Conflict-prone Genomes of Bacterial Pathogens During Infection

By

Kaitlyn Rose Browning

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

May 10th, 2024

Nashville, Tennessee

Approved:

Neil Osheroff, Ph.D.

Breann Brown, Ph.D.

Kathy Friedman, Ph.D.

Borden Lacy, Ph.D.

Houra Merrikh, Ph.D.

DEDICATION

To my family, who always knew I could do it,

and

To Lance, whose steadfastness sustained me.

ACKNOWLEDGEMENTS

This dissertation would be incomplete without first thanking those who helped make this work possible.

First and foremost, I would like to thank my mentor, Dr. Houra Merrikh, for energetically welcoming me into the lab all those years ago. Thank you for being the risk-taker I could never dream of being, and for showing me that uncertainty is a part of science – and life. Houra never doubted my steadiness, we both probably depended on it during difficult times, and I built confidence as a thinker and problem solver because of her appreciation for my level-headedness. Despite my stubbornness, she taught me the importance of flexibility and adaptability. Because of her philosophy, my talks improved as did my writing. I learned a great deal from her and grew as much as a person as I did a scientist.

Secondly, I'd like to thank all members of the Merrikh lab – past and present – whose contributions to my training are immeasurable. I learned the most from Kevin Lang, who not only arguably taught me every technical skill I used since day one, but who also believed in me to the point that I finally believed in myself. Ideas from my colleagues Juan Carvajal Garcia, Oyku Sensoy, Angel Joel Hernandez Viera, Dan Pers, Harrison Bracey, Sharath Narayan, and Taylor Thames undoubtedly improved my work and helped me achieve success. Thank you all for our discussions and for hearing me when I needed to vent. Lab was fun because of you all.

I also want to acknowledge those colleagues outside the Merrikh lab whose technical expertise truly made my experiments and data analysis possible. Advice from Kavi Mehta was seminal in the early development of my project and I deeply appreciate his help and insight. Anna Johnson sat with me at a computer for at least two months over one summer, setting aside her own work to help me write an entire ChIP sequencing data analysis pipeline from

scratch. Without her help, the process would have taken months longer and definitely would not have been as tolerable. I also want to thank Tyler Hansen and Lindsey Guerin for their technical expertise in writing and troubleshooting code and sequencing data analysis.

Much of this work would also not have been possible without the friendly, collaborative environment of the Biochemistry department. I appreciate every lab who shared equipment and resources throughout my project, especially the Cortez lab for sharing their Bioruptor sonicator that I used for every PIC-seq experiment. I want to also acknowledge the resources at Vanderbilt that contributed significantly to this project: VANTAGE and the Molecular Cell Biology Resource Core.

Aside from technical help, being a part of the Biochemistry Department was a highlight of my time in graduate school and I treasure the relationships I made. Thank you to all my friends who have made this a fun and exciting place to be and who never failed to check on me: Jessica Taylor, Jessica Collins, Kate Clowes, Jenny Tran, Iva Chitrakar, Tyler Hansen, Sam Lisy, Lily Campos, Tata Kavlashvili, Robert Mann, and so many more. I also thank fellow members of the Biochemistry Postdoc and Student Association for fostering such a thriving, inclusive culture and thank the faculty for being so supportive of trainees.

I also express gratitude to the members of my thesis committee, Drs. Neil Osheroff, Breann Brown, Kathy Friedman, and Borden Lacy for keeping me on my toes, expecting highly of me, giving me their time, and for their insightful suggestions. I sincerely appreciate the mentorship of Dr. Neil Osheroff and Dr. Kathy Friedman, who were both always willing to meet with me when I needed extra guidance and who both provided solid ground upon which to stand during hard times.

Thank you also to my previous science mentors, specifically Dr. Karen Buchmueller, who fostered my interest in scientific research and gave me the confidence and enthusiasm to

pursue this degree in the first place. It will never be lost on me the contribution each of my Furman chemistry professors made to getting me to this point. I wouldn't be the scientist I am today without that foundation.

Of course, I am indebted to my friends and family for their love and constant support. Thank you to two of the best friends I have ever known, Yelena Perevalova and Anna Johnson. I can say with confidence I could not have survived graduate school and become who I am today without your friendship, love, and pep talks. Yelena, your passion for science inspires me but your passion to be a good friend inspires me even more. Your encouragement, kindness, and true friendship are absolute treasures. Anna, your mind fascinates me and drives me to become a better, deeper, and wholistic thinker. I can discuss things with you in ways I cannot with anyone else and I have truly grown as a person because of your friendship.

To both of my parents, thank you for instilling in me the willpower and determination I needed to complete all this schooling, from start to finish. Thank you for never questioning my abilities and never doubting whether I could achieve whatever I set my mind to. I suppose I should also acknowledge you both for the stubbornness and hard-headedness that probably got me through my worst days – it finally came in handy! Thank you to both of my grandmothers who, whether they understood my work or not, understood when I needed someone to listen to my struggles and provided loving encouragement. Thank you to all my aunts, uncles, cousins, and my brother who all also never doubted my perseverance. I want to specifically thank Austin and Shelby for being so supportive upon our move to Nashville and providing much needed comic relief throughout both college and graduate school. I am also grateful to my loving in-laws, who took me in as their own and stayed engaged and invested in my progress from day one. I want to also thank my little sister, MaryBeth, who always knew when I needed a little extra love and encouragement and who I knew always believed in me.

Most of all, I would like to thank my husband, Lance. Not only did he never doubt my capability, tenacity, or potential, but he never let me doubt them either. His support was steadfast; he guided me through intense bouts of self-doubt and refused to let me give in to my worst feelings. He knew, despite my sarcasm, that he had to have enough optimism for the both of us. I depended on his perspective when I couldn't find my way. He makes life fun and there was always relief in the smiles he brought to my face. Thank you so much for each immeasurable way you helped me reach this accomplishment. My success is yours too.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CFU: colony forming units

ChIP: chromatin immunoprecipitation

d ratio: the normalized read counts mapping to a gene in the backtracking-prone state (Δ*greA* Δ*greB*) versus those mapping to the same gene in in WT

d' ratio: the ratio of *d* for infection and *d* for broth, $d' = d_{inf} / d_{broth}$

h p.i.: hour post infection

PIC-seq: post-infection ChIP followed by deep sequencing

RNAP: RNA polymerase

RpoB: the β subunit of RNAP

CHAPTER 1

INTRODUCTION

The Central Dogma[1]

Cells store and maintain their own genetic information and inherently possess the ability to use that genetic material to respond to their environments, communicate with each other, maintain homeostasis, and replicate themselves. This process, known as the central dogma of molecular biology, governs gene expression. It is comprised of three major steps. First, the genetic material, stored as DNA, is replicated to form identical daughter DNA molecules that are divided equally between the two daughter cells. Second, parts of that genetic material are read and transcribed into RNA. Finally, the genetic messages stored in those RNA molecules are translated into polypeptides with distinct amino acid sequences, called proteins, which ultimately carry out functions encoded in the genetic material (1). The following sections describe each process in detail.

DNA: storing and replicating the genetic material

The essence of the genetic information stored in DNA resides in the linear order of the four nucleotide building blocks that compose each strand of the double-stranded molecule. The four available nucleotide bases are adenine (A), thymine (T), guanosine (G), and cytosine (C). Thus, like long words composed from a four-letter alphabet, the order of these nucleotide bases

[1] Sections of Chapter 1 are adapted from the following publication: Browning, KR, Merrikh, H. 2023. Replication–Transcription Conflicts: A Perpetual War on the Chromosome. *Annu. Rev. Biochem*. In press.

conveys genetic codes, which are stored in the “book” that is the genome (the complete set of DNA in a cell). In most bacteria, the DNA is consolidated into a single, circular chromosome. (1).

The backbone of these two polynucleotide strands is composed of a string of deoxyribose sugars linked through phosphate groups. The strands are complementary, intertwined, and held together by hydrogen bonds between the bases: A pairs with T and G pairs with C. This configuration forms the double helix structure with which DNA is synonymous. The way the nucleotide bases are linked together along the backbone lends inherent polarity to the two DNA strands, such that the 5’ end is defined by an exposed phosphate group (linked through the 5’ position of the ribose ring) while the 3’ end is defined by a free hydroxyl group (the substituent on the 3’ position of the ribose ring). Importantly, many of the proteins that transact with DNA have defined polarity restrictions themselves and may only interact with a DNA strand in one direction or the other (1).

To pass these genetic codes to a daughter cell, the parental cell must replicate the DNA. This process is highly accurate, precise, and efficient. DNA is replicated semiconservatively, whereby the parental duplex DNA is separated, forming a replication fork structure, and each strand serves as the template for synthesis of a new, complementary strand. DNA synthesis during replication occurs in only one direction, 5’ to 3’ (Fig. 1.1). Accordingly, one strand of the duplex, known as the leading strand, is synthesized continuously, while the other strand, known as the lagging strand, is synthesized discontinuously. Synthesis of the leading strand proceeds in the same direction as the replication fork movement. In most bacteria, DNA synthesis initiates at a single site (*oriC*) and continues bidirectionally around the chromosome until both forks converge at the replication terminus site (*ter*) (1).

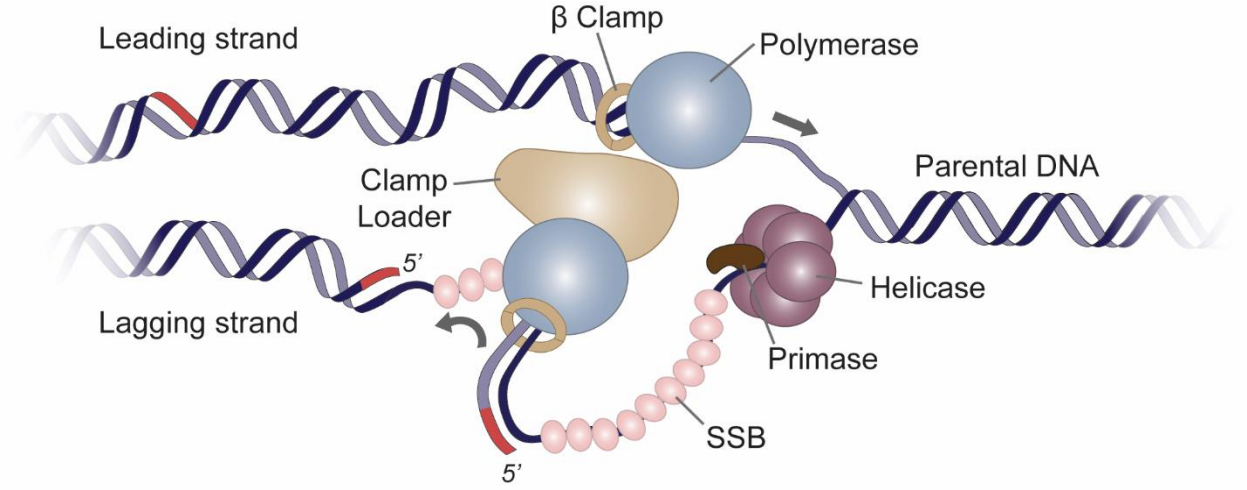

Figure 1.1. Schematic of a single bacterial replication fork. A helicase (magenta) unwinds the parental DNA duplex. Single stranded binding proteins (SSB, pink) protect single stranded DNA. The clamp loader complex (tan) loads the β clamp processivity factor (tan rings), one per polymerase. Two polymerases (blue) synthesize nascent DNA starting at RNA primers (red) which are laid down by primase (brown). The direction of polymerization is indicated by grey arrows.

Many proteins are involved in replication; these proteins collectively make up the 'replisome'. Included here are only the proteins involved in bacterial DNA replication, though eukaryotic DNA replication is very similar. Briefly, replication requires a helicase to unwind the DNA duplex, a β clamp and β clamp loader proteins to improve processivity, a primase to lay down RNA primers, polymerases to synthesize DNA, a ligase to connect pieces of newly synthesized DNA, and single-stranded (ss) binding proteins to protect ssDNA. Collectively, this complex of proteins is known as the replisome. There is one replisome per replication fork, each of which generally moves along DNA as a single unit (1).

RNA: transcribing the genetic material

A cell's DNA contains all its genetic information but is not directly the template for polypeptide synthesis. Instead, a second nucleic acid structure, messenger RNA, is synthesized from template DNA in a process known as transcription. This messgner RNA molecule not only contains genetic information, but also obtains its genetic specificity directly from the DNA (1).

Transcription is the first step in gene expression. The process is similar to DNA replication in that the duplex is separated and a polymerase, the RNA polymerase (RNAP) holoenzyme, synthesizes the messenger RNA molecule (Fig. 1.2). Each component of the bacterial holoenzyme (as determined in *Escherichia coli*) serves a distinct purpose; the β and β' subunits (encoded by the *rpoB* and *rpoC* genes, respectively) form the active site for polymerization. In contrast to DNA replication, transcription only selectively copies gene sequences into RNA, rather than the entire chromosome. Additionally, hundreds of RNA molecule copies of a gene may be synthesized at a given time. Accordingly, transcription is highly regulated. RNAP initiates transcription at a promoter sequence, synthesizes an RNA copy of the gene during elongation, and then terminates transcription. Like DNA replication, RNA polymerization only proceeds in the 5' to 3' direction. Because genes can be encoded on either strand of the DNA, RNAP can physically occupy either strand, depending on the gene (1).

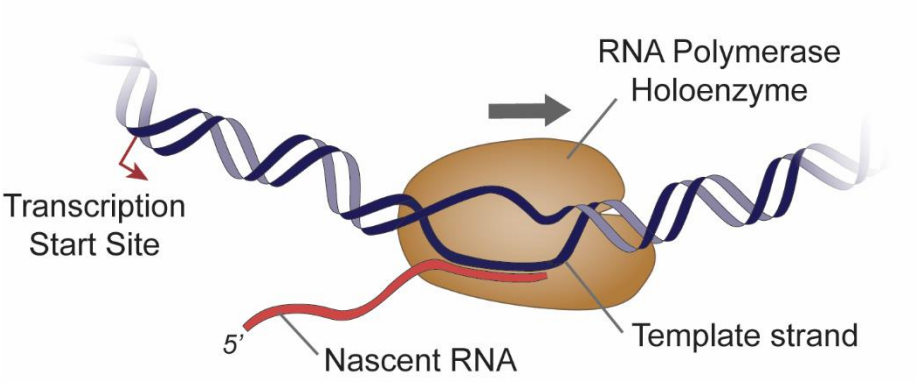

Figure 1.2. Schematic of the transcription bubble. Transcription initiates at a start site (red arrow). RNA is polymerized in the 5' to 3' direction along the gene and the nascent RNA (red) is extruded behind the polymerase

Protein: translating the genetic material

Cellular functions are dictated by the genetic information encoded within the cell. The final process of carrying out gene expression is the translation of this genetic information into proteins, which execute cellular functions. Messenger RNAs (mRNAs) containing gene-coding sequences must be interpreted to generate amino acid sequences that comprise proteins. The mRNA serves as a template for this polypeptide synthesis. The mRNA is translated three nucleotides at a time at ribosomes, with each triplet sequence specifying an amino acid. In this way, mRNA acts as the intermediary of the central dogma, using information from the genetic code to specify the amino acid sequence of a functional protein. Interestingly, there are other RNA molecules, such as ribosomal RNAs or transfer RNAs, that can fulfill biological functions without first having been translated into protein (1).

Overview of Replication-Transcription Conflicts in Bacteria

Efficient and accurate genome duplication is essential for all living cells. Effective and precise gene expression is equally necessary for cells to survive. However, these two processes do not occur independently from one another. Both the replisome and RNAPs use the same DNA substrate to carry out their functions at the same time. The rate of replication is more than 10 times faster than the rate of transcription (2), making encounters between the replisome and RNAPs inevitable. This lack of spatial and temporal separation leads to inevitable replication-transcription conflicts between the two machineries.

Because most bacteria have a single origin of replication, the orientation of each gene with respect to replication can be definitively determined. Genes encoded on the leading strand are transcribed such that the replication and transcription complexes move in the same direction, codirectionally along the DNA (Fig. 1.3). Although codirectional conflicts are disruptive (3, 4), they are less detrimental to cells than when the replisome meets RNAPs at lagging-strand genes, which results in head-on conflicts (Fig. 1.3).

The distinct differences in the outcomes of these two types of conflicts were identified early on using an inducible origin placed upstream or downstream of a heavily transcribed ribosomal RNA (rRNA) operon (5). The orientation-dependent effects of conflicts were later demonstrated to be a pervasive issue genome wide, not just within rRNA operons. Ultimately, the detrimental consequences of head-on conflicts were demonstrated to be much more severe, including replication stalling, replisome disassembly, and replication restart (5–16); pervasive DNA–RNA hybrid (R-loop) formation (11, 17); torsional stress at the conflict region (18); and increased mutagenesis (19, 20).

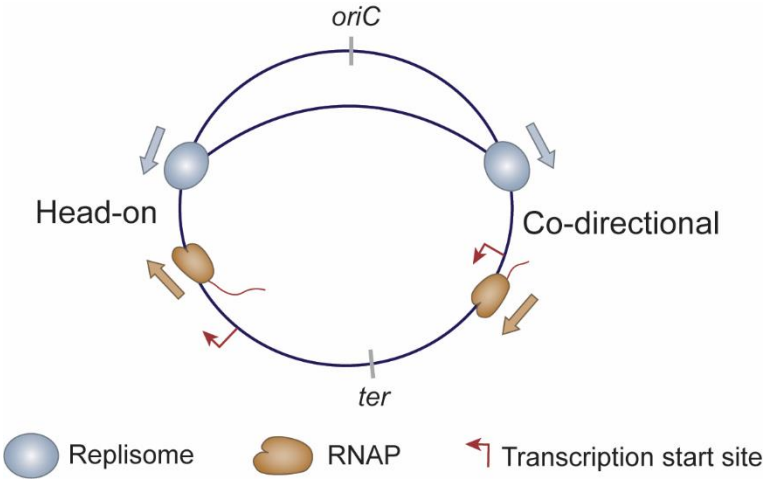

Figure 1.3. Replication-transcription conflicts in bacteria. Replication and transcription share the same DNA template substrate and often function at the same time. This lack of spatiotemporal separation in both prokaryotes and eukaryotes leads to conflicts between the two machineries, the replisome and RNAP. These macromolecular machines move in the same direction at leading-strand genes, resulting in codirectional conflicts, which are only slightly destabilizing to the genome. Head-on collisions between the replisome and RNAP occur at lagging-strand genes and lead to significant genome instability.

The Fate of Replication at Conflict Regions

Replication fork stalling

When the replisome meets an RNAP along DNA, especially one that is approaching from the opposing direction, it stalls. Early evidence from *in vitro* work with a reconstituted phage replication system demonstrated that the replication fork is completely blocked by a single head-on RNAP in the absence of additional helicases, and it is significantly slowed behind a codirectional RNAP (6, 21). These results are consistent with *in vivo* data: by placing an origin upstream or downstream of a heavily transcribed ribosomal RNA operon, French demonstrated that head-on conflicts drastically impede replication fork progression in *E. coli* (5).

This has been subsequently observed in other bacterial systems that are highly divergent from *E. coli* (9–11).

Numerous studies have attempted to identify the reason behind the severe stalling of replication forks at head-on conflict regions. Both replication and transcription require unwinding of the DNA template, generating positive supercoils ahead of the machineries (22, 23). This led to the model whereby buildup of positive supercoiling between the two machineries, rather than direct physical confrontation between macromolecular complexes, inherently stalls their progress and contributes to the severe consequences of head-on conflicts (5, 18, 24, 25). The majority of work now supports this model. A recent study in *Bacillus subtilis* demonstrated that two type II topoisomerases that resolve positive supercoiling, gyrase and topoisomerase IV (topo IV), preferentially associate with engineered head-on, but not codirectional, conflict regions (18). Furthermore, inhibiting these same topoisomerases led to further accumulation of the replicative helicase at the conflict regions when the gene was transcribed head-on but not codirectionally, suggesting that it is indeed the buildup of unresolved positive supercoiling during head-on conflicts that exacerbates replisome stalling (18). Accordingly, gyrase and topo IV are key conflict resolution factors.

The additive effects of positive supercoiling between the two machineries are seemingly minimized in codirectional conflicts. Whereas positive supercoils accumulate ahead of RNAP, negative supercoils are generated behind it (22). Accordingly, in a codirectional encounter between replication and transcription, the local positive supercoiling generated ahead of the replicative helicase would likely be neutralized by the negative supercoiling generated behind active RNAPs (5, 7, 18, 25). This is consistent with *in vitro* and *in vivo* data in bacteria that demonstrate codirectional conflicts do not stall the replisome to the same extent as head-on conflicts (3, 5–11).

Nevertheless, replication fork stalling is a significant consequence of replication-transcription conflicts and is often followed by replisome disassembly, recombination events, and eventual replication restart. Fork stalling and replication restart events outside of the origin of replication are unusual, except in terms of disruptions to replication such as replication-transcription conflicts. Therefore, these events, specifically as they occur outside of the origin of replication, provide an experimental opportunity to study conflicts, as has been done previously (3, 15, 26).

Replisome disassembly

Given that conflicts stall the replisome, the question remains as to what happens to the replication machinery during these encounters. One possibility is that after stalling, the replisome disassembles. Evidence from genetic and molecular studies supports this model. First, replication restart proteins that help rebuild the replisome outside of the origin independently of sequence are essential across diverse bacterial species (27–29), suggesting that, to some extent, the replisome frequently disassembles and reassembles during replication. Second, when this model was tested in living cells at single-molecule resolution, Mangiameli et al. demonstrated that the replisome stalls and disassembles several times in both gram-positive and gram-negative bacteria (13). In fact, it was estimated that at least five disassembly events occur during exponential growth. Notably, this study directly linked the observed discontinuity of replication to encounters with transcription: inhibition of transcription and/or unstable RNAP mutants eliminated the observed destabilization of the replicative helicase and clamp loader complexes and increased the rate of replication.

Replication restart

Stalled or collapsed forks must be restarted to maintain cell viability, especially in bacteria, as they have only a single replication origin. Indeed, it is well established that replication restart proteins are recruited to conflict regions, including essential Pri proteins, the replicative helicase (DnaB in *E. coli*, DnaC in *B. subtilis*), and helicase loader proteins (DnaC in *E. coli*, DnaB and DnaD in *B. subtilis*), as well as accessory helicases such as Rep and UvrD (3, 15, 26, 30). Growth defects in *B. subtilis* cells lacking PriA are rescued by mutations in RNAP that decrease the stability of this complex, supporting the model in which PriA enables fork resumption at conflict regions (31).

Recombination

In addition to the restart mechanisms just described, there is a growing body of work suggesting that recombination is intimately tied to replication restart and ultimate conflict resolution. RecA, a key bacterial recombinase, may be able to protect reversed forks from transcription-dependent damage (12, 15, 23). Though the exact role of RecA is debated, genetic and molecular studies in *B. subtilis* show that RecA functions in the same pathway as restart proteins at conflict regions, and that RecA is required for the recruitment of restart proteins to the conflict region (15). Nevertheless, it remains unclear whether the involvement of RecA is to facilitate recombination events, replication fork reversal, and/or replication restart.

Other recombination proteins have been implicated in replication restart and conflict resolution. The helicase–nuclease complex AddAB was found to be a prerequisite for replication restart in *B. subtilis* (13). Similarly, in *E. coli*, inversion of a highly expressed ribosomal operon rendered the double-strand repair protein complex RecBC (AddAB homolog) essential for viability, which could be alleviated by introducing destabilizing mutations in RNAP (12, 26).

Processing of the regressed fork by RecBCD/AddAB, RuvABC (Holliday junction resolvase), and RecG can create a DNA substrate for PriA-dependent replication restart (30, 32). PriA ultimately initiates loading of the replicative helicase, DnaB (DnaC in *B. subtilis*), after which the rest of the replisome is reassembled (33).

The Fate of Transcription at Conflicts Regions

RNAP clearance

The fate of RNAPs during an encounter with a replication fork has been the subject of intense research. Several lines of evidence in bacteria suggest that replication through a transcription unit results in RNAP eviction, regardless of the orientation of the encounter. This phenomenon was first observed in an early EM study in *E. coli*, where RNAP complexes transcribing in the same or opposite direction of replication were lost once the fork passed through the highly-transcribed region (5). Later studies using an *E. coli*-based reconstituted system demonstrated that the replisome itself is capable of evicting RNAP, likely by destabilizing the complex. This occurred in both the head-on (14) and codirectional orientations (10, 34). In the latter case, the replisome appears to be able to use the remaining abandoned mRNA still bound to the template DNA strand as a primer for leading-strand synthesis, at least *in vitro* (10, 34). This replisome-mediated RNAP removal is assisted by accessory factors (discussed below) and promotes replication progression beyond the conflict region.

