hallmarks of Cancer: New Dimensions', *Cancer Discovery*, 12(1), pp. 31–46. Available at: https://doi.org/10.1158/2159-8290.CD-21-1059.

Hanahan, D. and Weinberg, R.A. (2000) 'The Hallmarks of Cancer', *Cell*, 100(1), pp. 57–70. Available at: https://doi.org/10.1016/S0092-8674(00)81683-9.

Hanahan, D. and Weinberg, R.A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646–674. Available at: https://doi.org/10.1016/j.cell.2011.02.013.

Hartwell, L.H. and Weinert, T.A. (1989) 'Checkpoints: controls that ensure the order of cell cycle events', *Science*, 246(4930), pp. 629–634. Available at: https://doi.org/10.1126/science.2683079.

Hawkins, M., Dimude, J.U., Howard, J.A.L., Smith, A.J., Dillingham, M.S., Savery, N.J., Rudolph, C.J. and McGlynn, P. (2019) 'Direct removal of RNA polymerase barriers to replication by accessory replicative helicases', *Nucleic Acids Research*, 47(10), pp. 5100–5113. Available at: https://doi.org/10.1093/nar/gkz170.

Hayashi, C., Miyazaki, E., Ozaki, S., Abe, Y. and Katayama, T. (2020) 'DnaB helicase is recruited to the replication initiation complex via binding of DnaA domain I to the lateral surface of the DnaB N-terminal domain', *The Journal of Biological Chemistry*, 295(32), pp. 11131–11143. Available at: https://doi.org/10.1074/jbc.RA120.014235.

Hayflick, L. and Moorhead, P.S. (1961) 'The serial cultivation of human diploid cell strains', *Experimental Cell Research*, 25, pp. 585–621. Available at: https://doi.org/10.1016/0014-4827(61)90192-6.

Helgesen, E., Sætre, F. and Skarstad, K. (2021) 'Topoisomerase IV tracks behind the replication fork and the SeqA complex during DNA replication in Escherichia coli', *Scientific Reports*, 11, p. 474. Available at: https://doi.org/10.1038/s41598-020-80043-4.

Heller, R.C. and Marians, K.J. (2005) 'The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart', *Molecular Cell*, 17(5), pp. 733–743. Available at: https://doi.org/10.1016/j.molcel.2005.01.019.

Heller, R.C. and Marians, K.J. (2006) 'Replication fork reactivation downstream of a blocked nascent leading strand', *Nature*, 439(7076), pp. 557–562. Available at: https://doi.org/10.1038/nature04329.

Hernandez, A.J., Lee, S.-J. and Richardson, C.C. (2016) 'Primer release is the rate-limiting event in lagging-strand synthesis mediated by the T7 replisome', *Proceedings of the National Academy of Sciences of the United States of America*, 113(21), pp. 5916–5921. Available at: https://doi.org/10.1073/pnas.1604894113.

Hershey, A.D. and Chase, M. (1952) 'Independent functions of viral protein and nucleic acid in growth of bacteriophage', *The Journal of General Physiology*, 36(1), pp. 39–56. Available at: https://doi.org/10.1085/jgp.36.1.39.

Hetta, H.F., Sirag, N., Elfadil, H., Salama, A., Aljadrawi, S.F., Alfaifi, A.J., Alwabisi, A.N., AbuAlhasan, B.M., Alanazi, L.S., Aljohani, Y.A., Ramadan, Y.N., Abd Ellah, N.H. and Algammal, A.M. (2025) 'Artificial Sweeteners: A Double-Edged Sword for Gut Microbiome', *Diseases*, 13(4), p. 115. Available at: https://doi.org/10.3390/diseases13040115.

Hill, T.M. and Marians, K.J. (1990) 'Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro.', *Proceedings of the National Academy of Sciences*, 87(7), pp. 2481–2485. Available at: https://doi.org/10.1073/pnas.87.7.2481.

Hizume, K. and Araki, H. (2019) 'Replication fork pausing at protein barriers on chromosomes', *FEBS Letters*, 593(13), pp. 1449–1458. Available at: https://doi.org/10.1002/1873-3468.13481.

Ho, C.S., Wong, C.T.H., Aung, T.T., Lakshminarayanan, R., Mehta, J.S., Rauz, S., McNally, A., Kintses, B., Peacock, S.J., Fuente-Nunez, C. de la, Hancock, R.E.W. and Ting, D.S.J. (2025) 'Antimicrobial resistance: a concise update', *The Lancet Microbe*, 6(1). Available at: https://doi.org/10.1016/j.lanmic.2024.07.010.

Hogan, A.M. and Cardona, S.T. (2022) 'Gradients in gene essentiality reshape antibacterial research', *FEMS Microbiology Reviews*, 46(3), p. fuac005. Available at: https://doi.org/10.1093/femsre/fuac005.

Hooper, D.C. and Jacoby, G.A. (2016) 'Topoisomerase Inhibitors: Fluoroquinolone Mechanisms of Action and Resistance', *Cold Spring Harbor Perspectives in Medicine*, 6(9), p. a025320. Available at: https://doi.org/10.1101/cshperspect.a025320.

Hosea, R., Hillary, S., Naqvi, S., Wu, S. and Kasim, V. (2024) 'The two sides of chromosomal instability: drivers and brakes in cancer', *Signal Transduction and Targeted Therapy*, 9(1), p. 75. Available at: https://doi.org/10.1038/s41392-024-01767-7.

Howell, M., Daniel, J.J. and Brown, P.J.B. (2017) 'Live Cell Fluorescence Microscopy to Observe Essential Processes During Microbial Cell Growth', *Journal of Visualized Experiments : JoVE*, (129), p. 56497. Available at: https://doi.org/10.3791/56497.

Hsieh, P. and Yamane, K. (2008) 'DNA mismatch repair: molecular mechanism, cancer, and ageing', *Mechanisms of Ageing and Development*, 129(7–8), pp. 391–407. Available at: https://doi.org/10.1016/j.mad.2008.02.012.

Hu, Y. and Stillman, B. (2023) 'Origins of DNA Replication in Eukaryotes', *Molecular cell*, pp. S1097-2765(22)01203–5. Available at: https://doi.org/10.1016/j.molcel.2022.12.024.

Hu, Z. and Curtis, C. (2020) 'Looking backward in time to define the chronology of metastasis', *Nature Communications*, 11(1), p. 3213. Available at: https://doi.org/10.1038/s41467-020-16995-y.

Huang, C.-Y., Hsu, C.-H., Sun, Y.-J., Wu, H.-N. and Hsiao, C.-D. (2006) 'Complexed crystal structure of replication restart primosome protein PriB reveals a novel single-stranded DNA-binding mode', *Nucleic Acids Research*, 34(14), pp. 3878–3886. Available at: https://doi.org/10.1093/nar/gkl536.

Huang, R. and Zhou, P.-K. (2021) 'DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy', *Signal Transduction and Targeted Therapy*, 6(1), p. 254. Available at: https://doi.org/10.1038/s41392-021-00648-7.

Huang, Y.-H., Lin, M.-J. and Huang, C.-Y. (2013) 'DnaT is a single-stranded DNA binding protein', *Genes to Cells*, 18(11), pp. 1007–1019. Available at: https://doi.org/10.1111/gtc.12095.

Hutchison, C.A., Newbold, J.E., Potter, S.S. and Edgell, M.H. (1974) 'Maternal inheritance of mammalian mitochondrial DNA', *Nature*, 251(5475), pp. 536–538. Available at: https://doi.org/10.1038/251536a0.

Ilves, I., Petojevic, T., Pesavento, J.J. and Botchan, M.R. (2010) 'Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS proteins', *Molecular cell*, 37(2), pp. 247–258. Available at: https://doi.org/10.1016/j.molcel.2009.12.030.

International Human Genome Sequencing Consortium (2004) 'Finishing the euchromatic sequence of the human genome', *Nature*, 431(7011), pp. 931–945. Available at: https://doi.org/10.1038/nature03001.

Inusa, B.P.D., Hsu, L.L., Kohli, N., Patel, A., Ominu-Evbota, K., Anie, K.A. and Atoyebi, W. (2019) 'Sickle Cell Disease—Genetics, Pathophysiology, Clinical Presentation and Treatment', *International Journal of Neonatal Screening*, 5(2), p. 20. Available at: https://doi.org/10.3390/ijns5020020.

Ishikawa, F. and Naito, T. (1999) 'Why do we have linear chromosomes? A matter of Adam and Eve', *Mutation Research/DNA Repair*, 434(2), pp. 99–107. Available at: https://doi.org/10.1016/S0921-8777(99)00017-8.

Ivanova, D., Taylor, T., Smith, S.L., Dimude, J.U., Upton, A.L., Mehrjouy, M.M., Skovgaard, O., Sherratt, D.J., Retkute, R. and Rudolph, C.J. (2015a) 'Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin', *Nucleic Acids Research*, 43(16), pp. 7865–7877. Available at: https://doi.org/10.1093/nar/gkv704.

Ivanova, D., Taylor, T., Smith, S.L., Dimude, J.U., Upton, A.L., Mehrjouy, M.M., Skovgaard, O., Sherratt, D.J., Retkute, R. and Rudolph, C.J. (2015b) 'Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin', *Nucleic Acids Research*, 43(16), pp. 7865–7877. Available at: https://doi.org/10.1093/nar/gkv704.

Jacobs, A.C., Thompson, M.G., Black, C.C., Kessler, J.L., Clark, L.P., McQueary, C.N., Gancz, H.Y., Corey, B.W., Moon, J.K., Si, Y., Owen, M.T., Hallock, J.D., Kwak, Y.I., Summers, A., Li, C.Z., Rasko, D.A., Penwell, W.F., Honnold, C.L., Wise, M.C., Waterman, P.E., Lesho, E.P., Stewart, R.L., Actis, L.A., Palys, T.J., Craft, D.W. and Zurawski, D.V. (2014) 'AB5075, a Highly Virulent Isolate of Acinetobacter baumannii, as a Model Strain for the Evaluation of Pathogenesis and Antimicrobial Treatments', *mBio*, 5(3), pp. e01076-01014. Available at: https://doi.org/10.1128/mBio.01076-14.

Jameson, K.H., Rudolph, C.J. and Hawkins, M. (2021) 'Termination of DNA replication at Tus-ter barriers results in under-replication of template DNA', *Journal of Biological Chemistry*, 297(6). Available at: https://doi.org/10.1016/j.jbc.2021.101409.

Japaridze, A., Gogou, C., Kerssemakers, J.W.J., Nguyen, H.M. and Dekker, C. (2020) 'Direct observation of independently moving replisomes in Escherichia coli', *Nature Communications*, 11(1), p. 3109. Available at: https://doi.org/10.1038/s41467-020-16946-7.

Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2001) 'Crystal Structure of the Processivity Clamp Loader Gamma (γ) Complex of *E. coli* DNA Polymerase III', *Cell*, 106(4), pp. 429–441. Available at: https://doi.org/10.1016/S0092-8674(01)00463-9.

Joseph, A., Cointe, A., Mariani Kurkdjian, P., Rafat, C. and Hertig, A. (2020) 'Shiga Toxin-Associated Hemolytic Uremic Syndrome: A Narrative Review', *Toxins*, 12(2), p. 67. Available at: https://doi.org/10.3390/toxins12020067.

Jurcevic Zidar, B., Knezovic, Z., Pribisalic, A., Luetic, S., Jurcic, K., Knezovic, N. and Sutlovic, D. (2025) 'Consumer Perceptions of Artificial Sweeteners in Food Products, Consumption Frequency, and Body Mass Index: A Multivariate Analysis', *Nutrients*, 17(5), p. 814. Available at: https://doi.org/10.3390/nu17050814.

Kaguni, J.M. (2011) 'Replication initiation at the Escherichia coli chromosomal origin', *Current opinion in chemical biology*, 15(5), pp. 606–613. Available at: https://doi.org/10.1016/j.cbpa.2011.07.016.

Kaguni, J.M. and Kornberg, A. (1984) 'Replication initiated at the origin (oriC) of the E. coli chromosome reconstituted with purified enzymes', *Cell*, 38(1), pp. 183–190. Available at: https://doi.org/10.1016/0092-8674(84)90539-7.

Kaper, J.B., Nataro, J.P. and Mobley, H.L.T. (2004) 'Pathogenic Escherichia coli', *Nature Reviews Microbiology*, 2(2), pp. 123–140. Available at: https://doi.org/10.1038/nrmicro818.

Kaplan, D.L. (2006) 'Replication Termination: Mechanism of Polar Arrest Revealed', *Current Biology*, 16(17), pp. R684–R686. Available at: https://doi.org/10.1016/j.cub.2006.08.013.

Katayama, T., Ozaki, S., Keyamura, K. and Fujimitsu, K. (2010) 'Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC', *Nature Reviews Microbiology*, 8(3), pp. 163–170. Available at: https://doi.org/10.1038/nrmicro2314.

Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S. and Suzuki, H. (1990) 'New topoisomerase essential for chromosome segregation in E. coli', *Cell*, 63(2), pp. 393–404. Available at: https://doi.org/10.1016/0092-8674(90)90172-b.

Kciuk, M., Marciniak, B., Mojzych, M. and Kontek, R. (2020) 'Focus on UV-Induced DNA Damage and Repair—Disease Relevance and Protective Strategies', *International Journal of Molecular Sciences*, 21(19), p. 7264. Available at: https://doi.org/10.3390/ijms21197264.

