1

Introduction
1. Introduction

1.1. General introduction

Bacteria are masters of adaptation, occupying almost every environment on Earth. The key to their success lies in their ability to multiply rapidly. Organisms adapt to new environments gradually: on balance, offspring are better adapted to local environmental conditions than their parents are. Because they multiply rapidly, bacteria have incredibly short generation times and thus produce better-adapted offspring very quickly. Once a few bacterial cells become suitably adapted for life in a new environmental niche, they grow quickly and thus rapidly expand their range.

Changes in environmental conditions induce stress – a term used to describe mismatch between an organism’s capabilities and its environment. Exogenous stress can be a source of constant harm in bacterial cells. Such external pressures include starvation, abrupt changes in pH and temperature, DNA damage, and exposure to antibiotics. These conditions may affect the cell during its lifetime or result in accumulated effects felt over several generations. Over time, natural selection determines which organisms survive in a particular environment, and which do not.

Can environmental conditions also provoke changes in organisms? In many organisms, environmental stresses induce pathways that result in higher frequencies of genomic mutations. This stress-induced mutagenesis increases genetic diversity within populations of organisms that become maladapted to their environments. If stress-induced mutagenesis increases the number of offspring with the new physiological traits needed to succeed in the new conditions, it will accelerate the adaptation process \[1, 2\]. Stress-induced mutagenesis is one invocation of the broader theory of adaptive mutation. Under its broadest definition, the adaptive mutation theory posits that environmental conditions not only determine which organisms are selected for during the course of evolution, as in the classical neo-Darwinian description, but also plays a role in helping maladapted organisms mutate to better fit conditions \[3–5\].

The best-studied mechanism for stress-induced mutagenesis involves the SOS response – a global cellular response to DNA damage, typically brought about by exposure to external stress. In \textit{E. coli}, the SOS response leads to upregulation of more than 40 genes, including three different DNA polymerases \[6\]. These specialized polymerases (pols) have the ability to carry out translesion DNA synthesis (TLS), i.e. they can insert a nucleotide base opposite a damaged base, or lesion, on the template DNA. The primary purpose of TLS pols has long been thought to be rescue of replication complexes that have been stalled by lesions in the DNA \[7\]. Under this model, the TLS Pol temporarily replaces the replicative DNA polymerase at a stalled replication fork, synthesizes past the lesion and thus allows replication to continue. The lesions are left behind to be repaired by other mechanisms later. However, this ability to support TLS comes at a cost: these polymerases display a much higher error rate than the specialized replicative DNA polymerase and thus their use lead to higher rates of mutation. Why do cells maintain TLS pols, when other, less error-prone mechanisms for replication restart exist? One view is that the negative effects of increased mutation are offset by the positive effects on cell survival that their lesion bypass activity provides. An alternative view is that the
stress-induced mutation facilitated by TLS pols contributes to adaptive mutation, allowing cells carrying TLS pols to more readily adapt to new environments. In any case, TLS polymerases play an important role in balancing genetic diversity and genome maintenance [8–10].

The number of DNA pols produced by a species varies; the yeast Saccharomyces cerevisiae expresses 7 and humans produce at least 14 different DNA polymerases [7]. Despite belonging to several different sequence families, they share common features, such as an affinity for a processivity clamp (β₂ for bacteria and PCNA for eukaryotes and archaea). Furthermore, the DNA polymerases show conserved structural features. E. coli Pol IV, for example, has homologs in all domains of life [11, 12]. The vast knowledge accumulated on DNA replication and repair in E. coli and other model organisms allows us to hypothesize how DNA-damage responses, and stress-induced mutagenesis, might be regulated in different organisms.

Mechanisms to regulate TLS pols must exist to avoid their error-prone activity beyond what is minimally needed to rescue replication and/or to modulate mutation rates as part of an adaptive mutagenesis mechanism. How does E. coli decide which of its DNA polymerases will be active at any given moment? How does the cell coordinate which DNA polymerase replaces the replicative polymerase when replication is stalled? The predominant hypothesis in the field is that the transcriptional regulation (the SOS response) increases the cellular concentration of TLS pols and thus promotes mass action-driven switching of pols at stalled replication forks [7, 9, 13]. In this case, the thermodynamic properties of the system, such as affinities of the different polymerases for the replication machinery and their respective concentrations, should be the main determinants of which polymerase is present at the replication machinery.

Because E. coli displays short generation times and high genomic plasticity it represents an ideal model system to understand how organisms manage their genetic material and what factors influence their fitness. Its short generation time enables monitoring of damage response and the establishment of mutationally adapted progeny on a timescale that is easily experimentally accessible. Recently, the Austin group demonstrated that bacteria are able to evolve antibiotic resistance within ten hours [14]. These experiments and others are increasingly highlighting the relationship between antibiotic resistance and the activity of specialized but mutagenic polymerases [15]. Since several such mechanisms are conserved between large numbers of bacterial species and even higher organisms [11, 12], such knowledge improves our understanding of evolution mechanisms and may help to explain why bacteria develop resistance to antibiotics faster than we can develop new drugs and why cancer cells so often become resistant to treatment.

In this thesis, I describe how we have applied single-molecule microscopy tools to visualize the spatiotemporal dynamics of individual protein molecules in live E. coli cells to investigate regulatory mechanisms that control DNA polymerase V activity and to revisit the mass-action hypothesis. E. coli is a well-suited model organism to study how the specialized DNA pols are regulated and activated as a consequence of DNA damage [16]. These DNA pols play a key role in adaptive mutation; because mutagenesis rates are dependent on polymerase fidelity, low-fidelity TLS pols pro-
mote genetic diversity and consequently boost adaptation by producing genetic variety [17].