RNAP backtracking and conflicts

Not only can the replisome catch up to RNAP along the genome, necessitating the action of accessory helicases to remove the RNAP impediment (discussed below), but RNAP itself can stall along the genome, exacerbating conflicts with the replisome. It is well documented that obstacles on the DNA template such as damaged template bases or DNA-protein adducts stall transcription elongation complexes (35–38). Backtracking is the most common type of RNAP stalling (39). It can be brought on by DNA damage to the template strand and/or regulatory gene regions (24, 39). Backtracking involves the physical translocation of RNAP backward along the template DNA strand and extrusion of the 3′ end of the nascent transcript, resulting in an immobilized RNAP that must be enzymatically resolved for transcription to resume (39).

Due to the stability of the backtracked state, it is unsurprising that backtracking has been linked to replication-transcription conflicts (24, 40). In fact, backtracking may help explain some of the negative outcomes of conflicts: in *E. coli*, backtracking leads to replication-dependent and R-loop-dependent double stranded breaks (DSBs) at codirectional conflict regions (4, 10). Factors that resolve backtracked RNAPs are thus critical for conflict mitigation and resolution. Most bacteria possess two such direct antibacktracking factors, GreA and GreB, which promote the endonucleolytic cleavage of the RNA transcript and reactivation of backtracked RNAPs (41). Without such factors, *in vivo* replication within highly active transcription units is blocked (42). Not only does backtracking impede replication and lead to genome instability, but factors that rescue backtracked RNAPs or prevent stalls in the first place are critical for mitigating replication-transcription conflicts.

Conflicts and R-loops

Head-on replication-transcription conflicts are significantly more detrimental to cells than codirectional conflicts, suggesting there are important differences in how the replisome is disrupted when transcription proceeds in the same or the opposite orientation. R-loops are one of the key components of these differences. R-loops are three-stranded DNA:RNA hybrid structures formed when the nascent RNA transcript reanneals to the template DNA strand behind RNAP, leaving the non-template strand as single-stranded (ss) DNA (43).

Head-on conflicts promote the formation and accumulation of R-loops at the region of the encounter (11, 17). Without resolution, R-loops were found to completely block replication at head-on conflict regions and induce distinct DNA damage repair mechanisms (17, 18, 44). R-loop formation leads to chromosome rearrangements, recombination, double-stranded breaks (DSBs), and other markers of genome instability (45–47). R-loops may also promote mutagenesis within head-on conflict regions: cotranscriptional R-loops promote stress-induced mutagenesis through DSBs in *E. coli* (48), and removal of R-loops (by the overexpression of RNase H, the enzyme that degrades the RNA portion of the R-loops) relieves conflict-induced mutagenesis in *B. subtilis* (11). R-loop structures are a significant by-product of conflicts and are likely a critical factor that could explain the more severe effects of conflicts in the head-on orientation.

Conflict Mitigation and Resolution

Conflict avoidance: gene orientation bias

In contrast to eukaryotes, replication and transcription in bacteria are largely unable to be discretely separated in time or space (49). However, a major hallmark of most bacterial chromosomes is that the majority of genes are oriented such that transcription occurs codirectionally with respect to replication (i.e., they are encoded on the leading strand) (50, 51), thus minimizing harmful head-on conflicts. Across bacterial species, this codirectional bias is correlated with gene essentiality and expression level (51–53). For example, 94% and 97% of essential genes in *B. subtilis* and *E. coli*, respectively, are oriented codirectionally (52). Bacterial rRNA operons are also universally oriented codirectionally (50, 54). Despite this bias, a large percentage of genes (25–45% depending on the species) remain in the head-on orientation (encoded on the lagging strand) (50). It is possible that these genes have either not yet randomly inverted to the codirectional orientation, or that they exhibit conflicts that pose less of a fitness cost (55). However, evidence from GC-skew analyses suggests an alternative model. This type of analysis can identify genes whose orientation has inverted over evolutionary time, based on the abundance of guanosine bases within the gene, as guanosine is naturally more abundant on the leading strand of each chromosome arm (56). GC-skew analysis of divergent bacterial species shows that a large percentage of head-on genes were originally oriented codirectionally, suggesting that there may be some evolutionary advantage to the head-on configuration, which could be related to increased chances of mutagenesis and adaptive evolution (discussed below) (56, 57). Nevertheless, the persistent codirectional bias, especially for highly transcribed essential genes, lessens the potential for detrimental head-on conflicts and the genome instability with which they are associated. It is noteworthy that transcription through ribosomal genes regions (rDNA), though codirectional, still slows replication (3, 4).

Accessory helicases

Despite the ways in which replication-transcription conflicts can be potentially minimized, conflicts are inevitable, and cells must be equipped with tools to resolve them. Many of these mechanisms favor the continuation of replication at the expense of transcription, which is not surprising given that transcription can easily be reinitiated once the replisome is cleared. Several mechanisms to help resolve conflicts have been identified. Accessory helicases capable of removing RNAP impediments are one family of conflict resolution factors.

Two *E. coli* helicases, Rep and UvrD, were identified as playing potential roles in conflict resolution. These helicases move along the opposite strand from the replicative helicase (the leading strand) and translocate 3′ to 5′. Cells lacking either *rep* or *uvrD* are viable, but deletion of both genes is lethal (58). Interestingly, *E. coli* lacking Rep take twice as long to replicate their DNA as wild-type cells, even though they also contain more replication forks (59), implying that Rep facilitates replication fork movement. It was later demonstrated that both Rep and UvrD are capable of removing model protein roadblocks from the DNA *in vitro* (60). Growth defects in Rep and UvrD deletion strains could be suppressed by mutations within RNAP that destabilize the complex, strongly suggesting that Rep and/or UvrD promote fork progression through highly transcribed transcription units (61). Consistent with this hypothesis, direct evidence of Rep- and UvrD-mediated removal of RNAP fork blockages was observed using plasmid templates for reconstituted replication and transcription in *E. coli* (62). Under conditions that induced RNAP stalling, addition of Rep and UvrD separately resulted in the production of full-length leading-strand products and a compensatory loss of truncated strands, independent of transcription orientation. This suggested that Rep and/or UvrD can remove stalled, replication-blocking RNAPs.

Many gram-positive species lack Rep and UvrD. Instead, PcrA, a protein with C-terminal homology to UvrD that is found across gram-positive species, appears to serve a role equivalent

to the overlapping functions of Rep and UvrD (63). Accordingly, ectopic expression of PcrA in *E. coli* rescues the synthetic lethality of *rep* and *uvrD* deletions (64), revealing that PcrA likely performs similar functions as Rep and UvrD. Consistent with this hypothesis, PcrA was found to associate with highly transcribed tRNA and rRNA regions as well as some protein-coding genes (26). Furthermore, it was demonstrated that the helicase/ATPase function of PcrA is vital for replication progression and resolution of conflicts (26).

DinG is another accessory helicase in bacteria whose role in facilitating replication at regions of conflict is more ambiguous than those of Rep, UvrD, or PcrA. Belonging to a different class of helicases, DinG moves along DNA which in the opposite direction to that of Rep, UvrD, or PcrA – from 5′ to 3′ – and translocates on the same strand as the replicative helicase (the lagging strand) (62). This difference in polarity implies that DinG cannot resolve conflicts in the same manner as Rep or UvrD (62, 65). Despite these differences, DinG appears to share an overlapping role with Rep and UvrD in facilitating DNA replication, as inversion of rRNA operons in DinG-null cells results in transcription-dependent cell death (61). This lethality can be suppressed by destabilizing mutations within RNAP, much like in cells lacking Rep or UvrD. Interestingly, this growth defect can also be suppressed by overexpression of RNase H, which degrades R-loops, suggesting that DinG could resolve conflicts by way of removing R-loops *in vivo* (61). Despite the somewhat ambiguous mechanism of DinG-mediated conflict resolution, DinG is nevertheless an important conflict-resolving accessory helicase.

R-loop resolution factors

As R-loops have major implications for replication-transcription conflicts and their outcomes, factors that modulate the dynamics of their formation and resolution are also ultimately linked to resolution of conflicts. R-loops are mainly resolved by RNase H enzymes

and accessory helicases. The RNase H family of ribonucleases are found in various forms across bacterial species. Helicases presumably resolve R-loops by unwinding the RNA from the template DNA strand. These mechanisms are poorly characterized *in vivo* in bacteria, but these proteins may serve as important backups to RNase H-mediated R-loop degradation. In bacteria, concerted activity of the accessory helicases DinG, Rep, and/or UvrD (and PcrA in *B. subtilis*), may be capable of displacing R-loops and/or suppressing their formation, according to genetic studies (61, 66, 67).

The role of transcription-coupled repair in resolving conflicts

RNAP stalling, a known outcome of replication-transcription conflicts (discussed above), is an important determinant of and trigger for transcription-coupled repair (TCR). RNAP can be stalled by a DNA lesion and transcription can resume only once the lesion is repaired, which requires RNAP eviction (68). Therefore, TCR factors play a role in conflict resolution by modulating RNAP in the context of TCR. In bacteria, two different factors relevant to conflicts are involved in this context: Mfd and UvrD/PcrA. Mfd promotes the displacement of RNAP at the site of the lesion (69), while UvrD/PcrA is capable of dislodging the RNAP lesion and inducing backtracking (70).

Interestingly, there is a growing body of work that suggests the role of Mfd may not be limited to TCR (68). Notably, Mfd was found to facilitate replication restart at a head-on conflicts region in an *in vitro* system, which depended on its ATPase activity, suggesting that Mfd may help minimize conflicts by removing RNAPs ahead of an oncoming fork (14). It was also demonstrated that DSBs arising from codirectional conflicts could be relieved by Mfd (4). Deletion of *mfd* in cells containing inverted rRNA operons leads to only a marginal growth defect

(12), supporting the conclusion that conflict resolution mechanisms are largely redundant, with numerous available pathways.

The most profound effect of Mfd in cells, however, could be its role in promoting mutagenesis. In fact, lagging-strand genes accumulate more mutations than genes on the leading strand (discussed below) (19). According to work in *B. subtilis*, this is due to Mfd and its role in nucleotide excision repair (NER) (71). It was originally shown that Mfd functions in the same pathway as the NER protein UvrA as well as error-prone translesion synthesis polymerases at head-on conflict regions (71). However, it is unclear whether this activity is triggered directly by head-on conflicts, or indirectly by lesions inherently left behind by lagging-strand synthesis. More recently, it was determined that the mutagenic effect of Mfd lies in the same pathway as NER proteins, and that this pathway utilizes error-prone polymerases (72).

Pathological Consequences of Conflicts on Mutagenesis

As described, replication-transcription conflicts are a potent source of endogenous genomic instability. In bacteria, head-on conflicts increase mutagenesis in lagging-strand genes. This propensity of head-on conflict regions to accumulate mutations, however, may actually provide an evolutionary advantage to bacteria, for instance, by helping them accumulate mutations that contribute to the development of antibiotic resistance.

Head-on conflicts have been demonstrated to be significant drivers of mutagenesis in bacteria (19). This trend has been observed in different bacterial models using a variety of reporter genes inserted in either orientation at various chromosomal loci, suggesting that mutagenesis occurs at head-on conflict regions regardless of sequence or genomic context (19, 20, 73–75). In fact, head-on conflicts seemingly leave behind distinct mutation signatures,

characterized by indels and base substitutions (20). The mechanism by which mutagenesis occurs at head-on conflict regions is most likely through Mfd-mediated recruitment of NER factors, including error-prone translesion synthesis polymerases (19, 72).

Notably, the rate at which head-on genes across several divergent bacterial species accumulate nonsynonymous mutations exceeds the rate at which they accumulate synonymous mutations, suggesting that they are more frequently under positive selection (19, 56, 57). These data imply that the mutagenic potential of head-on conflicts could be beneficial. In other words, over evolutionary time, the head-on orientation of a subset of genes could be favored by positive selection (19, 49, 56, 57). This adaptive advantage of the head-on orientation is further supported by studies of gene inversion of divergent bacterial species existing in nature. In systematic analyses of different bacterial genomes over millions of years of divergence, Merrikh & Merrikh determined that almost all (89–96%) head-on genes in the investigated genomes were originally oriented codirectionally (56, 57), supporting the hypothesis that the head-on orientation may provide evolutionary advantages.

However, others argue that the presence of head-on genes is a result of inevitable, but rare, gene inversion events and that the distribution of genes in the codirectional versus head-on orientation is random (55). According to this model, genes that remain in the head-on orientation either have not yet been purified from the genome (negative selection) or exhibit replication-transcription conflicts that pose less of a fitness cost (55). Importantly, both models could be true, depending on the organism and the gene in question. It is likely that both negative selection and positive selection have shaped and continue to shape the landscape of lagging-strand transcripts in bacteria (57).

Interestingly, head-on genes in pathogenic bacterial species were found to be functionally enriched for antibiotic resistance, virulence, and stress responses (56). Most of these genes are transcribed only under certain conditions, generally minimizing potentially

damaging head-on conflicts. Assuming there is an adaptive advantage to the head-on orientation, which the above studies indicate, encoding key environmental response factors (such as virulence factors) in the head-on orientation could be a mechanism whereby cells are able to tweak their own evolution and facilitate adaptation.

On Replication-Transcription Conflicts in Eukaryotes

Although their molecular mechanisms are the same as that in bacteria, replication and transcription occur in a much different context within eukaryotic cells. In fact, the existence and frequency of replication-transcription conflicts in eukaryotic cells were initially debated, as both processes largely appear to be able to function separately in both time and space in eukaryotic systems. For example, unlike in most bacterial species, replication and transcription proceed at similar rates in eukaryotic cells (2), potentially minimizing collisions and their consequences. However, it has become clear that both codirectional and head-on conflicts also occur in eukaryotes (76–78). Specifically, there are situations where conflicts are inherently unavoidable regardless of the speed of the replication or transcription machineries, such as at genes that are too long to be transcribed in a single cell cycle (79) or in regions of high RNAP density (80). Transcription can disrupt replication in eukaryotic cells and induce recombination, a key biomarker of conflicts, much like it can in bacterial cells (7, 8, 81). Furthermore, recent evidence from various studies demonstrates that replication origins in human cells are enriched at transcription start sites (TSSs), all but guaranteeing that the replisome will encounter RNAP shortly after origin firing, possibly in both orientations (82, 83). Eukaryotic cells have similar mechanisms of conflict resolution to those in bacteria, including accessory helicases such as Rrm3 and Pif1 (84, 85), and R-loop modulating factors such as Sen1/sentaxin (86, 87). The

effect of conflicts on mutagenesis in eukaryotic cells is currently unclear, though dysfunctional conflict resolution does lead to pathological hallmarks of cancer (88). Conflicts are thus an unavoidable threat to cell viability and genome integrity in prokaryotes as well as eukaryotes.

Bacterial Pathogens and Conflicts

Despite extensive literature on conflicts in model bacteria such as *E. coli* and *B. subtilis* grown in laboratory conditions, there is a woeful lack of investigations into the presence of endogenous replication-transcription conflicts, the factors contributing to their detrimental effects, and their potentially dangerous consequences in non-model species residing in non-laboratory conditions. For example, bacterial pathogens such as *Listeria monocytogenes* and *Salmonella enterica* infect a variety of eukaryotic hosts, and It is unknown how this nutrient-deficient, stressful environment might affect the genome of the bacterial cell in terms of replication-transcription conflicts.

As discussed above, given that head-on (lagging-strand) genes were found to be enriched in functions related to stress response, virulence, and antibiotic resistance (56), and given that these genes rapidly accumulate mutations and are often under positive selection (provide evolutionary benefit to the cell), it is possible that head-on conflicts could increase the pathogenicity of bacterial cells through accelerating evolution. In other words, in the context of a bacterial pathogen, head-on virulence factors may drive their own evolution through mutagenic head-on conflicts. These conflicts would presumably become more frequent during infection, as head-on virulence factors, which are normally silent, are highly upregulated in response to a eukaryotic host. Currently, this model is untested, as it is unknown whether replication-transcription conflicts occur within a bacterial pathogen in the context of infection at all. How

molecular transactions along the genome of a bacterial pathogen are impacted by a eukaryotic host during infection has not been established. This is partially due to the lack of a precise method capable of detecting changes at the molecular level on the chromosome of pathogens while they are within host cells.

Scope of the Dissertation

Replication-transcription conflicts are an unavoidable threat to genome integrity. Although their detrimental effects have been demonstrated in model bacterial species grown in laboratory culture, as discussed earlier in this chapter, their prevalence and impact in bacterial species existing outside the laboratory are unknown. Due to their propensity to accelerate mutagenesis, conflicts could have severe consequences in terms of the evolution of bacterial pathogens. However, more work is needed to determine how pathogens and hosts interact on a molecular level, and if replication-conflicts are prevalent in this context.

The objective of my dissertation was to develop a method to study how eukaryotic hosts impact molecular transactions along the genome of a bacterial pathogen during infection. The development and execution of this method is detailed in Chapter 2. Using this method, I determined that RNAP occupancy patterns along bacterial genomes change during infection of a host. Most importantly, I discovered that RNAP backtracking is more prevalent during infection than outside infection. This is a significant step in our understanding of replication-transcription conflicts in bacterial pathogens during infection, as it suggests that the replisome could be disrupted by these persistent RNAP stalls. Additionally, this increase in backtracking events could have implications for mutagenesis. These results are discussed throughout Chapter 2.

Remaining open questions are considered in Chapter 3, with a focus on how the method I developed could be used to further investigate replication-transcription conflicts in bacterial pathogens during infection. Conclusions of this work are discussed in Chapter 4, followed by detailed methods in Chapter 5.

Ultimately, understanding how pathogens and hosts interact on a molecular level is key in the fight against antimicrobial resistance, which is an urgent threat to human health (89). Bacterial pathogens continue to evolve at a rate that outpaces almost all old and new therapeutics and resolving this problem requires a general understanding of how hosts and pathogens interact during infection.

CHAPTER 2

PATHOGENIC BACTERIA EXPERIENCE PERVASIVE RNA POLYMERASE BACKTRACKING DURING INFECTION

Summary[2]

Pathogenic bacteria and their eukaryotic hosts are in a constant arms race. Hosts have numerous defense mechanisms, including some that cause DNA damage. How these host defense mechanisms impact molecular transactions along the bacterial chromosome during infection is unclear. This is partially due to the lack of a method that can detect these events in pathogens while they are within host cells. We developed and optimized a system capable of mapping and measuring levels of bacterial proteins associated with the chromosome while they are actively infecting the host (referred to as PIC-seq). Here, we focused on the dynamics of RNAP movement and association with the chromosome in the pathogenic bacterium *S. enterica* as a model system during infection. Using PIC-seq, we found that RNAP association patterns with the chromosome change during infection genome wide, including at regions that encode for key virulence genes. Importantly, we found that infection of a host significantly increases RNAP backtracking on the bacterial chromosome. RNAP backtracking is the most common form of disruption to RNAP progress on the chromosome. Interestingly, we found that genes experiencing backtracking are downregulated. We observed that resolution of backtracked RNAPs via the anti-backtracking factors GreA and GreB is critical for pathogenesis and for regulation of gene expression during infection. Altogether, our findings have important

[2] Sections of Chapter 2 are adapted from the following publication: Browning, KR, Merrikh, H. 2023. Pathogenic bacteria experience pervasive RNA polymerase backtracking during infection. *mBio*. In press.

implications for both efficient transcription and stalled RNAP-driven mutagenesis, which can promote antimicrobial resistance and/or hypervirulence. These results suggest that the host can accelerate the evolution of bacterial pathogens, potentially at their own expense.

Introduction

Numerous studies over the past few decades have demonstrated how different pathogens reach, invade, and proliferate within their target hosts. However, how the host impacts the pathogen during infection at the molecular level, specifically the complexes that function on the bacterial chromosome, such as the replisome and RNAP, has not been thoroughly investigated.

Bacterial pathogens face significant challenges from the moment of initial contact with mammalian hosts. They must successfully survive harsh environments and regulate expression of key virulence factors like pathogenicity island-encoded secretion systems, flagella, ion transporters, and stress response genes. For enteric intracellular pathogens, ingested cells must respond to the acidic pH and the presence of reactive nitrogen species within the host stomach (90, 91). They must also overcome nutritional barriers caused by dense microbiota layers and withstand antimicrobial peptides released by host cells (92, 93). The bacteria are also heavily bombarded by oxidative stress (94–98). Although pathogenic bacteria are generally well equipped to respond to these host defense mechanisms, any number of these stresses during infection could lead to an accumulation of DNA damage, especially oxidative stress. Indeed, studies in *S. enterica* (99, 100), *Staphylococcus aureus* (101), and *Helicobacter pylori* (102) demonstrate the critical necessity of DNA repair during infection.

DNA damage could have numerous consequences for bacterial cells during infection. It is well documented that obstacles on the DNA template, such as damaged template bases or DNA-protein adducts, stall transcription elongation complexes (35–38). The resulting backtracking could have a profound impact on RNAP dynamics and transcription genome wide, with far-reaching consequences. For instance, movement of the replication forks may be impacted within chromosomal regions packed with stalled RNAPs. Disruptions to RNAP progression may perturb transcription and could exacerbate conflicts between replication and transcription, as previously discussed, leading to genome instability and mutagenesis (49). Backtracking could also increase mutagenesis by way of inducing transcription-coupled repair (72). In this way, it is possible that RNAP backtracking is increased within the pathogen during infection, thereby increasing the chances of generating hypervirulent strains and facilitating antimicrobial resistance (AMR) development.

Beyond these possibilities, any impediment to RNAP movement induced by the host is likely to significantly alter the transcriptional profile of the pathogen during infection. Gene expression programs that respond to the host environment are often highly intricate and tightly regulated (103, 104). Thus, any unresolved disruptions to RNAP progression during infection could be catastrophic to survival of the pathogen. In *S. enterica* grown in broth culture, the activity of anti-backtracking factors GreA and GreB in regulating transcriptional pauses is critical for expression of metabolic outputs that promote cellular survival under oxidative stress (105). Interestingly, both factors have also been demonstrated to be important for the expression of genes within *Salmonella* pathogenicity islands (SPIs) 1 and 2, even outside the host environment (105, 106). However, how a host impacts RNAP dynamics along the pathogen chromosome specifically during infection, and any consequences thereof, has not been investigated. The pervasiveness of RNAP backtracking along the pathogenic chromosome,

whether the host environment affects these levels, and if this renders Gre factors critical for virulence gene expression during infection remain open questions.

Here, we Investigated how the host Impacts RNAP movement on the chromosome of a bacterial pathogen, focusing our investigation on RNAP backtracking. We used the facultative intracellular pathogen *S. enterica* serovar Typhimurium (*S.* Typhimurium) as our model organism. We developed a modified chromatin immunoprecipitation (ChIP) method, Post-Infection ChIP followed by deep sequencing (PIC-seq), that we used to measure RNAP density and changes in RNAP movement on the bacterial chromosome while the pathogen is inside the host cell. Our data strongly suggest that RNAP backtracking is significantly more prevalent during infection compared to cells grown in broth culture. Furthermore, we find that these disruptions to RNAP progression occur across the genome, including at regions encoding for key *S.* Typhimurium virulence genes. Our results suggest that Gre factors regulate gene expression during infection. We also show that resolution of backtracking is critical to pathogenesis, highlighting a key role for anti-backtracking factors GreA and GreB in infections. Our results suggest that RNAP movement is substantially impacted by the host genome-wide, and that the resolution of these disruptions is critical for the survival of pathogens during infection.

Results

The development of a method to measure changes in bacterial protein association with the chromosome during infection

We hypothesized that RNAP backtracking is more prevalent during infection genome wide, and that resolution is critical to pathogenesis. To test this hypothesis, we needed a system that would allow us to map RNAP occupancy on the chromosome of the pathogen specifically

while the bacteria still resided within host cells. A successful method would allow isolation of bacterial protein-chromatin interactions without any signal loss or creation of artifacts during sample processing. To our knowledge, such a method did not exist. We developed and optimized a system whereby well-established *ex vivo* infection protocols are paired with chromatin immunoprecipitation (ChIP) of bacterial proteins, followed by deep sequencing of the immunoprecipitating DNA (Fig. 2.1A). We call this method Post-Infection ChIP-seq, or PIC-seq. By chemically fixing the protein-chromatin interactions while the bacterial pathogen still resides within the host cell, our method ensures that no fleeting molecular interactions are lost during sample processing and minimizes generation of artifacts, as could be the case if we first isolated the bacteria away from the host cells before fixation. In parallel, we performed ChIP-seq from cells grown in broth culture to isolate infection-specific effects. PIC-seq enabled us to determine RNAP dynamics at the molecular level specifically during infection.