Kelland, L.R. (2005) 'Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics--current status and future prospects', *European Journal of Cancer (Oxford, England: 1990)*, 41(7), pp. 971–979. Available at: https://doi.org/10.1016/j.ejca.2004.11.024.

Khatri, G.S., MacAllister, T., Sista, P.R. and Bastia, D. (1989) 'The replication terminator protein of E. coli is a DNA sequence-specific contra-helicase', *Cell*, 59(4), pp. 667–674. Available at: https://doi.org/10.1016/0092-8674(89)90012-3.

Kile, A.C., Chavez, D.A., Bacal, J., Eldirany, S., Korzhnev, D.M., Bezsonova, I., Eichman, B.F. and Cimprich, K.A. (2015) 'HLTF's Ancient HIRAN Domain Binds 3′ DNA Ends to Drive Replication Fork Reversal', *Molecular Cell*, 58(6), pp. 1090–1100. Available at: https://doi.org/10.1016/j.molcel.2015.05.013.

Killelea, T., Dimude, J.U., He, L., Stewart, A.L., Kemm, F.E., Radovčić, M., Ivančić-Baće, I., Rudolph, C.J. and Bolt, E.L. (2023) 'Cas1-Cas2 physically and functionally interacts with DnaK to modulate CRISPR Adaptation', *Nucleic Acids Research*, 51(13), pp. 6914–6926. Available at: https://doi.org/10.1093/nar/gkad473.

Kleinerman, R.A., Schonfeld, S.J. and Tucker, M.A. (2012) 'Sarcomas in hereditary retinoblastoma', *Clinical Sarcoma Research*, 2(1), p. 15. Available at: https://doi.org/10.1186/2045-3329-2-15.

Kohanski, M.A., Dwyer, D.J. and Collins, J.J. (2010) 'How antibiotics kill bacteria: from targets to networks', *Nature reviews. Microbiology*, 8(6), pp. 423–435. Available at: https://doi.org/10.1038/nrmicro2333.

Koo, M.M., Swann, R., McPhail, S., Abel, G.A., Elliss-Brookes, L., Rubin, G.P. and Lyratzopoulos, G. (2020) 'Presenting symptoms of cancer and stage at diagnosis: evidence from a cross-sectional, population-based study', *The Lancet Oncology*, 21(1), pp. 73–79. Available at: https://doi.org/10.1016/S1470-2045(19)30595-9.

de Kraker, M.E.A., Stewardson, A.J. and Harbarth, S. (2016) 'Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?', *PLoS Medicine*, 13(11), p. e1002184. Available at: https://doi.org/10.1371/journal.pmed.1002184.

Krishnan, A., Anahtar, M.N., Valeri, J.A., Jin, W., Donghia, N.M., Sieben, L., Luttens, A., Zhang, Y., Modaresi, S.M., Hennes, A., Fromer, J., Bandyopadhyay, P., Chen, J.C., Rehman, D., Desai, R., Edwards, P., Lach, R.S., Aschtgen, M.-S., Gaborieau, M., Gaetani, M., Palace, S.G., Omori, S., Khonde, L., Moroz, Y.S., Blough, B., Jin, C., Loh, E., Grad, Y.H., Saei, A.A., Coley, C.W., Wong, F. and Collins, J.J. (2025) 'A generative deep

learning approach to de novo antibiotic design', *Cell*, 0(0). Available at: https://doi.org/10.1016/j.cell.2025.07.033.

Kuchta, R.D. and Stengel, G. (2010) 'Mechanism and evolution of DNA primases', *Biochimica Et Biophysica Acta*, 1804(5), pp. 1180–1189. Available at: https://doi.org/10.1016/j.bbapap.2009.06.011.

Kunkel, T.A. and Bebenek, K. (2000) 'DNA replication fidelity', *Annual Review of Biochemistry*, 69, pp. 497–529. Available at: https://doi.org/10.1146/annurev.biochem.69.1.497.

Kurth, I. and O'Donnell, M. (2009) 'Replisome Dynamics during Chromosome Duplication', *EcoSal Plus*, 3(2), p. 10.1128/ecosalplus.4.4.2. Available at: https://doi.org/10.1128/ecosalplus.4.4.2.

Kuzminov, A. (1999) 'Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda', *Microbiology and molecular biology reviews: MMBR*, 63(4), pp. 751–813, table of contents. Available at: https://doi.org/10.1128/MMBR.63.4.751-813.1999.

Lalonde, M., Trauner, M., Werner, M. and Hamperl, S. (2021) 'Consequences and Resolution of Transcription–Replication Conflicts', *Life*, 11(7), p. 637. Available at: https://doi.org/10.3390/life11070637.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T.,

Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, C. for G.R., The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Department of Genome Analysis, I. of M.B., GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, Multimegabase Sequencing Center, T.I. for S.B., Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Cold Spring Harbor Laboratory, L.A.H.G.C., GBF—German Research Centre for Biotechnology:, *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):, Scientific management: National Human Genome Research Institute, U.N.I. of H., Stanford Human Genome Center:, University of Washington Genome Center:, Department of Molecular Biology, K.U.S. of M., University of Texas Southwestern Medical Center at Dallas:, Office of Science, U.D. of E., and The Wellcome Trust: (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. Available at: https://doi.org/10.1038/35057062.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E. and Merrikh, H. (2017) 'Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis',

Cell, 170(4), pp. 787-799.e18. Available at: https://doi.org/10.1016/j.cell.2017.07.044.

Lang, K.S. and Merrikh, H. (2018) 'The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria', *Annual Review of Microbiology*, 72, pp. 71–88. Available at: https://doi.org/10.1146/annurev-micro-090817-062514.

Langston, L.D. and O'Donnell, M. (2006) 'DNA replication: keep moving and don't mind the gap', *Molecular Cell*, 23(2), pp. 155–160. Available at: https://doi.org/10.1016/j.molcel.2006.05.034.

Laskar, R.S., Qu, C., Huyghe, J.R., Harrison, T., Hayes, R.B., Cao, Y., Campbell, P.T., Steinfelder, R., Talukdar, F.R., Brenner, H., Ogino, S., Brendt, S., Bishop, D.T., Buchanan, D.D., Chan, A.T., Cotterchio, M., Gruber, S.B., Gsur, A., Guelpen, B. van, Jenkins, M.A., Keku, T.O., Lynch, B.M., Marchand, L.L., Martin, R.M., McCarthy, K., Moreno, V., Pearlman, R., Song, M., Tsilidis, K.K., Vodička, P., Woods, M.O., Wu, K., Hsu, L., Gunter, M.J., Peters, U. and Murphy, N. (2024) 'Genome-wide association studies and Mendelian randomization analyses provide insights into the causes of early-onset colorectal cancer', *Annals of Oncology*, 35(6), pp. 523–536. Available at: https://doi.org/10.1016/j.annonc.2024.02.008.

Lederberg, J. (1952) 'Cell genetics and hereditary symbiosis', *Physiological Reviews*, 32(4), pp. 403–430. Available at: https://doi.org/10.1152/physrev.1952.32.4.403.

Lederberg, J. and Tatum, E.L. (1946) 'Gene Recombination in Escherichia Coli', *Nature*, 158(4016), pp. 558–558. Available at: https://doi.org/10.1038/158558a0.

Lee, S.Y., Kwak, M.J. and Kim, J.J. (2025) 'R-loops: a key driver of inflammatory responses in cancer', *Experimental & Molecular Medicine*, pp. 1–12. Available at: https://doi.org/10.1038/s12276-025-01495-0.

Lenoir, M., Serre, F., Cantin, L. and Ahmed, S.H. (2007) 'Intense sweetness surpasses cocaine reward', *PloS One*, 2(8), p. e698. Available at: https://doi.org/10.1371/journal.pone.0000698.

Leroux, M., Jani, N. and Sandler, S.J. (2017) 'A priA Mutant Expressed in Two Pieces Has Almost Full Activity in Escherichia coli K-12', *Journal of Bacteriology*, 199(17), pp. e00267-17. Available at: https://doi.org/10.1128/JB.00267-17.

Levine, A.J. and Oren, M. (2009) 'The first 30 years of p53: growing ever more complex', *Nature reviews. Cancer*, 9(10), pp. 749–758. Available at: https://doi.org/10.1038/nrc2723.

Li, S., Keenan, J.I., Shaw, I.C. and Frizelle, F.A. (2023) 'Could Microplastics Be a Driver for Early Onset Colorectal Cancer?', *Cancers*, 15(13), p. 3323. Available at: https://doi.org/10.3390/cancers15133323.

Li, Y. and Araki, H. (2013) 'Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication', *Genes to Cells*, 18(4), pp. 266–277. Available at: https://doi.org/10.1111/gtc.12040.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) 'Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland', *New England Journal of Medicine*, 343(2), pp. 78–85. Available at: https://doi.org/10.1056/NEJM200007133430201.

Lim, J.Y., Yoon, J.W. and Hovde, C.J. (2010) 'A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157', *Journal of microbiology and biotechnology*, 20(1), pp. 5–14.

Limas, J.C. and Cook, J.G. (2019) 'Preparation for DNA Replication: The Key to a Successful S phase', *FEBS letters*, 593(20), pp. 2853–2867. Available at: https://doi.org/10.1002/1873-3468.13619.

Lin, D.M., Koskella, B. and Lin, H.C. (2017) 'Phage therapy: An alternative to antibiotics in the age of multi-drug resistance', *World Journal of Gastrointestinal Pharmacology and Therapeutics*, 8(3), pp. 162–173. Available at: https://doi.org/10.4292/wjgpt.v8.i3.162.

Lin, Y.-L. and Pasero, P. (2017) 'Transcription-Replication Conflicts: Orientation Matters', *Cell*, 170(4), pp. 603–604. Available at: https://doi.org/10.1016/j.cell.2017.07.040.

Liu, G.-Y., Yu, D., Fan, M.-M., Zhang, X., Jin, Z.-Y., Tang, C. and Liu, X.-F. (2024) 'Antimicrobial resistance crisis: could artificial intelligence be the solution?', *Military Medical Research*, 11(1), p. 7. Available at: https://doi.org/10.1186/s40779-024-00510-1.

Liu, J.-H., Chang, T.-W., Huang, C.-Y., Chen, S.-U., Wu, H.-N., Chang, M.-C. and Hsiao, C.-D. (2004) 'Crystal structure of PriB, a primosomal DNA replication protein of Escherichia coli', *The Journal of Biological Chemistry*, 279(48), pp. 50465–50471. Available at: https://doi.org/10.1074/jbc.M406773200.

Llor, C. and Bjerrum, L. (2014) 'Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem', *Therapeutic Advances in Drug Safety*, 5(6), pp. 229–241. Available at: https://doi.org/10.1177/2042098614554919.

Loeb, L.A. and Kunkel, T.A. (1982) 'Fidelity of DNA Synthesis', *Annual Review of Biochemistry*, 51(Volume 51, 1982), pp. 429–457. Available at: https://doi.org/10.1146/annurev.bi.51.070182.002241.

Loeb, L.A., Loeb, K.R. and Anderson, J.P. (2003) 'Multiple mutations and cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), pp. 776–781. Available at: https://doi.org/10.1073/pnas.0334858100.

Ma, X., Nan, F., Liang, H., Shu, P., Fan, X., Song, X., Hou, Y. and Zhang, D. (2022) 'Excessive intake of sugar: An accomplice of inflammation', *Frontiers in Immunology*, 13, p. 988481. Available at: https://doi.org/10.3389/fimmu.2022.988481.

Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M.R. and Cebrat, S. (2004) 'Where does bacterial replication start? Rules for predicting the oriC region', *Nucleic Acids Research*, 32(13), pp. 3781–3791. Available at: https://doi.org/10.1093/nar/gkh699.

Makowska-Grzyska, M. and Kaguni, J.M. (2010) 'Primase Directs the Release of DnaC from DnaB', *Molecular cell*, 37(1), p. 90. Available at: https://doi.org/10.1016/j.molcel.2009.12.031.

Mancuso, G., Midiri, A., Gerace, E. and Biondo, C. (2021) 'Bacterial Antibiotic Resistance: The Most Critical Pathogens', *Pathogens*, 10(10), p. 1310. Available at: https://doi.org/10.3390/pathogens10101310.

Mani, K., Deng, D., Lin, C., Wang, M., Hsu, M.L. and Zaorsky, N.G. (2024) 'Causes of death among people living with metastatic cancer', *Nature Communications*, 15(1), p. 1519. Available at: https://doi.org/10.1038/s41467-024-45307-x.

Manyi-Loh, C., Mamphweli, S., Meyer, E. and Okoh, A. (2018) 'Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications', *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, 23(4), p. 795. Available at: https://doi.org/10.3390/molecules23040795.

Marians, K.J. (2000) 'PriA-directed replication fork restart in Escherichia coli', *Trends in Biochemical Sciences*, 25(4), pp. 185–189. Available at: https://doi.org/10.1016/s0968-0004(00)01565-6.

Marians, K.J. (2004) 'Mechanisms of replication fork restart in Escherichia coli.', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1441), pp. 71–77. Available at: https://doi.org/10.1098/rstb.2003.1366.