In this introductory chapter, I will briefly review key mechanistic properties of DNA pols in *Escherichia coli* and their roles in bacterial mutagenesis, including current models of DNA Pol switches at sites of replication. Further, I will discuss methodological aspects of single-molecule microscopy in live bacteria and how such approaches can address fundamental questions related to bacterial stress responses.

### 1.2. *E. coli* and its multiple DNA polymerases

*E. coli* contains five DNA polymerases (Table 1.1). DNA polymerase III (Pol III) is the main replicative polymerase of which at least two copies are present in the replisome, the multi-protein complex that is responsible for the coordinated unwinding of double-stranded DNA and production of two identical daughter duplexes. DNA polymerase I (Pol I) is responsible for a number of repair duties, a major one of which is the replacement of RNA with DNA at between Okazaki fragments on the lagging-strand product of replication. Both Pol I and Pol III are maintained at relatively constant expression levels. Three other specialized polymerases – Pol II, Pol IV and Pol V – are expressed as part of the SOS response. Each of these five polymerases differs in their structure, preference for substrates, and nature of the reactions they catalyse. For example, 3’–5’ exonuclease activity (proofreading) is present only on Pols I, II and III, and serves to enhance fidelity when copying an undamaged DNA template. Pols IV and V are error-prone polymerases able to replicate undamaged DNA, but their primary role is in carrying out TLS on damaged DNA.

#### 1.2.1. DNA Polymerae III and the replisone

The replisome is a dynamic multi-protein structure conserved across viruses, bacteria, archaea and eukaryotes (Table 1.2). Replisomes share a common structure that comprises a number of key parts: a helicase to unwind the double-stranded DNA; a polymerase subunit to synthesize the new strand; sliding clamps that encircle DNA and confer processivity to the replicative DNA polymerase; a clamp loader to load the sliding clamps and to organize the replisome on the DNA, a primase to synthesize RNA primers needed for polymerase activity, and single-stranded DNA-binding protein (SSB) to coat and protect transiently exposed single-stranded DNA.

In *E. coli*, the DNA Polymerase III Holoenzyme (Pol III HE) represents the core structure of the replisome and coordinates the activity of all components involved in unwinding and coordinated synthesis of the two daughter strands. Pol III HE consists of three main parts: Pol III core, the sliding clamp, and the clamp loader complex. Pol III core consists of three subunits, α, ε, and θ. The *dnaE* gene encodes the α-subunit has polymerase activity; the ε subunit (*dnaQ*) has 3’–5’ exonuclease activity; the θ subunit (*holE*) stimulates ε-proofreading activity. The β2 sliding clamp is a dimeric ring composed of two β subunits (*dnaN*) that encircles double-stranded
1.2. *E. coli* and its multiple DNA polymerases

Table 1.1 | The five DNA polymerases of *Escherichia coli* and few of their relevant properties.

<table>
<thead>
<tr>
<th></th>
<th>Pol I</th>
<th>Pol II</th>
<th>Pol III</th>
<th>Pol IV</th>
<th>Pol V</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-3’polymerase</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3’-5’ exonuclease</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’-3’ exonuclease</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Biological</td>
<td>Dna Replication</td>
<td>DNA Replication</td>
<td>DNA polymerase</td>
<td>TLS</td>
<td>TLS</td>
</tr>
<tr>
<td>functions</td>
<td>Okazaki fragment maturation DNA Repair (Backup polymerase) DNA Repair</td>
<td>TLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Gene</td>
<td>polA</td>
<td>polB</td>
<td>dnaE dnaQ holE</td>
<td>dinB</td>
<td>umuDC</td>
</tr>
<tr>
<td></td>
<td>PolIII</td>
<td>PolIII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–SOS</td>
<td>400</td>
<td>50 -75</td>
<td>10 -20</td>
<td>150 - 250</td>
<td>&lt; 15</td>
</tr>
<tr>
<td>+SOS</td>
<td>400</td>
<td>350 -1000</td>
<td>10 -20</td>
<td>1200 - 2500</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 1.2 | Replisomes and common characteristics across different groups.

<table>
<thead>
<tr>
<th></th>
<th>T7 Phage</th>
<th>T4 Phage</th>
<th><em>E. coli</em></th>
<th>Eukaryotes</th>
<th>Archaea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymerase</td>
<td>Gp5/TRX</td>
<td>Gp43</td>
<td>Pol III</td>
<td>Pol α/σ/ε</td>
<td>Pol</td>
</tr>
<tr>
<td>Clamp</td>
<td>–</td>
<td>Gp45</td>
<td>β</td>
<td>PCNA</td>
<td>PCNA</td>
</tr>
<tr>
<td>Clamp loader</td>
<td>–</td>
<td>Gp44/62</td>
<td>γ/τ complex</td>
<td>RFC</td>
<td>RFC</td>
</tr>
<tr>
<td>SSB</td>
<td>Gp2.5</td>
<td>Gp32</td>
<td>SSB</td>
<td>RPA</td>
<td>RPA</td>
</tr>
<tr>
<td>Primase</td>
<td>Gp4</td>
<td>Gp61</td>
<td>DnaG</td>
<td>Pol α-primase</td>
<td>primase</td>
</tr>
<tr>
<td>Helicase</td>
<td>Gp4</td>
<td>Gp41</td>
<td>DnaB</td>
<td>MCM2/CMG</td>
<td>MCM</td>
</tr>
</tbody>
</table>

DNA. It tethers Pol III to the DNA and ensures that the polymerase stays attached, allowing processive DNA synthesis. The clamp-loader complex \([τ\gamma_2σ'χψ]\) consists of a ring of three \(τ\) subunits (\(dnaX\)), one \(δ\) subunit (\(holA\)), and one \(δ'\) subunit (\(holB\)). Each of the \(τ\) subunits tethers a Pol III core complex to the replisome. A shorter \(dnaX\) product gives rise to the \(γ\) protein that can replace \(τ\) subunits in the clamp loader but is unable to bind to the Pol III core. Finally, the \(χ\) (\(holC\)) and \(ψ\) subunits (\(holD\)) interact with SSB and play a role in the coordination of the enzymatic events at the lagging strand [18, 19].