We infected HeLa epithelial cells with *S.* Typhimurium based on standard protocols in the literature (98, 107–109) at a multiplicity of infection (MOI) of approximately 100-to-1 (bacterial cells to HeLa cells) (Fig. 2.1B). At this MOI, we achieved an invasion efficiency of around 10% (% of bacterial inoculum recovered at 1 h post infection [p.i.]), in agreement with previously published results (107, 108). Many of the bacterial cells that successfully invade HeLa cells die quickly by 2.5 h p.i. (Fig. 2.1B), consistent with intracellular killing facilitated by the host cells (107, 108). In light of this lag, intracellular replication of *S.* Typhimurium is often monitored between 4.5 and 8 h p.i. (109). By 8 h p.i., each HeLa cell contains an average of approximately ten *S.* Typhimurium cells.

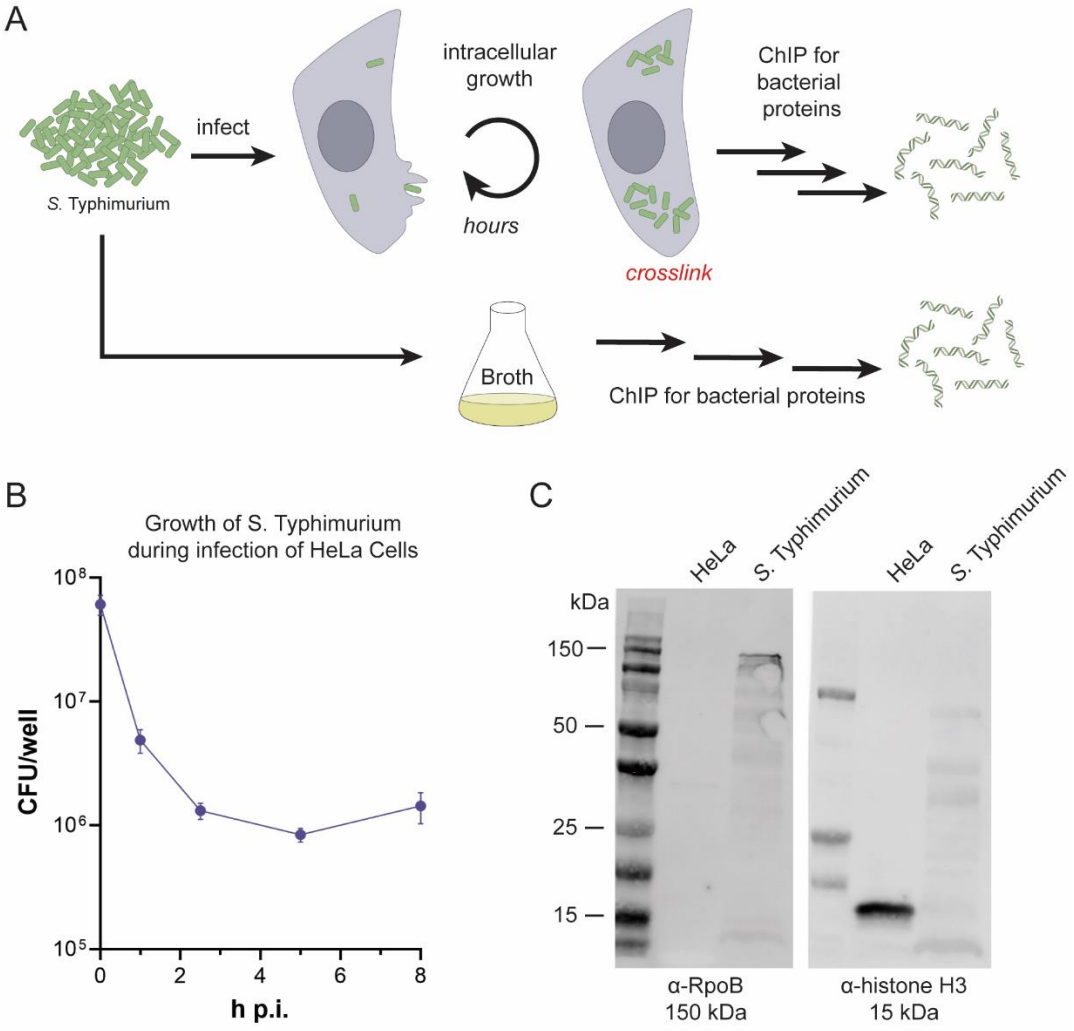

Figure 2.1. Development and optimization of PIC-seq method. (A) Schematic of PIC-seq in bacteria cells from infected HeLa. Eukaryotic host cells (HeLa) are infected with pathogenic bacteria cells (*S.* Typhimurium). Once the infection has progressed for a given time, the system is chemically fixed to lock protein-chromosome interactions in place in time. Subsequent immunoprecipitation can be performed. In parallel, ChIP-seq can be performed from bacteria cells grown in broth culture. (B) Average number of bacteria (colony forming units [CFU]) per well of HeLa cells at a given hour post infection (p.i.) as determined by standard gentamicin protection assay. Error bars indicate standard error of the mean for nine independent replicates. (C) Western blot showing that the native RpoB antibody used in these studies is specific to bacterial protein and does not cross-react to proteins specific to HeLa cells.

At defined timepoints post infection, bacterial proteins were chemically crosslinked to the DNA while the bacteria still resided within the host cell. Both bacterial and host cells were then lysed, and their DNA was fragmented by sonication. This step required significant optimization. Ultimately, the final lysis buffer formulation contains two different nonionic detergents, NP-40 and Triton, to facilitate dissolution of cellular membranes (both host and pathogen) and protein solubilization. Lysozyme is also added to facilitate lysis of the bacterial cell membrane. Additionally, whereas shearing the DNA of a bacterial broth culture to around 500 bp could be achieved using probe sonication, use of an ultrasonic water bath was required to shear the DNA of bacteria from an infected host cell to near 500 bp. This is likely due to the additional network of covalent crosslinks of the HeLa cells within the infected sample versus broth culture sample.

Following cell lysis and DNA sonication, RpoB, the β subunit of RNAP, was immunoprecipitated from the mixed lysate using a native antibody. Using Western blot analyses, we determined that the antibody is highly specific to RpoB and does not cross react with proteins specific to HeLa cells (Fig. 2.1C). The associated DNA was then extracted, purified, and deep sequenced. Any contaminating DNA from the host is naturally excluded from further analysis, as it does not map to the *S.* Typhimurium genome. On average, 12.4% of total reads mapped to the *S.* Typhimurium genome in our PIC-seq experiments, providing an average of approximately 80x coverage (approximately 2.6 million aligned reads per sample) (Table 2.1). This value was determined for the input samples at 8 h p.i.. In the IP samples, a greater percentage of the reads mapped to the *S.* Typhimurium genome as these immunoprecipitated samples are enriched for *S.* Typhimurium signal. On average, 22.3% of total reads from the IP samples mapped to the *S.* Typhimurium genome, providing approximately 150x coverage (Table 2.1).

	Replicate	Total Reads	Reads that did not align	Total aligned	%	Coverage
Input samples	1	20,690,370	17,161,361	3,529,009	17.06%	108x
	2	22,065,841	19,630,985	2,434,856	11.03%	74x
	3	19,541,523	17,762,178	1,779,345	9.11%	54x
IP samples	1	24,401,973	16,188,691	8,213,282	33.66%	252x
	2	18,353,186	15,524,416	2,828,770	15.41%	86x
	3	21,179,227	17,402,755	3,776,472	17.83%	116x

Table 2.1. Summary of PIC-seq mapped reads. Mapped reads were determined at 8 h p.i.. Coverage is calculated by multiplying the total aligned reads by read length (150 bp) and dividing by the *S.* Typhimurium genome size (4,878,012 bp).

We measured the association of RNAP with the pathogen genome using PIC-seq. To determine whether our signal was specific to the bacterial pathogen, we briefly treated cells with the antibiotic rifampicin during the infection. Rifampicin inhibits bacterial transcription initiation while allowing elongating RNAPs to complete transcription, thereby reducing RNAP signal that can be detected by ChIP. Indeed, our analysis revealed that RNAP occupancy signal was significantly reduced genome wide upon rifampicin treatment (Fig. 2.2). These findings demonstrate the specificity of PIC-seq for detection of bacterial protein-chromosome interactions.

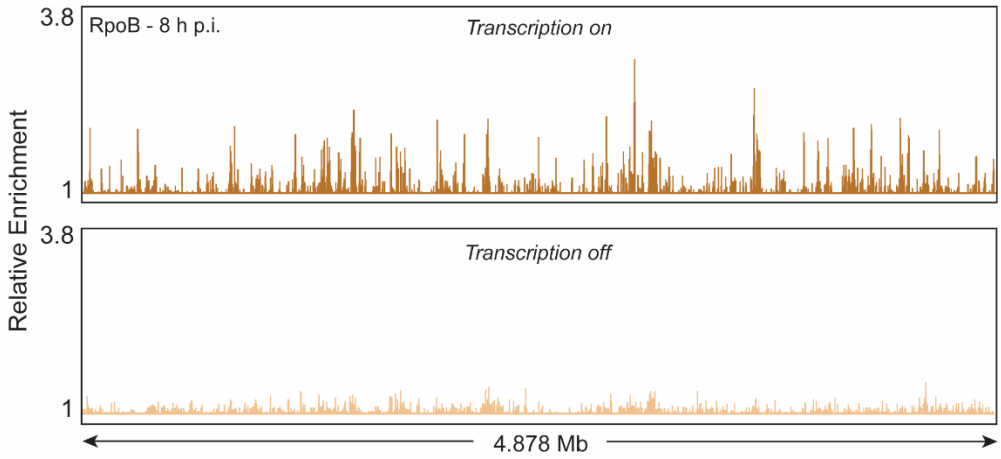

Figure 2.2. PIC-seq method can be used to measure RNAP occupancy during infection. Representative example of *S.* Typhimurium RNAP occupancy 8 h p.i. of HeLa cells as determined by PIC-seq of RpoB. Inhibiting transcription with rifampicin during the infection ("transcription off") abrogates RpoB (RNAP) enrichment genome wide. Relative enrichment is defined as the ratio of IP and input read counts normalized to total read depth. Total genome size is 4.878 Mb.

To rule out the possibility that this decrease in RNAP occupancy was caused by unintentional rifampicin-induced cell death, we treated cells with carbenicillin, a different antibiotic that inhibits cell wall synthesis rather than transcription, and measured RNAP occupancy using qPCR. Treatment with carbenicillin led to significant cell death during infection (Fig. 2.3A). Despite this pronounced effect, carbenicillin treatment did not abrogate RNAP occupancy (Fig. 2.3B). This indicates that the decrease in RNAP occupancy during treatment with rifampicin is specifically due to the inhibition of transcription, and not due to cell death. These findings further support our conclusion that PIC-seq can specifically isolate bacterial pathogen chromatin-protein complexes from an infection model.

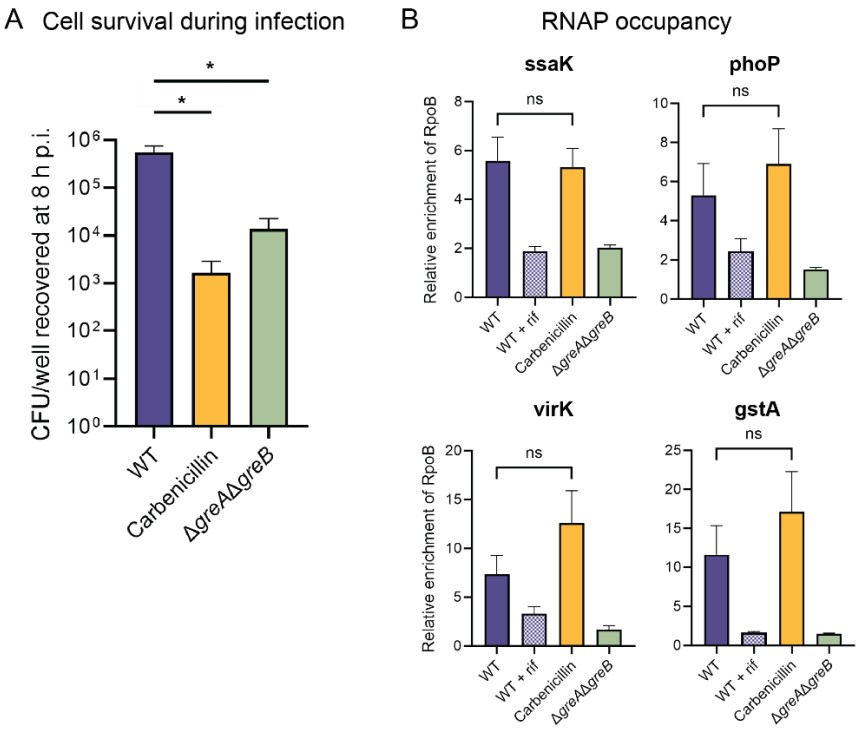

Figure 2.3. Changes to RNAP occupancy in the absence of Gre factors occur independently of loss in cell viability. (A) CFUs of WT S. Typhimurium, WT cells in the presence of carbenicillin, and S. Typhimurium cells lacking Gre factors at 8 h p.i. of HeLa cells as determined by gentamicin protection assay. Carbenicillin was added to a final concentration of 6.4 mg/mL at 1 h p.i. Data represent the means and standard deviation of at least six independent replicates. *p<0.0001, one-way ANOVA. (B) RNAP occupancy was determined by PIC-qPCR at the indicated genes at 8 h p.i.. Where indicated, the antibiotic rifampicin (rif) was added ten minutes prior to crosslinking (see methods). Relative enrichment is defined as the ratio of gene copy number of the indicated gene to gene copy number of eutN. Data represent the means and SEM of at least four independent replicates. Ns: not significant, one-way ANOVA.

To further confirm that PIC-seq can differentiate between cells grown in broth culture versus infection, we compared the RNAP occupancy profile of *S.* Typhimurium grown infection to that of cells grown in LB-Lennox (Fig 2.4A). Specifically, we analyzed the levels of RNAP occupancy with regions encoding virulence genes. As expected, RNAP is significantly enriched at regions of the chromosome containing multiple, successive virulence genes (*Salmonella* pathogenicity islands, "SPIs") during infection (Fig. 2.4B). Importantly, no significant RNAP signal was found at these genes in cells grown in broth (Fig. 2.4C). Similar trends were observed at a trio of virulence genes, *pipB2*, *virK*, and *mig-14*, where there is no RNAP enrichment in cells grown in broth, but clear enrichment during infection. Unsurprisingly, almost all genes with significant RNAP enrichment specifically during infection (not in broth culture) have a confirmed or putative role in virulence (Table 2.2). These results are consistent with previous RNA-seq datasets of wild type *S.* Typhimurium during infection (97, 110, 111), as well as the well-defined roles of SPI-1 and SPI-2 proteins in facilitating host cell invasion and promoting intracellular survival, respectively (112). In contrast, we observed similar levels of RNAP at housekeeping genes such as a variety of ribosomal components (*rpl*, *rpm*, and *rps* genes), transcriptional regulator operons (such as the transcription termination factor *rho*), and RNAP subunits (*rpoA* and *rpoH*), in both conditions (Fig. 2.4D, Table 2.3). The agreement between these datasets is unsurprising, given the essential nature of these genes. It also supports our conclusion that PIC-seq is an effective method for isolating bacterial protein-chromatin interactions from an infection. Taken together, our findings demonstrate that PIC-seq, when combined with ChIP-seq from broth controls, can specifically map and measure the levels of infection-specific RNAP association with the bacterial genome. Furthermore, the high levels of RNAP occupancy at virulence genes that we observed supports prior work and the conclusion that these genes are highly transcribed during infection.

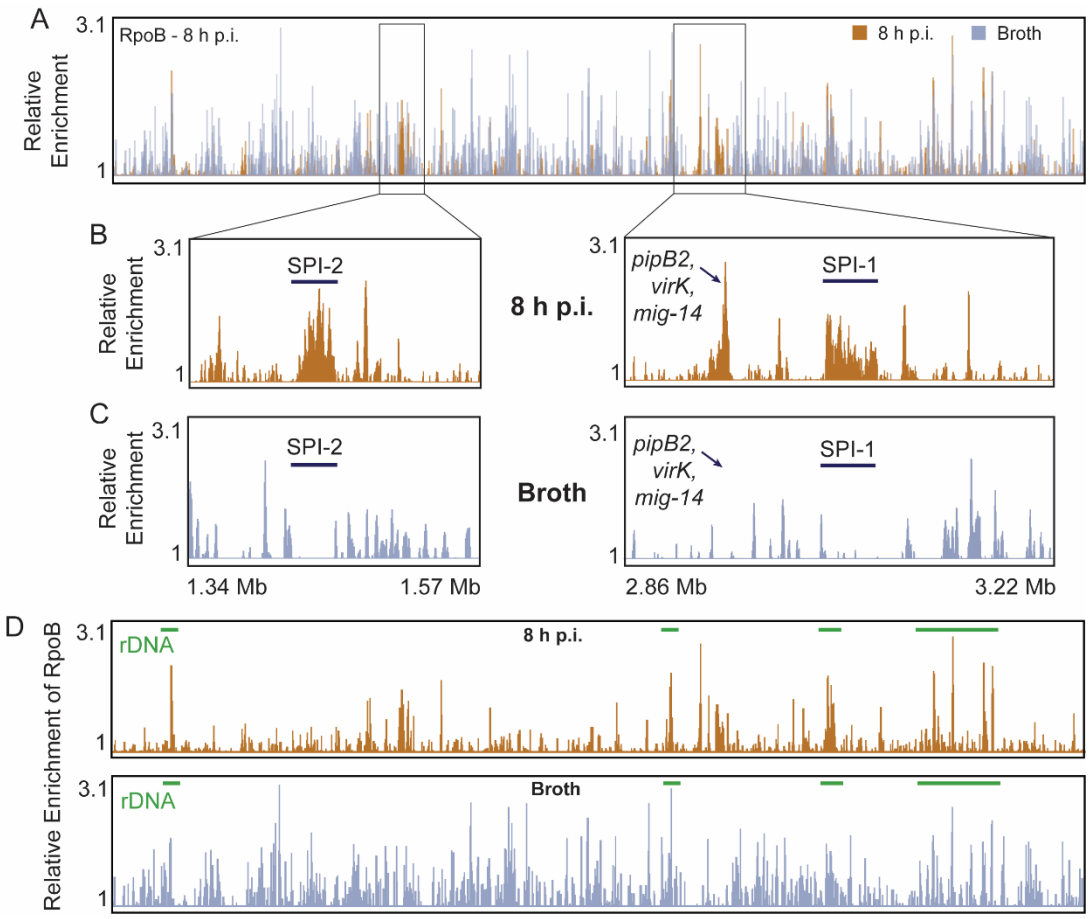

Figure 2.4. Specific detection of RNAP occupancy at *S.* Typhimurium virulence genes. (A) Representative overlay of *S.* Typhimurium RNAP occupancy as determined by PIC-seq of RpoB 8 h p.i. of HeLa cells or from cells grown in broth culture. Relative enrichment is defined as the ratio of IP and input read counts normalized to total read depth. (B) Reads mapped to the *S.* Typhimurium genome display areas of high RNAP enrichment during infection, notably within *Salmonella* pathogenicity islands (SPIs) and at a trio of virulence genes (*pipB2*, *virK*, and *mig-14*). (C) No significant RNAP signal was found at these genes in cells grown in broth culture. (D) Regions of ribosomal genes (rDNA) display high RNAP occupancy in both conditions (green bars).

Locus	Fold Enrichment	Localization	Function	Reference
mig-14	2.82	Chromosome	antimicrobial peptide resistance	(113)
virK	2.79	Chromosome	antimicrobial peptide resistance	(114)
sifB	2.70	Chromosome	putatively contributes to the formation of the intracellular replication niche	(115)
sitD	2.62	SPI-1	resistance to nitrosative stress	(116)
SL1344_1344	2.57	Chromosome	putative component of the type III secretion system (T3SS-2) apparatus	(117)
ssaK	2.57	SPI-2	T3SS-2 apparatus	(118)
sitC	2.53	SPI-1	resistance to nitrosative stress	(116)
pdgL	2.53	Chromosome	putative role in cell wall/membrane biogenesis	(119)
SL1344_2902	2.49	Chromosome	hypothetical protein	
rplC	2.49	Chromosome	ribosomal protein L3	
ssaH	2.46	SPI-2	putative role in T3SS-2 apparatus	(120)
ssaI	2.46	SPI-2	putative role in T3SS-2 apparatus	(120)
ssaJ	2.46	SPI-2	T3SS-2 apparatus	(118)
pipB2	2.41	Chromosome	kinesin accumulation in the SCV	(112)
ssaE	2.40	SPI-2	chaperone	(121)
SL1344_1530A	2.37	Chromosome	hypothetical protein	
ugtL	2.37	Chromosome	PhoP activation	(122)
sitA	2.36	SPI-1	resistance to nitrosative stress	(116)
sitB	2.36	SPI-1	resistance to nitrosative stress	(116)
sseB	2.34	SPI-2	T3SS-2 translocon	(123)
sscA	2.34	SPI-2	chaperone	(124)
sseD	2.33	SPI-2	T3SS-2 translocon	(123)
sseK/sseK3	2.27	Chromosome	promote intracellular survival	(125)
sseC	2.20	SPI-2	T3SS-2 translocon	(123)
ssaL	2.19	SPI-2	T3SS-2 apparatus	(118)
mgtC	2.18	SPI-3	maintenance of ATP homeostasis	(126)
ssaR	2.18	SPI-2	T3SS-2 apparatus	(118)
sseG	2.17	SPI-2	SIF biogenesis and SCV maintenance	(127)
iscS	2.15	Chromosome	tRNA synthesis	(128)
prgJ	2.15	SPI-1	T3SS-1	(129)
prgI	2.15	SPI-1	T3SS-1	(129)
ssaQ	2.10	SPI-2	T3SS-2 apparatus	(130)

Table 2.2. RNAP is significantly enriched at virulence genes during infection. Loci where RpoB was significantly enriched at least two-fold in *S.* Typhimurium 8 h p.i. as determined by PIC-seq. Each value represents the average fold enrichment for at least two independent replicates.

	Fold Enrichment	
Locus	**Broth**	**Infection**
yiaG	2.01	3.44
cspA	2.01	3.44
cueP	2.01	3.44
deaD	2.47	2.93
yrbN	2.47	2.93
nlpI	2.47	2.93
pnp	2.47	2.90
cspE (cspC)	2.23	2.84
yobF	2.23	2.84
SL1344_RS09185	2.23	2.82
mgrB (yobG)	2.23	2.82
yqcC	2.75	2.69
hdfR	2.33	2.50
trxA	2.07	2.35
rhoL	2.07	2.35
rho	2.07	2.35
rpsD	2.08	2.33
rpsK	2.08	2.33
rpsM	2.08	2.33
rpmJ	2.08	2.33
rplN	2.08	2.31
rplO	2.08	2.27
secY/prlA	2.08	2.26
rpoA/pez	2.08	2.18
rpsC	2.21	2.18
rplV	2.21	2.18
rplB	2.21	2.13
rpoH	2.07	2.12
rluD	2.37	2.08
pgeF	2.21	2.03

Table 2.3. RNAP is significantly enriched at some of the same genes in broth and infection. Loci where RpoB was significantly enriched at least two-fold in S. Typhimurium in cells grown in broth and 8 h p.i., as determined by PIC-seq. Each value represents the average fold enrichment for at least two independent replicates.

RNAP backtracking increases significantly during infection

RNAPs naturally backtrack, especially in the face of DNA damage or other obstacles along the DNA. We hypothesized that RNAP backtracking is more prevalent during infection genome wide because several host defense mechanisms can damage DNA (such as through oxidative stress), potentially inhibiting the movement of the transcription machinery along the genome and disrupting gene expression. In *S.* Typhimurium, the anti-backtracking factors GreA and GreB rescue backtracked RNAPs by stimulating the intrinsic endonucleolytic activity of RNAP to restore the active site (41, 131).

To test our hypothesis, we constructed an S*.* Typhimurium strain lacking GreA and GreB (4). If our hypothesis is correct, then cells lacking both Gre factors would have increased levels of stalled RNAPs during infection compared to growth in broth culture. To test this model, we measured the differences in RNAP occupancy levels across the genome in the presence (wild type, WT) and absence (Δ*greA* Δ*greB*) of these anti-backtracking factors. Since both Gre factors are directly and specifically involved in the resolution of backtracked RNAPs (41, 131), we deduced that sites where RNAP occupancy differs in the presence versus the absence of Gre factors are sites where pervasive backtracking is occurring. The prevalence of backtracking during infection can then be measured by determining how these RNAP occupancy differences change in cells grown in broth versus infection.