Markus, V. (2024) 'Artificial sweetener-induced dysbiosis and associated molecular signatures', *Biochemical and Biophysical Research Communications*, 735, p. 150798. Available at: https://doi.org/10.1016/j.bbrc.2024.150798.

Masai, H. and Arai, K. (2000) 'Regulation of DNA Replication During the Cell Cycle: Roles of Cdc7 Kinase and Coupling of Replication, Recombination, and Repair in Response to Replication Fork Arrest', *IUBMB Life*, 49(5), pp. 353–364. Available at: https://doi.org/10.1080/152165400410191.

de Massy, B., Patte, J., Louarn, J.M. and Bouché, J.P. (1984) 'oriX: a new replication origin in E. coli', *Cell*, 36(1), pp. 221–227. Available at: https://doi.org/10.1016/0092-8674(84)90092-8.

Mazumder, A. and Kapanidis, A.N. (2019) 'Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms', *Journal of Molecular Biology*, 431(20), pp. 3947–3959. Available at: https://doi.org/10.1016/j.jmb.2019.04.046.

McCutcheon, J.P. (2016) 'From microbiology to cell biology: when an intracellular bacterium becomes part of its host cell', *Current opinion in cell biology*, 41, pp. 132–136. Available at: https://doi.org/10.1016/j.ceb.2016.05.008.

McGlynn, P. and Guy, C.P. (2008a) 'Replication Forks Blocked by Protein–DNA Complexes Have Limited Stability *In Vitro*', *Journal of Molecular Biology*, 381(2), pp. 249–255. Available at: https://doi.org/10.1016/j.jmb.2008.05.053.

McGlynn, P. and Guy, C.P. (2008b) 'Replication forks blocked by protein-DNA complexes have limited stability in vitro', *Journal of Molecular Biology*, 381(2), pp. 249–255. Available at: https://doi.org/10.1016/j.jmb.2008.05.053.

McGlynn, P. and Lloyd, R.G. (2002) 'Recombinational repair and restart of damaged replication forks', *Nature Reviews. Molecular Cell Biology*, 3(11), pp. 859–870. Available at: https://doi.org/10.1038/nrm951.

McGlynn, P., Savery, N.J. and Dillingham, M.S. (2012) 'The conflict between DNA replication and transcription', *Molecular Microbiology*, 85(1), pp. 12–20. Available at: https://doi.org/10.1111/j.1365-2958.2012.08102.x.

McKenzie, A.M., Henry, C., Myers, K.S., Place, M.M. and Keck, J.L. (2022) 'Identification of genetic interactions with priB links the PriA/PriB DNA

replication restart pathway to double-strand DNA break repair in Escherichia coli', *G3: Genes|Genomes|Genetics*, 12(12), p. jkac295. Available at: https://doi.org/10.1093/g3journal/jkac295.

McPhail, S., Johnson, S., Greenberg, D., Peake, M. and Rous, B. (2015) 'Stage at diagnosis and early mortality from cancer in England', *British Journal of Cancer*, 112(1), pp. S108–S115. Available at: https://doi.org/10.1038/bjc.2015.49.

Menger, K.E., Chapman, J., Díaz-Maldonado, H., Khazeem, M.M., Deen, D., Erdinc, D., Casement, J.W., Di Leo, V., Pyle, A., Rodríguez-Luis, A., Cowell, I.G., Falkenberg, M., Austin, C.A. and Nicholls, T.J. (2022) 'Two type I topoisomerases maintain DNA topology in human mitochondria', *Nucleic Acids Research*, 50(19), pp. 11154–11174. Available at: https://doi.org/10.1093/nar/gkac857.

Merrikh, H., Zhang, Y., Grossman, A.D. and Wang, J.D. (2012) 'Replication-transcription conflicts in bacteria', *Nature Reviews. Microbiology*, 10(7), pp. 449–458. Available at: https://doi.org/10.1038/nrmicro2800.

Meselson, M. and Stahl, F.W. (1958) 'THE REPLICATION OF DNA IN ESCHERICHIA COLI*', *Proceedings of the National Academy of Sciences of the United States of America*, 44(7), pp. 671–682. Available at: https://doi.org/10.1073/pnas.44.7.671.

Messer, W., Blaesing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tüngler, D., Wegrzyn, G., Weigel, C., Welzeck, M. and Zakrzewska-Czerwinska, J. (1999) 'Functional domains of DnaA proteins', *Biochimie*, 81(8), pp. 819–825. Available at: https://doi.org/10.1016/S0300-9084(99)00215-1.

Michel, B. and Sandler, S.J. (2017a) 'Replication Restart in Bacteria', *Journal of Bacteriology*, 199(13), p. 10.1128/jb.00102-17. Available at: https://doi.org/10.1128/jb.00102-17.

Michel, B. and Sandler, S.J. (2017b) 'Replication Restart in Bacteria', *Journal of Bacteriology*, 199(13), pp. e00102-17. Available at: https://doi.org/10.1128/JB.00102-17.

Michel, B., Sinha, A.K. and Leach, D.R.F. (2018) 'Replication Fork Breakage and Restart in Escherichia coli', *Microbiology and Molecular Biology Reviews : MMBR*, 82(3), pp. e00013-18. Available at: https://doi.org/10.1128/MMBR.00013-18.

Midgley-Smith, S.L., Dimude, J.U. and Rudolph, C.J. (2019) 'A role for 3′ exonucleases at the final stages of chromosome duplication in Escherichia coli', *Nucleic Acids Research*, 47(4), pp. 1847–1860. Available at: https://doi.org/10.1093/nar/gky1253.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M. and Ding, W. (1994) 'A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1', *Science (New York, N.Y.)*, 266(5182), pp. 66–71. Available at: https://doi.org/10.1126/science.7545954.

Miller, K.D., Fidler-Benaoudia, M., Keegan, T.H., Hipp, H.S., Jemal, A. and Siegel, R.L. (2020) 'Cancer statistics for adolescents and young adults, 2020', *CA: A Cancer Journal for Clinicians*, 70(6), pp. 443–459. Available at: https://doi.org/10.3322/caac.21637.

Milne, R.L. and Antoniou, A.C. (2011) 'Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers', *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 22 Suppl 1, pp. i11-17. Available at: https://doi.org/10.1093/annonc/mdq660.

Mirkin, E.V. and Mirkin, S.M. (2005a) 'Mechanisms of transcription-replication collisions in bacteria', *Molecular and Cellular Biology*, 25(3), pp. 888–895. Available at: https://doi.org/10.1128/MCB.25.3.888-895.2005.

Mirkin, E.V. and Mirkin, S.M. (2005b) 'Mechanisms of transcription-replication collisions in bacteria', *Molecular and Cellular Biology*, 25(3), pp. 888–895. Available at: https://doi.org/10.1128/MCB.25.3.888-895.2005.

Mirkin, E.V. and Mirkin, S.M. (2007) 'Replication Fork Stalling at Natural Impediments', *Microbiology and Molecular Biology Reviews*, 71(1), pp. 13–35. Available at: https://doi.org/10.1128/mmbr.00030-06.

Moolman, M.C., Tiruvadi Krishnan, S., Kerssemakers, J.W.J., de Leeuw, R., Lorent, V., Sherratt, D.J. and Dekker, N.H. (2016) 'The progression of replication forks at natural replication barriers in live bacteria', *Nucleic Acids Research*, 44(13), pp. 6262–6273. Available at: https://doi.org/10.1093/nar/gkw397.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) 'A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes', *Proceedings of the National Academy of Sciences of the United States of America*, 85(18), pp. 6622–6626. Available at: https://doi.org/10.1073/pnas.85.18.6622.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K. and Bastia, D. (2001) 'Mechanism of termination of DNA replication of Escherichia coli involves helicase–contrahelicase interaction', *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), pp. 9569–9574. Available at: https://doi.org/10.1073/pnas.171065898.

Munita, J.M. and Arias, C.A. (2016) 'Mechanisms of Antibiotic Resistance', *Microbiology spectrum*, 4(2), p. 10.1128/microbiolspec.VMBF-0016–2015. Available at: https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

Murray, C.J.L., Ikuta, K.S., Sharara, F., Swetschinski, L., Aguilar, G.R., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Hamadani, B.H.K., Kumaran, E.A.P., McManigal, B., Achalapong, S., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Babin, F.-X., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Berkley, J.A., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Donatien, R.C., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., Luca, M.D., Dokova, K., Dramowski, A., Dunachie, S.J., Bich, T.D., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garcia, C., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Hsia, Y., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Jenney, A.W.J., Khorana, M., Khusuwan, S., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Lim, K., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Manoharan, A., Marks, F., May, J., Mayxay, M., Mturi, N., Munera-Huertas, T., Musicha, P., Musila, L.A., Mussi-Pinhata, M.M., Naidu, R.N., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Ochoa, T.J., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Ounchanum, P., Pak, G.D., Paredes, J.L., Peleg, A.Y., Perrone, C., Phe, T., Phommasone, K., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Rattanavong, S., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rosenthal, V.D., Rudd, K.E., Russell, N., Sader, H.S., Saengchan, W., Schnall, J., Scott, J.A.G., Seekaew, S., Sharland, M., Shivamallappa, M., Sifuentes-Osornio, J., Simpson, A.J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Tigoi, C., Turner, C., Turner, P., Doorn, H.R. van, Velaphi, S., Vongpradith, A., Vongsouvath, M., Vu, H., Walsh, T., Walson, J.L., Waner, S., Wangrangsimakul, T., Wannapinij, P., Wozniak, T., Sharma, T.E.M.W.Y., Yu, K.C., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C. and Naghavi, M. (2022) 'Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis', *The Lancet*, 399(10325), pp. 629–655. Available at: https://doi.org/10.1016/S0140-6736(21)02724-0.

Nagao, I. (1953) 'Comparative study of roseomycin, streptomycin, chloramphenicol and homosulfanilamide on the effect of experimental cholera infections', *The Tohoku Journal of Experimental Medicine*, 58(2), pp. 191–194. Available at: https://doi.org/10.1620/tjem.58.191.

Nass, M.M.K. and Nass, S. (1963) 'INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS : I. Fixation and Electron Staining Reactions', *Journal of Cell Biology*, 19(3), pp. 593–611. Available at: https://doi.org/10.1083/jcb.19.3.593.

Nathan, C. and Cars, O. (2014) 'Antibiotic Resistance — Problems, Progress, and Prospects', *New England Journal of Medicine*, 371(19), pp. 1761–1763. Available at: https://doi.org/10.1056/NEJMp1408040.

Negrini, S., Gorgoulis, V.G. and Halazonetis, T.D. (2010) 'Genomic instability — an evolving hallmark of cancer', *Nature Reviews Molecular Cell Biology*, 11(3), pp. 220–228. Available at: https://doi.org/10.1038/nrm2858.

Neylon, C., Kralicek, A.V., Hill, T.M. and Dixon, N.E. (2005) 'Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex', *Microbiology and molecular biology reviews: MMBR*, 69(3), pp. 501–526. Available at: https://doi.org/10.1128/MMBR.69.3.501-526.2005.

Niccum, B.A., Lee, H., MohammedIsmail, W., Tang, H. and Foster, P.L. (2018) 'The Spectrum of Replication Errors in the Absence of Error Correction Assayed Across the Whole Genome of Escherichia coli', *Genetics*, 209(4), pp. 1043–1054. Available at: https://doi.org/10.1534/genetics.117.300515.

Niedzwiedz, C.L., Knifton, L., Robb, K.A., Katikireddi, S.V. and Smith, D.J. (2019) 'Depression and anxiety among people living with and beyond cancer: a growing clinical and research priority', *BMC cancer*, 19(1), p. 943. Available at: https://doi.org/10.1186/s12885-019-6181-4.

Nishida, N., Yano, H., Nishida, T., Kamura, T. and Kojiro, M. (2006) 'Angiogenesis in cancer', *Vascular Health and Risk Management*, 2(3), pp. 213–219. Available at: https://doi.org/10.2147/vhrm.2006.2.3.213.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets,

A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H. and Phillippy, A.M. (2022) 'The complete sequence of a human genome', *Science (New York, N.Y.)*, 376(6588), pp. 44–53. Available at: https://doi.org/10.1126/science.abj6987.

Nurse, P., DiGate, R.J., Zavitz, K.H. and Marians, K.J. (1990) 'Molecular cloning and DNA sequence analysis of Escherichia coli priA, the gene encoding the primosomal protein replication factor Y', *Proceedings of the National Academy of Sciences of the United States of America*, 87(12), pp. 4615–4619. Available at: https://doi.org/10.1073/pnas.87.12.4615.

O'Donnell, M., Langston, L. and Stillman, B. (2013) 'Principles and concepts of DNA replication in bacteria, archaea, and eukarya', *Cold Spring Harbor Perspectives in Biology*, 5(7), p. a010108. Available at: https://doi.org/10.1101/cshperspect.a010108.

Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 mutations in human cancers: origins, consequences, and clinical use', *Cold Spring Harbor Perspectives in Biology*, 2(1), p. a001008. Available at: https://doi.org/10.1101/cshperspect.a001008.

O'Loughlin, E.V. and Robins-Browne, R.M. (2001) 'Effect of Shiga toxin and Shiga-like toxins on eukaryotic cells', *Microbes and Infection*, 3(6), pp. 493–507. Available at: https://doi.org/10.1016/S1286-4579(01)01405-8.