1.2.2. Coordination of DNA replication

*E. coli* has a single, circular chromosome of about 1.2 mm in length, containing 4.5 million base pairs (bp)[20] and more than 4200 protein-coding genes. With
two sister replisomes that each copy half of the chromosome at a rate of $\sim 1,000$ bp/s, the entire chromosome is duplicated in 40 minutes. Under optimal growth conditions, cells can divide as fast as one division each 20 minutes, which is achieved using nested cycles of replication [21, 22].

The replication of circular chromosomes occurs in three steps: initiation, extension and termination. Replication initiates at a single region called chromosomal origin of replication, oriC, and commences by the cooperative binding of the initiator protein DnaA to its multiple sites (DnaA boxes). Separation of the two DNA strands at the AT-rich DnaA boxes results in the local unwinding of DNA and enables the replicative DNA helicase (DnaB in \textit{E. coli}) to load. Subsequent binding of primase (DnaG) and assembly of two replisomes facing opposite directions ends the initiation step [23, 24](Fig. 1.1).

The elongation phase of replication, the processive and coordinated duplication of duplex DNA into two double-stranded daughter DNA strands, takes place in a bidirectional fashion, where replisomes travel in opposite directions around the circular chromosome. The synthesis of DNA in the replisome is a high-fidelity process mainly carried out by DNA Polymerase III (Pol III), although other polymerases may also play a role [25–29]. The DNA helicase (DnaB) separates double-stranded DNA ahead of the replisome, producing two complementary daughter strands as templates for the DNA polymerase III (Fig. 1.2). The leading strand is copied continuously with the direction of DNA synthesis the same as the direction of movement of the replisome. The lagging strand, however, needs to be synthesized in the opposite direction and as a result is copied as a series of 1-2 kb-long DNA fragments called Okazaki fragments. A DnaG primase is periodically recruited to the DnaB helicase to synthesize the RNA primers required to initiate the synthesis of every new Okazaki fragment.

1.2.3. Coordination of Multiple DNA polymerases

Our knowledge of DNA replication has increased substantially in recent years. Contrasting with older models of DNA replication, we now know that the replisome is a very dynamic structure that is able to dynamically exchange components such as the DNA polymerase while supporting accurate and processive replication [30–35].

First, I will describe the classical view of the coordination of DNA polymerases within the replisome and the switching of DNA polymerases in non-damaged cells. A more extensive discussion about polymerase switching in the context of heavy DNA damage follows later.

The traditional view of the bacterial replisome is that of a stable structure containing tightly bound DNA polymerases that are responsible for the processive synthesis of many hundreds of thousands of base pairs. The leading-strand polymerase synthesizes DNA directly on the ssDNA produced by the DNA helicase, allowing the 5’–3’ polymerase activity to work continuously at rates up to 1000 nucleotides per second [34, 36, 37]. The lagging strand, however, cannot be continuously copied and requires a mechanism that allows polymerases to switch from the end of one Okazaki fragment to the start of the next one. SsDNA newly produced by the he-
Bi-directional DNA replication in bacteria can be divided into three steps. To initiate the process, the origin of replication (oriC) is melted by the initiator protein DnaA, generating single-stranded DNA substrates for replication. The primosome (DnaB₆-DnaG₃) and DNA polymerase holoenzyme (Pol III HE) assembles at the melted origin. Each replisome contains two core polymerases, one for leading strand synthesis and one for lagging strand. In the elongation phase, replication proceed bidirectionally around the circular chromosome. The termination phase happens when replisomes meet at the Ter region and dissociate from the chromosome. The chromosomes are then separated. Top figure reproduced from [23].

With five DNA polymerases, E. coli requires mechanisms to regulate polymerase access to DNA templates, especially since each of the polymerase have particular characteristics and demonstrate different fidelities during nucleotide incorporation. DNA Pols II, IV and V are reported to switch with Pol III and gain access to the primer-DNA template [38, 39]. Studies using single-molecule microscopy revealed a highly dynamic polymerase exchange for T7 replication [40] and more recently a similar dynamic exchange of replicative polymerase has been observed for E. coli (Spenkelink, Lewis and van Oijen, unpublished results). Taken together, these results suggest that replisomes may not be the stable structures as envisioned before but instead allow dynamic exchange of their polymerases.

This revised view of a highly dynamic replisome raises new questions. How does the cell manage which polymerase will be active at any given moment? What mechanism triggers the switch between different polymerases? Selecting the incorrect
During DNA synthesis, *E. coli* DNA polymerase III holoenzyme (Pol III HE) can exchange polymerase subunits while simultaneously remaining attached at the replication fork. Each of the three \( \gamma \)-subunits in the clamp loader complex contains a binding site for a core polymerase (\( \alpha \epsilon \theta \)). Image reproduced from [23].

Polymerase to synthesize DNA in the replisome may have an immediate impact on mutation rates, which in turn affect cellular fitness.