We measured RNAP occupancy in WT and Δ*greA* Δ*greB* cells grown in broth and at 1 or 8 h p.i. using PIC-seq (Fig. 2.5). Indeed, deletion of both Gre factors led to pronounced changes in RNAP occupancy genome wide, especially at 8 h p.i. (Fig. 2.5). Given that inducing cell death during infection more than 100-fold by treatment with carbenicillin does not yield any significant difference in RNAP occupancy compared to WT cells (Fig. 2.3), the changes in RNAP occupancy we detect in the absence of Gre factors are unlikely to be due to any defects in cell survival. Additionally, there are no major differences in the total number of reads mapping to

features within the *S.* Typhimurium genome regardless of genotype, preventing any bias during comparative analyses (Table 2.4).

In general, we found that RNAP occupancy decreases in the absence of the two Gre factors. This was contrary to our initial prediction but is consistent with steady state conditions within the cell. This decrease in RNAP occupancy is likely due to the activity of accessory factors, such as helicases and other DNA translocases, that can remove stalled RNAPs from DNA (37, 62, 63, 69, 132–134). Backtracked RNAPs that are not rescued by the Gre factors are therefore expected to be removed by such helicases and other transcription terminators, leading to the decrease in RNAP occupancy that we observe.

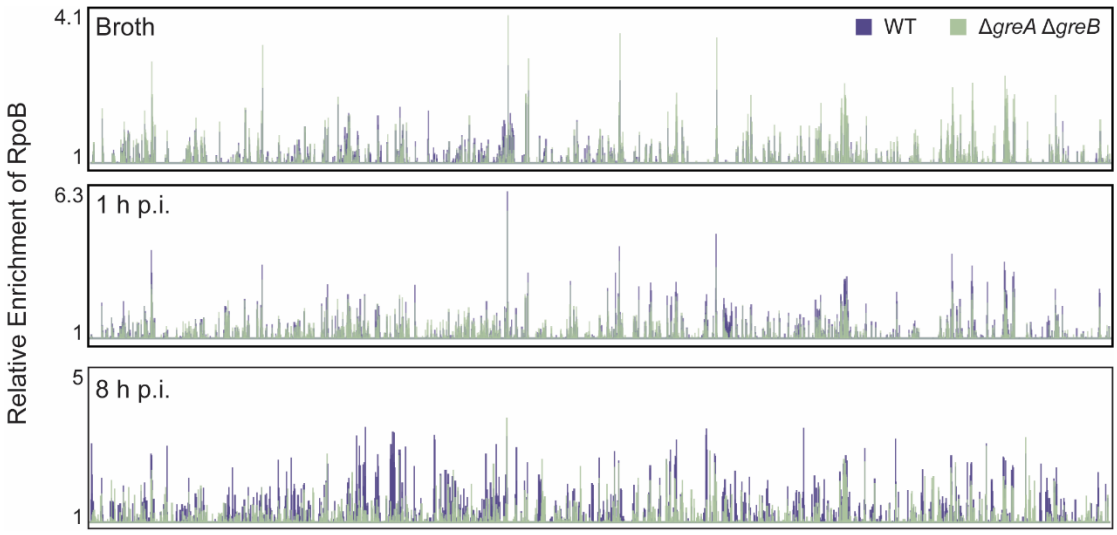

Figure 2.5. RNAP occupancy decreases in the absence of Gre factors. Representative RNAP occupancy profile for WT cells or cells lacking Gre factors (Δ*greA* Δ*greB*) grown in broth, 1 h p.i., or 8 h p.i. as determined by PIC-seq of RpoB. Relative enrichment is defined as the ratio of IP and input read counts normalized to total read depth.

		WT	Δ*greA* Δ*greB*
Broth	**IP**	2,949,076 ±237,211	2,386,700 ±661,479
	Input	3,631,482 ±730,566	3,600,436 ±272,131
1 h p.i.	**IP**	1,684,822 ±541,602	1,678,537 ±555,513
	Input	1,711,959 ±213,517	1,296,753 ±302,429
8 h p.i.	**IP**	1,283,399 ±239,135	1,076,680 ±668,727
	Input	1,264,645 ±300,494	832,641 ±606,166

Table 2.4. Read counts mapping to features within the *S.* Typhimurium genome. Values indicate average ± standard deviation of three independent replicates.

Having determined that RNAP occupancy changes upon deletion of both Gre factors, we identified the genes that exhibited differential RNAP occupancy in each condition (broth, 1 h p.i. and 8 h p.i.). However, because cells grown in these conditions have widely different gene expression profiles, we could not always directly compare backtracking in the same gene across each condition. In other words, we could not compare backtracking at a gene that is transcriptionally active in one condition but not transcriptionally active in the other. Therefore, we categorized the genes by relative transcription level within each condition. Relative transcription level was determined for each gene using the normalized number of reads mapping to that gene in the IP sample versus the input sample in the WT strain, averaged across three replicates. Genes were categorized independently for each condition (broth, 1 h p.i., and 8 h p.i.). This

resulted in 631 top-transcribed genes in broth, 645 at 1 h p.i., and 590 at 8 h p.i. (Table S2.1). A proportion of the top-transcribed genes (283 genes) were found across all three conditions (Table S2.1). Our results indicate that RNAP occupancy differs at genes in the presence and absence of Gre factors, not only within each independent condition, but also across conditions (Fig. 2.6A).

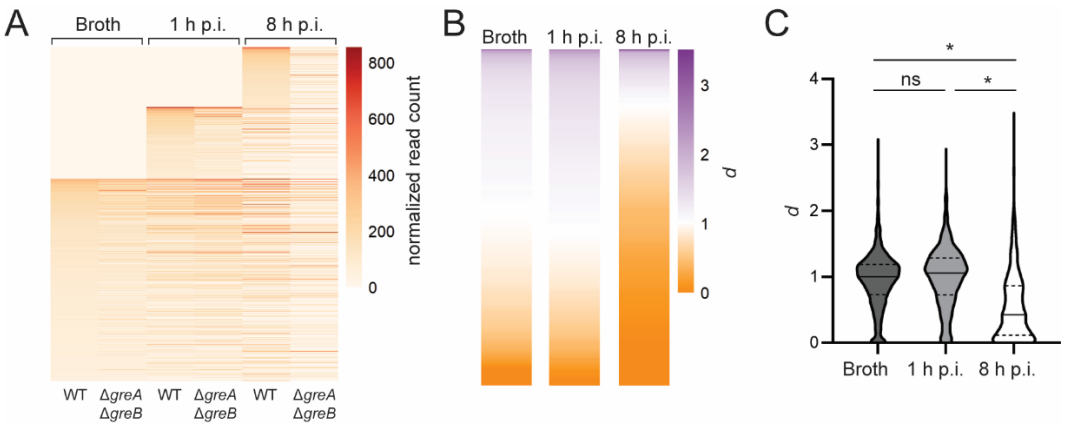

Figure 2.6. Backtracking is more prevalent later during infection. (A) RNAP occupancy changes visualized as a heat map, where every horizontal line represents the normalized read count for the same gene across each condition. All top-transcribed genes across all conditions are plotted (1043 genes total, see Table S2.1). Each value represents the average of three independent replicates. Zeros are plotted for genes that do not arise in one condition but do arise in one or both of the other conditions. (B) Ratio *d* of normalized read count in the absence of Gre factors (Δ*greA* Δ*greB*) versus in the presence of Gre factors (WT) for top-transcribed genes in each condition averaged across three independent replicates. (C) Quantification of (B) showing ratio *d* calculated for every top-transcribed gene, where the median is shown by a solid line and quartiles by dashed lines. *p<0.0001, ns: not significant, one-way ANOVA.

We performed a more quantitative analysis to determine whether these differences in RNAP occupancy change between cells grown during infection versus cells grown in broth culture. We defined a ratio for differential RNAP occupancy, *d*, as the normalized read counts mapping to a gene in the backtracking-prone state (Δ*greA* Δ*greB*) versus those mapping to the same gene in in WT. An equivalent ratio ($d \approx 1$) indicates that there is little difference in RNAP occupancy independent of the presence or absence of the Gre factors. We interpreted these regions to be those that are not experiencing backtracking. A ratio that deviates from one ($d < 1$ or $d > 1$) indicates that there is a Gre factor-dependent difference in RNAP occupancy at the gene, suggesting that the Gre factors are playing an important role in maintaining RNAP progression at that gene. We interpreted the regions with $d < 1$ to be those that are experiencing significant RNAP backtracking. A ratio $d > 1$ indicates that Gre factors may normally be involved in suppressing RNAP occupancy at that gene. This *d* ratio was calculated for every top-transcribed gene in each condition (Fig. 2.6B,C). For cells grown in broth and at 1 h p.i., the average ratio of RNAP occupancy in the absence of Gre factors versus in the presence is equivalent ($d \approx 1$) (Table 2.5, Fig. 2.6B,C), suggesting that, on average, cells do not depend on Gre factors to maintain proper RNAP occupancy under these conditions and that backtracking is likely minimal. In contrast, cells at 8 h p.i. do rely on the Gre factors to maintain RNAP occupancy (Table 2.5, Fig. 2.6B,C), suggesting that backtracking becomes pervasive as the infection progresses. Interestingly, this effect is independent of transcription level at 8 h p.i.; RNAP occupancy is abrogated in both lowly and highly transcribed genes.

We also determined the extent to which backtracking is more prevalent during infection than broth. We accomplished this by quantifying the average differences in RNAP occupancy changes in broth versus infection. We defined a new ratio, *d'*, as the ratio of *d* for infection versus *d* for broth, $d' = d_{inf} / d_{broth}$. Compared to broth, backtracking is not more prevalent early in infection (at 1 h p.i.), but it is more prevalent later in infection at 8 h p.i., $d' = 0.59$ (Table 2.5).

Transcription level	d_{broth}	$d_{1\ h\ p.i.}$	$d_{8\ h\ p.i.}$	*d'* 1 h p.i. vs. broth	*d'* 8 h p.i. vs. broth
++	0.96	1.01	0.56	1.06	0.59
+++	0.91	0.99	0.52	1.09	0.57
++++	0.91	0.94	0.48	1.03	0.53
All genes	0.94	1.00	0.55	1.06	0.59

Table 2.5. Gre factor-dependent RNAP occupancy changes in broth versus infection. Ratio *d* was calculated as the normalized read counts mapping to a gene in the backtracking-prone state (Δ*greA* Δ*greB*) versus in WT for each condition and each transcription level (++, +++, ++++). Ratio *d'* compares ratio *d* for 1 or 8 h p.i. to broth for each transcription level.

Backtracking occurs within key genes necessary for Salmonella pathogenesis

Having established that backtracking is more prevalent during infection genome wide, specifically at later time points, we endeavored to identify those regions where backtracking occurs and could potentially be problematic during infection. Thus, we focused on genes with the greatest changes in RNAP occupancy in the absence of Gre factors. These are the genes where RNAP activity is highly dependent on the function of the Gre factors, most likely due to backtracking. We used hierarchical clustering to identify these genes (Fig. 2.7, Table S2.2).

As backtracking is minimal in broth and at 1 h p.i. (Fig. 2.6, Table 2.5), we focused on how genes were clustered by RNAP occupancy differences at 8 h p.i. (Fig. 2.7, Table S2.2). Genes at 8 h p.i. fall into four clusters based on similarities in RNAP occupancy changes in the presence and absence of Gre factors: cluster one, 466 genes; cluster two, 70 genes; cluster three, 18 genes; and cluster four, 36 genes (Fig. 2.7A, Table S2.2). We categorized the RNAP occupancy at genes within each of these hierarchical clusters as either Gre factor dependent or independent, and identified enriched gene functions within each cluster.

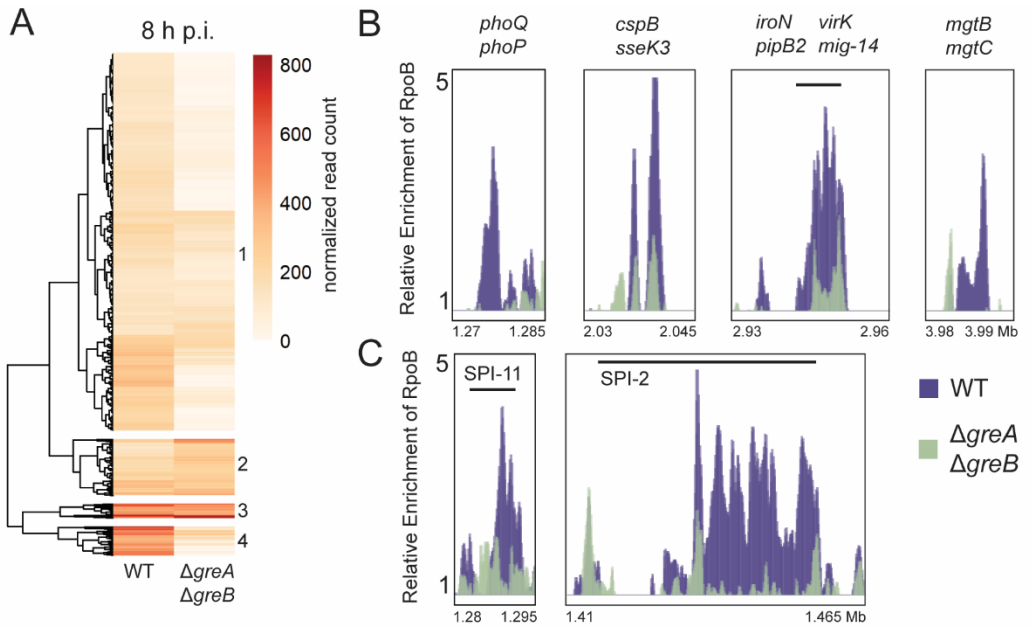

Figure 2.7. Backtracking is prevalent at key virulence genes during infection. (A) RNAP occupancy changes at 8 h p.i. in the absence of Gre factors visualized as a heat map. Each value represented is the average of three independent replicates. Hierarchical clustering (clustered numbered 1-4) was performed using the pheatmap function in Rstudio (see Table S2.2). (B) Key genes related to virulence in cluster 4 in (A) exhibit significant differences in RNAP occupancy ($p < 0.05$ for every gene listed, unpaired t test). Relative enrichment is defined as the ratio of IP and input read counts normalized to total read depth. (C) Genes belonging to SPIs that fall into cluster 1 in (A) and exhibit significant differences in RNAP occupancy.

In clusters four and one, RNAP occupancy is higher on average in the presence of Gre factors (WT) than in the absence (Δ*greA* Δ*greB*), indicating that the necessary levels of RNAP association with these genes depends on the presence of Gre factors. Genes in cluster four exhibit dramatic differences in RNAP occupancy depending on the presence of Gre factors, with an average *d* of 0.23. There was not enough statistical power to perform gene ontology (GO) functional enrichment analysis for genes in cluster four. However, manual annotation of these

gene functions reveals that out of 36 genes, at least 17 have been previously implicated in *S.* Typhimurium virulence and/or pathogenesis (Fig. 2.7B, Table S2.2). RNAP occupancy at genes in cluster one also depends on the presence of Gre factors (d = 0.43), despite genes in cluster one having lower RNAP occupancy levels overall (Fig. 2.7A). According to GO functional enrichment analysis, cluster one genes are enriched for functions related to cellular and macromolecular localization, membrane biogenesis, regulation of metabolic processes, secretion, and translation. This cluster also consists of numerous SPI genes that exhibit large differences in RNAP occupancy when Gre factors are absent (Fig. 2.7C, Table S2.2). Of the 25 top-transcribed SPI-2 genes in this dataset, 22 fall within cluster one. Similarly, all five of the top-transcribed SPI-11 genes in this dataset fall within cluster one. Efficient RNAP progression at these regions, and the virulence genes belonging to cluster four, would likely be critical to supporting the cell during infection, as these genes are required for cell survival throughout a significant proportion of the *S.* Typhimurium infection lifecycle (112). Our results suggest that infection of a host disrupts RNAP dynamics and that Gre factors are key to maintaining RNAP movement within these regions, and ultimately, supporting *S.* Typhimurium pathogenesis.

Genes in cluster two exhibit an average d ratio of 1.41, or greater RNAP occupancy in the absence than in the presence of Gre factors, indicating the RNAP occupancy at these genes does depend on the Gre factors, but in an opposite manner to those genes in clusters four and one. For instance, Gre factors may be necessary for interactions with a separate protein that negatively modulates RNAP levels at that gene, such that RNAP occupancy is higher in the absence of the Gre factors. Genes in cluster two are enriched in functions related to cellular nitrogen compound biosynthetic processes (GO analysis). Cluster three includes genes where RNAP occupancy does not change in the absence of Gre factors (d = 0.99), indicating that these are Gre factor-independent genes; RNAP presence and/or activity at these genes does not depend on the Gre factors. Manual annotation of genes in cluster three reveal functions

related to essential processes such as translation, carbon metabolism, and cell wall maintenance (Table S2.2).

Despite the functional enrichments we identified, the gene functions within each cluster vary. This is unsurprising, given our hypothesis that backtracking during infection is more prevalent due to stress from the host environment resulting in DNA damage within the pathogen. DNA damage, and subsequent RNAP backtracking would occur randomly throughout the genome during infection, but would likely be concentrated in actively transcribed genes, based on recent models of DNA damage accumulation (72). Therefore, it tracks that the set of genes in which RNAP occupancy depends on Gre factors would be characterized by a wide range of gene functions.

Additionally, we found that the prevalence of backtracking was not uniform genome wide. We reasoned that longer genes are more likely to accumulate DNA damage and therefore more likely to accumulate backtracked RNAPs. Accordingly, we found that the lengths of genes in cluster four, which exhibited the greatest differences in RNAP occupancy in a Gre factor-dependent manner, were significantly longer than the average length of all top-transcribed genes (Fig. 2.8). Overall, our results suggest that backtracking occurs at key virulence genes during infection, and that Gre factors are critical for RNAP progression at these regions.

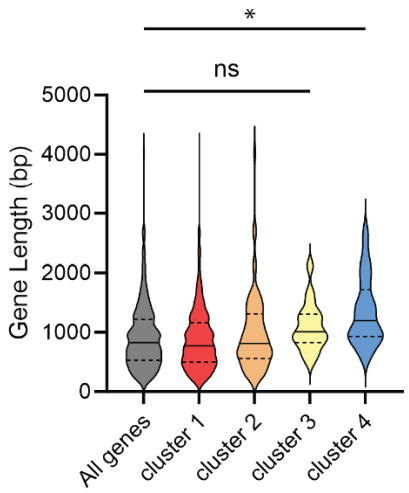

Figure 2.8. Longer genes experience more prevalent backtracking. All top-transcribed genes at 8 h p.i. were sorted by hierarchical clustering using the pheatmap function in Rstudio (590 genes total) (see Table S2.2). Gene length in bp was calculated for all genes in each cluster and plotted, where the median is shown by a solid line and quartiles by dashed lines. *p<0.001, ns: not significant, one-way ANOVA.

Backtracking alters gene expression during infection

The observation that backtracking is more prevalent during infection than in broth culture suggests that proper gene expression requires the resolution of backtracked RNAPs. In fact, prior work from others suggests that Gre factors are involved in the regulation of expression of a myriad of genes, including virulence genes, at least in broth culture (in minimal media (105) or in broth conditions that induce S. Typhimurium invasiveness (106)). However, whether this role is conserved and/or meaningful in the context of pathogenesis during infection of a host remained undefined. To examine whether the differential binding of RNAP we observe during infection due to the loss of both Gre factors ultimately correlates with altered gene expression, we compared the transcriptional profiles of WT cells and cells lacking both Gre factors during infection using RNA-seq. We enriched *S.* Typhimurium from infected HeLa cells at 8 h p.i. and isolated total bacterial RNA. In parallel, we performed RNA-seq of WT and Δ*greA* Δ*greB* cells grown in broth

(LB-Lennox). On average, approximately 14% of reads from each cDNA library from infection mapped to the S. Typhimurium genome, providing approximately 170x coverage (approximately 5.5 million aligned reads per sample). Of the reads that map to defined features within the *S.* Typhimurium genome, we did not observe major differences between genotypes in each condition, preventing any bias during comparative analyses (Table 2.6).

	WT	**Δ*greA* Δ*greB***
Broth	1,664,969 ±129,995	1,485,302 ±501,092
8 h p.i.	646,946 ±94,607	530,704 ±8,072

Table 2.6. Read counts mapping to features within the *S.* Typhimurium genome. Values indicate average ± standard deviation of three independent replicates for broth and two independent replicates for infection.

During infection, we observe that RNAP occupancy decreases genome wide in the absence of both Gre factors. Changes in RNAP occupancy likely coincide with changes in gene expression. Indeed, we found a positive correlation between RNAP occupancy (as determined by PIC-seq of RpoB) and expression level (as determined by RNA-seq) for both WT and Δ*greA* Δ*greB* cells during infection (Pearson coefficient 0.53 and 0.55, respectively) (Fig. 2.9), indicating that regions with low RNAP occupancy correlate with low gene expression.

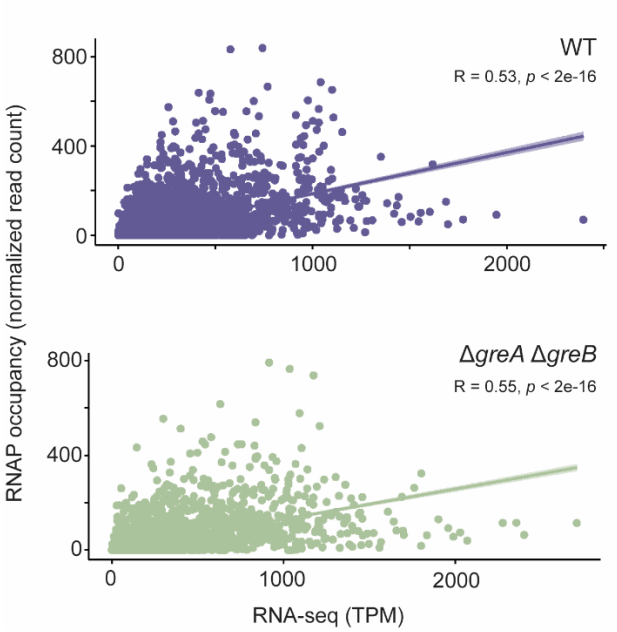

Figure 2.9. Gene expression correlates with RNAP occupancy. Linear regression analysis comparing gene expression (as determined by RNA-seq data) and RNAP occupancy (as determined by RpoB PIC-seq) during infection (at 8 h p.i.). Transcripts per million (TPM) and normalized read count for each gene were calculated as explained in the methods. Pearson's correlation coefficients I were determined to be 0.53 and 0.55 for WT and Δ*greA* Δ*greB*, respectively. The shaded area around each line indicates 95% confidence interval. Each dot represents the mean value from at least two independent replicates.

Although there were marked differences in gene expression in the presence and absence of Gre factors as revealed by principal component analysis (PCA), gene expression was more influenced by condition (broth versus infection) than genotype (WT versus Δ*greA greB*) (Fig 2.10A). These results had not been previously demonstrated but are unsurprising given the known gene expression programs that are activated during infection but are minimized otherwise (112). Interestingly, gene expression in cells lacking both Gre factors is affected more by a change in condition (infection versus broth culture) than is gene expression in WT cells (Fig. 2.10B). For WT cells, a large proportion of genes do not exhibit significant changes in expression when comparing infection to broth. However, for cells lacking both Gre factors, there

are very few genes whose expression does not change between conditions, suggesting that cells lacking both Gre factors are more sensitive to the change in condition than are WT cells.