Ornitz, D.M. and Legeai-Mallet, L. (2017) 'Achondroplasia: Development, pathogenesis, and therapy', *Developmental Dynamics*, 246(4), pp. 291–309. Available at: https://doi.org/10.1002/dvdy.24479.

Oser, B.L. (1985) 'Highlights in the history of saccharin toxicology', *Food and Chemical Toxicology*, 23(4), pp. 535–542. Available at: https://doi.org/10.1016/0278-6915(85)90148-6.

Parekh, V.J., Węgrzyn, G., Arluison, V. and Sinden, R.R. (2023) 'Genomic Instability of G-Quadruplex Sequences in Escherichia coli: Roles of DinG, RecG, and RecQ Helicases', *Genes*, 14(9), p. 1720. Available at: https://doi.org/10.3390/genes14091720.

Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C. and Göker, M. (2020) 'List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ', *International Journal of Systematic and Evolutionary Microbiology*, 70(11), pp. 5607–5612. Available at: https://doi.org/10.1099/ijsem.0.004332.

Paton, J.C. and Paton, A.W. (1998) 'Pathogenesis and Diagnosis of Shiga Toxin-Producing Escherichia coli Infections', *Clinical Microbiology Reviews*, 11(3), pp. 450–479. Available at: https://doi.org/10.1128/cmr.11.3.450.

Payne, A.N., Chassard, C. and Lacroix, C. (2012) 'Gut microbial adaptation to dietary consumption of fructose, artificial sweeteners and sugar alcohols: implications for host–microbe interactions contributing to obesity', *Obesity Reviews*, 13(9), pp. 799–809. Available at: https://doi.org/10.1111/j.1467-789X.2012.01009.x.

Payne, B.T.I., van Knippenberg, I.C., Bell, H., Filipe, S.R., Sherratt, D.J. and McGlynn, P. (2006a) 'Replication fork blockage by transcription factor-DNA complexes in Escherichia coli', *Nucleic Acids Research*, 34(18), pp. 5194–5202. Available at: https://doi.org/10.1093/nar/gkl682.

Payne, B.T.I., van Knippenberg, I.C., Bell, H., Filipe, S.R., Sherratt, D.J. and McGlynn, P. (2006b) 'Replication fork blockage by transcription factor-DNA complexes in Escherichia coli', *Nucleic Acids Research*, 34(18), pp. 5194–5202. Available at: https://doi.org/10.1093/nar/gkl682.

Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M. and Vitale, L. (2019) 'On the length, weight and GC content of the human genome', *BMC Research Notes*, 12(1), p. 106. Available at: https://doi.org/10.1186/s13104-019-4137-z.

Pires, D., Melo, L., Vilas Boas, D., Sillankorva, S. and Azeredo, J. (2017) 'Phage therapy as an alternative or complementary strategy to prevent and control biofilm-related infections', *Current Opinion in Microbiology*, 39, pp. 48–56. Available at: https://doi.org/10.1016/j.mib.2017.09.004.

Pommier, Y., Nussenzweig, A., Takeda, S. and Austin, C. (2022) 'Human topoisomerases and their roles in genome stability and organization', *Nature Reviews Molecular Cell Biology*, 23(6), pp. 407–427. Available at: https://doi.org/10.1038/s41580-022-00452-3.

Pommier, Y., Sun, Y., Huang, S.N. and Nitiss, J.L. (2016) 'Roles of eukaryotic topoisomerases in transcription, replication and genomic stability', *Nature reviews. Molecular cell biology*, 17(11), pp. 703–721. Available at: https://doi.org/10.1038/nrm.2016.111.

Poon, A., Davis, B.H. and Chao, L. (2005) 'The Coupon Collector and the Suppressor Mutation', *Genetics*, 170(3), pp. 1323–1332. Available at: https://doi.org/10.1534/genetics.104.037259.

Possoz, C., Filipe, S.R., Grainge, I. and Sherratt, D.J. (2006) 'Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo', *The EMBO Journal*, 25(11), pp. 2596–2604. Available at: https://doi.org/10.1038/sj.emboj.7601155.

Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D. and Cozzarelli, N.R. (2001) 'Topological challenges to DNA replication: Conformations at the fork', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8219–8226. Available at: https://doi.org/10.1073/pnas.111006998.

Prescott, D.M. and Kuempel, P.L. (1972) 'Bidirectional Replication of the Chromosome in Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*, 69(10), pp. 2842–2845.

Prestinaci, F., Pezzotti, P. and Pantosti, A. (2015) 'Antimicrobial resistance: a global multifaceted phenomenon', *Pathogens and Global Health*, 109(7), p. 309. Available at: https://doi.org/10.1179/2047773215Y.0000000030.

Price, R. (2016) 'O'Neill report on antimicrobial resistance: funding for antimicrobial specialists should be improved', *European Journal of Hospital Pharmacy*, 23(4), pp. 245–247. Available at: https://doi.org/10.1136/ejhpharm-2016-001013.

Projan, S.J. (2003) 'Why is big Pharma getting out of antibacterial drug discovery?', *Current Opinion in Microbiology*, 6(5), pp. 427–430. Available at: https://doi.org/10.1016/j.mib.2003.08.003.

Rechkoblit, O., Gupta, Y.K., Malik, R., Rajashankar, K.R., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2016) 'Structure and mechanism of human PrimPol, a DNA polymerase with primase activity', *Science Advances*, 2(10), p. e1601317. Available at: https://doi.org/10.1126/sciadv.1601317.

Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P.C., Garcillán-Barcia, M.P. and de la Cruz, F. (2020) 'Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids', *Nature Communications*, 11(1), p. 3602. Available at: https://doi.org/10.1038/s41467-020-17278-2.

Renwick, A.G. (1985) 'The disposition of saccharin in animals and man—A review', *Food and Chemical Toxicology*, 23(4), pp. 429–435. Available at: https://doi.org/10.1016/0278-6915(85)90136-X.

Renwick, A.G. (2006) 'The intake of intense sweeteners - an update review', *Food Additives and Contaminants*, 23(4), pp. 327–338. Available at: https://doi.org/10.1080/02652030500442532.

Reuber, M.D. (1978) 'Carcinogenicity of saccharin.', *Environmental Health Perspectives*, 25, pp. 173–200.

Reyes-Lamothe, R., Sherratt, D.J. and Leake, M.C. (2010) 'Stoichiometry and architecture of active DNA replication machinery in Escherichia coli', *Science (New York, N.Y.)*, 328(5977), pp. 498–501. Available at: https://doi.org/10.1126/science.1185757.

Reyes-Lamothe, R., Wang, X. and Sherratt, D. (2008) 'Escherichia coli and its chromosome', *Trends in Microbiology*, 16(5), pp. 238–245. Available at: https://doi.org/10.1016/j.tim.2008.02.003.

Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A. and Cohen, M.L. (1983) 'Hemorrhagic colitis associated with a rare Escherichia coli serotype', *The New England Journal of Medicine*, 308(12), pp. 681–685. Available at: https://doi.org/10.1056/NEJM198303243081203.

Rocha, E.P.C. (2004) 'The replication-related organization of bacterial genomes', *Microbiology (Reading, England)*, 150(Pt 6), pp. 1609–1627. Available at: https://doi.org/10.1099/mic.0.26974-0.

Rocha, E.P.C. (2018) 'Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics', *Molecular Biology and Evolution*, 35(6), pp. 1338–1347. Available at: https://doi.org/10.1093/molbev/msy078.

Romero Romero, M.L., Landerer, C., Poehls, J. and Toth-Petroczy, A. (2022) 'Phenotypic mutations contribute to protein diversity and shape protein evolution', *Protein Science : A Publication of the Protein Society*, 31(9), p. e4397. Available at: https://doi.org/10.1002/pro.4397.

Rosselló-Mora, R. and Amann, R. (2001) 'The species concept for prokaryotes', *FEMS Microbiology Reviews*, 25(1), pp. 39–67. Available at: https://doi.org/10.1111/j.1574-6976.2001.tb00571.x.

Rozwandowicz, M., Brouwer, M.S.M., Fischer, J., Wagenaar, J.A., Gonzalez-Zorn, B., Guerra, B., Mevius, D.J. and Hordijk, J. (2018) 'Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae', *Journal of Antimicrobial Chemotherapy*, 73(5), pp. 1121–1137. Available at: https://doi.org/10.1093/jac/dkx488.

Rudolph, C.J., Corocher, T.-A., Grainge, I. and Duggin, I.G. (2019) 'Termination of DNA Replication in Prokaryotes', *eLS*. John Wiley & Sons, Ltd, pp. 1–15. Available at: https://doi.org/10.1002/9780470015902.a0001056.pub3.

Rudolph, C.J., Dhillon, P., Moore, T. and Lloyd, R.G. (2007) 'Avoiding and resolving conflicts between DNA replication and transcription', *DNA repair*, 6(7), pp. 981–993. Available at: https://doi.org/10.1016/j.dnarep.2007.02.017.

Rudolph, C.J., Upton, A.L., Briggs, G.S. and Lloyd, R.G. (2010) 'Is RecG a general guardian of the bacterial genome?', *DNA Repair*, 9(3), pp. 210–223. Available at: https://doi.org/10.1016/j.dnarep.2009.12.014.

Rudolph, C.J., Upton, A.L. and Lloyd, R.G. (2009) 'Replication fork collisions cause pathological chromosomal amplification in cells lacking RecG DNA translocase', *Molecular Microbiology*, 74(4), pp. 940–955. Available at: https://doi.org/10.1111/j.1365-2958.2009.06909.x.

Rudolph, C.J., Upton, A.L., Stockum, A., Nieduszynski, C.A. and Lloyd, R.G. (2013) 'Avoiding chromosome pathology when replication forks collide', *Nature*, 500(7464), p. 10.1038/nature12312. Available at: https://doi.org/10.1038/nature12312.

Ruiz, N. and Silhavy, T.J. (2022) 'How Escherichia coli Became the Flagship Bacterium of Molecular Biology', *Journal of Bacteriology*. Edited by G. O'Toole, 204(9), pp. e00230-22. Available at: https://doi.org/10.1128/jb.00230-22.

Sakiyama, Y., Nagata, M., Yoshida, R., Kasho, K., Ozaki, S. and Katayama, T. (2022) 'Concerted actions of DnaA complexes with DNA-unwinding sequences within and flanking replication origin oriC promote DnaB helicase loading', *The Journal of Biological Chemistry*, 298(6), p. 102051. Available at: https://doi.org/10.1016/j.jbc.2022.102051.

Salam, Md.A., Al-Amin, Md.Y., Salam, M.T., Pawar, J.S., Akhter, N., Rabaan, A.A. and Alqumber, M.A.A. (2023) 'Antimicrobial Resistance: A Growing Serious Threat for Global Public Health', *Healthcare*, 11(13), p. 1946. Available at: https://doi.org/10.3390/healthcare11131946.

Sandler, S.J. (2000) 'Multiple genetic pathways for restarting DNA replication forks in Escherichia coli K-12', *Genetics*, 155(2), pp. 487–497. Available at: https://doi.org/10.1093/genetics/155.2.487.

Sandler, S.J., Leroux, M., Windgassen, T.A. and Keck, J.L. (2021) 'Escherichia coli K-12 has two distinguishable PriA-PriB replication restart pathways', *Molecular*

Microbiology, 116(4), pp. 1140–1150. Available at: https://doi.org/10.1111/mmi.14802.

Sandler, S.J. and Marians, K.J. (2000) 'Role of PriA in replication fork reactivation in Escherichia coli', *Journal of Bacteriology*, 182(1), pp. 9–13. Available at: https://doi.org/10.1128/JB.182.1.9-13.2000.

Sandler, S.J., McCool, J.D., Do, T.T. and Johansen, R.U. (2001) 'PriA mutations that affect PriA-PriC function during replication restart', *Molecular Microbiology*, 41(3), pp. 697–704. Available at: https://doi.org/10.1046/j.1365-2958.2001.02547.x.

Sandler, S.J., Samra, H.S. and Clark, A.J. (1996) 'Differential suppression of priA2::kan phenotypes in Escherichia coli K-12 by mutations in priA, lexA, and dnaC', *Genetics*, 143(1), pp. 5–13. Available at: https://doi.org/10.1093/genetics/143.1.5.

Sayyed, H.E., Chat, L.L., Lebailly, E., Vickridge, E., Pages, C., Cornet, F., Lagomarsino, M.C. and Espéli, O. (2016) 'Mapping Topoisomerase IV Binding and Activity Sites on the E. coli Genome', *PLOS Genetics*, 12(5), p. e1006025. Available at: https://doi.org/10.1371/journal.pgen.1006025.

Schaaper, R.M. (1993) 'Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli', *The Journal of Biological Chemistry*, 268(32), pp. 23762–23765.

Schatz, A., Bugle, E. and Waksman, S.A. (1944) 'Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*†', *Proceedings of the Society for Experimental Biology and Medicine*, 55(1), pp. 66–69. Available at: https://doi.org/10.3181/00379727-55-14461.

Schekman, R., Weiner, J.H., Weiner, A. and Kornberg, A. (1975) 'Ten proteins required for conversion of phiX174 single-stranded DNA to duplex form in vitro. Resolution and reconstitution', *The Journal of Biological Chemistry*, 250(15), pp. 5859–5865.