The dominant mechanistic hypothesis for how for polymerase switching occurs at the replisome relies on the \( \beta_2 \) sliding clamp acting as a “tool belt”. This model, originally proposed by Pages and Fuchs, is based on the ability of multiple polymerases to bind to a single \( \beta \)-sliding clamp [13]. Depending on the expression levels of the various polymerases and their relative affinities for the sliding clamp, mass action dictates which polymerases are associated with the clamp at what time [13]. The importance of mass action in controlling polymerase exchange has been demonstrated in a number of studies. For example, DNA polymerases have a common binding site at the C-terminus of the \( \beta \)-binding clamp (\( \beta_2 \)). Different DNA polymerases may exchange with Pol III and with two copies of this binding site in the dimeric clamp, different polymerases may even interact simultaneously with \( \beta_2 \) [41]. Another study used single-molecule tools to visualise the simultaneous binding of Pol III and Pol IV to the \( \beta \) clamp [42]. The authors showed that an increase in Pol IV concentration directly reduced Pol III processivity, supportive of
a polymerase exchange driven by mass-action exchange. The strength of the interaction between pols and the sliding clamp depends on whether the clamp is bound to DNA [32]. If the sliding clamp is not attached to DNA, the polymerase cannot form a stabilizing interaction with the primer-template and the dissociation rates are high. While the dimeric β clamp provides two binding sites for polymerases, one per monomer, it has been shown that polymerase interactions are preferred to take place on, or even limited to, only one half of the clamp [32]. In such a scenario, other interactions and contacts must be present to drive and regulate polymerase switching. For example, studies of the replacement of Pol III by Pol IV on β have shown that additional contacts between Pol III and the clamp-loader protein τ trigger polymerase release [43]. This process suggests that additional mechanisms must be present to regulate switching between the various polymerases. In a simplified picture, Pol III is a very efficient, fast and accurate polymerase; it should not switch with other polymerases unless its path is blocked. Since lagging-strand synthesis involves frequent interruptions of Pol III synthesis, it is expected that Pols II, IV and V gain sporadic access to the DNA on the lagging strand [27](Fig. 1.3).

**Figure 1.3 | Stalled bacterial replisome.** Bacteria replisome stalled at a damaged site (red cross) and DNA polymerases available for switching.
1.3. DNA-repair mechanisms

Endogenous and exogenous sources of damage constantly introduce lesions to genomic DNA, resulting in mutations if not resolved. The frequency of generation of spontaneous mutations within the genome is found to be very similar within a broad range of different organisms [44, 45]. In *E. coli*, for example, the mutation rate is as low as $10^{-3}$ per genome per generation [44, 46]. While baker’s yeast (*S. cerevisiae*) and the filamentous fungus *N. crassa* have genome sizes 3 times and 9 times larger than that of *E. coli*, respectively, their mutation rates per genome are very similar [44]. Maintaining such low mutation rates suggests the presence of tightly controlled repair mechanisms.

DNA damage can be divided into two main categories: endogenous, occurring as a consequence of cellular metabolism; and exogenous, introduced by environmental factors such as antibiotics, chemical agents or UV light. Several pathways in the cell are available to remove lesions from genomic DNA [39]. Here, I will briefly discuss three fundamental repair pathways: Base excision repair, nucleotide excision repair, and mismatch repair.

Base excision repair (BER) is the main pathway responsible for removing a broad spectrum of small base lesions, especially lesions caused by oxidation, alkylation and deamination [47, 48]. The first example of BER was provided by Tomas Lindahl describing the activity of a uracil DNA glycosylase from *E. coli* [49]. DNA glycosylases recognise DNA lesions and cleave the base-deoxyribose bonds to remove the base while leaving the phosphodiester backbone intact. Several specific DNA glycosylases exist to deal with different classes of lesions. Subsequent nuclease, polymerase, and ligase activity is responsible for restoring the DNA proximal to the lesion to its original state [48, 49].

Nucleotide excision repair (NER) is the mechanism responsible for recognizing and removing bulky lesions such as thymine dimers and 6-4-photoproducts. These lesions, predominantly caused by UV radiation, are repaired in bacteria by the UvrABC endonuclease complex. UvrA scans the DNA searching for distortions in the DNA helix. When found, UvrA is replaced by UvrB, which melts the dsDNA and recruits UvrC. UvrC makes an incision away from the lesion, removes the lesion with DNA helicase II, leaving a stretch of ssDNA. This gap is then filled by Pol I and DNA ligase [50].

Mismatch repair (MMR) is responsible for removing lesions that escape the proof-reading activity of DNA polymerases and form a non-Watson-Crick base pair[51]. Briefly, MMR in *E. coli* is initiated by MutS searching for and recognizing the lesion. MutS then recruits MutL to trigger a cascade of reactions that results in the removal of the lesion and synthesis of correctly base-paired DNA. Its ability to repair mistakes made by the replicative DNA polymerases makes MMR essential to avoid mutations becoming permanent after cell division[51].

Deficiency in these repair mechanisms has major impact on cellular functions. In humans, many genetic diseases are related to malfunctions in repair mechanisms. For example, deficiencies in NER are related to xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy, diseases mainly related to deficiencies in deal-
1.3. DNA-repair mechanisms

ing with UV-dependent lesions [52]. Further, the accumulation of lesions caused by a deficiency in BER is reported to be related to cellular aging and cancer [48] and MMR deficiencies are related to neurological diseases [51] and a variety of cancers, such as hereditary colorectal cancer [53]. However, a direct link between the deficiency in DNA repair and a particular disease is often not clear. For instance, Huntington’s disease is reported to be associated with both BER and MMR [48, 51].

Other mechanisms such as repair mediated by homologous recombination (HR) exist to supplement the activity of the DNA-repair mechanisms discussed above. Homologous recombination is essential for accurate repair of breaks of dsDNA caused by ionizing radiation and by anti-tumour agents in humans [54]. The non-homologous end joining (NHEJ) mechanism repairs double-strand breaks by direct ligation. Many species of bacteria, including *E. coli*, lack the ability to perform NHEJ and instead have to rely on homologous recombination to repair double-stranded breaks. Figure 1.4 summarize these mechanisms.