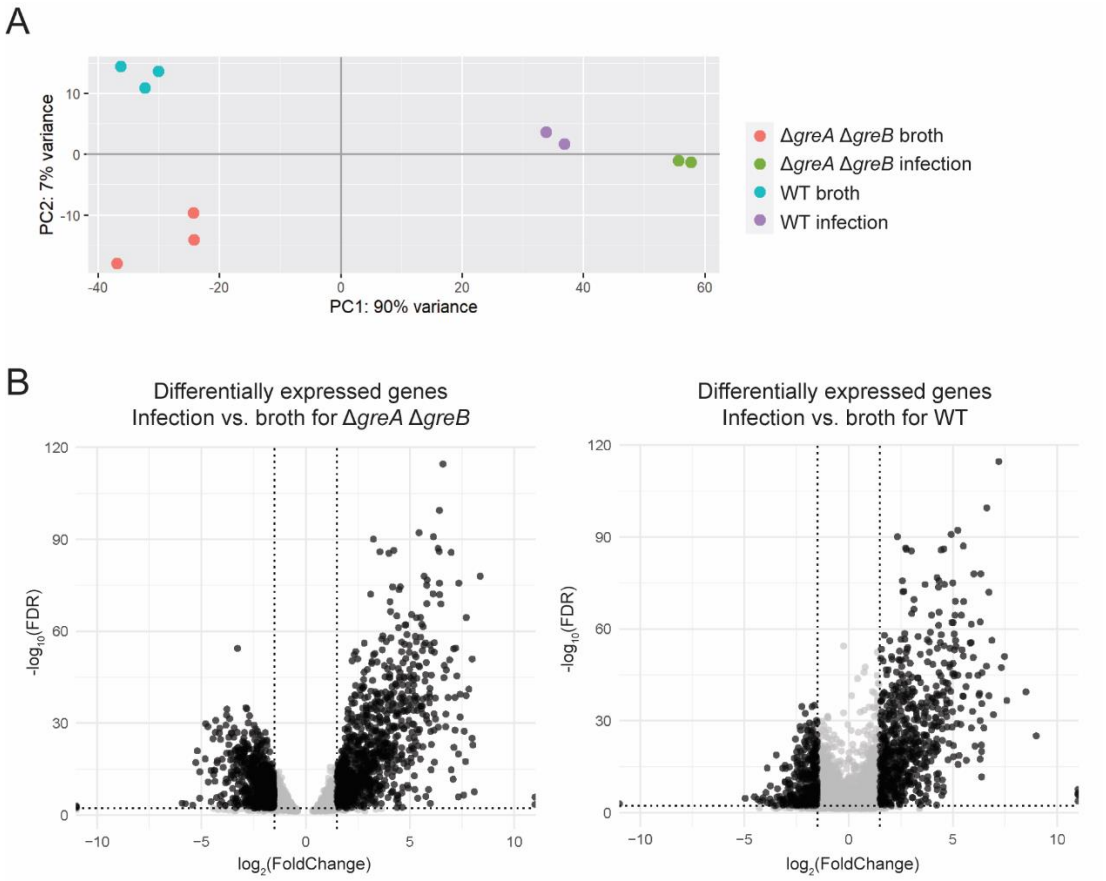

Figure 2.10. Principal component analysis. Principal component analysis of RNA-seq data obtained from WT and Δ*greA* Δ*greB* cells grown in broth and at 8 h p.i.. PC1 is condition (broth and infection) and PC2 is genotype (WT and Δ*greA* Δ*greB*). (B) Volcano plot of the distribution of differentially expressed genes in infection versus broth for Δ*greA* Δ*greB* cells (left) or for WT cells (right) upon Deseq2 analysis. Each dot represents the mean value of one gene from two (infection) or three (broth) independent replicates. Black color indicates significant results (FDR < 0.05, $\log_2$FC > |1.5|).

Differential expression analysis revealed that 1183 genes were differentially expressed in the absence versus the presence of Gre factors during infection (FDR-corrected $p < 0.05$) (Fig. 2.11A, Table S2.3). Expression levels for 713 of these genes differed by more than two-fold ($\log_2$ of the fold change ≥ |1|) (Table S2.3). Of these genes, 369 were downregulated and 344 were upregulated (Table S2.3). GO functional enrichment analysis of these differentially expressed genes revealed that expression of genes related to metabolic processes is significantly impacted by the Gre factors during infection, as these functions are overrepresented and underrepresented in downregulated and upregulated genes, respectively (Fig. 2.11B). GO analysis did not reveal an overrepresentation of any functions in upregulated genes.

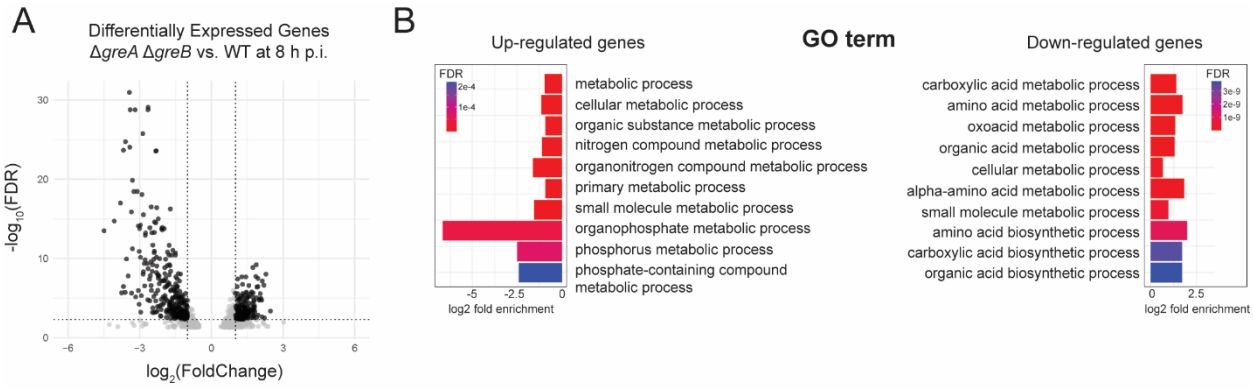

Figure 2.11. Differential gene expression during infection in the absence of Gre factors. (A) Volcano plot summary of the distribution of the 1183 differentially expressed genes in Δ*greA* Δ*greB* cells versus WT during infection after Deseq2 analysis. FDR = false discovery rate (*p* value corrected for multiple testing). Each dot represents the mean value from three independent replicates of one gene. Black color indicates significant results (FDR < 0.05, $\log_2$FC > |1|) (see Table S2.3). (B) Top ten GO terms overrepresented in downregulated genes and underrepresented in upregulated genes. Fold enrichment is defined as the ratio of the number of genes observed in the dataset versus the number of genes expected to belong to that GO term based on the reference list. Fold enrichment was then $\log_2$ transformed. GO terms with values greater than zero represent overrepresented terms, and those with values less than zero represent underrepresented terms.

Interestingly, on average, SPI-1 genes were poorly expressed during infection in cells lacking Gre factors (Table S2.3), which is consistent with previous gene expression data from Δ*greA* Δ*greB* cells grown in broth culture (105, 106). This indicates that backtracking resolution facilitated by the Gre factors is likely required for efficient expression of these genes during infection, as was previously hypothesized (106). However, in contrast to previously reported results from cells grown in broth culture (105, 106), our data indicate that SPI-2 genes were moderately upregulated during infection in cells lacking Gre factors (Table S2.3). This contrast with previous studies is likely due to the difference in growth conditions, as multiple distinct pathways exist for SPI-2 induction that are triggered by different environmental conditions (135), such as those differences between broth culture and infection. Overall, our data suggest that the potential role of the Gre factors in regulating gene expression also extends to infection.

We hypothesized that genes experiencing RNAP backtracking during infection (a low *d* ratio) would also exhibit low gene expression. To test this, we again focused on genes at 8 h p.i., when backtracking is most prevalent genome wide. We categorized genes into three groups according to differential RNAP occupancy ($d \leq 0.75$, $0.75 < d \leq 1.25$, and $d > 1.25$) and determined the average expression level (mean of the log2 of the fold change) of significantly differentially expressed genes (FDR < 0.05) within each group. Indeed, genes that experience severe backtracking ($d \leq 0.75$) are significantly downregulated in the absence of the Gre factors compared to genes with little-to-no backtracking ($0.75 < d \leq 1.25$), or those where the absence of Gre factors increases RNAP occupancy compared to WT ($d > 1.25$) (Fig. 2.12). These data suggest that, on average, the expression of genes experiencing backtracking during infection in the absence of Gre factors is significantly decreased. Thus, not only does backtracking influence gene expression, but if backtracked RNAPs are not resolved, regulation of gene expression during infection suffers.

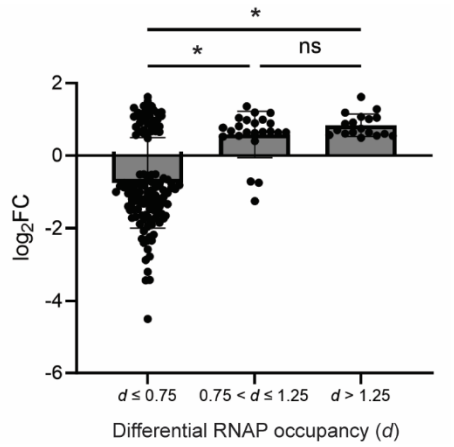

Figure 2.12. Genes experiencing backtracking are downregulated during infection. All top-transcribed genes at 8 h p.i. (see Table S2.1) were sorted into three groups by the value of their *d* ratio. The average change in gene expression (as indicated by $\log_2$ of the fold change) for each group was calculated. *p<0.0001, ns: not significant.

Resolution of backtracking is critical to pathogenesis

Knowing that backtracking is prevalent at key virulence genes during infection and that genes experiencing backtracking are downregulated, we tested whether disruption to RNAP progression is problematic for pathogenesis in general. A prior study demonstrated that the ability of *S.* Typhimurium lacking both Gre factors to invade human host cells or colonize a mouse spleen was ablated compared to WT (106). However, how backtracking affected growth of *S.* Typhimurium over the course of the infection, beyond initial host invasion, was not examined.

To determine the impact of pervasive backtracking on *S.* Typhimurium pathogenicity throughout the course of infection, we infected HeLa cells with backtracking-prone *S.* Typhimurium lacking one or both Gre factors. We measured the number of viable intracellular *S.* Typhimurium over the course of the infection using a standard gentamicin protection assay (Fig. 2.13). Deletion of *greB* alone did not significantly affect the growth of *S.* Typhimurium inside

HeLa cells over the course of the infection, while deletion of *greA* led to a modest, yet statistically significant loss in intracellular growth. However, deletion of both Gre factors significantly abrogated intracellular growth of S. Typhimurium throughout the duration of the infection (Fig. 2.13A). Interestingly, at 1 h p.i., there are no differences in the viability of cells in the presence or absence of Gre factors, supporting our earlier finding that backtracking is not pervasive at this point during the infection. The negative effect of depleting cells of both Gre factors is specific to infection, as there were no significant differences in growth of these strains compared to WT when cultured in LB (Fig. 2.13B). Taken together, our data indicate that pervasive backtracking during infection disrupts gene expression, and resolution of that backtracking is critical to intracellular survival.

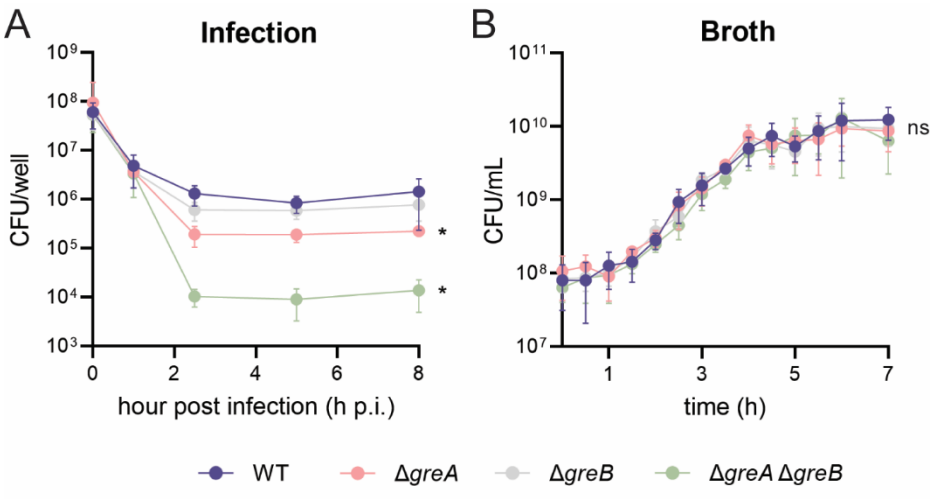

Figure 2.13. Resolution of backtracking during infection is key to pathogenesis. Average number of CFUs of backtracking-prone *S.* Typhimurium either (A) isolated during infection or (B) cultured in LB-Lennox culture at 37°C. Error bars indicate standard deviation of at least six independent replicates. *p<0.005, ns: not significant, one-way ANOVA with Bonferroni correction.

Discussion

During infection, bacterial pathogens must survive numerous host defense mechanisms that are not only damaging to the cell itself but can also lead to DNA damage capable of stalling RNAP (35–38). Pathogens must quickly sense and respond to the host environment, and any disruption to these virulence gene programs at the transcriptional level would be significant. Our work suggests that infection of a host significantly impacts bacterial RNAP dynamics and gene expression, causing pervasive backtracking of these machineries along the bacterial chromosome that must be resolved to ensure successful pathogenesis.

Our data show that disruptions to RNAP movement increase as the infection progresses. As evidenced by the lack of difference in RNAP occupancy in the presence and absence of Gre factors at 1 h p.i. (Fig. 2.6, Table 2.5), it appears that early during infection, RNAPs are not yet perturbed by interactions with the host. This is further emphasized by growth of cells at 1 h p.i.: cells at this time point do not require Gre factors to survive during infection (Fig. 2.13A). It is possible, however, that RNAP disruptions at 1 h p.i. are underestimated in our experiments due to detection thresholds of PIC-seq as a population assay and/or the heterogeneity of the cell population at this early point post infection (136). Cells at 8 h p.i., however, have endured stress from the host environment for much longer than those at 1 h p.i., potentially increasing the likelihood of DNA damage accumulation that can stall RNAPs and lead to backtracking. Accordingly, cells at this later time point exhibit large differences in RNAP occupancy and disruptions to gene expression (Figs. 2.6, 2.7, 2.11, 2.12). These observations, combined with the lack of intracellular viability observed in the absence of Gre factors (2.13A), suggest that cells require backtracking resolution for intracellular proliferation, most significantly later during infection. These problems are likely occurring at earlier timepoints; however, they are most apparent at later timepoints.

Our findings show for the first time that backtracking occurs genome wide in a bacterial pathogen during infection to a greater extent than in the same cells grown outside the host environment (LB-Lennox broth culture, in our case). Thus, backtracking resolution facilitated by the Gre factors is likely to be even more important during infection due to the increased prevalence of these RNAP disruptions. Consistent with this conclusion, even though Gre factor-dependent expression of a variety of genes had been identified in broth culture models (i.e. outside of infection) (105, 106), our data demonstrate that the role of the Gre factors in regulating gene expression does indeed extend to cells grown within the host environment during infection. Specifically, our data show the downregulation of genes that experience backtracking. Interestingly, in *S*. Typhimurium grown in minimal broth culture, Gre factors had been previously shown to facilitate transcriptional elongation of central metabolic genes that resist oxidative stress (105). Our data extend these findings and reveal that backtracking occurs within three of these metabolic genes (*nuoA*, *eno*, *gapA*) at 8 h p.i. (Table S2.1), suggesting that, during infection, the ability of the Gre factors to facilitate transcriptional elongation of key metabolic genes may extend to infection as well. In light of these previously published results, our data are not mutually exclusive with the possibility that some of the changes in RNAP occupancy that we observe during infection are due to altered metabolism within the pathogen itself, and not solely due to damaging host defense mechanisms. In other words, in much the same way that different broth conditions alter *S.* Typhimurium metabolism (105), infection of a host may lead to changes in endogenous *S*. Typhimurium metabolism (such as generation of reactive oxygen species) that could affect RNAP backtracking. Regardless of the origin of any increased damage to the DNA in *S*. Typhimurium during infection, our data nonetheless suggest that RNAP backtracking is worsened during infection as a result.

The pervasive, genome-wide backtracking during infection that we identified likely impacts a multitude of gene functions. Therefore, our work supports the possibility that Gre

factors are required for backtracking resolution within a wide variety of genes and determinants of pathogenesis, some of which may have yet to be defined. Overall, our results highlight an emerging appreciation for the role of the Gre factors in bacterial pathogens, specifically during infection. This class of proteins promote survival and guarantee proper gene expression during infection by regulating and modulating RNAP backtracking during transcription elongation. It would be interesting to test other bacterial pathogen-eukaryotic host infection models for similarly increased levels of backtracking. In fact, similar impacts of the Gre factors on virulence phenotypes have also been shown in other, highly divergent bacterial species, highlighting the conservation of the importance of Gre factors for pathogenesis (137–139).

Precise regulation of gene expression during infection is critical to the survival of bacterial pathogens. Our data strongly suggest that host defense mechanisms, such as those resulting in DNA damage, threaten this precision. Any disruptions to RNAP progression arising during infection must be resolved. Altogether, our results support the model that backtracking is prevalent during infection and that Gre factors are required to ensure proper gene regulation and survival during infection.

CHAPTER 3

CONCLUSIONS AND REMAINING OPEN QUESTIONS

Summary of Findings

Accurate genome duplication and efficient gene expression are the fundamental building blocks of life. Genetic information stored in DNA must be replicated, transcribed, and translated for cells to live, grow, and reproduce (1). Because these processes happen ubiquitously and simultaneously along the entire genome, physical confrontations between their machineries are inevitable (49). These replication-transcription conflicts lead to numerous outcomes that threaten genome integrity. Though codirectional conflicts are not benign (3), head-on conflicts are more detrimental to cells, largely due to their topological constraints and propensity to form R-loops (11). As such, the replisome stalls, disassembles, and must restart at sites of conflicts, all of which threaten genome stability (49). Most notably, mutation rates are accelerated due to head-on conflicts (19). Considering these negative consequences, bacteria have evolved various ways to mitigate the occurrence of conflicts and have adopted multiple strategies to resolve or tolerate them when they inevitably arise (49).

Little is known about replication-transcription conflicts that occur naturally within bacteria grown outside of a laboratory culture. Whether conflicts arise endogenously along the genomes of bacteria inhabiting natural niches has not been tested. Furthermore, consequences of these conflicts have only been studied under laboratory conditions, which do not necessarily reflect natural processes. Given that conflicts are mutagenic, especially those that are head-on, the impact of conflicts arising in an endogenous setting may be profound. For instance, should conflicts occur in bacterial pathogens during infection of a host, it is possible that they promote

mutagenesis within head-on genes. Bioinformatics investigations reveal that head-on genes are broadly enriched in functions related to environmental response, including virulence (56), suggesting that, in fact, this hypothesis could be true.

Testing the above model requires a way to follow molecular transactions with the bacterial genome during infection of a host. Such a technique did not exist prior to my dissertation research. As discussed in Chapter 2, I developed and optimized a protocol by which bacterial proteins could be selectively immunoprecipitated from an infection system to answer open questions such as those discussed above. I used the pathogen *S. enterica* serovar Typhimurium to infect HeLa cells as the model system.

Conflicts between replication and transcription can, of course, only occur within regions of active transcription. These sites were identified in *S.* Typhimurium during infection by mapping RNAP occupancy. This analysis revealed that RNAP occupancy changes significantly during infection of a host, with pervasive backtracking being much more prevalent compared to cells grown outside of infection. Changes to RNAP progression had profound impacts on gene expression: genes that experience more severe backtracking were downregulated. Interestingly, backtracking resolution is key for *S.* Typhimurium pathogenesis.

Not only do these results demonstrate a novel relationship between infection of a host and molecular transactions along bacterial genomes, but this pervasive backtracking also likely exacerbates replication-transcription conflicts. Increased RNAP occupancy along the genome, especially in the stable backtracking state, inherently increases the likelihood that the replisome will meet an RNAP obstacle. Such consequences could be profound, namely in the possibility of accelerated mutagenesis as discussed throughout this dissertation. Whether this model is true, along with the numerous open questions posed below, requires further investigation.

Remaining Open Questions

The above model, that conflicts within head-on virulence could lead to increased mutagenesis during infection, is currently untested and investigating it is predicated on determining whether replication-transcription conflicts occur in bacterial pathogens during infection. This remains a significant open question.

Though the extent to which conflicts are present in this context is currently unclear, *S.* Typhimurium cells lacking key conflict resolution factors UvrD or Rep helicases and antibacktracking factors GreA/GreB do exhibit significant growth defects during infection of HeLa cells (Fig. 3.1), suggesting that conflicts could be an issue for pathogens during infection. Additionally, the results discussed in Chapter 2 show that the prevalence of RNAP backtracking is increased genome wide during infection; such an increase likely worsens the occurrence of conflicts between transcription and an oncoming replication fork (4).

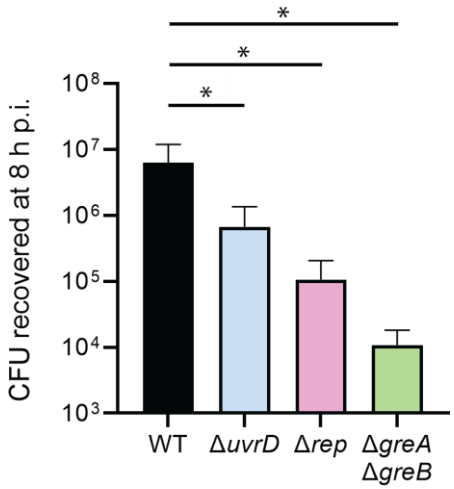

Figure 3.1. Cells lacking key conflict resolution factors exhibit significant growth defects during infection. HeLa cells were infected with S. enterica cells lacking the indicated genes. Bacteria were harvested at 8 h p.i. and plated for CFU enumeration. *p<0.05, one-way ANOVA.

Immunoprecipitation of proteins relevant to conflict formation or resolution can be a powerful method to identify conflict regions. For instance, regions of replication restart due to conflicts in *B. subtilis* have been identified by ChIP of the DnaC helicase and helicase loader (3, 26). The accessory helicase PcrA in *B. subtilis* was implicated in conflict resolution also using ChIP (26), as were the topoisomerases gyrase and Topo IV (18). Transcription has been followed using ChIP of RNAP holoenzyme components, RpoB and RpoC (26). In each of these studies, either a native antibody was used to immunoprecipitate the protein of interest, or the protein was modified with an epitope tag and immunoprecipitated with the corresponding antibody. Immunoprecipitation of DNA:RNA hybrids using the S9.6 antibody (DRIP-seq) was also critical in the discovery that head-on conflicts promote R-loops formation (11, 17). Similar experiments could be carried out in bacterial pathogens during infection using PIC-seq. Applied to the right experiment, PIC-seq would enable conclusions to be drawn about the prevalence of conflicts during infection, whether by way of mapping replication stalling, replication restart, R-loop formation, RNAP stalling, or sites of conflict resolution factor enrichment.

Throughout my dissertation research, I made several attempts to specifically identify conflict regions in *S.* Typhimurium during infection using PIC-seq. Mainly, I focused on mapping enrichment of various replisome proteins and conflict resolution factors as well as mapping R-loop formation. Though my efforts were unsuccessful, they provide valuable insight into what experimental conditions should be considered, what strategic improvements could be made, and what approaches should be avoided in the future. The following sections describe each attempted approach and highlight the ways in which PIC-seq could be altered and/or further optimized to investigate conflicts in bacterial pathogens during infection. Specific future experiments are recommended, and alternative approaches are discussed.

Mapping Replisome Stalling

Conflicts with transcription significantly disrupt DNA replication, in part, by stalling the replisome (5, 61). Therefore, I reasoned that conflicts arising during infection could be identified by mapping transcription-dependent replisome stalling. To accomplish this, DnaN (the β processivity clamp) and DnaB (the replicative helicase) were immunoprecipitated from *S.* Typhimurium cells during infection using PIC-seq. These two proteins were chosen because there are multiple copies of each protein at every replication fork in the cell: DnaN exists as a homodimer, with 3-6 dimers (6-12 DnaN proteins) per fork; DnaB exists as a homohexamer, with one hexamer (6 DnaB proteins) per fork (13, 140–142). Also, DnaB appears to stay stably associated with the fork longer than other replisome components in the case of replisome stalling (141).

Because there were no native antibodies available for these two proteins, epitope tags were incorporated. DnaN was N-terminally tagged by incorporation of Ypet, a GFP derivative, into the *dnaN* gene at its endogenous location on the chromosome. The *dnaB* gene is essential, so the gene, along with its putative endogenous promoter, were cloned into a plasmid where a 6x His tag would be incorporated into the C-terminus of the protein. The endogenous *dnaB* copy remained unaltered. *S.* Typhimurium maintained this tagged DnaB expression plasmid during the infection, even without selection. Both tagged proteins could be visualized by Western blotting (Fig. 3.2). To exacerbate replisome stalling due to conflicts with transcription in an attempt to increase detectable signal, DnaB was immunoprecipitated from cells lacking either RNase HI or GreA.

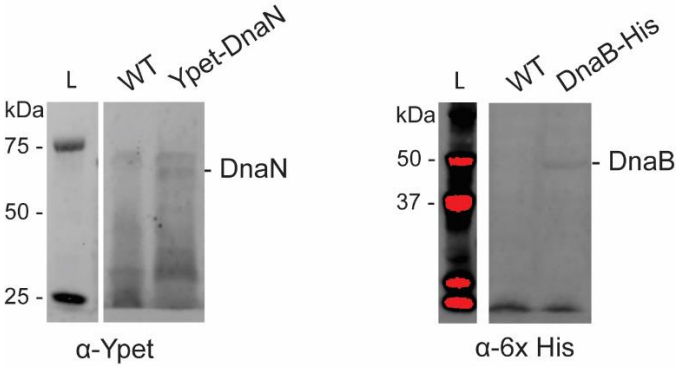

Figure 3.2. Western blot showing detection of tagged DnaN and DnaB.