Scheuermann, R., Tam, S., Burgers, P.M., Lu, C. and Echols, H. (1983) 'Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication', *Proceedings of the National Academy of Sciences of the United States of America*, 80(23), pp. 7085–7089. Available at: https://doi.org/10.1073/pnas.80.23.7085.

Scheuermann, R.H. and Echols, H. (1984) 'A separate editing exonuclease for DNA replication: the epsilon subunit of Escherichia coli DNA polymerase III

holoenzyme', *Proceedings of the National Academy of Sciences of the United States of America*, 81(24), pp. 7747–7751. Available at: https://doi.org/10.1073/pnas.81.24.7747.

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M.O., Bartels, D., Bekel, T., Beyer, S., Bode, E., Bode, H.B., Bolten, C.J., Choudhuri, J.V., Doss, S., Elnakady, Y.A., Frank, B., Gaigalat, L., Goesmann, A., Groeger, C., Gross, F., Jelsbak, Lars, Jelsbak, Lotte, Kalinowski, J., Kegler, C., Knauber, T., Konietzny, S., Kopp, M., Krause, L., Krug, D., Linke, B., Mahmud, T., Martinez-Arias, R., McHardy, A.C., Merai, M., Meyer, F., Mormann, S., Muñoz-Dorado, J., Perez, J., Pradella, S., Rachid, S., Raddatz, G., Rosenau, F., Rückert, C., Sasse, F., Scharfe, M., Schuster, S.C., Suen, G., Treuner-Lange, A., Velicer, G.J., Vorhölter, F.-J., Weissman, K.J., Welch, R.D., Wenzel, S.C., Whitworth, D.E., Wilhelm, S., Wittmann, C., Blöcker, H., Pühler, A. and Müller, R. (2007) 'Complete genome sequence of the myxobacterium Sorangium cellulosum', *Nature Biotechnology*, 25(11), pp. 1281–1289. Available at: https://doi.org/10.1038/nbt1354.

Schwob, E. (2004) 'Flexibility and governance in eukaryotic DNA replication', *Current Opinion in Microbiology*, 7(6), pp. 680–690. Available at: https://doi.org/10.1016/j.mib.2004.10.017.

Sclafani, R.A. and Holzen, T.M. (2007) 'Cell Cycle Regulation of DNA Replication', *Annual review of genetics*, 41, pp. 237–280. Available at: https://doi.org/10.1146/annurev.genet.41.110306.130308.

Sender, R., Fuchs, S. and Milo, R. (2016) 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLoS biology*, 14(8), p. e1002533. Available at: https://doi.org/10.1371/journal.pbio.1002533.

Serres, M.H., Gopal, S., Nahum, L.A., Liang, P., Gaasterland, T. and Riley, M. (2001) 'A functional update of the Escherichia coli K-12 genome', *Genome Biology*, 2(9), p. RESEARCH0035. Available at: https://doi.org/10.1186/gb-2001-2-9-research0035.

Shariati, A., Noei, M. and Chegini, Z. (2023) 'Bacteriophages: The promising therapeutic approach for enhancing ciprofloxacin efficacy against bacterial infection', *Journal of Clinical Laboratory Analysis*, 37(9–10), p. e24932. Available at: https://doi.org/10.1002/jcla.24932.

Shay, J.W. and Keith, W.N. (2008) 'Targeting telomerase for cancer therapeutics', *British Journal of Cancer*, 98(4), pp. 677–683. Available at: https://doi.org/10.1038/sj.bjc.6604209.

Shen, Z. and Prasanth, S.G. (2012) 'Emerging players in the initiation of eukaryotic DNA replication', *Cell Division*, 7, p. 22. Available at: https://doi.org/10.1186/1747-1028-7-22.

Sherr, C.J. (2004) 'Principles of tumor suppression', *Cell*, 116(2), pp. 235–246. Available at: https://doi.org/10.1016/s0092-8674(03)01075-4.

Shi, Z., Wu, X., Yu, S., Huynh, M., Jena, P.K., Nguyen, M., Wan, Y.-J.Y. and Hwang, S.T. (2020) 'Short-Term Exposure to a Western Diet Induces Psoriasiform Dermatitis by Promoting Accumulation of IL-17A–Producing γδ T Cells', *Journal of Investigative Dermatology*, 140(9), pp. 1815–1823. Available at: https://doi.org/10.1016/j.jid.2020.01.020.

Shintani, M., Sanchez, Z.K. and Kimbara, K. (2015) 'Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy', *Frontiers in Microbiology*, 6, p. 242. Available at: https://doi.org/10.3389/fmicb.2015.00242.

Shioi, S., Ose, T., Maenaka, K., Shiroishi, M., Abe, Y., Kohda, D., Katayama, T. and Ueda, T. (2005) 'Crystal structure of a biologically functional form of PriB from Escherichia coli reveals a potential single-stranded DNA-binding site', *Biochemical and Biophysical Research Communications*, 326(4), pp. 766–776. Available at: https://doi.org/10.1016/j.bbrc.2004.11.104.

Shiomi, D., Mori, H. and Niki, H. (2009) 'Genetic mechanism regulating bacterial cell shape and metabolism', *Communicative & Integrative Biology*, 2(3), pp. 219–220. Available at: https://doi.org/10.4161/cib.2.3.7930.

Shulman, S.T., Friedmann, H.C. and Sims, R.H. (2007) 'Theodor Escherich: The First Pediatric Infectious Diseases Physician?', *Clinical Infectious Diseases*, 45(8), pp. 1025–1029. Available at: https://doi.org/10.1086/521946.

Simonsen, S., Søgaard, C.K., Olsen, J.G., Otterlei, M. and Kragelund, B.B. (2024) 'The bacterial DNA sliding clamp, β-clamp: structure, interactions, dynamics and drug discovery', *Cellular and Molecular Life Sciences: CMLS*, 81(1), p. 245. Available at: https://doi.org/10.1007/s00018-024-05252-w.

Singh, N., Singh Lubana, S., Arora, S. and Sachmechi, I. (2020) 'A Study of Artificial Sweeteners and Thyroid Cancer Risk', *Journal of Clinical Medicine Research*, 12(8), pp. 492–498. Available at: https://doi.org/10.14740/jocmr4258.

Skarstad, K. and Katayama, T. (2013) 'Regulating DNA Replication in Bacteria', *Cold Spring Harbor Perspectives in Biology*, 5(4), pp. a012922–a012922. Available at: https://doi.org/10.1101/cshperspect.a012922.

Smillie, C., Garcillán-Barcia, M.P., Francia, M.V., Rocha, E.P.C. and de la Cruz, F. (2010) 'Mobility of Plasmids', *Microbiology and Molecular Biology Reviews : MMBR*, 74(3), pp. 434–452. Available at: https://doi.org/10.1128/MMBR.00020-10.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A. and Cimprich, K.A. (2014) 'Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability', *Molecular Cell*, 56(6), pp. 777–785. Available at: https://doi.org/10.1016/j.molcel.2014.10.020.

Soultanas, P. (2005) 'The Bacterial Helicase-Primase Interaction: A Common Structural/Functional Module', *Structure (London, England : 1993)*, 13(6), pp. 839–844. Available at: https://doi.org/10.1016/j.str.2005.04.006.

Srinivas, N., Rachakonda, S. and Kumar, R. (2020) 'Telomeres and Telomere Length: A General Overview', *Cancers*, 12(3), p. 558. Available at: https://doi.org/10.3390/cancers12030558.

Srivatsan, A., Tehranchi, A., MacAlpine, D.M. and Wang, J.D. (2010) 'Co-Orientation of Replication and Transcription Preserves Genome Integrity', *PLOS Genetics*, 6(1), p. e1000810. Available at: https://doi.org/10.1371/journal.pgen.1000810.

Stehelin, D., Varmus, H.E., Bishop, J.M. and Vogt, P.K. (1976) 'DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA', *Nature*, 260(5547), pp. 170–173. Available at: https://doi.org/10.1038/260170a0.

Stillman, B. (2015) 'Reconsidering DNA Polymerases at the Replication Fork in Eukaryotes', *Molecular cell*, 59(2), pp. 139–141. Available at: https://doi.org/10.1016/j.molcel.2015.07.004.

Stoneking, M. and Soodyall, H. (1996) 'Human evolution and the mitochondrial genome', *Current Opinion in Genetics & Development*, 6(6), pp. 731–736. Available at: https://doi.org/10.1016/S0959-437X(96)80028-1.

Stoy, H., Zwicky, K., Kuster, D., Lang, K.S., Krietsch, J., Crossley, M.P., Schmid, J.A., Cimprich, K.A., Merrikh, H. and Lopes, M. (2023) 'Direct visualization of transcription-replication conflicts reveals post-replicative DNA:RNA hybrids', *Nature Structural & Molecular Biology*, 30(3), pp. 348–359. Available at: https://doi.org/10.1038/s41594-023-00928-6.

Strathdee, S.A., Hatfull, G.F., Mutalik, V.K. and Schooley, R.T. (2023) 'Phage therapy: From biological mechanisms to future directions', *Cell*, 186(1), pp. 17–31. Available at: https://doi.org/10.1016/j.cell.2022.11.017.

Strycharska, M.S., Arias-Palomo, E., Lyubimov, A.Y., Erzberger, J.P., O'Shea, V., Bustamante, C.J. and Berger, J.M. (2013) 'Nucleotide and partner-protein control of bacterial replicative helicase structure and function', *Molecular cell*, 52(6), pp. 844–854. Available at: https://doi.org/10.1016/j.molcel.2013.11.016.

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C.A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., Kuperman, Y., Harmelin, A., Kolodkin-Gal, I., Shapiro, H., Halpern, Z., Segal, E. and Elinav, E. (2014) 'Artificial sweeteners induce glucose intolerance by altering the gut microbiota', *Nature*, 514(7521), pp. 181–186. Available at: https://doi.org/10.1038/nature13793.

Sun, J. and Kong, D. (2010) 'DNA replication origins, ORC/DNA interaction, and assembly of pre-replication complex in eukaryotes', *Acta Biochimica et Biophysica Sinica*, 42(7), pp. 433–439. Available at: https://doi.org/10.1093/abbs/gmq048.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA: a cancer journal for clinicians*, 71(3), pp. 209–249. Available at: https://doi.org/10.3322/caac.21660.

Surh, Y.-J. (2021) 'The 50-Year War on Cancer Revisited: Should We Continue to Fight the Enemy Within?', *Journal of Cancer Prevention*, 26(4), pp. 219–223. Available at: https://doi.org/10.15430/JCP.2021.26.4.219.

Sutton, M.D. and Kaguni, J.M. (1997) 'The Escherichia coli dnaA gene: four functional domains', *Journal of Molecular Biology*, 274(4), pp. 546–561. Available at: https://doi.org/10.1006/jmbi.1997.1425.

Syeda, A.H., Dimude, J.U., Skovgaard, O. and Rudolph, C.J. (2020) 'Too Much of a Good Thing: How Ectopic DNA Replication Affects Bacterial Replication Dynamics', *Frontiers in Microbiology*, 11. Available at: https://doi.org/10.3389/fmicb.2020.00534.

Szostak, J.W. and Blackburn, E.H. (1982) 'Cloning yeast telomeres on linear plasmid vectors', *Cell*, 29(1), pp. 245–255. Available at: https://doi.org/10.1016/0092-8674(82)90109-x.

Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S. and Hirota, Y. (1983) 'The 245 base-pair oriC sequence of the E. coli chromosome directs bidirectional replication at an adjacent region.', *Nucleic Acids Research*, 11(9), pp. 2617–2626.

Tahmasebi, H., Arjmand, N., Monemi, M., Babaeizad, A., Alibabaei, F., Alibabaei, N., Bahar, A., Oksenych, V. and Eslami, M. (2025) 'From Cure to Crisis: Understanding

the Evolution of Antibiotic-Resistant Bacteria in Human Microbiota', *Biomolecules*, 15(1), p. 93. Available at: https://doi.org/10.3390/biom15010093.

Takahashi, D.T., Gadelle, D., Agama, K., Kiselev, E., Zhang, H., Yab, E., Petrella, S., Forterre, P., Pommier, Y. and Mayer, C. (2022) 'Topoisomerase I (TOP1) dynamics: conformational transition from open to closed states', *Nature Communications*, 13(1), p. 59. Available at: https://doi.org/10.1038/s41467-021-27686-7.

Tenaillon, O., Skurnik, D., Picard, B. and Denamur, E. (2010) 'The population genetics of commensal Escherichia coli', *Nature Reviews Microbiology*, 8(3), pp. 207–217. Available at: https://doi.org/10.1038/nrmicro2298.

Tomasetti, C., Li, L. and Vogelstein, B. (2017) 'Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention', *Science (New York, N.Y.)*, 355(6331), pp. 1330–1334. Available at: https://doi.org/10.1126/science.aaf9011.

Tougu, K. and Marians, K.J. (1996) 'The Interaction between Helicase and Primase Sets the Replication Fork Clock*', *Journal of Biological Chemistry*, 271(35), pp. 21398–21405. Available at: https://doi.org/10.1074/jbc.271.35.21398.