![Figure 1.4 | DNA damage and repair mechanisms.](image)

Figure 1.4 | DNA damage and repair mechanisms. DNA damage and repair mechanisms. Exemplification of DNA lesions caused by a common DNA damaging agents and the pathways responsible for damage removal. Figure reproduced from [55].

1.3.1. Replisome and DNA lesions

The DNA-repair mechanisms discussed above act on lesions outside the context of DNA replication. In such a scenario, DNA damage may have occurred well before the replisome encounters the damage site, thus leaving ample time for repair mechanisms to deal with the lesion before the replisome arrives at the damage site. As
discussed before, Pol III is highly efficient when replicating undamaged DNA, but is unable to synthesize past most lesions. When a replisome encounters a lesion before DNA repair pathways have taken care of it, the replication machinery will stall and may collapse entirely. In UV-damaged cells, for example, replisomes stalls for at least 15 minutes having serious consequences on genome stability and the cell fate [56]. Therefore, it is of crucial importance that replication is resumed as rapidly as possible [57–60]. Many mechanisms exist to restart replication in an error-free manner. These pathways include helicase-mediated (PriA RecA-mediated restart) [58, 59] and the replisome skipping over the lesion, reengaging DNA downstream of the lesion site [60].

Excessive levels of DNA damage, however, surpass the cell’s basal ability to restart replication and lead to the induction of other repair pathways, such as the RpoS and SOS response. The RpoS and SOS response are complementary mechanisms and both are activated in response to various sources of stresses [57, 61]. The RpoS response is based on the RNA polymerase subunit $\sigma^{\mathrm{RpoS}}$. The transcription factor $\sigma^{\mathrm{RpoS}}$ is involved in initiating transcription of up to 10% of E. coli’s genome and is upregulated by environmental changes, such as pH change or lack of nutrients. Like the SOS response, RpoS response is not detectable in fast-growing cells [62]. The SOS response is a general response to DNA damage and lies at the heart of bacterial mutagenesis. Since it also represents the main focus of this thesis, it will be discussed in much greater detail below.

### 1.3.2. Regulation of the SOS response

The SOS response is a global response to DNA damage that once triggered arrests the cell cycle and starts a number of DNA-repair and mutagenesis pathways. The process is started by the accumulation of single-stranded DNA (ssDNA) inside the cell. This accumulation of ssDNA is typically attributed to continued unwinding of the parental DNA after Pol III has stalled at a leading strand lesion [63]. However, ssDNA gaps left in the wake of the replisome likely also contribute to induction of the SOS response [64]. Such gaps are thought to arise frequently, each time Pol III encounters a lesion on the lagging strand, but are also produced under the leading strand lesion-skipping mechanism. The SOS response and its regulatory mechanisms that control the upregulation of various genes are well understood in E. coli, although the subsequent regulatory pathways involving the expressed proteins are far from clear. In this section, the discussion will focus on the mechanisms for which experimental support is in good agreement; work presented later in this thesis will address aspects of the regulatory pathways that have been more poorly understood.

The long stretches of ssDNA produced when Pol III encounters lesions represent the binding substrate for RecA, a DNA-dependent ATPase that forms helical filaments on ssDNA. RecA is constitutively expressed at a level of 10,000 monomers per cell to facilitate its role in homologous recombination [65], a copy number sufficiently high to ensure filament formation. To selectively form RecA filaments only on long stretches of ssDNA and not on the transiently exposed ssDNA in an active
1.3. DNA-repair mechanisms

1.3.1. DNA-repair mechanisms

Figure 1.5 | The SOS response in *Escherichia coli*. Only a subset of the 43 known genes included in this system is shown. Under normal growth conditions, the SOS genes are repressed by the dimeric transcriptional repressor, LexA. Upon exposure to DNA damage, such as ultraviolet radiation, the LexA repressor is autocatalytically cleaved in a reaction involving the RecA protein, and the SOS genes are induced. Figure reproduced from [38].

Replisome, RecA binds as nucleoprotein filaments in two steps. The first, and slowest, step involves nucleation of a small number of RecA monomers on the DNA and is followed by a rapid, cooperative extension of the filaments in the 5′–3′ direction [66]. Single-molecule experiments have shown that two monomers of RecA are sufficient for nucleation and that slow growth takes place in both directions, although faster in the 5′→ 3′ direction on the DNA [67].

RecA also functions as a co-protease. Rapidly after formation of the RecA nucleoprotein filament (denoted as RecA*), it catalyses auto cleavage of the LexA protein, the repressor that controls expression of the SOS genes.

At a rate that depends on the amount of RecA* present in the cell after damage, the concentration of intact LexA repressor will gradually decrease. With the different affinities of LexA to the repressor sites of the various genes, the expression of these genes is carefully timed, with some genes transcribed well before others. Pols II and IV, for example, are expressed in the first minutes of the SOS response while *umuD* and *umuC*, the Pol V genes, are not expressed until ∼ 30 min after damage [6, 65].

1.3.3. SOS-inducible DNA polymerases

**Pol II**
DNA polymerase II is encoded by the gene *polB (dinA)* and is one of the SOS-inducible polymerases [6, 68, 69]. Its promoter displays the weakest affinity to the LexA repressor of all those controlling the DNA polymerase genes, as predicted by
the Heterologous Index (HI) [70, 71]. The HI for polB is close to that for dinB, the gene encoding for Pol IV, but is much higher than that for umuDC (HI_{polB}=12.55, HI_{dinB}=9.40, HI_{umuDC}=2.12). Lower HI values indicate tighter binding to LexA repressor and therefore a later expression in the SOS response. Courcelle and collaborators show in their microarray experiments that E. coli cells exposed to UV damage promptly de-repress polB and dinB while umuDC keeps repressed for at least 30 minutes [6].