However, upon sequencing the immunoprecipitated DNA, there were no regions of DnaN or DnaB enrichment over background detected (data not shown). There are several reasonable explanations for this negative data. First, although both DnaN and DnaB exist in multiple copies per replication fork, it is possible that these multiple stoichiometries are insufficient for effective signal detection, given that there are only two replication forks per cell. Furthermore, DnaN may exhibit a short half-life at arrested forks (140). Similarly, not every cell will contain a stalled replication fork, and for those that do, those forks are likely to be stalled randomly throughout the genome. This lack of homogeneity could dilute detectable signal of replication fork enrichment at any particular site, due to the nature of PIC-seq as a population assay. The best way to overcome these problems may be to introduce an engineered conflict system similar to the one that has been widely used in *B. subtilis* studies (11, 19, 26). Transcription of the reporter could be controlled, where high transcription would be expected to stall replication, which could then be measured at that site using the above technique. Although this method would fall short of detecting endogenous conflicts, it would provide an intermediate step to mapping replisome stalling in bacterial pathogens during infection.

Another possible explanation for the lack of detectable replisome stalling is that the rate of *S.* Typhimurium growth during infection is slow, even for WT cells (Fig. 2.1B). Therefore, it cannot be ruled out that there are not enough detectable replication forks to surpass the detection threshold of PIC-seq at the current scale. To address this, the scale of the infections could be increased, or compounds such as wortmannin could be added to increase the replication rate of the rapidly-dividing cytosolic population of *S.* Typhimurium inside HeLa cells (136, 143). Alternatively, a completely different bacterial infection system could be used for replisome protein PIC-seq, where the bacteria cells have a faster growth rate during infection. *Listeria monocytogenes*, for example, exhibits a multiplication rate of 40 min in a murine macrophage model of infection (144), much faster than the apparent rate of *S.* Typhimurium replication in this work.

Lastly, there are obvious weaknesses to consider when using epitope-tagged proteins. For instance, DnaN forms dimers and DnaB forms hexamers. This quaternary structural organization could make a tag inaccessible to the antibody, regardless of which terminus is tagged. In the case of DnaB, since the endogenous copy of the gene is still intact, 6x His-tagged DnaB monomers expressed from the plasmid may be excluded from the hexamer due to altered structure. Even if the tagged DnaB monomers can be incorporated into the hexamer, untagged monomers will dilute the detectable signal. It would be best to use native antibodies in the future to avoid these shortcomings.

Mapping R-loop Formation

Conflicts between replication and transcription promote the formation of R-loops (11, 17). Mapping sites of R-loop enrichment, therefore, could be used as a proxy to locate conflict regions. R-loops can be detected by DRIP-seq (145).

I immunoprecipitated R-loops from *S.* Typhimurium using the S9.6 antibody. However, this antibody is specific to DNA:RNA hybrids, regardless of whether they are prokaryotic or eukaryotic in origin. Therefore, HeLa host cell lysate would first need to be washed away to increase antibody selectivity for bacterial hybrids. However, in doing so, the quantity of eukaryotic lysate and nucleic acids remained orders of magnitude greater than that of the infecting *S.* Typhimurium cells, such that I was unable to sufficiently isolate the bacterial cells or detect hybrids contained within them. In other words, there was enough contaminating nucleic acids from the host cells to presumably saturate the S9.6 antibody, precluding it from recognizing bacterial hybrids to any detectable degree. It is possible that this method could be improved using techniques like those implemented in the RNA-seq experiments described in Chapter 2. For instance, differential lysis with stringent detergents and centrifugation may improve bacterial isolation from infected host cells (146). However, these steps are time consuming, during which time R-loops may be altered from their original state when the cell resided within the host cell. Crosslinking the infection system before immunoprecipitation (as in traditional ChIP techniques) is an option to avoid this issue; however, proteins crosslinked to the chromatin may interfere with S9.6 recognition of the hybrids. Additionally, the scale of these experiments would have to increase significantly even from the scale of the above RNA-seq experiments, since DNA:RNA hybrids are rare in contrast to omnipresent RNA molecules.

Due to the shortcomings of using DRIP-seq to detect R-loops during infection, I adopted an alternative approach to identify R-loops. RNase HI, encoded by the *rnhA* gene in *S.* Typhimurium, is the ribonuclease that digests the RNA moiety of DNA:RNA hybrids, thereby

resolving R-loops at conflict regions. Finding sites of RNase HI enrichment using PIC-seq may indicate where R-loops, and, by extension, conflicts, are occurring, much like mapping R-loop formation can be a proxy for identifying conflict regions. To accomplish this, the endogenous copy of *rnhA* was replaced with a mutated version that encodes a catalytically inactive enzyme. According to the literature, this enzyme retains its high affinity for nucleic acids, but cannot process R-loops (147, 148). Consistent with this, growth of cells expressing the inactive form of *rnhA* during infection mimics that of cells lacking *rnhA* (Fig. 3.3). An 6x His tag on the N-terminus of the protein facilitates immunoprecipitation.

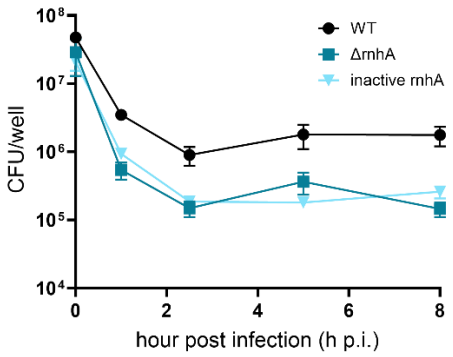

Figure 3.3. RNase HI is required for growth during infection. Average number of CFUs of *S*. Typhimurium lacking *rnhA* or expressing an inactive form of *rnhA* isolated during infection. Error bars indicate standard error of the mean of at least three independent replicates.

Enrichment of inactive RNase HI along the *S.* Typhimurium genome during infection was determined by PIC-qPCR. Multiple target genes were tested, including *cspA*, *gstA*, *mgrB*, and *virK*. Each of these genes were chosen because they contain high RNAP occupancy as determined by PIC-seq of RpoB (see Chapter 2). I reasoned that areas of high RNAP occupancy were prone to replication-transcription conflicts and/or R-loop formation, thus, they

may be areas of inactive RNase HI enrichment. However, no enrichment of the inactivated enzyme was found at any of the target genes.

There are multiple reasonable explanations for lack of detectable signal, none of which are mutually exclusive. First, as discussed in the previous section, using epitope tags to effectively immunoprecipitate a protein of interest is often an imperfect practice. Despite the success of using epitope tags in this method in eukaryotic systems (149–151), there was little precedent regarding epitope tagging of bacterial RNase HI for the purpose of immunoprecipitation. It is also possible that the specificity between the 6x His tag and the antibody was not enough or was not optimized to a suitable degree to attain detectable signal. Though unlikely since a 6x His tag is relatively small, it is possible that the tag could interfere with the ability of the inactive enzyme to recognize and/or associate with nucleic acid structures. Using a native antibody would minimize these specific issues. Lastly, there may be other factors present that can resolve R-loops, such as DinG or RNase HII to an extent, preempting association of the tagged, inactive RNase HI. PIC-seq for inactive RNase HI should be done in the absence of other R-loop resolution factors to increase association of the enzyme with the chromosome.

One possible explanation for the negative outcomes of both methods, DRIP-seq from infection and PIC-seq of inactive RNase HI, is the same one described in the section above: an insufficient signal-to-noise ratio. Even if the experiments are technically sound, there may not be enough R-loops or inactive RNase HI proteins in the cells to surpass the detection threshold of PIC-seq. Again, incorporating an engineered conflict system with inducible transcription onto the chromosome may be the most straight forward strategy to successfully measure R-loop levels or RNase HI enrichment during infection. Attaining detectable signal this way would aid in optimizing any epitope tag-antibody pairs as well.

Different Applications of PIC-seq

Although we successfully used PIC-seq to immunoprecipitate RNAP from a *Salmonella* infection model of HeLa, PIC-seq can, in principle, be adapted to examine any protein of interest with a specific antibody (or an epitope tag/antibody pairing) in any host-pathogen combination. These investigations also do not have to be limited to mapping replication or transcription in the context of conflicts. For instance, PIC-seq could be used to probe regulation of global gene expression during infection. One could use PIC-seq to map transcription factor (TF) binding sites during infection to reveal how TF binding is altered in response to a host (152). Temporal resolution could also be attained by performing these experiments over a time course. In conjunction with expression data, PIC-seq could also be used to decipher virulence gene regulons, which often involve multiple regulator factors with both condition- and time-sensitive responses (129, 153, 154). The sensitivity of PIC-seq could also be harnessed to probe protein-protein interactions, for instance, in a co-immunoprecipitation scheme. Such an approach would likely require some optimization, as PIC-seq is specifically designed for isolating chromatin.

Alternative Methodologies and Future Directions

In addition to the above approaches that capitalize the potential of PIC-seq, there are numerous alternative methodologies to map conflict regions in bacterial pathogens during infection and to study their effects. There are also open questions remaining regarding the increased levels of RNAP backtracking during infection that we discovered and the impact it may have. In addition to these, PIC-seq could be used

Mapping conflict resolution factor enrichment

In a similar vein to mapping RNase HI signal to identify conflict regions during infection, as described in the previous section, there are numerous other conflict resolution factors that could be immunoprecipitated using PIC-seq. Native antibodies should be used, when possible, to avoid the weaknesses of epitope tagging discussed above. Also, immunoprecipitations should be performed in strains missing other resolution factors to exacerbate conflicts and increase signal. It may be helpful to start in a strain carrying an engineered conflict system, as a type of positive control, to facilitate optimization of pulldown conditions and scale of the infection.

Replisome velocity analysis

One of the hallmarks of replication-transcription conflicts is slowed and/or stalled replication. Such events can be detected by accumulation of replisome proteins outside the origin, as explained in the previous section. However, a new, high-resolution method has been developed in which *in vivo* replisome dynamics can be robustly and quantitatively characterized (155). This method, known as lag-time analysis, can be used to determine replisome velocity by comparing genome copy number in newer versus older cells over time (155). This method could be adopted for use in bacterial pathogens during infection to determine how infection impacts replication dynamics. Regions where conflicts arise during infection can be identified by measuring replication velocity as a function of transcription level/RNAP occupancy (which can be modulated using rifampicin, a transcription inhibitor).

Mutagenesis

Replication-transcription conflicts are mutagenic (19, 56). Should they arise in virulence genes in bacterial pathogens during infection, they have the potential to significantly increase adaptive variation of pathogenesis (56). Therefore, in addition to identifying conflict regions, it is critical to determine whether these conflicts induce mutagenesis in this context. This could be done by flipping a virulence gene into the opposite orientation on the chromosome and mapping mutation accumulation over the course of an infection with duplex sequencing. An interesting possibility would be to perform this analysis on genes with varying levels of transcription. Alternatively, a reporter gene could be incorporated in either orientation under an inducible promoter, creating a more controlled environment for measuring mutation accumulation (forward genetic assay). This approach would also avoid potentially deleterious effects of modifying endogenous virulence gene expression programs. Luria-Delbruck fluctuation assays could also be used to measure mutation rates and compare across conditions (156). Such mutagenesis has the potential to accelerate the evolution of bacterial pathogens specifically during infection, allowing them to adapt to the host environment and possibly develop antimicrobial resistance.

In addition to exacerbating conflicts, increased RNAP backtracking during infection could, on its own, have significant impacts on genome integrity. Previous work has shown that cells that are prone to backtracking accumulate mutations faster (4) and have higher rates of recombination than cells with both Gre factors (157). Additionally, stalled RNAPs trigger downstream DNA repair pathways that can be mutagenic (35, 71, 72, 132, 158, 159). Although pathogenic cells normally carry both Gre factors and will thus experience less backtracking under regular conditions, the fact that these RNAP stalling events are much more prevalent during infection inherently increases the chances of mutagenic events occurring. Therefore, though the host is fighting the bacteria with DNA damaging mechanisms such as oxidative stress, the host may also be enabling bacteria to attain mutations that could lead to adaptive

evolution. This model requires further investigation. However, it is certainly interesting to investigate in future studies as this dichotomy could perpetuate the development of AMR and/or hypervirulence.

Determining the impact of oxidative stress during infection

The data discussed in Chapter 2 strongly suggest that RNAP backtracking is more prevalent during infection than in broth. Backtracking can occur because of damaged DNA templates (35–38), which can arise from both endogenous and exogenous sources (160). Oxidative stress is a primary defense mechanism leveraged by eukaryotic host cells against bacterial invaders and is likely to be a significant source of damage to pathogenic DNA during infection (94–98). However, whether this damage indeed contributes to the pervasive RNAP backtracking observed during infection is an outstanding question. Similar experiments to those in Chapter 2 could be performed in the presence and absence of antioxidants or eukaryotic NADPH oxidase inhibitors to determine if this is the case. Additionally, levels of backtracking could be compared between HeLa cell and macrophage infection models, as the latter imposes greater oxidative burden onto bacterial pathogens (98). It would also be interesting to perform these experiments using *L. monocytogenes* as the bacterial pathogen, since it is generally more tolerant to oxidative stress and often uses cues from an oxidative environment to establish its virulence (161, 162). It is also of interest to determine how altered metabolism within the pathogen during infection contributes to endogenous oxidative stress, DNA damage, and ultimate RNAP backtracking.

Mapping exact backtracking sites with RNET-seq

The data presented here strongly suggest that backtracking is prevalent at key virulence genes during infection, though these studies do not reveal the exact location within the gene where this backtracking takes place. A novel nascent elongation transcript sequencing followed by RNase digestion method known as RNET-seq could be used to identify backtracked RNAPs and reveal the exact sequence upon which they are backtracked (163–165). Such an analysis would be key for determining the degree to which oxidatively damaged DNA bases disrupt RNAP and where RNAP backtracking occurs during infection, with single nucleotide resolution. This type backtracking mapping may also reveal consensus sequences within the DNA that lead to backtracking, in addition to those RNAPs stalled by DNA damage. For instance, it has been demonstrated that RNAP consistently backtracks in the 3'UTR of the *S.* Typhimurium *hilD* gene, though the mechanism is unknown (106). Overall, RNET-seq is the next step for higher resolution mapping of RNAP backtracking in bacterial pathogens during infection, enabling deeper investigations into their causes, dynamics, and impacts.

CHAPTER 4

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

S. enterica serovar Typhimurium SL1344 was the wild type (WT) strain used in these studies. Derivative mutant strains are listed in Table 4.1. Bacteria were grown on LB-Lennox (5 g/L yeast extract, 10 g/L tryptone, 5 g/L NaCl) agar plates at 37°C with the following antibiotics when appropriate: 25 μg/mL chloramphenicol (Cat), 50 μg/mL kanamycin (Kan), 100 μg/mL carbenicillin (Carb), or 10 μg/mL tetracycline (Tet). Single colonies were used to inoculate liquid cultures in LB-Lennox and grown at 37°C with aeration (260 rpm).

Construction of Chromosomal Deletion Mutants

The bacteriophage λ Red recombination system was used for construction of chromosomal deletion mutants (166–168). The plasmid pSIM27 was transformed into WT SL1344 for expression of the Red recombinase system, which yielded strain HM4324. Gene deletions were made as described below using primers that contained 40 nucleotides homologous to the regions immediately flanking each gene (including the ATG start codon, Table 4.2). Each gene deletion and mutation was confirmed by Sanger sequencing.

Strain	Genotype/Features	Identifier	Reference/Source
SL1344	WT *S.* Typhimurium strain SL1344	HM4315	This study
SL1344 + pSIM27	WT *S.* Typhimurium carrying pSIM27 for λ Red recombination	HM4324	This study
T-SACK	W3110 araD<>tetA-sacB amp fliC<>cat argG::Tn5	HM4177	(166)
Δ*greB*	*greB*::Cat	HM4525	This study
Δ*greA*	*greA*::Kan	HM4527	This study
Δ*greA* Δ*greB*	*greA*::Kan *greB*::Cat	HM4529	This study
Δ*uvrD*	*uvrD*::Cat	HM4481	This study
Δ*rep*	*rep*::Cat	HM4480	This study
Ypet-DnaN	*dnaN*::Kan-ypet-*dnaN*	HM4383	This study
DnaB-6x His	Expresses *dnaB* with C-terminal 6x His tag under endogenous promoter (plasmid pHM720)	HM4474	This study
Δ*rnhA*	*rnhA*::Kan	HM4378	This study
Inactive *rnhA*	*rnhA*::6x His-inactive rnhA (D10R and E48R)	HM4462	This study

Table 4.1. Strains used in this study.

Primer	Sequence	Description
HM6042	CATTGCCCCCTACAGGAATGTTCAAGAGG-TATAACAAATGTATGGACAGCAAGCGAACCG	For recombineering *greA* with the Kan cassette from T-SACK
HM6043	TTTACAATACACCAACAATTTGCGTATTGAG-TACTGCTTATCAGAAGAACTCGTCAAGAAG	For recombineering *greA* with the Kan cassette from T-SACK
HM6040	GTGTGCGCAATATCGACAGCAAAGGTAAAT-CAACGAGATGTGTGACGGAAGATCACTTCG	For recombineering *greB* with the Cat cassette from T-SACK
HM6041	TGCCAGCCATCAGCGGGGGCTTAGGATTC-TTCTTGTCTTAACCAGCAATAGACATAAGCG	For recombineering *greB* with the Cat cassette from T-SACK
HM6034	GCGTTTTTAGTTTTCATGTTTTTACGCGGC-GGTGCCAATGTGTGACGGAAGATCACTTCG	For recombineering *uvrD* with the Cat cassette from T-SACK
HM6035	CCGGATAAGACGCGCCATCCGGCACTATG-AGAGTCGGTCAACCAGCAATAGACATAAGCG	For recombineering *uvrD* with the Cat cassette from T-SACK
HM6036	CAATCCCTCCCCCGTTCGAAGATTGAGCAC-TATTCCCATGTGTGACGGAAGATCACTTCG	For recombineering *rep* with the Cat cassette from T-SACK
HM6037	GTAGTTTGATTGCCGGATGACGCTTCGCTT-ATCCGGCCTAACCAGCAATAGACATAAGCG	For recombineering *rep* with the Cat cassette from T-SACK
HM6028	CCGTTGGCGCCACCCGGCAATATCGCAAAC-CGGATGGCTATATGGACAGCAAGCGAACCG	For recombineering *rnhA* with the Kan cassette from T-SACK
HM6029	TTTGTTATCGATTTCAATTACAGGAAGTCTA-CCAGAGATGCTCAGAAGAACTCGTCAAGAAG	For recombineering *rnhA* with the Kan cassette from T-SACK
HM6438	CAATCAGGCGCTTCATCGTG	For fusing *dnaN* to an N-terminal Ypet tag by recombineering
HM6439	GCCCTGCAAAATAACGCCAT	For fusing dnaN to an N-terminal Ypet tag by recombineering
HM6768	GGATCCCCACCATGCTTGTGAAATTC	Amplify *dnaB* + 99 bp promoter from SL1344, adds a BamHI cut site at 5' end of gene
HM6769	GATATCTTAGTGATGGTGATGGTGATGCTC-ATCATCGTATTGCGG	Amplify dnaB 3' end, adds C terminal 6x His tag and a EcoRV cute site at the 3' end of the gene
HM6648	ATCGGTGAAAATTTCTACCTGTTTAAGCATT-GACATTAACCTATAAAAATAGGCGTATCA	Gibson primer to amplify pGCS for *rnhA* insertion
HM6649	GAAGATAGTGGCTACCAGGCGGAAGCTTA-GTGATAATAATGGTTTCTTAGACGTCAGGTG	Gibson primer to amplify pGCS for *rnhA* insertion
HM6688	CCCCAGGCAAGAGCCGCGGGTGAAAATT-TCTAC	Quick Change primer to make D10R mutation in *rnhA* gene
HM6689	GTAGAAATTTTCACCCGCGGCTCTTGCCT-GGGG	Quick Change primer to make D10R mutation in *rnhA* gene
HM6690	GATCGCCGCCATCAGGCGCATACGGTTAT-TGG	Quick Change primer to make E48R mutation in *rnhA* gene

HM6691	CCAATAACCGTATGCGCCTGATGGCGGC-GATC	Quick Change primer to make E48R mutation in *rnhA* gene
HM6706	TTGTTATCGATTTCAATTACAGGAAGTCTA-CCAGAGATGCATGCATCACCATCACCATC-ACCTTAAACAGGTAGAAAT	Amplify inactive *rnhA* from pGCS for recombineering into Δ*rnhA* KO, also adds 6x His tag to 5' end of gene
HM6708	CAGACCGTTGGCGCCACCCGGCAATATCG-CAAACCGATGGTTACCAATGCTTAATCAGTG	Amplify inactive *rnhA* + AmpR from pGCS for recombineering into Δ*rnhA* (3' end of gene)

Table 4.2. Primers used in this study to construct chromosomal deletion strains.

Construction of greA and greB knockouts

Kanamycin and chloramphenicol resistance cassettes were amplified with Phusion High-Fidelity polymerase (Thermo) from T-SACK gDNA (166) using primers HM6042/HM6043 and HM6040/HM6041 for deletion of *greA* and *greB*, respectively. Amplicons were electroporated separately into HM4324 and cells were grown on selective LB agar to yield single knockout strains Δ*greA* (*greA*::Kan, HM4527) and Δ*greB* (*greB*::Cat, HM4525). The Δ*greB* strain was also transformed with pSIM27 in order to construct the double knockout strain Δ*greA* Δ*greB* (HM4529). The *greA* gene was replaced with the kanamycin resistance cassette in the Δ*greB* + pSIM27 strain using the same steps as above.

Construction of rep, uvrD, and rnhA knockouts

Kanamycin and chloramphenicol resistance cassettes were amplified with Phusion High-Fidelity polymerase (Thermo) from T-SACK gDNA using primers HM6034/HM6035, HM6036/HM6037, and HM6028/6029 for deletion of *uvrD*, *rep*, and *rnhA*, respectively. Amplicons were electroporated separately into HM4324 and cells were grown on selective LB

agar to yield single knockout strains Δ*uvrD* (*uvrD*::Cat, HM4481), Δ*rep* (*rep*::Cat, HM4480), and Δ*rnhA* (*rnhA*::Kan, HM4378).

Construction of Ypet-DnaN strain

The N-terminal fusion of Ypet to the *Salmonella dnaN* gene was amplified with Phusion High-Fielity polymerase (Thermo) from HM3756 gDNA using primers HM6438/6439. The amplicon was electroporated into HM4324 and cells were grown on selective LB agar to yield an N-terminal fusion of Ypet to DnaN.

Construction of DnaB-6x His expression plasmid

Primers HM6768 and 6769 were used to amplify the *dnaB* gene, including 99 bp upstream of the start codon of the gene, from SL1344 gDNA and add a *Bam*HI and *Eco*RV restriction sites. The resulting amplicon was digested with *Bam*HI/*Eco*RV and ligated into *Bam*HI/*Eco*RV-digested pWSK29. This plasmid, pHM720, was transformed into SL1344 and maintained with carbenicillin.

Construction of inactive rnhA strain

The Δ*rnhA* strain was transformed with pSIM27 for recombineering. Primers HM6650/6651 were used to amplify the WT *rnhA* gene from SL1344 gDNA and primers HM6648/6649 were used to amplify the pGCS plasmid. The *rnhA* gene was inserted into the pGCS backbone by Gibson assembly. The resulting plasmid was amplified with primers HM6688/HM6689 to introduce D10R mutation. The resulting plasmid was purified and subsequently amplified with primers HM6690/6691 to introduce the E48R mutation. Both

mutations were confirmed by Sanger sequencing. The inactive *rnhA* gene was amplified from this plasmid using primers HM6706/HM6708, which also added a 6x His tag to the 5' end of the gene. The amplicon was electroporated into Δ*rnhA* + pSIM27 and cells were grown on selective LB agar to yield an N-terminally tagged inactive RNase H1.

Mammalian Cell Culture

HeLa cells (ATCC) were grown in high glucose Dulbecco's minimal essential medium (DMEM, Gibco – 11995065) supplemented with 10% heat inactivated fetal bovine serum (FBS, R&D Systems), 4 mM L-glutamine (Gibco), and 1X Penicillin/Streptomycin (Gibco). Antibiotic-free media of the same formulation was used for bacterial infections. Cells were maintained at 37°C with 5% CO_2 and passaged following ATCC guidelines.