Towse, A., Hoyle, C.K., Goodall, J., Hirsch, M., Mestre-Ferrandiz, J. and Rex, J.H. (2017) 'Time for a change in how new antibiotics are reimbursed: Development of an insurance framework for funding new antibiotics based on a policy of risk mitigation', *Health Policy (Amsterdam, Netherlands)*, 121(10), pp. 1025–1030. Available at: https://doi.org/10.1016/j.healthpol.2017.07.011.

Trautinger, B.W. and Lloyd, R.G. (2002) 'Modulation of DNA repair by mutations flanking the DNA channel through RNA polymerase', *The EMBO Journal*, 21(24), pp. 6944–6953. Available at: https://doi.org/10.1093/emboj/cdf654.

Uddin, T.M., Chakraborty, A.J., Khusro, A., Zidan, B.R.M., Mitra, S., Emran, T.B., Dhama, K., Ripon, Md.K.H., Gajdács, M., Sahibzada, M.U.K., Hossain, Md.J. and Koirala, N. (2021) 'Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects', *Journal of Infection and Public Health*, 14(12), pp. 1750–1766. Available at: https://doi.org/10.1016/j.jiph.2021.10.020.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan,

M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Deslattes Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351. Available at: https://doi.org/10.1126/science.1058040.

Ventola, C.L. (2015) 'The Antibiotic Resistance Crisis', *Pharmacy and Therapeutics*, 40(4), pp. 277–283.

Verma, S.C., Qian, Z. and Adhya, S.L. (2019) 'Architecture of the Escherichia coli nucleoid', *PLOS Genetics*, 15(12), p. e1008456. Available at: https://doi.org/10.1371/journal.pgen.1008456.

de Visser, K.E. and Joyce, J.A. (2023) 'The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth', *Cancer Cell*, 41(3), pp. 374–403. Available at: https://doi.org/10.1016/j.ccell.2023.02.016.

Volff, J.N. and Altenbuchner, J. (2000) 'A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution', *FEMS microbiology letters*, 186(2), pp. 143–150. Available at: https://doi.org/10.1111/j.1574-6968.2000.tb09095.x.

Vousden, K.H. and Prives, C. (2009) 'Blinded by the Light: The Growing Complexity of p53', *Cell*, 137(3), pp. 413–431. Available at: https://doi.org/10.1016/j.cell.2009.04.037.

Wadolkowski, E.A., Sung, L.M., Burris, J.A., Samuel, J.E. and O'Brien, A.D. (1990) 'Acute renal tubular necrosis and death of mice orally infected with Escherichia coli strains that produce Shiga-like toxin type II', *Infection and Immunity*, 58(12), pp. 3959–3965. Available at: https://doi.org/10.1128/iai.58.12.3959-3965.1990.

Wang, J.D., Berkmen, M.B. and Grossman, A.D. (2007) 'Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp. 5608–5613. Available at: https://doi.org/10.1073/pnas.0608999104.

Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G. and Sherratt, D.J. (2011) 'Replication and segregation of an Escherichia coli chromosome with two replication origins', *Proceedings of the National Academy of Sciences*, 108(26), pp. E243–E250. Available at: https://doi.org/10.1073/pnas.1100874108.

Wang, X., Llopis, P.M. and Rudner, D.Z. (2013) 'Organization and segregation of bacterial chromosomes', *Nature reviews. Genetics*, 14(3), p. 10.1038/nrg3375. Available at: https://doi.org/10.1038/nrg3375.

Wang, Z. (2021) 'Regulation of Cell Cycle Progression by Growth Factor-Induced Cell Signaling', *Cells*, 10(12), p. 3327. Available at: https://doi.org/10.3390/cells10123327.

Warner, D.J. (2008) 'Ira Remsen, Saccharin, and the Linear Model', *Ambix*, 55(1), pp. 50–61. Available at: https://doi.org/10.1179/174582308X255415.

Watson, J.D. and Crick, F.H. (1953) 'Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid', *Nature*, 171(4356), pp. 737–738. Available at: https://doi.org/10.1038/171737a0.

Waxler, R.P. (2016) 'Moving Into Stage IV: Pancreatic Cancer', *Journal of Pain & Palliative Care Pharmacotherapy*, 30(4), pp. 308–320. Available at: https://doi.org/10.1080/15360288.2016.1231741.

Wells, J.G., Davis, B.R., Wachsmuth, I.K., Riley, L.W., Remis, R.S., Sokolow, R. and Morris, G.K. (1983) 'Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare Escherichia coli serotype', *Journal of Clinical Microbiology*, 18(3), pp. 512–520. Available at: https://doi.org/10.1128/jcm.18.3.512-520.1983.

Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A. and Berger, J.M. (2012) 'The Structure of DNA-Bound Human Topoisomerase II Alpha: Conformational Mechanisms for Coordinating Inter-Subunit Interactions with DNA Cleavage', *Journal of Molecular Biology*, 424(3), pp. 109–124. Available at: https://doi.org/10.1016/j.jmb.2012.07.014.

Wessel, S.R., Cornilescu, C.C., Cornilescu, G., Metz, A., Leroux, M., Hu, K., Sandler, S.J., Markley, J.L. and Keck, J.L. (2016) 'Structure and Function of the PriC DNA Replication Restart Protein', *The Journal of Biological Chemistry*, 291(35), pp. 18384–18396. Available at: https://doi.org/10.1074/jbc.M116.738781.

Wessel, S.R., Marceau, A.H., Massoni, S.C., Zhou, R., Ha, T., Sandler, S.J. and Keck, J.L. (2013) 'PriC-mediated DNA Replication Restart Requires PriC Complex Formation with the Single-stranded DNA-binding Protein *', *Journal of Biological Chemistry*, 288(24), pp. 17569–17578. Available at: https://doi.org/10.1074/jbc.M113.478156.

Windgassen, T.A., Wessel, S.R., Bhattacharyya, B. and Keck, J.L. (2018) 'Mechanisms of bacterial DNA replication restart', *Nucleic Acids Research*, 46(2), pp. 504–519. Available at: https://doi.org/10.1093/nar/gkx1203.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) 'Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya', *Proceedings of the National Academy of Sciences of the United States of America*, 87(12), pp. 4576–4579. Available at: https://doi.org/10.1073/pnas.87.12.4576.

Woodford, N. and Ellington, M.J. (2007) 'The emergence of antibiotic resistance by mutation', *Clinical Microbiology and Infection*, 13(1), pp. 5–18. Available at: https://doi.org/10.1111/j.1469-0691.2006.01492.x.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C. and Micklem, G. (1995) 'Identification of the breast cancer

susceptibility gene BRCA2', *Nature*, 378(6559), pp. 789–792. Available at: https://doi.org/10.1038/378789a0.

Wu, W., Hickson, I.D. and Liu, Y. (2020) 'The prevention and resolution of DNA replication–transcription conflicts in eukaryotic cells', *Genome Instability & Disease*, 1(3), pp. 114–128. Available at: https://doi.org/10.1007/s42764-020-00012-z.

Wu, Z., Xia, F. and Lin, R. (2024) 'Global burden of cancer and associated risk factors in 204 countries and territories, 1980–2021: a systematic analysis for the GBD 2021', *Journal of Hematology & Oncology*, 17(1), p. 119. Available at: https://doi.org/10.1186/s13045-024-01640-8.

Yao, N.Y. and O'Donnell, M. (2012) 'The RFC Clamp Loader: Structure and Function', *Sub-cellular biochemistry*, 62, pp. 259–279. Available at: https://doi.org/10.1007/978-94-007-4572-8_14.

Yao, Y. and Dai, W. (2014) 'Genomic Instability and Cancer', *Journal of carcinogenesis & mutagenesis*, 5, p. 1000165. Available at: https://doi.org/10.4172/2157-2518.1000165.

Zatyka, M. and Thomas, C.M. (1998) 'Control of genes for conjugative transfer of plasmids and other mobile elements', *FEMS Microbiology Reviews*, 21(4), pp. 291–319. Available at: https://doi.org/10.1111/j.1574-6976.1998.tb00355.x.

Zavitz, K.H. and Marians, K.J. (1992) 'ATPase-deficient mutants of the Escherichia coli DNA replication protein PriA are capable of catalyzing the assembly of active primosomes', *The Journal of Biological Chemistry*, 267(10), pp. 6933–6940.

Zeman, M.K. and Cimprich, K.A. (2014) 'Causes and consequences of replication stress', *Nature Cell Biology*, 16(1), pp. 2–9. Available at: https://doi.org/10.1038/ncb2897.

Zheng, L. and Shen, B. (2011) 'Okazaki fragment maturation: nucleases take centre stage', *Journal of Molecular Cell Biology*, 3(1), pp. 23–30. Available at: https://doi.org/10.1093/jmcb/mjq048.

Appendix

Appendix A: Python Scripts for Data Processing (Chapter 3)

The following scripts were developed by **Dr. Lewis Frame (University of York)** and utilized for the automated processing and statistical analysis of *E. coli* microscopy data as described in Section 2.7.

A.1: Cell Segmentation Pipeline (CpnResNeXt)

```
import matplotlib.pyplot as plt
import cv2
import numpy as np
print(np.version.version)
import os
import torch
import celldetection as cd
from scipy import ndimage
import nd2
import shutil
from skimage import measure, morphology

#from colicoords import Data, Cell, CellPlot
import csv

os.environ["KMP_DUPLICATE_LIB_OK"] = "TRUE"
os.environ["OMP_NUM_THREADS"] = "1"
os.environ["MKL_NUM_THREADS"] = "1"
```

```
torch.set_num_threads(1)
cv2.setNumThreads(0)
directory = "."
model = None
device = 'cuda' if torch.cuda.is_available() else 'cpu'
print(device)

model = cd.fetch_model('ginoro_CpnResNeXt101UNet-fbe875f1a3e5ce2c',
check_hash=True,progress=True).to(device)
#print(model)
model.eval()

for img_file in os.listdir("./Images"):
   if img_file.split(".")[-1] == "nd2":
      #img_path = 'priA300 lacO34 0_1 ara IPTG 0 min011.nd2'
      img_path = img_file
      print(f'Processing: {img_path}')
      folder_name = f'./{img_path[:-4]}'

      try:
         os.makedirs(folder_name)
      except:
         shutil.rmtree(folder_name)
         os.makedirs(folder_name)

      outfile = open(f'{folder_name}/results.tsv','w',newline='')
      fieldnames = ["cell_num", "eccentricity","length_skl","area_ski","a","edge"]
```

```
    writer = csv.DictWriter(outfile, fieldnames=fieldnames, delimiter='\t')
    writer.writeheader()

    input_image = nd2.imread(f'./Images/{img_path}')
#       if input_image.ndim > 0:
#         input_image = input_image[0,:,:]
    min_val = np.min(input_image)
    max_val = np.max(input_image)
    img_8bit_1channel = np.uint8(((input_image - min_val) / (max_val - min_val))
* 255)
    print(input_image.shape)

    img_8bit_3channel = cv2.cvtColor(img_8bit_1channel,
cv2.COLOR_GRAY2RGB)


    with torch.no_grad():
      x = cd.to_tensor(img_8bit_3channel, transpose=True, device=device,
dtype=torch.float32)
      x = x / 255  # ensure 0..1 range
      x = x[None]  # add batch dimension: Tensor[3, h, w] -> Tensor[1, 3, h, w]
      y = model(x)

    print(y)
    contours = y['contours'][0]
    np.save(f'{folder_name}/celldetection_output_contours.npy',
cd.asnumpy(contours))
    cd.imshow_row(x, x, figsize=(16, 9), titles=('input', 'contours'))
```

```
    cd.plot_contours(contours,contour_line_width=1)
    plt.savefig(f'{folder_name}/celldetection_output_img.png',dpi=300)
    plt.close()
    print(contours.shape)

    full_mask = np.zeros(input_image.shape,dtype=np.uint8)

    for index,contour in enumerate(contours):
        np_contour = cd.asnumpy(contour)
        #print(np_contour.shape)
        mask = np.zeros(input_image.shape, dtype=np.uint8)
        np_contour = np_contour.reshape(-1, 1, 2).astype(np.int32) # Reshape and
ensure int32 type
        mask = cv2.drawContours(mask, [np_contour], -1, color=255,
thickness=cv2.FILLED)
        mask = mask.astype(np.uint8)

        #plt.imshow(mask,cmap='binary_r')
        #plt.show()

        label_im, nb_labels = ndimage.label(mask)
        print(f'Number of cells: {nb_labels}')

        object_features = measure.regionprops(label_im) #,input_image)
        properties = ['area', 'eccentricity', 'perimeter', 'intensity_mean']

        label = 0
        label_i = object_features[label].label
```

```
        im = (label_im == label_i)
        im = np.asarray(im,dtype=np.uint8)

        eccentricity = getattr(object_features[label], 'eccentricity')
        area_ski = getattr(object_features[label], 'area')

        # Find the coordinates of the non-zero pixels in the mask
        rows, cols = np.where(im > 0)
        margin = 10
        # Calculate bounding box with margin
        min_row = max(0, np.min(rows) - margin)
        max_row = min(im.shape[0] - 1, np.max(rows) + margin)
        min_col = max(0, np.min(cols) - margin)
        max_col = min(im.shape[1] - 1, np.max(cols) + margin)
        if max_row == im.shape[0] - 1 or min_row == 0 or max_col == im.shape[1]
or min_col == 0:
            edge = True
        else:
            edge = False
        # Crop the image
        im = im[min_row:max_row + 1, min_col:max_col+1] # +1 for right and
bottom

        masked_indices = im > 0

        plt.imshow(im,cmap='binary_r', interpolation='nearest')
        #plt.show()
        plt.close()
```

```
    skeleton = morphology.skeletonize(im).astype(np.float32)

    y_skl,x_skl = np.where(skeleton > 0)

    try:
       a,b,c = np.polyfit(x_skl, y_skl, 2)
    except Exception as e:
       print(f"An unexpected error occurred: {e}. Skipping to next item.")
       continue
    print(a,b,c)
    # Generate points for the fitted curve
    x_fit = np.linspace(0, im.shape[1], 1000)  # More points for smooth curve
    #x_fit = x_fit[masked_indices]
    y_fit = a * x_fit**2 + b * x_fit + c

    # Apply the mask to the fitted line (more efficient)
    y_fit_masked = []
    x_fit_masked = []
    for i, x_val in enumerate(x_fit):
       #print(x_val)
       y_val = a * (x_val**2) + b * (x_val) + c
       #print(y_val)
       y_val_round = int(round(y_val))
       x_val_round = int(round(x_val))
       if 0 <= y_val_round < im.shape[0] and 0 <= x_val_round < im.shape[1] and
im[y_val_round, x_val_round] > 0:
          y_fit_masked.append(y_val)
```

```
            x_fit_masked.append(x_val)

        x_fit_masked = np.array(x_fit_masked)
        y_fit_masked = np.array(y_fit_masked)

        #length_skl = np.sqrt((max(x_fit_masked)-min(x_fit_masked))**2 +
(max(y_fit_masked)-min(y_fit_masked))**2)
        curve = np.column_stack((x_fit_masked,y_fit_masked))
        length_skl = np.sum(np.sqrt(np.sum((curve[:-1]-curve[1:])**2,axis=1)))

        #print(np.sum(skeleton))

        plt.imshow(im,cmap='binary_r', interpolation='nearest')
        plt.imshow(skeleton, cmap=plt.cm.Reds, alpha=0.5)
        plt.plot(x_fit_masked, y_fit_masked, c='blue', label='Quadratic Fit',
linewidth=2)
        plt.title(f'Area: {round(area_ski,3)}, Eccentricity: {round(eccentricity,3)},
Length: {round(length_skl,3)}, a: {round(a,5)}')
        plt.savefig(f'{folder_name}/cell_{index}_skeleton.png')
        #plt.show()
        plt.close()

        cv2.imwrite(f'{folder_name}/cell_{index}_mask_img.tif',mask)

        stats = {"cell_num":index, "eccentricity":eccentricity, "length_skl":
length_skl, "area_ski": area_ski, "a": a, "edge": edge}
        writer.writerow(stats)
        full_mask = full_mask + mask
```

```
    outfile.close()
    full_mask[full_mask > 255] = 255

cv2.imwrite(f'{folder_name}/CpnResNeXt101UNet_Prediction.tif',full_mask.asty
pe(np.uint8))
```