Pol II is a member of the B family of polymerases and is the only TLS polymerase with the 3′–5′ exonuclease activity that enables proofreading. As such, it replicates DNA with relatively higher affinity than other TLS polymerases [72]. In normal conditions (no SOS), Pol II is reported to be present in the cell with 30-50 copies [73]. Upon induction of the SOS response, Pol II levels increase about 10-fold, to a level of 350 – 1000 copies per cell [7].

The precise role of Pol II remained obscure for years after its discovery in 1970. Today, Pol II is thought to participate in a number of pathways, such as bypass of abasic sites, repair of interstrand cross links, and repair of UV-induced lesions [74, 75]. In the context of SOS response and polymerase switching in the replisome, Pol II has been reported to be a key protein in replication restart, allowing a replisome stalled on UV-damaged DNA template to restart in an error-free manner. This activity could explain why Pol II confers greater fitness to fast growing E. coli cells [76, 77]. Despite its proofreading activity, Pol II displays error-prone TLS at certain lesions. The first report of Pol II TLS activity involved the bypass of a single N-2-acetylaminofluorene (AFF) guanine located within the NarI hotspot and resulting in -2 frameshift [75].

These aspects taken together, and many others that would require a deeper discussion that is outside the scope of this thesis, suggest a picture that Pol II acts as a backup for Pol III, exchanging places in the replisome via a polymerase switch mechanism. Such a model is supported by the observation that stalled Pol III replisomes are prone to disengage from the DNA template, favouring association of Pol II [27].

**Pol IV**

DNA Polymerase IV (Pol IV) and DNA Polymerase V are the two members of the Y family of TLS polymerases found in E. coli [27, 78, 79]. Y-family polymerases have the ability to accommodate bulky lesions and are able to replicate past damaged sites, thanks to their more open catalytic site. However, when copying undamaged templates Pol IV and Pol V display higher error rates than other polymerases. As a result, these polymerases play a major role in induced mutagenesis and stress-induced adaptive modification by allowing replication to continue at the cost of increased mutagenesis [79, 80].

Pol IV is encoded by the damage-inducible dinB gene and is constitutively expressed in high amounts in the cell (~ 200 copies) [81], increasing up to ten-fold upon induction of the SOS response. In contrast to the other polymerases constitutively expressed by the cell, Pol IV lacks 3′–5′ proofreading exonuclease activity and, as a result, displays relatively low fidelity [78]. The high Pol IV concentration in the cell, even without damage, raises several questions about its role in
1.3. DNA-repair mechanisms

DNA replication and cellular fitness [27]. Cells lacking Pol IV do not display a clear phenotype and even grow as efficiently as wild-type cells in rich medium. In the context of DNA damage, Pol IV activity is mostly related to -1 frameshift mutations, accurate bypass of N2-dG adducts, and TLS alkylation damage [79]. Upregulation of Pol IV allows it to compete and switch with Pol III HE, reducing significantly DNA replication velocity and increasing significantly mutagenic rates [82].

**Pol V**

DNA polymerase V (Pol V) is a heterotrimeric complex, composed of the UmuC and UmuD gene products. The active form, UmuD’2C, contains the catalytic subunit (UmuC) and two copies of the gene product UmuD’. UmuD’ is the product of RecA-mediated auto cleavage of UmuD. Pol V also lacks proofreading activity and is clearly an error-prone polymerase with TLS activity across a large number of different types of lesions [27]. Pol V is expressed in the SOS response later than the other TLS polymerases, with the transcription of *umuDC* being delayed for at least 20 minutes and Pol V achieving maximum cellular concentration around 45 min after induction [6]. Pol V is the major TLS polymerase in *E. coli*, able to bypass a much wider variety of lesions than Pol II and Pol IV. Pol V requires a number of activation steps before it acquires TLS activity. These activation mechanisms will be discussed in great detail in Chapter 3 [83].
1.4. Single-molecule in live cells

“Seeing is believing” is at the core of the pioneering work of Antonie van Leeuwenhoek, which fundamentally changed biology by allowing us to visualize microscopic structures and life forms. Since the first reported observations of microorganisms (animacula) in the late 17th century, the microscope as a scientific tool has changed drastically: simple lenses have been replaced by objectives containing dozens of lenses; excitation sources evolved from a candle to high-intensity lasers; and the human eye cannot compete with the sensitivity and precision of modern electron-multiplying CCDs. A full historical review is outside the scope of this thesis; instead, I would like to refer the reader to a recent publication that nicely summarizes the evolution of microscopy over the last 300 years [84].

The observation of molecular processes inside living cells is key to our understanding of how biological systems work. For many centuries, Abbe’s diffraction limit has posed a challenge to microscopy. Diffraction theory states that the wave-like nature of light limits the resolution of a microscope. With typical optics, the smallest observable object has a size equal to half of the wavelength of the light used to observe it (Fig. 1.7).