Seeding and Bacterial Invasion of Mammalian Cells

HeLa cells at low passage number were seeded at the following densities 16-18 hours before infection: 1.5×10^5 cells per well (24-well dish) or 1.14×10^7 cells per plate (15-cm plate). Immediately prior to infection, HeLa cells were washed with 1X PBS (Gibco), and antibiotic-free media was added to the cells. Cultures of *S*. Typhimurium were inoculated in LB from single colonies and grown overnight at 37°C while shaking. The next day, the precultures were diluted back to OD_{600} = 0.05 in LB and grown at 37°C while shaking until OD_{600} reached 0.6 (approximately 2 h). Bacteria were collected by centrifugation, washed in 1X PBS, resuspended in antibiotic-free media, and used immediately to infect HeLa cells at an MOI of ~100:1. Bacteria

were allowed to invade for 1 h at 37°C with 5% CO_2. Bacteria were removed after invasion, and HeLa cells were washed once with 1X PBS and fresh antibiotic-free media was added. Thirty minutes later (1.5 h p.i.), gentamicin (Gibco) was added to a final concentration of 50 µg/mL for the duration of the experiment. Infected HeLa cells were maintained at 37°C with 5% CO_2 until times indicated.

Gentamicin Protection Assays

HeLa cells were seeded in 24-well dishes and infected with bacteria as described above. Extracellular growth of *S*. Typhimurium was inhibited by gentamicin as described above. At indicated time points, infected HeLa cells were washed once with 1X PBS and lysed with ice-cold 1% Triton X-100 in H_2O. Viable bacteria were enumerated by plating on LB agar and grown at 37°C overnight.

Western Blotting

HeLa whole cell lysate was prepared in RIPA buffer and diluted into Laemmli sample buffer. *S.* Typhimurium whole cell lysate was prepared in lysis buffer (10 mM Tris-HCl pH 7, 10 mM EDTA, 1X protease inhibitor, 0.1 mg/mL lysozyme) and diluted into Laemmli sample buffer. Lysates were boiled for 10 min and loaded onto a 12% Mini-PROTEAN TGX gel (Bio-Rad). Proteins were subsequently transferred to a PVDF nitrocellulose membrane. The membrane was blocked for 1 h at room temperature with Intercept PBS blocking buffer (LI-COR) before being immunoblotted with anti-RpoB (8RB13, Thermo), anti-Histone H3 (PA5-16183, Thermo),

anti-GFP (A-11122, Thermo), or anti-6x His (MA1-21315, Thermo) antibodies overnight at 4°C. The blot was incubated with secondary antibodies IRDye-680RD and 800CW (LI-COR) for 30 min at room temperature before being imaged on a ChemiDoc MP imager (Bio-Rad).

Chromatin Immunoprecipitation from Infection

HeLa cells were seeded in 15-cm plates and infected with bacteria as described above. For treatment with rifampicin, 1 mg/mL (final concentration) was added to the media ten minutes prior to each time point. At indicated time points, media was removed from infected HeLa cells, the cells were washed once with 1X PBS and then crosslinked in 1% methanol-free formaldehyde (Thermo) in 1X PBS for 10 min at room temperature. Crosslinking was subsequently quenched with 0.5 M glycine. Crosslinked infected HeLa cells were washed twice with 1X PBS, then dislodged in 1X PBS by scraping. Cells from two 15-cm plates were combined for each replicate and collected by centrifugation. Cell pellets were stored at -80°C for future processing. Pellets were thawed on ice and resuspended in 2.5 mL ice-cold NPT lysis buffer (169) (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5 % NP-40, 0.1 % Triton X-100, complete protease inhibitor cocktail [Roche] added fresh) for 10 min on ice. Lysozyme was added to 10 mg/mL and the lysate was incubated at 37°C for 30 min. Lysates were sonicated for 5 cycles of 30 s on/off (2.5 min total sonication time) at 4°C in a Bioruptor Plus sonication system (Diagenode) and pelleted by centrifugation at 8000 RPM for 15 min at 4°C. A 40 µL aliquot was taken from the lysate supernatant as the input control. For the immunoprecipitation, 6 µL RpoB monoclonal antibody (8RB13, Thermo), 6 µL anti-GFP antibody (A-11122, Thermo), or 6 µL anti-6x His antibody (MA1-21315, Thermo) was added to the lysate supernatant and rotated overnight at 4°C. The next day, 90 µL of a 50% protein A Sepharose bead slurry (GE)

was added, and IPs were incubated for 1 h at room temperature with gentle rotation. Beads were pelleted by centrifugation at 2000 RPM for 1 min. The supernatant was discarded, and the beads were washed six times for 3 min each in wash buffer (50 mM Tris-HCl pH 7.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100), followed by one wash with TE pH 8.0. The elution was carried out at 65°C for 10 min in 200 uL elution buffer I (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS). Beads were pelleted at 5000 RPM for 1 min and the supernatant saved. The beads were washed with 150 uL elution buffer II (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.67% SDS) and pelleted at 7000 RPM for 1 min. The second supernatant was combined with the first eluate. The combined eluates and the accompanying input controls were de-crosslinked overnight by incubation at 65°C. The following day, the eluates and input controls were treated with proteinase K (0.4 mg/mL) at 37°C for 2 h. Sodium acetate was added and the DNA purified by phenol:chloroform:isoamyl alcohol extraction. The DNA was precipitated in 100% ethanol at -20°C and pelleted before being resuspended in resuspension buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

Chromatin Immunoprecipitation from Broth

Cultures of *S*. Typhimurium were inoculated in LB from single colonies and grown overnight at 37°C while shaking. The next day, the precultures were diluted back to OD_{600} = 0.05 in LB and grown at 37°C while shaking until OD_{600} reached 0.6 (approximately 2 h). Bacteria were crosslinked with 1% formaldehyde for 20 min at room temperature before being quenched with 0.5 M glycine. Cells were collected by centrifugation and washed once in cold 1X PBS. Cell pellets were resuspended in 1.5 mL Solution A (10 mM Tris-HCl pH 8.0, 20% w/v sucrose, 50 mM NaCl, 10 mM EDTA, 10 mg/ml lysozyme, 1 mM AEBSF) and incubated at 37° C for 30 min.

After incubation, 1.5 mL of 2X IP buffer (100 mM Tris pH 7.0, 10 mM EDTA, 2% Triton X-100, 300 mM NaCl and 1 mM AEBSF, added fresh) was added and lysates were chilled on ice for 30 min. Lysates were sonicated four times for 10 s (40 s total sonication time) at 30% amplitude and pelleted by centrifugation at 8000 RPM for 15 min at 4°C. A 40 uL aliquot was taken from the lysate supernatant as the input control. For the immunoprecipitation, 2 uL of RpoB monoclonal antibody (clone 8RB13, Thermo), 2 µL anti-GFP antibody (A-11122, Thermo), or 2 µL anti-6x His antibody (MA1-21315, Thermo) was added to 1 mL of the lysate supernatant and rotated overnight at 4°C. The next day, 30 uL of 50% protein A Sepharose bead slurry (GE) was added, and IPs were incubated for 1 h at room temperature with gentle rotation. Beads were pelleted by centrifugation at 2000 RPM for 1 min. The supernatant was discarded, and the beads were washed six times for 3 min each in wash buffer (50 mM Tris-HCl pH 7.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100), followed by one wash with TE pH 8.0. The elution was carried out at 65°C for 10 min in 100 uL elution buffer I (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS). Beads were pelleted at 5000 RPM for 1 min and the supernatant saved. The beads were washed with 150 uL elution buffer II (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.67% SDS) and pelleted at 7000 RPM for 1 min. The second supernatant was combined with the first elution. The combined eluates and the accompanying input controls were de-crosslinked overnight by incubation at 65°C. The following day, the eluates and input controls were treated with proteinase K (0.4 mg/mL) at 37°C for 2 hr. Sodium acetate was added and the DNA purified by phenol:chloroform:isoamyl alcohol extraction. The DNA was precipitated in 100% ethanol at -20°C for 1 hr and pelleted, before being resuspended in resuspension buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

PIC Deep Sequencing and Data Processing

Sequencing libraries were prepared using a Nextera XT DNA Library Preparation kit (Illumina). Libraries were deep sequenced by the Vanderbilt Technology for Advanced Genomics (VANTAGE) sequencing core (Vanderbilt University) on an Illumina NovaSeq platform, resulting in approximately 24M x 150 bp paired-end reads per sample. Raw reads were trimmed (Trimmomatic v0.39 (170)) then mapped to the S. enterica serovar Typhimurium strain SL1344 genome (GenBank: FQ312003.1) using Bowtie2 v2.2.5 (171). Both PCR and optical duplicates were removed using Picard v1.3 (Broad Institute) and bam files were sorted and indexed using SAMtools v1.13 (172).

The number of reads mapping to each gene in both the IP and input samples was quantified using featureCounts v2.0.3 (173) and then normalized to total mapped reads. For each gene, the normalized number of reads in the input sample was subtracted from the normalized number of reads in the IP sample and averaged across three independent replicates to calculate “normalized read count.” For any gene where this calculation resulted in a negative number, the normalized read count was redefined as 0 for that gene. At this point, rDNA genes were excluded from further analysis. Genes were then categorized into five groups based on transcription level using K-means clustering for each individual condition. Transcription level was defined by the normalized read count in the wild-type sample for each condition. The three groups containing the most transcribed genes for each condition were used for further analysis. Ratio *d* was calculated for these genes by dividing the normalized read count for the Δ*greA* Δ*greB* strain by the normalized read count in the wild-type strain. Ratio *d’* was calculated by dividing d from the 1 h p.i. or 8 h p.i. condition by the *d* from the broth condition. Heatmaps were plotted and hierarchical clustering was performed using the pheatmap package in Rstudio. Gene ontology functional enrichment analysis was performed using the PANTHER v.14 (174) pipeline through The Gene Ontology Resource (175, 176).

For enrichment analysis, peaks were called from processed bam files using macs2 v.2.2.7.1[132] and assigned to genome features using the “closest” tool from the BEDTools suite v.2.30.0 (177).

To visualize relative enrichment, bam files were normalized to the total number of mapped reads and the ratio of IP versus input read depths was calculated using the bamCompare tool from the deepTools suite v3.5.1 (178) to generate bedgraph files. These bedgraph files were visualized on the Integrated Genomics Viewer (igv) platform v2.12.1 (179). All statistical tests were performed in GraphPad Prism v9.

Quantitative PCR

Quantitative PCRs (qPCRs) were performed using SsoAdvanced SYBR Green master mix (Bio-Rad) and the CFX96 Touch Real-Time PCR system (Bio-Rad). Data were normalized to gene copy number by the ratios of input to IP samples. Relative enrichment was determined by the ratio of gene copy number for the target gene to *eutN* (primers listed in Table 4.3). Statistical analysis was performed in GraphPad Prism v9.

Primer	Sequence	Description
HM5726	TCGACCGGTGATAATTCGCT	*S.* Typhimurium *eutN* qPCR
HM5727	CGCTATCGACAGTATCGGGG	*S.* Typhimurium *eutN* qPCR
HM6792	GCCCGAGTTGGATCGTCTTC	*S.* Typhimurium *ssaK* qPCR
HM6793	CTGCCGTTTTTGCCTGTCAT	*S.* Typhimurium *ssaK* qPCR
HM6788	CGGTTGCTATGCCAATCTGC	*S.* Typhimurium *gstA* qPCR
HM6789	CGGTTATGTCGCCAACCTGA	*S.* Typhimurium *gstA* qPCR
HM6784	AATCCAGCGGCCCTATTTGT	*S.* Typhimurium *virK* qPCR
HM6785	TCCGGAAAGGACTGAACGAA	*S.* Typhimurium *virK* qPCR
HM7065	CGAACGCCGTGAGTTTGATG	*S.* Typhimurium *phoP* qPCR
HM7066	TAATGCGCCGTAATAGCGGT	*S.* Typhimurium *phoP* qPCR
HM6780	TTTTCCCCGCCCGATTACTG	*S.* Typhimurium *cspA* qPCR
HM6781	CGTCAAACCGTTGAGCACAT	*S.* Typhimurium *cspA* qPCR
HM6794	TCTATCTGCTCGTTGGCACG	*S.* Typhimurium *mgrB* qPCR
HM6795	GGTGCGTTTTCTCTTGCATCT	*S.* Typhimurium *mgrB* qPCR

Table 4.3. Primers used for qPCR amplification in this study.

Bacterial RNA Isolation for RNA-seq

HeLa cells were seeded and infected with *S*. Typhimurium (12 15-cm plates for WT and 24 15-cm plates for Δ*greA* Δ*greB*) as described above. Bacterial RNA was isolated from infected HeLa cells essentially as described previously (110). Briefly, HeLa cells were washed once with pre-chilled 1X PBS after 8 h p.i. and lysed on ice in 15 mL lysis buffer (0.1% SDS, 1% acidic phenol, 19% ethanol in water, ice cold), which also served to stabilize the RNA. Lysates from each plate were collected by scraping and pooled together. Bacteria were collected from the lysates by centrifugation at 3,300 x g for 30 min at 4°C. The pellet was washed three times in wash buffer (0.1% acidic phenol, 19% ethanol in water, ice cold) and centrifuged at 3,800 x g for 30 min at 4°C each time. After the final wash, the bacteria were resuspended in <1 mL wash buffer, moved to a 1.5 mL microcentrifuge tube (bacteria for each *S*. Typhimurium genotype were combined at this point), and collected by centrifugation at 16,000 x g for 2 min at 4°C. The

pellet was resuspended in 1 mL TRIzol on ice by pipetting up and down ~ 60 times. The resuspended pellet was stored at -80°C. Total RNA was extracted as previously described (180) and stored at -80°C prior to DNase treatment (below). The infection process and RNA isolation for library preparation/deep sequencing was performed in duplicate, yielding two independent replicates for both WT and Δ*greA* Δ*greB* strains.

In parallel, RNA was extracted from WT and Δ*greA* Δ*greB S*. Typhimurium grown in broth culture (LB-Lennox) as a control. Overnight cultures grown from single colonies were diluted back to OD_{600} = 0.05 in LB-Lennox and grown at 37°C while shaking. Once the culture reached an OD_{600} of 0.6., 1.5 mL of cells were mixed with 1.5 mL 100% ice-cold methanol and incubated on ice for 10 min. Bacteria were collected by centrifugation at 4,000 rpm for 5 min and immediately used for total RNA extraction using the GeneJET RNA Purification Kit (Thermo) according to the manufacturer's instructions. Purified RNA was stored at -80°C prior to DNase treatment (below). RNA was isolated for library preparation/deep sequencing from three independent replicates for both WT and Δ*greA* Δ*greB* strains.

A 1 µg sample of RNA from each condition (two replicates/infection, three replicates/broth) was treated with RNase-free DNase I for 40 min at 37°C. The digestion was quenched by addition of EDTA and incubation at 65°C for 10 minutes. RNA was stored at -80°C prior to library preparation.

cDNA Library preparation, Deep Sequencing, and Data Processing

Ribo-depletion, library preparation, and deep sequencing were all performed by the VANTAGE Sequencing Core (Vanderbilt University). DNase-treated total RNA samples (500 ng) were depleted of rRNA using NEBNext rRNA Depletion Kit (New England Biolabs [NEB]).

Libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) with indices from NEBNext Multiplex Oligos for Illumina (NEB). Libraries were deep sequenced on an Illumina NovaSeq platform, resulting in approximately 50M x 150 bp paired-end reads per sample.

Raw reads were trimmed (Trimmomatic v0.39 (170)) and mapped to the *S. enterica* serovar Typhimurium strain SL1344 genome (GenBank: FQ312003.1) using Bowtie2 v2.2.5 (171). Both PCR and optical duplicates were removed using Picard v1.3 (Broad Institute) and bam files were sorted and indexed using SAMtools v1.13 (172). The number of reads mapping to each gene was quantified using featureCounts v2.0.3 (173), from which transcripts per million (TPM) was calculated by hand. Differential expression analysis of loci with more than 10 mapped reads total across all samples was performed with Deseq2 v1.38.3 (181) in Rstudio v4.2.3. Plots were created using ggplot2 v3.4.3 in Rstudio or GraphPad Prism v9. Gene ontology functional enrichment analysis was performed using the PANTHER v.14 pipeline (174) through The Gene Ontology Resource (175, 176).

APPENDIX A

SUPPLEMENTAL TABLE LEGENDS

Table S2.1. Top transcribed genes in cells grown in broth, 1 h p.i., and 8 h p.i. List of the top transcribed genes as determined by k-means clustering for each condition. Values represent the average normalized read count and the ratio of the normalized read counts of three independent replicates. This table also lists the 283 top transcribed genes that arise in all three conditions.

Table S2.2. RNAP occupancy changes for the top transcribed genes as categorized by hierarchical clustering. Lists the top transcribed genes for each condition that fall into each hierarchical cluster, as determined by pheatmap function in R. Values represent the average normalized read count and the ratio of the normalized read counts of three independent replicates. Also lists the manually annotated functions of genes in clusters three and four for the 8 h p.i. condition.

Table S2.3. RNA-seq summary. Tab 1 shows raw read counts per gene for each individual replicate for both conditions (broth and 8 hours p.i.). An average of 588,000 reads mapped to features in the infection condition, approximately 1.5 million reads mapped to features in the broth condition. Tab 2 shows TPM calculated for every gene per individual replicate for both conditions (see methods). Tab 3 shows differential expression as determined by DEseq2 for ΔgreA ΔgreB cells versus WT cells at 8 h p.i. Only genes with more than ten mapped reads are included (total = 4504 genes). Tab 4 shows differential expression of genes in SPI-1 and SPI-2. Tab 5 shows differential expression analysis of only those genes where differential expression was statistically significant (FDR < 0.05, total = 1184 genes). Tab 6 shows differential expression analysis of only those genes where the differential expression (as determined by log2 of the fold change) differed by more than two-fold (log2FoldChange ≥ |1|, total = 713 genes). Tab 7 shows differential expression analysis of genes from Tab 6 that are upregulated (344 genes). Tab 8 shows differential expression analysis of genes from Tab 7 that are downregulated (369 genes).

REFERENCES

1. Watson J, Gann A, Baker T, Levine M, Bell S, Losick R. 2013. *Molecular Biology of the Gene*. Pearson. 7th ed.
2. Helmrich A, Ballarino M, Nudler E, Tora L. 2013. Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol*. 20(4):412–18
3. Merrikh H, Machón C, Grainger WH, Grossman AD, Soultanas P. 2011. Co-directional replication-transcription conflicts lead to replication restart. *Nature*. 470(7335):554–57
4. Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E. 2011. Linking RNA Polymerase Backtracking to Genome Instability. *Cell*. 146(4):533–43
5. French S. 1992. Consequences of replication fork movement through transcription units in vivo. *Science*. 258(5086):1362–66
6. Liu B, Alberts BM. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science*. 267(5201):1131–37
7. Deshpande AM, Newlon CS. 1996. DNA Replication Fork Pause Sites Dependent on Transcription. *Science*. 272(5264):1030–33
8. Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J*. 24(6):1267–76
9. Mirkin EV, Mirkin SM. 2005. Mechanisms of Transcription-Replication Collisions in Bacteria. *Mol Cell Biol*. 25(3):888–95
10. Pomerantz RT, O'Donnell M. 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature*. 456(7223):762–66
11. Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, et al. 2017. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell*. 170(4):787-799.e18
12. De Septenville AL, Duigou S, Boubakri H, Michel B. 2012. Replication Fork Reversal after Replication–Transcription Collision. *PLoS Genet*. 8(4):e1002622
13. Mangiameli SM, Merrikh CN, Wiggins PA, Merrikh H. Transcription leads to pervasive replisome instability in bacteria. *eLife*. 6:e19848
14. Pomerantz RT, O'Donnell M. 2010. Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. *Science*. 327(5965):590–92
15. Million-Weaver S, Samadpour AN, Merrikh H. 2015. Replication Restart after Replication-Transcription Conflicts Requires RecA in *Bacillus subtilis*. *J Bacteriol*. 197(14):2374–82
16. Chappidi N, Nascakova Z, Boleslavska B, Zellweger R, Isik E, et al. 2020. Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. *Molecular Cell*. 77(3):528-541.e8

17. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. 2017. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell*. 170(4):774-786.e19

18. Lang KS, Merrikh H. 2021. Topological stress is responsible for the detrimental outcomes of head-on replication-transcription conflicts. *Cell Rep*. 34(9):108797

19. Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. 2013. Accelerated gene evolution via replication-transcription conflicts. *Nature*. 495(7442):10.1038/nature11989

20. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The Nature of Mutations Induced by Replication-Transcription Collisions. *Nature*. 535(7610):178–81

21. Bedinger P, Hochstrasser M, Victor Jongeneel C, Alberts BM. 1983. Properties of the T4 bacteriophage DNA replication apparatus: The T4 *dda* DNA helicase is required to pass a bound RNA polymerase molecule. *Cell*. 34(1):115–23

22. Wu HY, Shyy SH, Wang JC, Liu LF. 1988. Transcription generates positively and negatively supercoiled domains in the template. *Cell*. 53(3):433–40

23. Hiasa H, Marians KJ. 1996. Two Distinct Modes of Strand Unlinking during θ-Type DNA Replication. *Journal of Biological Chemistry*. 271(35):21529–35

24. García-Muse T, Aguilera A. 2016. Transcription–replication conflicts: how they occur and how they are resolved. *Nat Rev Mol Cell Biol*. 17(9):553–63

25. Brewer BJ. 1988. When polymerases collide: Replication and the transcriptional organization of the *E. coli* chromosome. *Cell*. 53(5):679–86

26. Merrikh CN, Brewer BJ, Merrikh H. 2015. The *B. subtilis* Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. *PLoS Genet*. 11(6):e1005289

27. Polard P, Marsin S, McGovern S, Velten M, Wigley DB, et al. 2002. Restart of DNA replication in Gram-positive bacteria: functional characterisation of the *Bacillus subtilis* PriA initiator. *Nucleic Acids Res*. 30(7):1593–1605

28. Sandler SJ, Marians KJ. 2000. Role of PriA in Replication Fork Reactivation in *Escherichia coli*. *J Bacteriol*. 182(1):9–13

29. Sandler SJ, Marians KJ, Zavitz KH, Coutu J, Parent MA, Clark AJ. 1999. *dnaC* mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. *Mol Microbiol*. 34(1):91–101

30. Heller RC, Marians KJ. 2006. Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol*. 7(12):932–43

31. Mahdi AA, Buckman C, Harris L, Lloyd RG. 2006. Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. *Genes Dev*. 20(15):2135–47

32. Xu L, Marians KJ. 2003. PriA Mediates DNA Replication Pathway Choice at Recombination Intermediates. *Molecular Cell*. 11(3):817–26

33. Duckworth AT, Ducos PL, McMillan SD, Satyshur KA, Blumenthal KH, et al. 2023. Replication fork binding triggers structural changes in the PriA helicase that govern DNA replication restart in *E. coli*. *Nat Commun*. 14:2725

34. Brüning J-G, Marians KJ. 2021. Bypass of complex co-directional replication-transcription collisions by replisome skipping. *Nucleic Acids Res*. 49(17):9870–85

35. Agapov A, Olina A, Kulbachinskiy A. 2022. RNA polymerase pausing, stalling and bypass during transcription of damaged DNA: from molecular basis to functional consequences. *Nucleic Acids Res*. 50(6):3018–41

36. Lass-Napiorkowska A, Heyduk T. 2016. Real-time observation of backtracking by bacterial RNA polymerase. *Biochemistry*. 55(4):647–58

37. Epshtein V, Toulmé F, Rahmouni AR, Borukhov S, Nudler E. 2003. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J*. 22(18):4719–27

38. Walter W, Kireeva ML, Studitsky VM, Kashlev M. 2003. Bacterial polymerase and yeast polymerase II use similar mechanisms for transcription through nucleosomes. *J Biol Chem*. 278(38):36148–56

39. Nudler E. 2012. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell*. 149(7):1438–45

40. Kemiha S, Poli J, Lin Y-L, Lengronne A, Pasero P. 2021. Toxic R-loops: Cause or consequence of replication stress? *DNA Repair*. 107:103199

41. Toulmé F, Mosrin-Huaman C, Sparkowski J, Das A, Leng M, Rachid Rahmouni A. 2000. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J*. 19(24):6853–59

42. Tehranchi AK, Blankschien MD, Zhang Y, Halliday J, Srivatsan A, et al. 2010. The Transcription Factor DksA Prevents Disruption of DNA Replication upon Nutritional Stress. *Cell*. 141(4):595–605

43. Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A*. 73(7):2294–98

44. Zardoni L, Nardini E, Brambati A, Lucca C, Choudhary R, et al. 2021. Elongating RNA polymerase II and RNA:DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. *Nucleic Acids Res*. 49(22):12769–84

45. Wellinger RE, Prado F, Aguilera A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and Cellular Biology*. 26(8):3327–34

46. Gan W, Guan Z, Liu J, Gui T, Shen K, et al. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*. 25(19):2041–56

47. Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA-DNA hybrids from generating genome instability. *Mol Cell*. 44(6):978–88