A.2: Data Processing and Statistical Analysis

```
import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

import os

import seaborn as sns

from scipy.stats import gaussian_kde

import re

from scipy import stats

import warnings


warnings.filterwarnings("ignore")


#No_induction_exp_1 =
pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])


total_data = {}


px_size = 0.065 #um


##################
```

```
dataset_name = "lacO34 ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)
```

```
# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')
```

```
# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
```

```
counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)
```

```
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()




plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()




#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)
```

```
total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



####################


dataset_name = "lacO34 IPTG ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]
```

```
        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)

# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
```

```
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []
```

```
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
plt.close()

plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()
```

```
#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



####################


dataset_name = "lacO34 PriB IPTG ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "priB" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]
```

```
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')
```

```
# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
```

```
axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed
bin_width = bins[1]-bins[0]
kde = gaussian_kde(length_data)
kde_x = np.linspace(min(length_data), max(length_data), 1000)
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')
plt.show()
plt.close()
plt.close()
```

```
sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()
```

```
plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


############################


dataset_name = "lacO34 PriB ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "priB" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]
```

```
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')
```

```
# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
```

```
axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed
bin_width = bins[1]-bins[0]
kde = gaussian_kde(length_data)
kde_x = np.linspace(min(length_data), max(length_data), 1000)
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')
plt.show()
plt.close()
plt.close()
```

```
sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()
```

```
plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


##################


dataset_name = "lacO34 PriC IPTG ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "priC" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]
```

```
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')
```

```
# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
```

```
axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()
```

```
sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()



plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
```

```
plt.show()

plt.close()




#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



######################


dataset_name = "lacO34 PriC ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "priC" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')
```

```
        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')
        # print(df)
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)

# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
```

```
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
```

```
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as
needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')
```

```
#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')
```

```
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



##################


dataset_name = "120_lacO34 ara Exp 1"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp1/'
for folder in os.listdir(folder_path):
    # print(folder)
```

```
    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "arabinose":
        print(f'{dataset_name}: {folder_path+folder}')
        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')
        # print(df)
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]


        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)




# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']
```

```
fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
```

```
axes[1, 0].set_title(f'{dataset_name}: Length distribution')

axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)
```

```
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')
plt.show()
plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation
density = kde(length_data)       # Calculate density at each data point
jitter_x = []
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
```

```
plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)

total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



####################

dataset_name = "120_lacO34 IPTG ara Exp 1"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
```

```
# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]



        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.

# print(final_df)

eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']
```

```
length_data = final_df['length_skl']


fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns


# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')


# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')
```

```
sns.violinplot(length_data, ax=axes[1,0])

axes[1, 0].set_title(f'{dataset_name}: Length distribution')

axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as
needed
```

```
bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



####################
```

```
dataset_name = "120_lacO34 PriB IPTG ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "priB" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)
```

```
# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
```

```
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()
```

```
counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
```

```
plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


############################
```

```
dataset_name = "120_lacO34 PriB ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "priB" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)
```

```
# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
```

```
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()
```

```
counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
```

```
plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


##################
```

```
dataset_name = "120_lacO34 PriC IPTG ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "priC" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)
```

```
# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')
```

```
# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
```

```
counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)
```

```
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()




plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()




#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)
```

```
total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



####################


dataset_name = "120_lacO34 PriC ara Exp 1"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp1/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "priC" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]
```

```
        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)

# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
```

```
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []
```

```
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
plt.close()

plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()
```

```
#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



##################


# Exp2


##################


dataset_name = "lacO34 ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')
```

```
    df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')
    # print(df)
    df = df[df['edge'] == False]
    df = df[df['area_ski'] > 100]
    df = df[df['eccentricity'] > 0.8]
    df = df[df['a'] < 1]
    df = df[df['a'] > -1]

    # print(df)
    final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
```

```
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')
```

```
# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed
bin_width = bins[1]-bins[0]
kde = gaussian_kde(length_data)
kde_x = np.linspace(min(length_data), max(length_data), 1000)
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")
plt.ylabel('Length')
plt.title(f'{dataset_name}')
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()



plt.boxplot(length_data)
```

```
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)

total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



######################

dataset_name = "lacO34 IPTG ara Exp 2"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp2/'
for folder in os.listdir(folder_path):
```

```
    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)


# Extract the 'eccentricity' column.

# print(final_df)

eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']

length_data = final_df['length_skl']


fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns
```

```
# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
```

```
axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as
needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)
```

```
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()
```

```
plt.close()


plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



dataset_name = "lacO34 PriB IPTG ara Exp 2"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp2/'
```

```
for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "PriB" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.

# print(final_df)

eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']

length_data = final_df['length_skl']
```

```
fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')
```

```
sns.violinplot(length_data, ax=axes[1,0])

axes[1, 0].set_title(f'{dataset_name}: Length distribution')

axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as
needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)
```

```
kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()
```

```
plt.close()
plt.close()

plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})

############################

dataset_name = "lacO34 PriB ara Exp 2"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp2/'
```

```
for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "PriB" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.

# print(final_df)

eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']

length_data = final_df['length_skl']
```

```
fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')
```

```
sns.violinplot(length_data, ax=axes[1,0])

axes[1, 0].set_title(f'{dataset_name}: Length distribution')

axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as
needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)
```

```
kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()
```

```
plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


##################


dataset_name = "lacO34 PriC IPTG ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'
```

```
for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "PriC" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)




# Extract the 'eccentricity' column.

# print(final_df)

eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']
```

```
length_data = final_df['length_skl']


fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns


# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')


# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')
```

```
sns.violinplot(length_data, ax=axes[1,0])

axes[1, 0].set_title(f'{dataset_name}: Length distribution')

axes[1, 0].set_xlabel('Eccentricity')

axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram

axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]
```

```
kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
```

```
plt.show()

plt.close()



plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)

total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



######################
```

```
dataset_name = "lacO34 PriC ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "lacO34" and folder.split(" ")[1] == "PriC" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)


# Extract the 'eccentricity' column.

# print(final_df)
```

```
eccentricity_data = final_df['eccentricity']

area_data = final_df['area_ski']

length_data = final_df['length_skl']


fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns


# Plot 1: Eccentricity histogram

axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')

axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')

axes[0, 0].set_xlabel('Eccentricity')

axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram

axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')

axes[0, 1].set_title(f'{dataset_name}: Area Distribution')

axes[0, 1].set_xlabel('Area')

axes[0, 1].set_ylabel('Frequency')


# Plot 3: Scatter plot of eccentricity vs. area

# axes[1, 0].scatter(eccentricity_data, area_data)

# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
```

```
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()
```

```
counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')
```

```
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
plt.close()

plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()


#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)

total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})
```

```
##################


dataset_name = "120_lacO34 ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)

        df = df[df['edge'] == False]

        df = df[df['area_ski'] > 100]

        df = df[df['eccentricity'] > 0.8]

        df = df[df['a'] < 1]

        df = df[df['a'] > -1]


        # print(df)

        final_df = pd.concat([final_df, df], ignore_index=True)
```

```
# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
```

```
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')

# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
```

```
plt.show()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
```

```
    jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()



plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()
```

```
#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



####################


dataset_name = "120_lacO34 IPTG ara Exp 2"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp2/'
for folder in os.listdir(folder_path):
    # print(folder)
    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "IPTG":
        print(f'{dataset_name}: {folder_path+folder}')
        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')
        # print(df)
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
```

```
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)

# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')

# Plot 2: Area histogram
```

```
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')
```

```
# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation
```

```
density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()


plt.boxplot(length_data)

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()
```

```
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})



######################


dataset_name = "120_lacO34 PriB IPTG ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "PriB" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)
```

```
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
```

```
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
```

```
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()
```

```
plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation
density = kde(length_data)       # Calculate density at each data point
jitter_x = []
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
plt.close()


plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


############################


dataset_name = "120_lacO34 PriB ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "PriB" and folder.split(" ")[2] == "arabinose":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)
```

```
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
```

```
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
```

```
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()

plt.close()


counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()

plt.close()
```

```
plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation
density = kde(length_data)       # Calculate density at each data point
jitter_x = []
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()
plt.close()


plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
```

```
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')

plt.show()

plt.close()

plt.close()



total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})

print({dataset_name}, {len(length_data)})


##################


dataset_name = "120_lacO34 PriC IPTG ara Exp 2"

final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])

# print(final_df.keys())

folder_path = 'Exp2/'

for folder in os.listdir(folder_path):

    # print(folder)

    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "PriC" and folder.split(" ")[2] == "IPTG":

        print(f'{dataset_name}: {folder_path+folder}')

        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')

        # print(df)
```

```
    df = df[df['edge'] == False]
    df = df[df['area_ski'] > 100]
    df = df[df['eccentricity'] > 0.8]
    df = df[df['a'] < 1]
    df = df[df['a'] > -1]

    # print(df)
    final_df = pd.concat([final_df, df], ignore_index=True)



# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns

# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
```

```
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')


# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
axes[1, 0].set_ylabel('Area')

# Plot 4: Length histogram
```

```
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')

axes[1, 1].set_title(f'{dataset_name}: Length Distribution')

axes[1, 1].set_xlabel('Length')

axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels

plt.tight_layout()

plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')

plt.show()

plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed

bin_width = bins[1]-bins[0]

kde = gaussian_kde(length_data)

kde_x = np.linspace(min(length_data), max(length_data), 1000)

plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

plt.show()
```

```
plt.close()



sns.violinplot(length_data,color='skyblue')  # Kernel density estimation
density = kde(length_data)       # Calculate density at each data point
jitter_x = []
for i,x in enumerate(length_data):
    jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))
    jitter_x.append(jitter)
plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for size
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')
plt.show()
plt.close()



plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
```

```
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()



#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)

total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})



######################

dataset_name = "120_lacO34 PriC ara Exp 2"
final_df = pd.DataFrame(columns=['cell_num','eccentricity','length_skl','area_ski','a','edge'])
# print(final_df.keys())
folder_path = 'Exp2/'
for folder in os.listdir(folder_path):
    # print(folder)
```

```
    if os.path.isdir(folder_path+folder) and folder.split(" ")[0] == "120_lacO34" and
folder.split(" ")[1] == "PriC" and folder.split(" ")[2] == "arabinose":
        print(f'{dataset_name}: {folder_path+folder}')
        df = pd.read_csv(folder_path+folder+"/results.tsv", sep='\t')
        # print(df)
        df = df[df['edge'] == False]
        df = df[df['area_ski'] > 100]
        df = df[df['eccentricity'] > 0.8]
        df = df[df['a'] < 1]
        df = df[df['a'] > -1]

        # print(df)
        final_df = pd.concat([final_df, df], ignore_index=True)

# Extract the 'eccentricity' column.
# print(final_df)
eccentricity_data = final_df['eccentricity']
area_data = final_df['area_ski']
length_data = final_df['length_skl']

fig, axes = plt.subplots(2, 2, figsize=(10, 8))  # 2 rows, 2 columns
```

```
# Plot 1: Eccentricity histogram
axes[0, 0].hist(eccentricity_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 0].set_title(f'{dataset_name}: Eccentricity Distribution')
axes[0, 0].set_xlabel('Eccentricity')
axes[0, 0].set_ylabel('Frequency')



# Plot 2: Area histogram
axes[0, 1].hist(area_data, color='skyblue',bins=50, edgecolor='black')
axes[0, 1].set_title(f'{dataset_name}: Area Distribution')
axes[0, 1].set_xlabel('Area')
axes[0, 1].set_ylabel('Frequency')

# Plot 3: Scatter plot of eccentricity vs. area
# axes[1, 0].scatter(eccentricity_data, area_data)
# axes[1, 0].set_title(f'{dataset_name}: Eccentricity vs. Area')
# axes[1, 0].set_xlabel('Eccentricity')
# axes[1, 0].set_ylabel('Area')

sns.violinplot(length_data, ax=axes[1,0])
axes[1, 0].set_title(f'{dataset_name}: Length distribution')
axes[1, 0].set_xlabel('Eccentricity')
```

```
axes[1, 0].set_ylabel('Area')


# Plot 4: Length histogram
axes[1, 1].hist(length_data, color='skyblue', edgecolor='black')
axes[1, 1].set_title(f'{dataset_name}: Length Distribution')
axes[1, 1].set_xlabel('Length')
axes[1, 1].set_ylabel('Frequency')


# Adjust layout to prevent overlapping titles/labels
plt.tight_layout()
plt.savefig(f'{dataset_name.replace(" ","_")}_all_plots.png')
plt.show()
plt.close()
plt.close()



counts,bins,_ = plt.hist(length_data, color='skyblue', edgecolor='black')  # Adjust bins as needed
bin_width = bins[1]-bins[0]
kde = gaussian_kde(length_data)
kde_x = np.linspace(min(length_data), max(length_data), 1000)
plt.plot(kde_x,kde(kde_x)*bin_width*len(length_data),color="blue")
```

```
plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_hist.png')

#plt.show()

plt.close()

plt.close()


sns.violinplot(length_data,color='skyblue')  # Kernel density estimation

density = kde(length_data)       # Calculate density at each data point

jitter_x = []

for i,x in enumerate(length_data):

   jitter = np.random.normal(0, 0.05 * (density[i] / max(density)))

   jitter_x.append(jitter)

plt.scatter(jitter_x, length_data, color='blue', alpha=0.5, s=10) # alpha for transparency, s for
size

plt.title('Distribution of Eccentricity')

plt.ylabel('Length')

plt.title(f'{dataset_name}')

plt.savefig(f'{dataset_name.replace(" ","_")}_length_vp.png')

plt.show()

plt.close()

plt.close()
```

```
plt.boxplot(length_data)
plt.title('Distribution of Eccentricity')
plt.ylabel('Length')
plt.title(f'{dataset_name}')
plt.savefig(f'{dataset_name.replace(" ","_")}_length_bp.png')
plt.show()
plt.close()
plt.close()


#No_induction_exp_1 = pd.concat([No_induction_exp_1, final_df], ignore_index=True)


total_data.update({str(dataset_name.replace(" ","_")):length_data*px_size})
print({dataset_name}, {len(length_data)})


##################


total_data = pd.DataFrame(total_data)
print(total_data)
```

```
for i, column_name in enumerate(total_data.columns):
    count = total_data[column_name].count()
    print(f"Column {i+1}: '{column_name}': {count} ")

df_melt = total_data.melt(var_name=' ', value_name='Length')
plt.figure(figsize=(12, 6))  # Adjust figure size as needed
sns.boxplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability
plt.title('Boxplots of Measurements for Each Sample')
plt.ylabel("Length ($\mu m$)")
plt.tight_layout()
plt.show()
plt.close()
plt.close()

# Create the violin plot
plt.figure(figsize=(10, 6))
sns.violinplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels if needed
plt.title('Violin Plot of All Data')
#plt.xlabel("Sample")
plt.ylabel("Length ($\mu m$)")
```

```
plt.tight_layout()

plt.show()

plt.close()

plt.close()


##

for col in range(0, int(len(total_data.columns))):

   for col_2 in range(0, int(len(total_data.columns))):

      if col_2 != col:

         data_1 = total_data.iloc[:,col].dropna()

         #print(data_1)

         data_2 = total_data.iloc[:, col_2].dropna()

         t_statistic, p_value = stats.ttest_ind(data_1, data_2, equal_var=False)

         print(f'T-test between columns {total_data.columns[col]} (N={len(data_1)},
median={np.median(data_1)}) and {total_data.columns[col_2]} (N={len(data_2)},
median={np.median(data_2)}) is: {p_value}')

##


data_1 = total_data.filter(regex=r'1$')  # Columns ending in '1'

data_2 = total_data.filter(regex=r'2$')  # Columns ending in '2'


df_melt = data_1.melt(var_name=' ', value_name='Length')
```

```
plt.figure(figsize=(12, 6))  # Adjust figure size as needed
sns.boxplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability
plt.title('Boxplots of Exp 1')
plt.tight_layout()
plt.savefig("./boxplots_exp1.svg")
plt.show()
plt.close()
plt.close()

# Create the violin plot
plt.figure(figsize=(10, 6))
sns.violinplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels if needed
plt.title('Violin Plot of Exp 1')
#plt.xlabel("Sample")
plt.ylabel("Length ($\mu m$)")
plt.tight_layout()
plt.savefig("./violin_exp1.svg")
plt.show()
plt.close()
plt.close()
```

```
df_melt = data_2.melt(var_name=' ', value_name='Length')
plt.figure(figsize=(12, 6))  # Adjust figure size as needed
sns.boxplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability
plt.title('Boxplots of Exp 2')
plt.ylabel("Length ($\mu m$)")
plt.tight_layout()
plt.savefig("./boxplots_exp2.svg")
plt.show()
plt.close()
plt.close()

# Create the violin plot
plt.figure(figsize=(10, 6))
sns.violinplot(x=' ', y='Length', data=df_melt)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels if needed
plt.title('Violin Plot of Exp 2')
#plt.xlabel("Sample")
plt.ylabel("Length ($\mu m$)")
plt.tight_layout()
plt.savefig("./violin_exp2.svg")
```

```
plt.show()

plt.close()

plt.close()


# 1. Group columns by base name (removing the 1 or 2)

grouped_data = {}

for col in total_data.columns:

    base_name = re.sub(r'_[12]$|_Exp_[12]$', '', col)  # Remove 1 or 2 from end

    if base_name not in grouped_data:

        grouped_data[base_name] = []

    grouped_data[base_name].append(col)


print(grouped_data)


# 2. Combine data for each group

combined_data = pd.DataFrame()

for base_name, cols in grouped_data.items():

    if len(cols) > 0:  # Check if there are columns for this base name

        group_data = total_data[cols]

        # Melt the data for this group

        group_data_melted = group_data.melt(var_name='Condition', value_name='Length')

        # Add a "Group" column
```

```
        group_data_melted['Group'] = base_name
        combined_data = pd.concat([combined_data, group_data_melted], ignore_index=True)
        num_combined_points = group_data_melted['Length'].count()
        print(f"{base_name}: {num_combined_points}")

print(combined_data)

# Pivot the DataFrame to create separate columns for each group
pivoted_df = combined_data.pivot_table(index=combined_data.index, columns='Group',
values='Length', aggfunc='first')

# Display the pivoted DataFrame
print(pivoted_df)
print(pivoted_df.columns)

##
for col in range(0, int(len(pivoted_df.columns))):
    for col_2 in range(0, int(len(pivoted_df.columns))):
        if col_2 != col:
            data_1 = pivoted_df.iloc[:,col].dropna()
            #print(data_1)
            data_2 = pivoted_df.iloc[:, col_2].dropna()
```

```
        t_statistic, p_value = stats.ttest_ind(data_1, data_2, equal_var=False)

        print(f'T-test between columns {pivoted_df.columns[col]} (N={len(data_1)}, median={np.median(data_1)}) and {pivoted_df.columns[col_2]} (N={len(data_2)}, median={np.median(data_2)}) is: {p_value}')

##




# 3. Create the violin plot

plt.figure(figsize=(12, 8))  # Adjust figure size as needed

sns.boxplot(x='Group', y='Length', data=combined_data)

plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability

#plt.title('Boxplots of Measurements for Each Sample')

plt.ylabel("Length ($\mu m$)")

plt.tight_layout()

plt.savefig("./boxplots_combined_priBC.svg")

plt.show()

plt.close()


plt.figure(figsize=(12, 6))  # Adjust figure size as needed

sns.violinplot(x='Group', y='Length', data=combined_data)

plt.xticks(rotation=45, ha='right')

#plt.title('Violin Plot of Combined Data')
```

```
plt.xlabel(" ")

plt.ylabel("Length ($\mu m$)")

plt.tight_layout()

plt.savefig("./violin_combined_priBC.svg")

plt.show()

plt.close()

plt.close()
```