![Figure 1.7 | The diffraction limit. Abbe was the first to describe that the smallest observable object in a microscope has a size of no less than 0.2 µm when using visible light. Figure reproduced from Johan Jarnestad/The Royal Swedish Academy of Science.](image)

The ability to visualize the fluorescence of individual molecules has catalysed the development of methods that overcome the resolution barrier. Boris Rotman developed methods to sensitively detect the activity of β-galactosidase by fluorescence and demonstrated the potential for the observation of single molecules in an aqueous environment [85]. Not until 1989, W.E. Moerner observed for the first time the optical absorption profile of an individual molecule at liquid-helium (4 K) temperatures [86]. In 1995, Eric Betzig published a theoretical proposal to use similar single-molecule imaging as a basis for a method to visualize structures at a length scale smaller than the diffraction limit [87, 88]. In combination with the development of the ability to switch the fluorescence of a single molecule on and off [89], this approach led to the birth of a number of very powerful imaging approaches that to overcome the diffraction limit [90–92]. In parallel, Stefan Hell developed a method based on stimulated emission to quench the fluorescence of molecules sur-
1.4. Single-molecule in live cells

... rounding the center of a laser spot. This technique, later called stimulated emission depletion (STED), removes the fluorescence of the outside of a diffraction-limited spot and results in a spot size that is much smaller than the diffraction limit [93]. Moerner, Hell and Betzig received the 2014 Nobel Prize in Chemistry for their contributions to the development of super-resolution microscopy.

1.4.1. Ensemble versus single molecule studies

Single-molecule based super-resolution microscopy directly exploits the observation of individual molecules and provides astonishing pictures of biological phenomena. Single-molecule microscopes allow us to observe and follow an individual molecule in the crowded interior of a cell. The ability to follow an individual molecule inside a living cell opens up tremendous possibilities in the study of the molecular processes that underlie life.

During decades of advances in biochemistry, molecular biology, cell biology, and genetics, we have accumulated enormous amounts of data and molecular insight. Much of this knowledge is obtained by “ensemble” approaches. In ensemble experiments, a measurement of a molecular property represents the average behaviour of billions and billions of individual contributions. Capturing averaged behaviour comes at the cost of hiding molecular heterogeneity and thus not completely representing the system. For instance, when assuming that solutions are in equilibrium and stable states are prevalent, ensemble experiments are often not sensitive to subpopulations corresponding to states with a short lifetimes and thus providing a biased observation.

Single-molecule studies represent an approach in a way opposite to ensemble measurements. The goal is to disentangle the heterogeneity and follow individual molecules throughout the experiment. By doing so, subpopulations can be separated and information about molecular heterogeneity and synchronicity be obtained. Furthermore, biomolecules often work in the energetic regime corresponding to a few $k_B T$. This means that Brownian motion has a significant effect that can directly influence molecular behaviour and induce heterogeneity. A good example is the F1-ATPase molecular motor, which acts as the smallest rotary engine in the cell and works with efficiencies close to 100% by making use of thermally induced random effects [94]. The direct observation of such molecules in action at the single-molecule level is critical to their mechanistic understanding.

1.4.2. Test tube versus live cells

Single-molecule imaging as a research field is relatively young, especially applied to the visualization of biomolecular processes in live cells. The next step in the biophysical and biochemical tools that aim to study single molecules is to perform such measurements inside living organisms. *In vitro* experiments based on aqueous solutions of purified proteins are often significantly better controlled, usually provide better signal-to-noise ratio, less contamination and are easier to handle. Experiments involving live cells (*in vivo* studies) provide a much more complex environment for reactions and cells cannot be faithfully reduced as test tubes yet.
Single-molecule imaging in cells adds many uncontrolled factors: variation in the subpopulations, asynchronicities in the cell cycle, uncertainty in the amount of reactants taken up by the cell, and high sensitivity of the imaging to contamination.

The first in vivo single-molecule imaging was reported in 2000. Yanagida and colleagues investigated signal transduction from the epidermal growth effect receptor (EGFR) in human carcinoma cell and suggested that EGFR could exist in different states [95]. At a similar time, Schindler and colleagues reported 3D imaging of ion channels using fluorescently labelled toxin ligands at the plasma membrane of T-cells with 40 nm resolution [96]. A few years later, the bacterial actin homolog tagged with a yellow fluorescent protein was imaged at the single-molecule level [97].

The fast development of single-molecule techniques is changing the way we visualize cells and study biomolecules [23, 98]. Nevertheless, there are still fundamental challenges to overcome. Many studies exploit the fusing fluorescent proteins (FPs) to the protein of interest with well-developed molecular biology methods [83, 99–101]. The main advantage of such an approach is that bacteria can directly synthesize the tag within the target protein, removing the need of fixation followed by staining. More recently, tools have become available that allow the introduction of unnatural amino acids with fluorescent properties, but major changes in the cellular background are needed [102].

When imaging single molecules in live cells, three main categories of molecular behaviour are expected: molecules bound to structures in the cytosol, molecules diffusing through the cytosol, and molecules in the membrane. These different behaviours are identified by determining the diffusion properties of the tracked molecule (Fig. 1.8). Even though the precise identification of single-molecules trajectories in live cells remains a challenge due to photo physical limitations [103], this classification represents the major methodological framework within which the single-molecule imaging described in this thesis has been done.

1.4.3. In vivo single-molecule studies of DNA replication

Genome processing lies at the core of cellular metabolism and represents a rich area for single-molecule studies. DNA replication, for example, involves several protein components that are found in relatively low copy in cells [99] and thus is ideally suited to the high sensitivity of single-molecule fluorescence imaging. The events corresponding to these protein factors binding to and unbinding from DNA result in starkly different diffusive behaviours and the spatial geometry of the bacterial nucleoid enables the visualization of replication complexes moving through the cell.