48. Wimberly H, Shee C, Thornton PC, Sivaramakrishnan P, Rosenberg SM, Hastings PJ. 2013. R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat Commun*. 4:2115

49. Lang KS, Merrikh H. 2018. The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. *Annu Rev Microbiol*. 72:71–88

50. Merrikh H. 2017. Spatial and temporal control of evolution through replication-transcription conflicts. *Trends Microbiol*. 25(7):515–21

51. Rocha EPC. 2004. The replication-related organization of bacterial genomes. *Microbiology*. 150(6):1609–27

52. Rocha EPC, Danchin A. 2003. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet*. 34(4):377–78

53. Zheng W-X, Luo C-S, Deng Y-Y, Guo F-B. 2015. Essentiality drives the orientation bias of bacterial genes in a continuous manner. *Sci Rep*. 5:16431

54. Rocha EPC. 2002. Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends in Microbiology*. 10(9):393–95

55. Schroeder JW, Sankar TS, Wang JD, Simmons LA. 2020. The roles of replication-transcription conflict in mutagenesis and evolution of genome organization. *PLoS Genet*. 16(8):e1008987

56. Merrikh CN, Merrikh H. 2018. Gene inversion potentiates bacterial evolvability and virulence. *Nat Commun*. 9:4662

57. Merrikh H, Merrikh C. 2022. Reply to: Testing the adaptive hypothesis of lagging-strand encoding in bacterial genomes. *Nat Commun*. 13(1):2627

58. Abdel-Monem M, Arthur HM, Benz I, Hoffmann-Berling H, Reygers U, et al. 1984. Functions of DNA Helicases in the DNA Metabolism of Escherichia Coli. In *Proteins Involved in DNA Replication*, ed U Hübscher, S Spadari, pp. 385–93. Boston, MA: Springer US

59. David Lane HE, Denhardt DT. 1975. The rep mutation: IV. Slower movement of replication forks in Escherichia coli rep strains. *Journal of Molecular Biology*. 97(1):99–112

60. Guy CP, Atkinson J, Gupta MK, Mahdi AA, Gwynn EJ, et al. 2009. Rep Provides a Second Motor at the Replisome to Promote Duplication of Protein-Bound DNA. *Mol Cell*. 36(4):654–66

61. Boubakri H, de Septenville AL, Viguera E, Michel B. 2010. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J*. 29(1):145–57

62. Hawkins M, Dimude JU, Howard JAL, Smith AJ, Dillingham MS, et al. 2019. Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. *Nucleic Acids Res*. 47(10):5100–5113

63. Gwynn EJ, Smith AJ, Guy CP, Savery NJ, McGlynn P, Dillingham MS. 2013. The Conserved C-Terminus of the PcrA/UvrD Helicase Interacts Directly with RNA Polymerase. *PLoS One*. 8(10):e78141

64. Petit M-A, Ehrlich D. 2002. Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J*. 21(12):3137–47

65. Brüning J-G, Howard JL, McGlynn P. 2014. Accessory Replicative Helicases and the Replication of Protein-Bound DNA. *Journal of Molecular Biology*. 426(24):3917–28

66. Moreno-del Álamo M, Carrasco B, Torres R, Alonso JC. 2021. *Bacillus subtilis* PcrA Helicase Removes Trafficking Barriers. *Cells*. 10(4):935

67. Urrutia-Irazabal I, Ault JR, Sobott F, Savery NJ, Dillingham MS. Analysis of the PcrA-RNA polymerase complex reveals a helicase interaction motif and a role for PcrA/UvrD helicase in the suppression of R-loops. *eLife*. 10:e68829

68. Bharati BK, Gowder M, Zheng F, Alzoubi K, Svetlov V, et al. 2022. Essential Role and Mechanism of Transcription-coupled DNA Repair in Bacteria. *Nature*. 604(7904):152–59

69. Selby CP, Sancar A. 1993. Molecular Mechanism of Transcription-Repair Coupling. *Science*. 260(5104):53–58

70. Epshtein V, Kamarthapu V, McGary K, Svetlov V, Ueberheide B, et al. 2014. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature*. 505(7483):372–77

71. Million-Weaver S, Samadpour AN, Moreno-Habel DA, Nugent P, Brittnacher MJ, et al. 2015. An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*. 112(10):E1096-1105

72. Carvajal-Garcia J, Samadpour AN, Hernandez Viera AJ, Merrikh H. 2023. Oxidative stress drives mutagenesis through transcription-coupled repair in bacteria. *Proceedings of the National Academy of Sciences*. 120(27):e2300761120

73. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM. 1998. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc Natl Acad Sci U S A*. 95(17):10020–25

74. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. 2010. Co-Orientation of Replication and Transcription Preserves Genome Integrity. *PLoS Genet*. 6(1):e1000810

75. Merrikh CN, Weiss E, Merrikh H. 2016. The Accelerated Evolution of Lagging Strand Genes Is Independent of Sequence Context. *Genome Biol Evol*. 8(12):3696–3702

76. Wang J, Rojas P, Mao J, Mustè Sadurnì M, Garnier O, et al. 2021. Persistence of RNA transcription during DNA replication delays duplication of transcription start sites until G2/M. *Cell Rep*. 34(7):108759

77. St Germain CP, Zhao H, Sinha V, Sanz LA, Chédin F, Barlow JH. 2022. Genomic patterns of transcription–replication interactions in mouse primary B cells. *Nucleic Acids Res*. 50(4):2051–73

78. Martin MM, Ryan M, Kim R, Zakas AL, Fu H, et al. 2011. Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res*. 21(11):1822–32

79. Helmrich A, Ballarino M, Tora L. 2011. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell*. 44(6):966–77

80. Saponaro M. 2022. Transcription-Replication Coordination. *Life (Basel)*. 12(1):108

81. Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol Cell*. 34(6):722–34

82. Chen Y-H, Keegan S, Kahli M, Tonzi P, Fenyö D, et al. 2019. Transcription shapes DNA replication initiation and termination in human cells. *Nat Struct Mol Biol*. 26(1):67–77

83. Bhowmick R, Mehta KPM, Lerdrup M, Cortez D. 2023. Integrator facilitates RNAPII removal to prevent transcription-replication collisions and genome instability. *Molecular Cell*

84. Tran PLT, Pohl TJ, Chen C-F, Chan A, Pott S, Zakian VA. 2017. PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nat Commun*. 8:15025

85. Schauer GD, Spenkelink LM, Lewis JS, Yurieva O, Mueller SH, et al. 2020. Replisome bypass of a protein-based R-loop block by Pif1. *Proc Natl Acad Sci U S A*. 117(48):30354–61

86. Aiello U, Challal D, Wentzinger G, Lengronne A, Appanah R, et al. 2022. Sen1 is a key regulator of transcription-driven conflicts. *Molecular Cell*. 82(16):2952-2966.e6

87. Hasanova Z, Klapstova V, Porrua O, Stefl R, Sebesta M. 2023. Human senataxin is a bona fide R-loop resolving enzyme and transcription termination factor. *Nucleic Acids Res*. 51(6):2818–37

88. Patel PS, Algouneh A, Krishnan R, Reynolds JJ, Nixon KCJ, et al. 2023. Excessive transcription-replication conflicts are a vulnerability of *BRCA1*-mutant cancers. *Nucleic Acids Research*. 51(9):4341–62

89. Shallcross LJ, Howard SJ, Fowler T, Davies SC. 2015. Tackling the threat of antimicrobial resistance: from policy to sustainable action. *Philos Trans R Soc Lond B Biol Sci*. 370(1670):20140082

90. Foster JW, Hall HK. 1991. Inducible pH homeostasis and the acid tolerance response of *Salmonella typhimurium*. *J Bacteriol*. 173(16):5129–35

91. Henard CA, Vázquez-Torres A. 2011. Nitric Oxide and *Salmonella* Pathogenesis. *Front Microbiol*. 2:84

92. Rogers AWL, Tsolis RM, Bäumler AJ. 2021. *Salmonella* versus the Microbiome. *Microbiol Mol Biol Rev*. 85(1):e00027-19

93. Ouellette AJ. 2011. Paneth Cell α-Defensins in Enteric Innate Immunity. *Cell Mol Life Sci*. 68(13):2215–29

94. Fang FC, Frawley ER, Tapscott T, Vazquez-Torres A. 2016. Bacterial Stress Responses during Host Infection. *Cell Host Microbe*. 20(2):133–43

95. Fu J, Qi L, Hu M, Liu Y, Yu K, et al. 2017. Salmonella proteomics under oxidative stress reveals coordinated regulation of antioxidant defense with iron metabolism and bacterial virulence. *Journal of Proteomics*. 157:52–58

96. Garai P, Gnanadhas DP, Chakravortty D. 2012. *Salmonella enterica* serovars Typhimurium and Typhi as model organisms. *Virulence*. 3(4):377–88

97. Imlay JA. 2013. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat Rev Microbiol*. 11(7):443–54

98. Hautefort I, Thompson A, Eriksson-Ygberg S, Parker ML, Lucchini S, et al. 2008. During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell Microbiol*. 10(4):958–84

99. Buchmeier NA, Libby SJ, Xu Y, Loewen PC, Switala J, et al. 1995. DNA repair is more important than catalase for *Salmonella virulence* in mice. *J Clin Invest*. 95(3):1047–53

100. Suvarnapunya AE, Lagassé HAD, Stein MA. 2003. The role of DNA base excision repair in the pathogenesis of *Salmonella enterica* serovar Typhimurium. *Molecular Microbiology*. 48(2):549–59

101. Ha KP, Clarke RS, Kim G-L, Brittan JL, Rowley JE, et al. 2020. Staphylococcal DNA Repair Is Required for Infection. *mBio*. 11(6):e02288-20

102. O'Rourke EJ, Chevalier C, Pinto AV, Thiberge JM, Ielpi L, et al. 2003. Pathogen DNA as target for host-generated oxidative stress: Role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. *Proceedings of the National Academy of Sciences*. 100(5):2789–94

103. Colgan AM, Kröger C, Diard M, Hardt W-D, Puente JL, et al. 2016. The Impact of 18 Ancestral and Horizontally-Acquired Regulatory Proteins upon the Transcriptome and sRNA Landscape of *Salmonella enterica* serovar Typhimurium. *PLOS Genetics*. 12(8):e1006258

104. Rhen M, Dorman CJ. 2005. Hierarchical gene regulators adapt *Salmonella enterica* to its host milieus. *Int J Med Microbiol*. 294(8):487–502

105. Kant S, Till JKA, Liu L, Margolis A, Uppalapati S, et al. 2023. Gre factors help *Salmonella* adapt to oxidative stress by improving transcription elongation and fidelity of metabolic genes. *PLOS Biology*. 21(4):e3002051

106. Gaviria-Cantin T, El Mouali Y, Le Guyon S, Römling U, Balsalobre C. 2017. Gre factors-mediated control of *hilD* transcription is essential for the invasion of epithelial cells by *Salmonella enterica* serovar Typhimurium. *PLoS Pathog*. 13(4):e1006312

107. Klein JA, Powers TR, Knodler LA. 2017. Measurement of *Salmonella enterica* Internalization and Vacuole Lysis in Epithelial Cells. *Methods Mol Biol*. 1519:285–96

108. Steele-Mortimer O, Méresse S, Gorvel J-P, Toh B-H, Finlay BB. 1999. Biogenesis of *Salmonella* typhimurium-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cellular Microbiology*. 1(1):33–49

109. Steele-Mortimer O. 2008. Infection of Epithelial Cells With *Salmonella enterica*. In *Bacterial Pathogenesis: Methods and Protocols*, ed FR DeLeo, M Otto, pp. 201–11. Totowa, NJ: Humana Press

110. Powers TR, Haeberle AL, Predeus AV, Hammarlöf DL, Cundiff JA, et al. 2021. Intracellular niche-specific profiling reveals transcriptional adaptations required for the cytosolic lifestyle of *Salmonella enterica*. *PLoS Pathog*. 17(8):e1009280

111. Canals R, Hammarlöf DL, Kröger C, Owen SV, Fong WY, et al. 2019. Adding function to the genome of African *Salmonella* Typhimurium ST313 strain D23580. *PLOS Biology*. 17(1):e3000059

112. Fàbrega A, Vila J. 2013. *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin Microbiol Rev*. 26(2):308–41

113. Brodsky IE, Ernst RK, Miller SI, Falkow S. 2002. *mig-14* Is a *Salmonella* Gene That Plays a Role in Bacterial Resistance to Antimicrobial Peptides. *J Bacteriol*. 184(12):3203–13

114. Detweiler CS, Monack DM, Brodsky IE, Mathew H, Falkow S. 2003. *virK*, *somA* and *rcsC* are important for systemic *Salmonella enterica* serovar Typhimurium infection and cationic peptide resistance. *Mol Microbiol*. 48(2):385–400

115. Freeman JA, Ohl ME, Miller SI. 2003. The *Salmonella enterica* Serovar Typhimurium Translocated Effectors SseJ and SifB Are Targeted to the *Salmonella*-Containing Vacuole. *Infect Immun*. 71(1):418–27

116. Yousuf S, Karlinsey JE, Neville SL, McDevitt CA, Libby SJ, et al. 2020. Manganese import protects *Salmonella enterica* serovar Typhimurium against nitrosative stress†. *Metallomics*. 12(11):1791–1801

117. Mazé A, Glatter T, Bumann D. 2014. The Central Metabolism Regulator EIIA Glc Switches *Salmonella* from Growth Arrest to Acute Virulence through Activation of Virulence Factor Secretion. *Cell Reports*. 7(5):1426–33

118. Hensel M, Shea JE, Raupach B, Monack D, Falkow S, et al. 1997. Functional analysis of *ssaJ* and the *ssaK/U* operon, 13 genes encoding components of the type III secretion apparatus of *Salmonella* Pathogenicity Island 2. *Molecular Microbiology*. 24(1):155–67

119. Yoon H, McDermott JE, Porwollik S, McClelland M, Heffron F. 2009. Coordinated Regulation of Virulence during Systemic Infection of *Salmonella enterica* Serovar Typhimurium. *PLoS Pathog*. 5(2):e1000306

120. Chakravortty D, Rohde M, Jäger L, Deiwick J, Hensel M. 2005. Formation of a novel surface structure encoded by *Salmonella* Pathogenicity Island 2. *EMBO J*. 24(11):2043–52

121. Miki T, Shibagaki Y, Danbara H, Okada N. 2009. Functional Characterization of SsaE, a Novel Chaperone Protein of the Type III Secretion System Encoded by *Salmonella* Pathogenicity Island 2. *Journal of Bacteriology*. 191(22):6843–54

122. Choi J, Salvail H, Groisman EA. 2021. RNA chaperone activates *Salmonella* virulence program during infection. *Nucleic Acids Research*. 49(20):11614–28

123. Nikolaus T, Deiwick J, Rappl C, Freeman JA, Schröder W, et al. 2001. SseBCD proteins are secreted by the type III secretion system of *Salmonella* pathogenicity island 2 and function as a translocon. *J Bacteriol*. 183(20):6036–45

124. Zurawski DV, Stein MA. 2004. The SPI2-encoded SseA chaperone has discrete domains required for SseB stabilization and export, and binds within the C-terminus of SseB and SseD. *Microbiology*. 150(7):2055–68

125. Günster RA, Matthews SA, Holden DW, Thurston TLM. 2017. SseK1 and SseK3 Type III Secretion System Effectors Inhibit NF-κB Signaling and Necroptotic Cell Death in *Salmonella*-Infected Macrophages. *Infect Immun*. 85(3):e00010-17

126. Lee J-W, Lee E-J. 2015. Regulation and function of the *Salmonella* MgtC virulence protein. *J Microbiol*. 53(10):667–72

127. Knuff-Janzen K, Tupin A, Yurist-Doutsch S, Rowland JL, Finlay BB. 2020. Multiple *Salmonella*-pathogenicity island 2 effectors are required to facilitate bacterial establishment of its intracellular niche and virulence. *PLoS One*. 15(6):e0235020

128. Lundgren HK, Björk GR. 2006. Structural alterations of the cysteine desulfurase IscS of *Salmonella enterica* serovar Typhimurium reveal substrate specificity of IscS in tRNA thiolation. *J Bacteriol*. 188(8):3052–62

129. Lou L, Zhang P, Piao R, Wang Y. 2019. *Salmonella* Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. *Frontiers in Cellular and Infection Microbiology*. 9:

130. Yu X-J, Liu M, Matthews S, Holden DW. 2011. Tandem translation generates a chaperone for the *Salmonella* type III secretion system protein SsaQ. *J Biol Chem*. 286(41):36098–107

131. Laptenko O, Lee J, Lomakin I, Borukhov S. 2003. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J*. 22(23):6322–34

132. Ragheb MN, Merrikh C, Browning K, Merrikh H. 2021. Mfd regulates RNA polymerase association with hard-to-transcribe regions in vivo, especially those with structured RNAs. *Proc Natl Acad Sci U S A*. 118(1):e2008498118

133. Park J-S, Marr MT, Roberts JW. 2002. *E. col*i Transcription Repair Coupling Factor (Mfd Protein) Rescues Arrested Complexes by Promoting Forward Translocation. *Cell*. 109(6):757–67

134. Kamarthapu V, Nudler E. 2015. Rethinking Transcription Coupled DNA Repair. *Curr Opin Microbiol*. 24:15–20

135. Martínez LC, Banda MM, Fernández-Mora M, Santana FJ, Bustamante VH. 2014. HilD Induces Expression of *Salmonella* Pathogenicity Island 2 Genes by Displacing the Global Negative Regulator H-NS from *ssrAB*. *Journal of Bacteriology*. 196(21):3746–55

136. Knodler LA, Nair V, Steele-Mortimer O. 2014. Quantitative Assessment of Cytosolic *Salmonella* in Epithelial Cells. *PLOS ONE*. 9(1):e84681

137. Cui G, Wang J, Qi X, Su J. 2018. Transcription Elongation Factor GreA Plays a Key Role in Cellular Invasion and Virulence of *Francisella tularensis* subsp. novicida. *Sci Rep*. 8(1):6895

138. Feng S, Liu Y, Liang W, El-Sayed Ahmed MAE-G, Zhao Z, et al. 2020. Involvement of Transcription Elongation Factor GreA in *Mycobacterium* Viability, Antibiotic Susceptibility, and Intracellular Fitness. *Front Microbiol*. 11:413

139. Flashner Y, Mamroud E, Tidhar A, Ber R, Aftalion M, et al. 2004. Generation of *Yersinia pestis* Attenuated Strains by Signature-Tagged Mutagenesis in Search of Novel Vaccine Candidates. *Infect Immun*. 72(2):908–15

140. Su'etsugu M, Errington J. 2011. The Replicase Sliding Clamp Dynamically Accumulates behind Progressing Replication Forks in *Bacillus subtilis* Cells. *Molecular Cell*. 41(6):720–32

141. Spinks RR, Spenkelink LM, Stratmann SA, Xu Z-Q, Stamford NPJ, et al. 2021. DnaB helicase dynamics in bacterial DNA replication resolved by single-molecule studies. *Nucleic Acids Research*. 49(12):6804–16

142. Reyes-Lamothe R, Sherratt DJ, Leake MC. 2010. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science*. 328(5977):498–501

143. Steele-Mortimer O, Brumell JH, Knodler LA, Méresse S, Lopez A, Finlay BB. 2002. The invasion-associated type III secretion system of *Salmonella enterica* serovar Typhimurium is necessary for intracellular proliferation and vacuole biogenesis in epithelial cells. *Cellular Microbiology*. 4(1):43–54

144. de Chastellier C, Berche P. 1994. Fate of *Listeria monocytogenes* in murine macrophages: evidence for simultaneous killing and survival of intracellular bacteria. *Infection and Immunity*. 62(2):543–53

145. Sanz LA, Chédin F. 2019. High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nat Protoc*. 14(6):1734–55

146. Hinton JC, Hautefort I, Eriksson S, Thompson A, Rhen M. 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Current Opinion in Microbiology*. 7(3):277–82

147. Tsunaka Y, Haruki M, Morikawa M, Kanaya S. 2001. Strong nucleic acid binding to the *Escherichia coli* RNase HI mutant with two arginine residues at the active site. *Biochim Biophys Acta*. 1547(1):135–42

148. Katayanagi K, Okumura M, Morikawa K. 1993. Crystal structure of *Escherichia coli* RNase HI in complex with Mg2+ at 2.8 A resolution: proof for a single Mg(2+)-binding site. *Proteins*. 17(4):337–46

149. Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, et al. 2017. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell*. 68(4):745-757.e5

150. Chen J-Y, Zhang X, Fu X-D, Chen L. 2019. R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. *Nat Protoc*. 14(5):1661–85

151. Chédin F, Hartono SR, Sanz LA, Vanoosthuyse V. 2021. Best practices for the visualization, mapping, and manipulation of R-loops. *The EMBO Journal*. 40(4):e106394

152. Myers KS, Park DM, Beauchene NA, Kiley PJ. 2015. Defining bacterial regulons using ChIP-seq. *Methods*. 86:80–88

153. Johansson J, Freitag NE. 2019. Regulation of *Listeria monocytogenes* Virulence. *Microbiology Spectrum*

154. Jenul C, Horswill AR. 2018. Regulation of *Staphylococcus aureus* virulence. *Microbiol Spectr*. 6(1):10.1128/microbiolspec.GPP3-0031–2018

155. Huang D, Johnson AE, Sim BS, Lo TW, Merrikh H, Wiggins PA. 2023. The in vivo measurement of replication fork velocity and pausing by lag-time analysis. *Nat Commun*. 14:1762

156. Luria SE, Delbrück M. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. 28(6):491–511

157. Poteete AR. 2011. Recombination Phenotypes of *Escherichia coli greA* Mutants. *BMC Molecular Biology*. 12(1):12

158. Gómez-Marroquín M, Martin HA, Pepper A, Girard ME, Kidman AA, et al. 2016. Stationary-Phase Mutagenesis in Stressed *Bacillus subtilis* Cells Operates by Mfd-Dependent Mutagenic Pathways. *Genes*. 7(7):33

159. Ross C, Pybus C, Pedraza-Reyes M, Sung H-M, Yasbin RE, Robleto E. 2006. Novel Role of *mfd*: Effects on Stationary-Phase Mutagenesis in *Bacillus subtilis*. *J Bacteriol*. 188(21):7512–20

160. Poetsch AR. 2020. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Computational and Structural Biotechnology Journal*. 18:207–19

161. Zhang Y, Guo Q, Fang X, Yuan M, Hu W, et al. 2023. Destroying glutathione peroxidase improves the oxidative stress resistance and pathogenicity of *Listeria monocytogenes*. *Frontiers in Microbiology*. 14:

162. Mains DR, Eallonardo SJ, Freitag NE. 2021. Identification of *Listeria monocytogenes* Genes Contributing to Oxidative Stress Resistance under Conditions Relevant to Host Infection. *Infect Immun*. 89(4):e00700-20

163. Imashimizu M, Takahashi H, Oshima T, McIntosh C, Bubunenko M, et al. 2015. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol*. 16(1):98

164. Sun Z, Yakhnin AV, FitzGerald PC, McIntosh CE, Kashlev M. 2021. Nascent RNA sequencing identifies a widespread sigma70-dependent pausing regulated by Gre factors in bacteria. *Nat Commun*. 12:906

165. Yakhnin AV, FitzGerald PC, McIntosh C, Yakhnin H, Kireeva M, et al. 2020. NusG controls transcription pausing and RNA polymerase translocation throughout the *Bacillus subtilis* genome. *Proc Natl Acad Sci U S A*. 117(35):21628–36

166. Li X, Thomason LC, Sawitzke JA, Costantino N, Court DL. 2013. Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res*. 41(22):e204

167. Thomason LC, Costantino N, Li X, Court DL. 2023. Recombineering: Genetic Engineering in *Escherichia coli* Using Homologous Recombination. *Current Protocols*. 3(2):e656

168. Yu D, Ellis HM, Lee E-C, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 97(11):5978–83

169. Leman AR, Noguchi E. 2014. Chromatin Immunoprecipitation to Investigate Origin Association of Replication Factors in Mammalian Cells. *Methods Mol Biol*. 1170:539–47

170. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–20

171. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 9(4):357

172. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078–79

173. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30(7):923–30

174. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*. 47(D1):D419–26

175. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet*. 25(1):25–29

176. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, et al. 2021. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*. 49(D1):D325–34

177. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6):841–42

178. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, et al. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*. 44(W1):W160–65

179. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. 2011. Integrative Genomics Viewer. *Nat Biotechnol*. 29(1):24–26

180. Srikumar S, Kröger C, Hébrard M, Colgan A, Owen SV, et al. 2015. RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of *Salmonella* Typhimurium. *PLoS Pathog*. 11(11):e1005262

181. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15(12):550