Appendix B: R Statistical Analysis and Visualization (Chapter 4)

The following R script was independently developed by the candidate to perform the final data visualization (Violin/Boxplot hybrid distributions) and formal non-parametric statistical testing (Wilcoxon rank-sum) for the results presented in Chapter 4.

```
# ================================================================================
# PHD THESIS: ANALYSIS OF CAS1-CAS2 FOCI

# Features: BOLD Nomenclature, Superscripts, Subscripts

# ================================================================================
```

```
# 1. LOAD LIBRARIES
library(ggplot2)
library(dplyr)
library(tidyr)
library(ggpubr)
library(rstatix)
library(janitor)

# 2. DATA IMPORT
df_raw <- read.csv("foci_data.csv", skip = 3, check.names = FALSE)

# 3. CLEANING & RENAMING
foci_data <- df_raw %>%
 clean_names() %>%
 select(
   "Control" = contains("ori_z") & !contains("pri"),
   "priA300" = contains("pri_a300"),
   "priB"    = contains("pri_b"),
   "priC"    = contains("pri_c")
 ) %>%
 pivot_longer(
   cols = everything(),
```

```
  names_to = "Strain",
  values_to = "Foci_Count"
 ) %>%
 drop_na(Foci_Count)

# 4. SET ORDER FOR X-AXIS
foci_data$Strain <- factor(foci_data$Strain, levels = c("Control", "priA300", "priB", "priC"))

# 5. STATISTICAL TESTING
stat_test <- foci_data %>%
 wilcox_test(Foci_Count ~ Strain, ref.group = "Control") %>%
 add_significance(p.col = "p", output.col = "p.signif") %>%
 add_xy_position(x = "Strain")

# --- NEW: CALCULATE Y-AXIS LIMITS ---
max_foci <- max(foci_data$Foci_Count, na.rm = TRUE)
upper_limit <- ceiling(max_foci / 2) * 2 + 4 # Rounds up to even and adds space for stars

# 6. VISUALIZATION
thesis_plot <- ggviolin(
 foci_data, x = "Strain", y = "Foci_Count",
 fill = "Strain",
```

```
  palette = c("#999999", "#D55E00", "#56B4E9", "#009E73"),
  add = "boxplot",
  add.params = list(fill = "white", width = 0.1),
  trim = TRUE
) +
  geom_jitter(alpha = 0.05, size = 0.5, width = 0.15, color = "grey50") +

  # --- NEW: DYNAMIC 2-BY-2 Y-AXIS ---
  scale_y_continuous(
    breaks = seq(0, upper_limit, by = 2),
    limits = c(0, upper_limit),
    expand = expansion(mult = c(0, 0.05))
  ) +

  stat_pvalue_manual(
    stat_test,
    label = "p.signif",
    hide.ns = FALSE,
    tip.length = 0.01,
    bracket.nudge.y = 1.0,
    fontface = "bold" # Makes the asterisks/stars bold
  ) +
```

```
  # Applying BOLD labels for X-axis

  scale_x_discrete(labels = c(

    "Control" = expression(bolditalic(oriC)^bold("+") ~ bolditalic(oriZ)^bold("+")),

    "priA300" = expression(bolditalic(oriC)^bold("+") ~ bolditalic(oriZ)^bold("+") ~
bolditalic(priA) * bold("300")),

    "priB"    = expression(bolditalic(oriC)^bold("+") ~ bolditalic(oriZ)^bold("+") ~
bold(Delta)*bolditalic(priB)),

    "priC"    = expression(bolditalic(oriC)^bold("+") ~ bolditalic(
```