The first single-molecule visualization studies of the bacterial replisome within live cells were reported in 2008, where the authors tagged replisomal components with the bright fluorescent protein YPet. They observed that replisomes always assemble and initiate replication at the oriC region in the bacterial chromosome, irrespective of where oriC is located in the cell. This finding changed the view of replication happening in “replication factories” in E. coli, which were considered fixed structures for DNA replication [104]. The authors showed that the replisome
1.4. Single-molecule in live cells

**Figure 1.8 | Imaging single molecules in live bacteria.** A – Profile of the flow-cell device used in the experiments reported in the chapters 4 and 5 of this thesis. Bacteria attach to a glass surface that is treated with (3-aminopropyl)triethoxysilate (APTES, Sigma). The quartz piece on top of the flow cell allows for in situ UV irradiation as a means to induce DNA damage. B – Bacteria adhere to the treated coverslip surface by electrostatic interactions. C – Three main molecular behaviours are expected: freely diffusing proteins cannot be precisely localized and give rise to a blurred signal; proteins binding to cellular structures are much more static on the timescale of the frame duration and thus yield low diffusion coefficients; and proteins interacting with the cellular membrane also giving rise to reduced diffusion coefficients.

Position is highly dynamic and that the movements of the bacterial chromosome during cell cycle are linked to the replication process. These findings argue against an existing structure restricting the replication machinery to a certain region of the cell [104]. Further, a precise characterization of the replisome subunit stoichiometry was obtained with single-molecule microscopy. This work strongly supports a model in which three copies of Pol III are associated with the clamp loader, in contrast to the long-standing model of two polymerases, one for leading-strand synthesis and other for lagging-strand synthesis [99].

Aspects of DNA repair have also recently been investigated by single-molecule microscopy. The visualization of individual DNA Polymerase I inside single *E. coli* cells enabled the characterization of the mechanisms underlying the binding to DNA primers, the identification of miss-incorporated nucleotides, and the synthesis of DNA synthesis by Pol I [100]. Even though precise reaction kinetics are yet to be fully characterized *in vivo*, the authors investigated the spatial organization of and interactions between Pol I and DNA ligase by attaching fluorescent proteins to each protein and performing two-color single-molecule imaging.

The studies described above have played an important role in the development of single-molecule microscopy tools to study the dynamics of DNA polymerases inside living cells and have paved the road for a further development of these techniques. Single-molecule microscopy in live cells has the potential to enormously improve our current understanding of molecular processes and even change our thinking on how bacteria behave. Although relatively young, single-molecule microscopy is rapidly establishing itself as an important field, providing contributions that challenge our current understanding about biological processes [23, 98, 105].
1.5. Aims of the project

This thesis focuses on establishing methodology for single-molecule imaging of TLS polymerases in live *E. coli* cells, with a particular focus on Pol V. At the outset of my thesis work, a number of basic properties of Pol V were poorly understood, such as expression levels and localization upon DNA damage. With the ability to produce *E. coli* strains expressing UmuC-mKate2 (Pol V) and DnaX-YPet (τ-subunit of the replisome), we set out to answer the following questions:

1. Can we observe and monitor at the single-molecule level the behaviour of TLS polymerases inside living cells? As reported, DNA Polymerase V is hardly present in undamaged cells [6, 65]. Can we visualize the behaviour of individual Pol V molecules upon induction of DNA damage and the SOS response?

2. Previous reports of the cellular concentration of Pol V after DNA damage are based on Western blots and other ensemble measurements [65], which hide cell-to-cell variations. Can we use single-molecule imaging to arrive at an accurate assessment of Pol V expression levels at the single-cell levels? What are the factors that influence Pol V cellular localization? Pol V is dependent on UmuD' and RecA to be active [38]. Does Pol V cellular localization depend on its activation state?

3. Can we co-localize Pol V and the bacterial replisome? Does Pol V access the DNA independently of the replisome?

4. Can we observe the effects of polymerase competition, as reported by Fuchs and Fujii [30] inside living cells? How does the presence of Pol II and Pol IV affect the localization/behaviour of Pol V?

5. The field of live-cell single-molecule imaging is young and large part of the methodology not well developed. What tools do we need to develop to acquire and analyse the large amounts of fluorescence images needed for the single-molecule studies? How do we implement existing single-molecule image processing strategies to cope with the extra requirements of single-molecule imaging in live cells? (i.e. hierarchical organization of data; subdivision of images into cells; cellular movement and growth);

1.6. Scope of this thesis

The research underlying this thesis was conducted in the Single-Molecule Biophysics group at the University of Groningen. The research focus is twofold: the development of single-molecule imaging techniques for imaging of live *Escherichia coli* cells and the use of these tools for the mechanistic investigation of TLS polymerases, particularly DNA Polymerase V, in the context of the SOS response to DNA damage.

A brief summary of the content of this thesis is presented below:
1. Chapter 2 presents the development of an open-source software platform developed as a plugin for ImageJ. iSBatch (in Singulo Biology Batch tools) aims to aid image processing and storage of highly hierarchical datasets. Most of the analysis performed in this thesis took advantage of this software package.

2. Chapter 3 addresses a protein localization artefact found when imaging Red Fluorescent Proteins (RFPs) at the single-molecule level inside *E. coli*. We found that lower concentrations of RFPs induce association with the membrane, leading to biased localization and potential for miss interpretation of biological phenomena.

3. Chapter 4 presents a new finding in the regulation of DNA Polymerase V mutagenesis activity. We identified a spatial and temporal localization in bacteria and discuss the consequences of the Pol V sequestration by the inner membrane for the cell;

4. Chapter 5 challenges the current model of DNA polymerase exchange. We investigate how the deletion of TLS polymerase influences Pol V localization and Pol V access to DNA. Current models focus on protein competition based on mass-action principles while we demonstrate that although present, mass-action is a secondary actor in the exchange of polymerases, at least in relation to Pol V.

5. Chapter 6 provides a general conclusion of this thesis and a brief outlook on the implications of our findings.
References


References


